

# ADVANCES IN OIL CROPS RESEARCH – CLASSICAL AND NEW APPROACHES TO ACHIEVE SUSTAINABLE PRODUCTIVITY

EDITED BY: Dragana Miladinović, Johann Vollmann, Leire Molinero-Ruiz  
and Mariela Torres

PUBLISHED IN: *Frontiers in Plant Science*







# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88963-068-4

DOI 10.3389/978-2-88963-068-4

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)



# ADVANCES IN OIL CROPS RESEARCH – CLASSICAL AND NEW APPROACHES TO ACHIEVE SUSTAINABLE PRODUCTIVITY

Topic Editors:

**Dragana Miladinović**, Institute of Field and Vegetable Crops, Serbia

**Johann Vollmann**, University of Natural Resources and Life Sciences Vienna, Austria

**Leire Molinero-Ruiz**, Instituto de Agricultura Sostenible, Spain

**Mariela Torres**, Instituto Nacional de Tecnología Agropecuaria, Argentina



Image: Ilija Radeka, IFVCNS.

**Citation:** Miladinović, D., Vollmann, J., Molinero-Ruiz, L., Torres, M., eds. (2019). Advances in Oil Crops Research – Classical and New Approaches to Achieve Sustainable Productivity. Lausanne: Frontiers Media.  
doi: 10.3389/978-2-88963-068-4



# Table of Contents

- 06 Editorial: Advances in Oil Crops Research—Classical and New Approaches to Achieve Sustainable Productivity**  
Dragana Miladinović, Johann Vollmann, Leire Molinero-Ruiz and Mariela Torres
- 10 Allelic Variation and Distribution of the Major Maturity Genes in Different Soybean Collections**  
Jegor Miladinović, Marina Čeran, Vuk Đorđević, Svetlana Balešević-Tubić, Kristina Petrović, Vojin Đukić and Dragana Miladinović
- 18 OliveCan: A Process-Based Model of Development, Growth and Yield of Olive Orchards**  
Álvaro López-Bernal, Alejandro Morales, Omar García-Tejera, Luca Testi, Francisco Orgaz, J. P. De Melo-Abreu and Francisco J. Villalobos
- 29 Association Mapping for Important Agronomic Traits in Safflower (*Carthamus tinctorius* L.) Core Collection Using Microsatellite Markers**  
Heena Ambreen, Shivendra Kumar, Amar Kumar, Manu Agarwal, Arun Jagannath and Shailendra Goel
- 44 Genetics, Host Range, and Molecular and Pathogenic Characterization of *Verticillium dahliae* From Sunflower Reveal Two Differentiated Groups in Europe**  
Alberto Martín-Sanz, Sandra Rueda, Ana B. García-Carneros, Sara González-Fernández, Pedro Miranda-Fuentes, Sandra Castuera-Santacruz and Leire Molinero-Ruiz
- 57 Fruit Phenolic Profiling: A New Selection Criterion in Olive Breeding Programs**  
Ana G. Pérez, Lorenzo León, Carlos Sanz and Raúl de la Rosa
- 71 Using Wild Olives in Breeding Programs: Implications on Oil Quality Composition**  
Lorenzo León, Raúl de la Rosa, Leonardo Velasco and Angelina Belaj
- 80 Divergent N Deficiency-Dependent Senescence and Transcriptome Response in Developmentally Old and Young Brassica napus Leaves**  
Vajiheh Safavi-Rizi, Jürgen Franzaring, Andreas Fangmeier and Reinhard Kunze
- 95 Metarhizium brunneum (Ascomycota; Hypocreales) Treatments Targeting Olive Fly in the Soil for Sustainable Crop Production**  
Meelad Yousef, Carmen Alba-Ramírez, Inmaculada Garrido Jurado, Jordi Mateu, Silvia Raya Díaz, Pablo Valverde-García and Enrique Quesada-Moraga
- 106 Sunflower Hybrid Breeding: From Markers to Genomic Selection**  
Aleksandra Dimitrijevic and Renate Horn
- 126 Genetic Architecture of Capitule Glandular Trichome Density in Florets of Domesticated Sunflower (*Helianthus annuus* L.)**  
Qing-Ming Gao, Nolan C. Kane, Brent S. Hulke, Stephan Reinert, Cloe S. Pogoda, Silas Tittes and Jarrad R. Prasifka



- 141 ***An Oleuropein  $\beta$ -Glucosidase From Olive Fruit is Involved in Determining the Phenolic Composition of Virgin Olive Oil***  
David Velázquez-Palmero, Carmen Romero-Segura, Rosa García-Rodríguez, María L. Hernández, Fabián E. Vaistij, Ian A. Graham, Ana G. Pérez and José M. Martínez-Rivas
- 153 ***Variability in Susceptibility to Anthracnose in the World Collection of Olive Cultivars of Cordoba (Spain)***  
Juan Moral, Carlos J. Xaviér, José R. Viruega, Luis F. Roca, Juan Caballero and Antonio Trapero
- 164 ***Olive Cultivation in the Southern Hemisphere: Flowering, Water Requirements and Oil Quality Responses to New Crop Environments***  
Mariela Torres, Pierluigi Pierantozzi, Peter Searles, M. Cecilia Rousseaux, Georgina García-Inza, Andrea Miserere, Romina Bodoira, Cibeles Contreras and Damián Maestri
- 176 ***Genomic Prediction of Sunflower Hybrids Oil Content***  
Brigitte Mangin, Fanny Bonnafous, Nicolas Blanchet, Marie-Claude Boniface, Emmanuelle Bret-Mestries, Sébastien Carrère, Ludovic Cottret, Ludovic Legrand, Gwenola Marage, Prune Pegot-Espagnet, Stéphane Munos, Nicolas Pouilly, Felicity Vear, Patrick Vincourt and Nicolas B. Langlade
- 188 ***Identification and Functional Annotation of Genes Differentially Expressed in the Reproductive Tissues of the Olive Tree (*Olea europaea* L.) Through the Generation of Subtractive Libraries***  
Adoración Zafra, Rosario Carmona, José A. Traverso, John T. Hancock, Maria H. S. Goldman, M. Gonzalo Claros, Simon J. Hiscock and Juan D. Alche
- 206 ***Genetic Tracing of *Jatropha curcas* L. From its Mesoamerican Origin to the World***  
Haiyan Li, Suguru Tsuchimoto, Kyuya Harada, Masanori Yamasaki, Hiroe Sakai, Naoki Wada, Atefeh Alipour, Tomohiro Sasai, Atsushi Tsunekawa, Hisashi Tsujimoto, Takayuki Ando, Hisashi Tomemori, Shusei Sato, Hideki Hirakawa, Victor P. Quintero, Alfredo Zamarripa, Primitivo Santos, Adel Hegazy, Abdalla M. Ali and Kiichi Fukui
- 217 ***Development of Highly Informative Genome-Wide Single Sequence Repeat Markers for Breeding Applications in Sesame and Construction of a Web Resource: SisatBase***  
Komivi Dossa, Jingyin Yu, Boshou Liao, Ndiaga Cisse and Xiurong Zhang
- 227 ***The First Molecular Identification of an Olive Collection Applying Standard Simple Sequence Repeats and Novel Expressed Sequence Tag Markers***  
Soraya Mousavi, Roberto Mariotti, Luca Regni, Luigi Nasini, Marina Bufacchi, Saverio Pandolfi, Luciana Baldoni and Primo Proietti
- 237 ***The Emerging Oilseed Crop *Sesamum indicum* Enters the "Omics" Era***  
Komivi Dossa, Diaga Diouf, Linhai Wang, Xin Wei, Yanxin Zhang, Mareme Niang, Daniel Fonceka, Jingyin Yu, Marie A. Mmadi, Louis W. Yehouessi, Boshou Liao, Xiurong Zhang and Ndiaga Cisse
- 253 ***Recent Genetic Gains in Nitrogen Use Efficiency in Oilseed Rape***  
Andreas Stahl, Mara Pfeifer, Matthias Frisch, Benjamin Wittkop and Rod J. Snowdon



**266 Use of Blue-Green Fluorescence and Thermal Imaging in the Early Detection of Sunflower Infection by the Root Parasitic Weed *Orobanche cumana* Wallr.**

Carmen M. Ortiz-Bustos, María L. Pérez-Bueno, Matilde Barón and Leire Molinero-Ruiz

**276 Genome-Wide Identification of the MIKC-Type MADS-Box Gene Family in *Gossypium hirsutum* L. Unravels Their Roles in Flowering**

Zhongying Ren, Daoqian Yu, Zhaoen Yang, Changfeng Li, Ghulam Qanmber, Yi Li, Jie Li, Zhao Liu, Lili Lu, Lingling Wang, Hua Zhang, Quanjia Chen, Fuguang Li and Zuoren Yang

**293 Modeling Allometric Relationships in Leaves of Young Rapeseed (*Brassica napus* L.) Grown at Different Temperature Treatments**

Tian Tian, Lingtong Wu, Michael Henke, Basharat Ali, Weijun Zhou and Gerhard Buck-Sorlin





# Editorial: Advances in Oil Crops Research—Classical and New Approaches to Achieve Sustainable Productivity

Dragana Miladinović<sup>1\*</sup>, Johann Vollmann<sup>2</sup>, Leire Molinero-Ruiz<sup>3</sup> and Mariela Torres<sup>4</sup>

<sup>1</sup> Sunflower Department, Institute of Field and Vegetable Crops, Novi Sad, Serbia, <sup>2</sup> Department of Crop Sciences, University of Natural Resources and Life Sciences Vienna, Vienna, Austria, <sup>3</sup> Departamento de Protección de Cultivos, Instituto de Agricultura Sostenible, Córdoba, Spain, <sup>4</sup> Estación Experimental Agropecuaria San Juan, Instituto Nacional de Tecnología Agropecuaria, Buenos Aires, Argentina

**Keywords:** oil crops, breeding, quality, production, diversity

## Editorial on the Research Topic

### Advances in Oil Crops Research—Classical and New Approaches to Achieve Sustainable Productivity

The world production of main oil crops is steadily increasing, mainly due to population growth and increased use of oil crops in bio-fuel production and in edible vegetable oils. From the perspective of sowing area in the world, oil crops are only preceded by cereals in importance. Edible or industrial oils are extracted from seeds, fruits or mesocarp, and nuts of both annual and perennial species. Oil can be obtained from about 40 different crops, but soybean, sunflower, olive tree, and rapeseed have a major importance in the total world trade. The purpose of the Research Topic “Advances in Oil Crops Research—Classical and New Approaches to Achieve Sustainable Productivity” is to provide the reader compiled information of the latest research results about different aspects of oil crops.

This research topic incorporates 23 publications including 19 research papers, three review articles, and one perspective.

## OPEN ACCESS

### Edited and reviewed by:

Sergio J. Ochatt,  
INRA UMR1347 Agroécologie, France

### \*Correspondence:

Dragana Miladinović  
draganavas@yahoo.com

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 28 February 2019

**Accepted:** 31 May 2019

**Published:** 18 June 2019

### Citation:

Miladinović D, Vollmann J,  
Molinero-Ruiz L and Torres M (2019)  
Editorial: Advances in Oil Crops  
Research—Classical and New  
Approaches to Achieve Sustainable  
Productivity. *Front. Plant Sci.* 10:791.  
doi: 10.3389/fpls.2019.00791

## DIVERSITY CHARACTERIZATION

Germplasm collections are very important for conservation of oil crop diversity. In olive tree, only few collections have been fully genotyped and characterized. Mousavi et al. screened previously uncharacterized olive trees from the olive collection of Perugia University with standard simple sequence repeats (SSR) and new expressed sequence tag (EST)-SSR markers. They found the new OLea Expressed Sequence Tags (OLEST) SSR markers to be as effective as other commonly used markers. OLEST-SSR markers are very useful to create a common database from worldwide collections of olive cultivars.

Jatropha has been recognized as one of the best candidates for future biodiesel production. Attempts for commercial cultivation of the species in Africa and Asia have failed due to low productivity. Thus, better yielding commercial cultivars must be developed on the basis of an increased knowledge on the species genetic diversity. Li et al. analyzed the genetic diversity of 246 Jatropha accessions collected in Africa, Asia, and Mesoamerica by means of SSR and retrotransposon-based insertion polymorphism markers. The authors identified that two accessions from Veracruz (Mesoamerica) were the source of all Jatropha of Africa and Asia, and hypothesized that human selection caused low Jatropha productivity. They have also suggested that Jatropha of Africa and Asia can be improved through the implementation of breeding strategies.

## ENVIRONMENT AND AGRONOMY

Olive is a crop very well-adapted to temperature and precipitation regimes typical of regions in the Mediterranean Basin. The increasing international demand for olive oil has led to the expansion of olive cultivation in new production regions where environmental conditions differ from those of the Mediterranean Basin and there is little information about olive adaptation to these new environments. Torres et al. reviewed the scientific literature on olive cultivation in non-Mediterranean environments, with focus on chilling requirements for flowering, water requirements, and irrigation management, as well as environmental effects on fruit oil content and quality. Their information could be useful to determine whether environmental conditions in new growing regions are appropriate for achieving sustainable olive and oil productions. Several simulation models of the olive crop have been formulated so far, but none of them is capable to account for the impacts of environmental conditions and management practices on growth and productivity in the absence of nutrient deficiencies, diseases, and pests. López-Bernal et al. described OliveCan, a process-oriented model formulated using previous models by the group, that enables the assessment of combined effects of management operations, soil traits, and weather over crop performance both under unstressed and water deficit conditions.

Functional-structural plant modeling (FSPM) is a dynamic method for prediction of plant growth under varying environmental conditions. As the temperature is a primary factor affecting growth and development of rapeseed, Tian et al. tested three different temperature treatments by using a FSPM of seedlings with a growth function used for leaf extension and biomass accumulation implemented by combining measurements with literature data.

## BIOTIC STRESSES

Effective, economically viable, and environmentally friendly methods for pest control are needed in the framework of sustainable agriculture in Europe. These novel methods are particularly important when dealing with insect pests that have developed insecticide resistance, such as olive fruit fly. Yousef et al. tested the efficacy of soil treatment with the entomopathogenic fungus *Metarhizium brunneum* for controlling the olive fly and found that, during spring, the density of pest population emerging from the soil was reduced by up to 70%. They also studied the retention of conidia as a function of soil type and rain amount.

Several secondary metabolites can be produced by capitate glandular trichomes (CGT). Some of these secondary metabolites provide durable resistance to insect pests. The specialist pest sunflower moth is combated by host resistance based on CGT. Gao et al. have identified two major QTLs controlling CGT density in sunflower florets. They constructed a genetic linkage map from genotyping-by-sequencing data using SNP markers. The obtained results will advance the understanding of CGT-mediated resistance to insect pests in sunflower. The authors

also provide a resource for marker-assisted selection for insect resistance in this oil crop species.

Anthraxnose is an important disease that causes fruit rot and branch dieback of olive trees. The appearance of symptoms and their evolution in time are highly dependent on cultivar susceptibility and environmental conditions. Moral et al. looked for potential sources of resistance to *Colletotrichum acutatum*, causal agent of anthracnose, in the World Olive Germplasm Bank located in Córdoba (Spain). The authors also described the original methodology that could be also used in the evaluation of olive cultivar responses to other aerial fungi affecting leaves and fruit.

*Verticillium dahliae* is a soil-borne fungus that causes wilt and leaf mottle in sunflower. The incidence of this disease of sunflower has increased in recent years consequently affecting sunflower oil production. Martin-Sanz et al. described the population structure of *V. dahliae* affecting sunflowers in Europe. According to the results of genetic, molecular, and pathogenic analyses of particular traits of the isolates, they identified two groups. One was diverse and included *V. dahliae* isolates from Eastern Europe. The other was highly homogeneous and clustered isolates from Western Europe. The authors recommend taking into account the existence of these two groups of *V. dahliae* when looking for sources of resistance to this disease in European environments.

*Orobanche cumana* L. is one of the main limiting factors in sunflower production. Infection of this parasitic weed occurs early after sowing and affects host physiology from that moment and during underground parasite stages. Ortiz-Bustos et al. used the blue-green fluorescence (BGF) emission and thermal imaging of leaves for the detection of the infection of sunflower by *O. cumana* during underground parasite development. Lower BGF emissions and higher temperatures were detected in leaves of infected plants as compared to those of healthy plants, indicating that BGF imaging and thermography could be used for fast and non-destructive screening of lines of sunflower from breeding programs for resistance to this parasitic weed.

## PRODUCT QUALITY

The high genetic diversity of wild olives can be used for introduction of some important agronomic traits into cultivated varieties. However, some of the traits introduced from wild relatives can have negative effects on some other desirable traits such as the composition of virgin olive oil (VOO). As VOO from olive cultivars is highly appreciated for its fatty acid composition, León et al. compared the fatty acid profiles of VOO from wild olives and olive cultivars. They found that the use of wild germplasm in olive breeding programs does not have a negative impact on fatty acid composition, tocopherol content, and tocopherol and phytosterol profiles if the selection for these traits is conducted from early generations of crossings.

Phenolic compounds that are present in olive oil have a strong antioxidant activity and are one of the factors responsible for health benefits associated with VOO consumption. However, phenolic composition is not a common selection criterion in olive

breeding programs. A simplified procedure for phenolic profiling of the olive fruits avoiding the previous step of oil extraction was developed by Pérez et al. When applying it in the detection of phenolic content variability in different olive genotypes, a high genotypic variance of fruit phenol content was found in the tested genotypes. The authors concluded that the observed high genotypic variance together with the simplified method for fruit phenol evaluation can be useful in olive tree breeding for improved phenolic profile of the oil. Velázquez-Palmero et al. studied genes and enzymes responsible for the phenolic composition of VOO, such as  $\beta$ -glucosidases. They found that olive *GLU* gene expression is cultivar-dependent and regulated by temperature, light, and water regime, as well as its transcription in olive fruit spatially and temporally regulated. Their study is a further step in elucidating the factors involved in biosynthesis of the major phenolic compounds in VOO. It can also help to develop molecular markers for the selection of cultivars with improved oil quality characteristics.

## PLANT NUTRITION

Nitrogen fertilization allows farmers to improve plant productivity. However, it has to be carefully managed to avoid harmful environmental impacts due to nitrogen loss. Increasing nitrogen use efficiency (NUE), or the amount of N fertilizer taken up and used by the crop, is vital to solve the conflict between productivity and the protection of the environment. The NUE of rapeseed is low, although it has a high uptake capacity for inorganic N. The development of cultivars with improved NUE is a major goal in rapeseed breeding. A collection of 30 elite rapeseed varieties registered between 1989 and 2014 was tested by Stahl et al. under two different fertilization regimes in a 2-year experiment in 10 different environments for changes in seed yield and seed quality traits. Their data revealed that genetic improvement in combination with reduced N fertilizer inputs has a tremendous potential to increase NUE of rapeseed.

Using a transcriptome approach, Safavi-Rizi et al. found an opposing effect of N depletion on gene regulation: some genes were up- or downregulated in N-depleted lower canopy leaves, contrarily to their regulation in upper leaves under ample N supply. Also, some genes were expressed as associated to senescence in rapeseed, but not in *Arabidopsis*. They hypothesized that some of these genes may have rapeseed-specific functions in nitrogen remobilization during N-deficiency induced leaf senescence. Some genes also seem to contribute to differences in senescence and N resources mobilization in upper and lower canopy leaves.

## PLANT GROWTH AND DEVELOPMENT

Production of olive trees relies on the successful achievement of sexual reproduction. Due to self-incompatibility in the species, successful fertilization is highly favored by the presence of pollen grains from a different cultivar. Zafra et al. studied the reproductive biology of the olive through identification of

key gene products involved in pollen and pistil physiology. Aiming at elucidating the biological processes occurring during the courtship, pollen grain germination, and fertilization in olive, they constructed SSH libraries using pistil and pollen that allowed the identification of transcripts with important roles in stigma physiology.

In the MADS-Box gene family, MADS intervening keratin-like and MIKC-type are found only in plants. They encode transcription factors that have major roles in plant growth and developmental processes. However, a comprehensive analysis of this gene family in *G. hirsutum*, concerning characterization and functions, has not yet been reported. Ren et al. analyzed MIKC-type MADS genes in the tetraploid cotton, which is the most widely cultivated cotton species. A total of 110 *GhMIKC* genes were identified, located in the genome, and phylogenetically classified into 13 subfamilies. Since most of them were highly expressed in floral organs, the study provides useful information of the involvement of *GhMIKCs* regulation in cotton flowering.

## MOLECULAR TOOLS FOR BREEDING

In their review, Dimitrijević and Horn elaborated present and past achievements in sunflower breeding and biotechnology, as well as the future perspectives of using modern molecular tools in this research area. They emphasized that the key for further successes in sunflower breeding is in integrative approaches to the genetics of complex quantitative traits and physiological and biochemical mechanisms involved, a challenge that must be faced with new high-throughput technologies in combination with new genomic-based breeding strategies.

Genomic selection (GS) models can predict characters performance after being trained on phenotypic and genotypic information for those characters, and their most important advantage is that they do not have to include all hybrid parents. Mangin et al. compared the accuracy of sunflower oil content prediction of commonly used general combining ability (GCA) and genomic predictions. Their results show that GS provides more accurate predictions compared to the classical predictions based on GCA with at least one parent is untested or not well-characterized, thus increasing breeding speediness and efficiency.

The lack of identified marker-trait associations is a major limitation toward development of successful marker-assisted breeding programs in safflower. Ambreen et al. tested a safflower panel (CartAP) comprising of 124 accessions for its suitability for association mapping. They have detected several marker-trait associations, either in a majority or in some environments. The results of this study will facilitate wider use of marker-assisted breeding in safflower, as well as identification of genetic determinants of trait variability.

Soybean flowering time and maturity are controlled by *E* genes whose different allelic combinations determine soybean adaptation to a certain latitude. Miladinović et al. have described the first attempt to assess adaptation of soybean genotypes developed at Institute of Field and Vegetable Crops, Novi Sad, Serbia based on *E* gene variation, as well as to comparatively



assess *E* gene variation in North-American, Chinese and European genotypes. As *e1-as/e2/E3/E4* was the most common genotype and was present in the top performing genotypes, this specific allele combination was proposed as optimal for the environments of Central-Eastern Europe.

Despite its economic and nutritional importance, genetic improvement of sesame lags behind that of the other oil crops. In their review, Dossa et al. described the evolution of research in sesame genetics and highlighted the recent advances in the—omics area. They have also discussed the future prospects for genetic improvement and better expansion of sesame. In another study, Dossa et al. developed SisatBase, an user-friendly database providing free access to SSRs data of sesame. It is as well an integrated platform for functional analyses of this oil crop.

## CONCLUSION

The paraphrased last sentence from the review of Dimitrijević and Horn is a perfect conclusion message for the topic and a recommendation for future work on oil crops:

“... Only a combined effort of the oil crop research community can make oil crops more competitive to other crops. The new high-throughput technologies combined with new genomic-based breeding strategies give us the opportunity, as never before, to understand and mine genetic variation and to use it for improvement of oil crops...”.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2019 Miladinović, Vollmann, Molinero-Ruiz and Torres. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Allelic Variation and Distribution of the Major Maturity Genes in Different Soybean Collections

Jegor Miladinović<sup>1\*</sup>, Marina Čeran<sup>1</sup>, Vuk Đorđević<sup>1</sup>, Svetlana Balešević-Tubić<sup>1</sup>, Kristina Petrović<sup>1</sup>, Vojin Đukić<sup>1</sup> and Dragana Miladinović<sup>2</sup>

<sup>1</sup> Soybean Department, Institute of Field and Vegetable Crops, Novi Sad, Serbia, <sup>2</sup> Industrial Crops Department, Institute of Field and Vegetable Crops, Novi Sad, Serbia

## OPEN ACCESS

### Edited by:

Rafael Lozano,  
University of Almería, Spain

### Reviewed by:

Gurvant Baliram Patil,  
University of Minnesota, United States  
Matthew Nicholas Nelson,  
Royal Botanic Gardens, Kew,  
United Kingdom

### \*Correspondence:

Jegor Miladinović  
jegor.miladinovic@ifvcns.ns.ac.rs

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 09 October 2017

**Accepted:** 16 August 2018

**Published:** 04 September 2018

### Citation:

Miladinović J, Čeran M, Đorđević V, Balešević-Tubić S, Petrović K, Đukić V and Miladinović D (2018) Allelic Variation and Distribution of the Major Maturity Genes in Different Soybean Collections. *Front. Plant Sci.* 9:1286. doi: 10.3389/fpls.2018.01286

Soybean time of flowering and maturity are genetically controlled by *E* genes. Different allelic combinations of these genes determine soybean adaptation to a specific latitude. The paper describes the first attempt to assess adaptation of soybean genotypes developed and realized at Institute of Field and Vegetable Crops, Novi Sad, Serbia [Novi Sad (NS) varieties and breeding lines] based on *E* gene variation, as well as to comparatively assess *E* gene variation in North-American (NA), Chinese, and European genotypes, as most of the studies published so far deal with North-American and Chinese cultivars and breeding material. Allelic variation and distribution of the major maturity genes (*E1*, *E2*, *E3*, and *E4*) has been determined in 445 genotypes from soybean collections of NA ancestral lines, Chinese germplasm, and European varieties, as well as NS varieties and breeding lines. The study showed that allelic combinations of *E1–E4* genes significantly determined the adaptation of varieties to different geographical regions, although they have different impacts on maturity. In general, each collection had one major *E* genotype haplogroup, comprising over 50% of the lines. The exceptions were European varieties that had two predominant haplogroups and NA ancestral lines distributed almost evenly among several haplogroups. As *e1-as/e2/E3/E4* was the most common genotype in NS population, present in the best-performing genotypes in terms of yield, this specific allele combination was proposed as the optimal combination for the environments of Central-Eastern Europe.

**Keywords:** photoperiodism, *E* genes, soybean, maturity, allelic variation, SNP

## INTRODUCTION

Photoperiodism is reaction of a plant to the length of day and night. The discovery of photoperiodism in soybean and tobacco (Garner and Allard, 1930), and subsequently in numerous other plant species, paved the way for investigation of mechanisms which control seasonal rhythm in plants, such as flowering, growth and reproduction (Putterill et al., 2009). Besides plant growth, photoperiod also affects other aspects of plant development, including changes of developmental phases and overall length of vegetation period. Knowledge of photoperiodic response of cultivated crops is one of the vital aspects of modern-day plant production, as shifting seasonal timing of reproduction became a major goal of plant breeding programs in their effort to produce novel varieties better adapted to local environments and variable climate conditions. This also stands for

soybean [*Glycine max* (L.) Merr.] since it is a photoperiod-sensitive short-day plant, i.e., soybean transition from vegetative to reproductive stage directly depends on the length of day (Miladinović and Đorđević, 2011).

## SOYBEAN AND PHOTOPERIODICITY

Estimation of plant phenological stage of development, as a function of specific environment, is the key factor in any attempt of modifying plant growth, adaptation, or productivity (Jones and Laing, 1978; Davidson and Campbell, 1983; Hodges and French, 1985; Wang et al., 1987; Colson et al., 1995; Miladinović and Đorđević, 2011). Soybean phenology is difficult to predict, as it depends on photoperiod (Garner and Allard, 1930; Seddigh et al., 1989), temperature (Egli and Wardlaw, 1980; Board and Hall, 1984), and amount of plant-available water (Meckel et al., 1984; Miladinović et al., 2006).

Dependence of soybean on the length of the day resulted in its geographic distribution into 13 maturity group (MG) zones, from MG000 comprised of varieties which can thrive in higher geographic latitudes, to MGX which includes varieties grown in lower latitudes (Hartwig, 1973). By definition, MGs are the result of classification of soybeans based on their growth and development. The difference between MGs in a certain area are caused by photoperiodic requirements of a variety, while the difference in maturity date between two adjacent MGs ranges from 10 to 18 days. In soybeans, the critical photoperiod, or the definite day length light period above or below which the plant never blooms, decreases progressively from higher to lower geographic latitudes. Photoperiodic requirements thus reduce soybean growing area to a narrow latitude range of around 200 km (Scott and Aldrich, 1983). If grown in higher geographic latitudes compared to its optimal growing area, a soybean variety will flower and mature later, while a variety grown in lower geographic latitudes will flower and mature earlier, resulting in lower vegetative mass and lower yields. Hence, there is an optimal MG for each soybean growing region (Miladinović and Đorđević, 2011).

## SOYBEAN AND *E* GENES

Time of flowering and maturity in soybean are genetically controlled by *E* genes that have different roles in maturity and photoperiod sensitivity. Their allelic combinations contribute to fine adaptation of genotype to certain latitude and climate conditions.

Up to date, 11 major loci (*E1–E10* and *J*) affecting time of flowering and maturity have been identified in soybean (Samanfar et al., 2017). The dominant allele of *E* genes usually confers later flowering and later maturity, except for *E6*, *E9*, and *J* (Ray et al., 1995; Bonato and Vello, 1999; Kong et al., 2014), while photoperiod sensitivity decreases with numbers of recessive alleles. Owen (1927) described the effect of *E1* gene in soybean in his inheritance studies, while Bernard (1971) described two major genes (*E1* and *E2*) which control flowering

time and maturity. Buzzell (1971) and Kilen and Hartwig (1971) investigated soybean flowering response to fluorescent daylength conditions and concluded that *E3* gene determines late flowering, while *e3* causes insensitivity to fluorescent light. Buzzell and Voldeng (1980) studied inheritance of insensitivity to extended daylength and found that *E4* allele confers late flowering and sensitivity to extended daylength, while *e4* allele influences early flowering and insensitivity to extended daylength. McBlain and Bernard (1987) described *E5* gene and its genetic effect similar to *E2*. However, in the study conducted in order to map *E5* locus, the results obtained from different F2 mapping populations expected to segregate for *E5* were not consistent, and no candidate QTL was found, indicating that a unique *E5* gene may not exist (Dissanayaka et al., 2016). Bonato and Vello (1999) described the dominant allele of *E6* gene which confers early flowering and maturity in soybeans. Cober and Voldeng (2001) described *E7* - soybean maturity and photoperiod-sensitivity locus - and its link to *E1* and *T*. Cober et al. (2010) mapped *E8* gene, with the dominant allele resulting in later maturity and recessive allele conferring early maturity. The *E9* locus was subjected to detailed molecular analysis and gene *FT2a* was identified, with the recessive allele delaying flowering (Kong et al., 2014; Zhao et al., 2016). Samanfar et al. (2017) identified *E10*, new maturity locus in soybean, with *FT4* being the predicted functional gene underlying this locus.

Maturity genes *E1* (Xia et al., 2012), *E2* (Watanabe et al., 2011), *E3* (Watanabe et al., 2009), *E4* (Liu et al., 2008), *E9* (Zhao et al., 2016), and *J* (Yue et al., 2017) have recently been characterized at the molecular level. Tsubokura et al. (2014) estimated that maturity genes *E1–E4* contribute to 62–66% of the flowering time variation. Association analysis and the study of genetic architecture and networks underlying agronomical traits revealed that *E1* and *E2* loci have pleiotropic effects across the traits related to yield and seed quality, as the key nodes in the regulation of different traits (Fang et al., 2017). *E1* gene was located on chromosome 6 and identified as legume-specific transcription factor which functions as the flowering repressor with putative nuclear localization signal and B3-related domain (Xia et al., 2012). *E1* allele is functional, *e1-as* allele with missense mutation is not fully functional, while *e1-fs* with a frameshift mutation and *e1-nl* with deletion of entire *E1* gene are both non-functional (Xia et al., 2012). Mutant *e1-nl* allele was identified in MG000 European (Swedish) cultivar Fiskeby III, along with the major mutant alleles for other maturity genes (*E2* and *E3*) (Valliyodan et al., 2016). Out of the known *E* loci in soybean, *E1* gene is considered to have the largest effect on determination of flowering time under field conditions (Xia et al., 2012). For improvement of varietal adaptability, the function of *E1* gene should be reduced due to its strong impact on delaying maturity. Besides *E1*, allelic variability of other *E* genes is also important in adaptation and can provide significant genetic plasticity that could enable cultivation of soybeans in wider geographic areas (Cober et al., 1996). Located on chromosome 10, *E2* gene is an ortholog of *Arabidopsis* flowering gene *GIGANTEA* involved in the circadian rhythm and flowering time pathway (Watanabe et al., 2011). The dominant *E2* allele is functional, while *e2* with nonsense mutation presents the non-functional



allele with an early-flowering phenotype (Watanabe et al., 2011). *E3* (GmPhyA3) and *E4* (GmPhyA2) are located on chromosomes 19 and 20, respectively, and they represent phytochrome A genes. *E3-Ha* and *E3-Mi* are functional alleles, while *e3-tr*, *e3-ns*, and *e3-fs* are non-functional with mutations causing truncated proteins (Watanabe et al., 2009). *E3* delays flowering under long-day conditions which ultimately affects maturity. *E4* gene has a functional allele *E4* and five non-functional alleles (*e4-SORE-1*, *e4-oto*, *e4-tsu*, *e4-kam*, and *e4-kes*) (Liu et al., 2008; Tsubokura et al., 2013). Being photoperiodically insensitive, varieties with non-functional alleles are mostly adapted to high latitudes, which indicates the importance of *e4* allele for high latitude adaptation. *E1* and *E2* genes have significant impact on vegetative development, while loci *E3* and *E4* might affect post-flowering reproductive development by increasing pod filling duration, and the number of nodes and pods by up-regulating the expression of growth habit gene *Dt1* (Xu et al., 2013).

Higher temperatures at reproductive stages were found to affect plant vigor and overall yield (Egli et al., 2005; Ren et al., 2009). It has also been reported that N and P concentrations in mature seeds increase with the increment of day/night temperatures (Thomas, 2001). This observation implicates that there might be some loci which show optimum activity of N uptake and transport at elevated temperatures (Patil et al., 2017). Protein content was also positively correlated with higher temperature (Wolf et al., 1982). Higher protein content was observed in late maturity group lines (MGV-MGX) compared to early maturity group lines (MG000-MGII). Since genetic variation in seed composition is more affected by geographic regions (~5%) than by MG (~2%) (Bandillo et al., 2015), it can be concluded that the effect of lower temperatures on seed composition at higher latitudes might be overcome by the selection of more diverse parental lines in early MGs (Piper and Boote, 1999). Hence, achieving higher protein content without affecting yield or oil content could be obtained by exploring genetic variants of *E* genes.

## SOYBEAN AND *E* GENES ALLELIC VARIATION

Study of allelic variation and diversity of soybean flowering-time and maturity genes may enhance soybean breeding for particular environments, since achieving appropriate maturity in a target environment maximizes crop yield potential (Langewisch et al., 2017). However, there are not many studies focused on soybean maturity and genes affecting it. This especially stands for European soybean varieties, as most of the studies published so far deal with North-American and Chinese genotypes (Tsubokura et al., 2013, 2014; Jiang et al., 2014; Langewisch et al., 2014, 2017; Zhai et al., 2014; Abugalieva et al., 2016; Valliyodan et al., 2016; Kurasch et al., 2017; Li et al., 2017; Wolfigang and An, 2017).

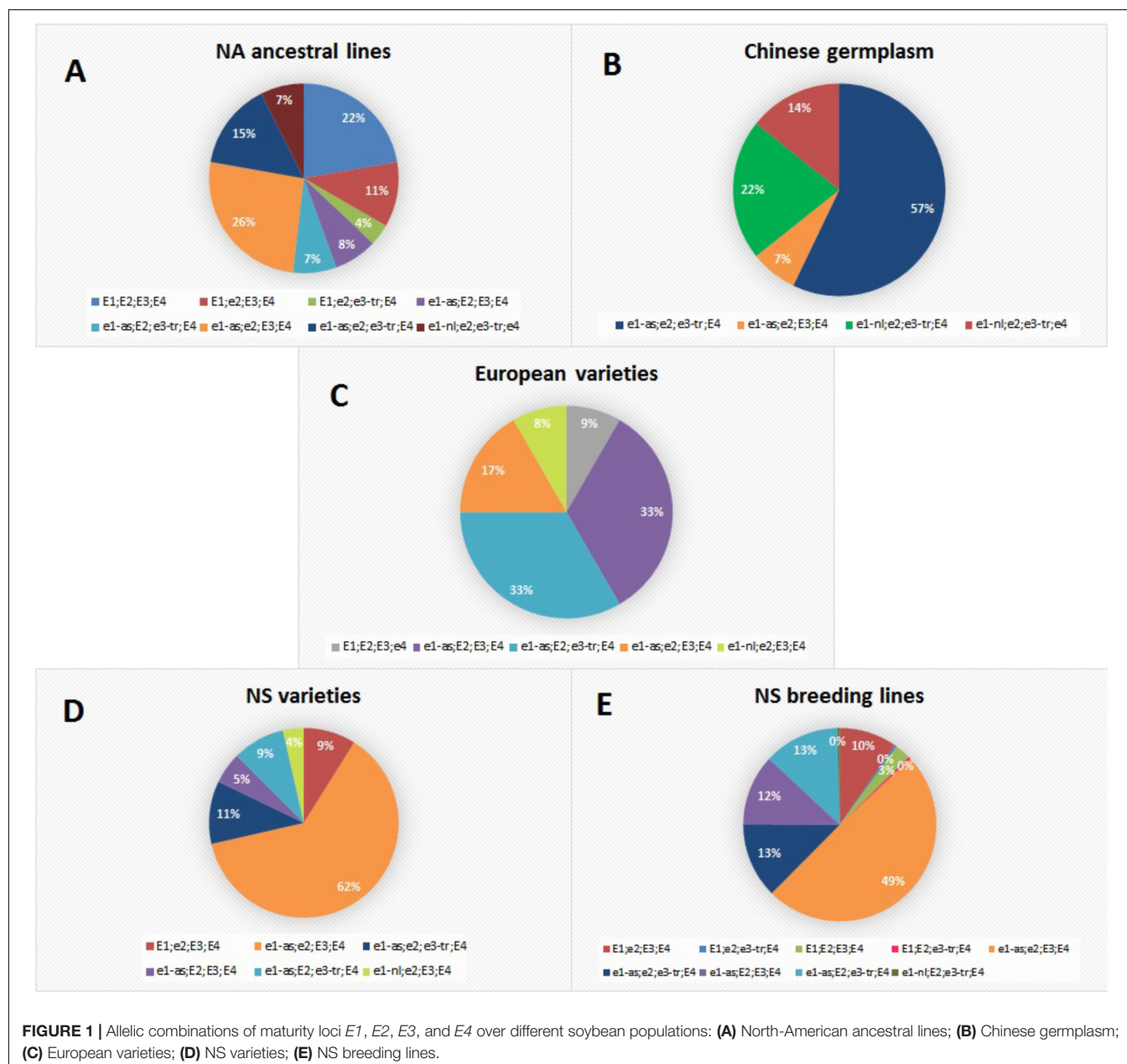
Soybean collection maintained at Institute of Field and Vegetable Crops Novi Sad, Serbia (IFVCNS), consists of more than 1200 cultivars and lines originating from America (50.1%), Asia (15.6%), and Europe (34.4%). Most of the genotypes belong to MG0 and MGI, but the collection also includes genotypes

from MG000 to MGVI (Hrustić and Miladinović, 2011). In order to characterize allelic variation and distribution of the major maturity genes (*E1–E4*) we conducted the study on 445 genotypes from soybean collections of North-American (NA) ancestral lines (27), Chinese germplasm (14), European varieties (12), genotypes developed and released at IFVCNS - NS varieties (56), NS breeding lines (229), 57 high-protein and 50 high-oil genotypes from NS collection. Our research was mainly focused on identification of the most prominent *E* allelic combinations in NS varieties and breeding lines, with the aim of applying the obtained results for the improvement and targeted breeding of soybean varieties adapted to the environments of Central-Eastern Europe. MGs of the analyzed soybean genotypes ranged from MG000 to MGIII, including several genotypes from NA ancestral population which belong to later MGs MGV-MGVII. All 445 soybean genotypes were characterized using genotyping-by-sequencing approach (Elshire et al., 2011). Sequenced libraries produced a total of  $145 \times 10^7$  raw 100-bp DNA reads that were aligned against soybean reference genome Williams 82. Reads aligned in the same genomic region were used for SNP calling, producing initially more than 85,000 SNP markers, where mean sequencing depth per SNP locus was 5. After filtering and application of a combination of criteria, such as minor allele frequency (MAF < 0.05) and the percentage of missing values (PMV < 20%), the obtained marker dataset was used for prediction of the allelic variation of four maturity genes (*E1/e1-as*, *E2/e2*, *E3/e3-tr*, and *E4/e4*). Genes *E1* and *E2* possess SNP that gives rise to alleles *e1-as* (G/C; Glyma.W82.a2: 20,207,322 bp) and *e2* (A/T; Glyma.W82.a2: 45,310,781 bp), while *E3* and *E4* genes do not have a functional SNP, causing corresponding occurrence of variant alleles *e3-tr* and *e4* correspondingly. The determinant SNP markers for genes *E1* and *E2* were not present in our dataset, as marker analysis was performed with reduced representation genotyping approach. Thus we examined groups of markers positioned in specific genomic regions which include entire *E* genes (positions given in **Supplementary Table S1**), by characterizing haplotypes aiming to identify the causative haplotype for different alleles, functional, and variant (**Supplementary Table S1**). SNPviz tool was used for haplotype visualization and comparison, in order to provide information on genetic neighborhood in which variable alleles appeared, as described by Langewisch et al. (2014). By comparing the obtained haplotypes with previously published profiles of *E* genes for NA ancestral lines and European varieties (Kurasch et al., 2017; Langewisch et al., 2017), we managed to identify the groups of haplotypes for each allele. Within these regions, lines with known different alleles were distinctively separated to specific haplotypes, which allows assignment of alleles of each gene to specific haplotypes. Moreover, it was made possible to further specify the identified haplotypes and select groups of diagnostic SNP markers able to determine different alleles of *E* genes (**Supplementary Table S1** and **Supplementary Figures S1–S4**). Four SNP markers were identified for *E1* gene which could be distinguished among alleles *E1*, *e1-as*, and *e1-nl* **Supplementary Figure S1**. Two SNP markers were defined for genes *E2* and *E3*, causing the difference between *E2* and *e2* alleles, as well as between *E3* and *e3-tr* **Supplementary**

**Figures S2, S3.** According to the five identified SNP markers, lines could be classified as *E4* or *e4* allele (**Supplementary Figure S4**). The obtained results were applied for further allele scoring of *E* genes of previously non-genotyped lines, such as NS-varieties and NS-breeding lines. Predictions did not include the non-functional *e1-fs* allele, and no distinction could be made between *E3-Ha* and *E3-Mi* alleles, or *e4-oto*, *e4-tsu*, *e4-kam*, and *e4-kes*.

All examined genotypes were classified into 12 *E* haplogroups which represent a unique combination of *E1*, *E2*, *E3*, and *E4* alleles (**Supplementary Table S2**). Each collection had one major *E* haplogroup which included most of the lines ( $\geq 50\%$ ) (**Figure 1**). The exception were European varieties with

two predominant groups and NA ancestral lines which were distributed almost evenly among several haplogroups. The most abundant NA ancestral haplogroup was *e1-as/e2/E3/E4* with 26% of genotypes. Over 20% of genotypes had *E1/E2/E3/E4*, which coincides with the high number of late-maturing genotypes included in the NA collection. This is concordant with molecular maturity model proposed by Langewisch et al. (2017) where early-maturing lines had non-functional *e1/e2/e3-tr* and late-maturing lines had *E1/E2/E3*. None of the lines from the Chinese collection carried functional photoperiod-sensitive *E1* allele, while two photoperiod-insensitive alleles (*e1-nl* and *e1-as*) were identified at *E1* locus. Contrary to previous reports, where most of the tested accessions of breeding collections

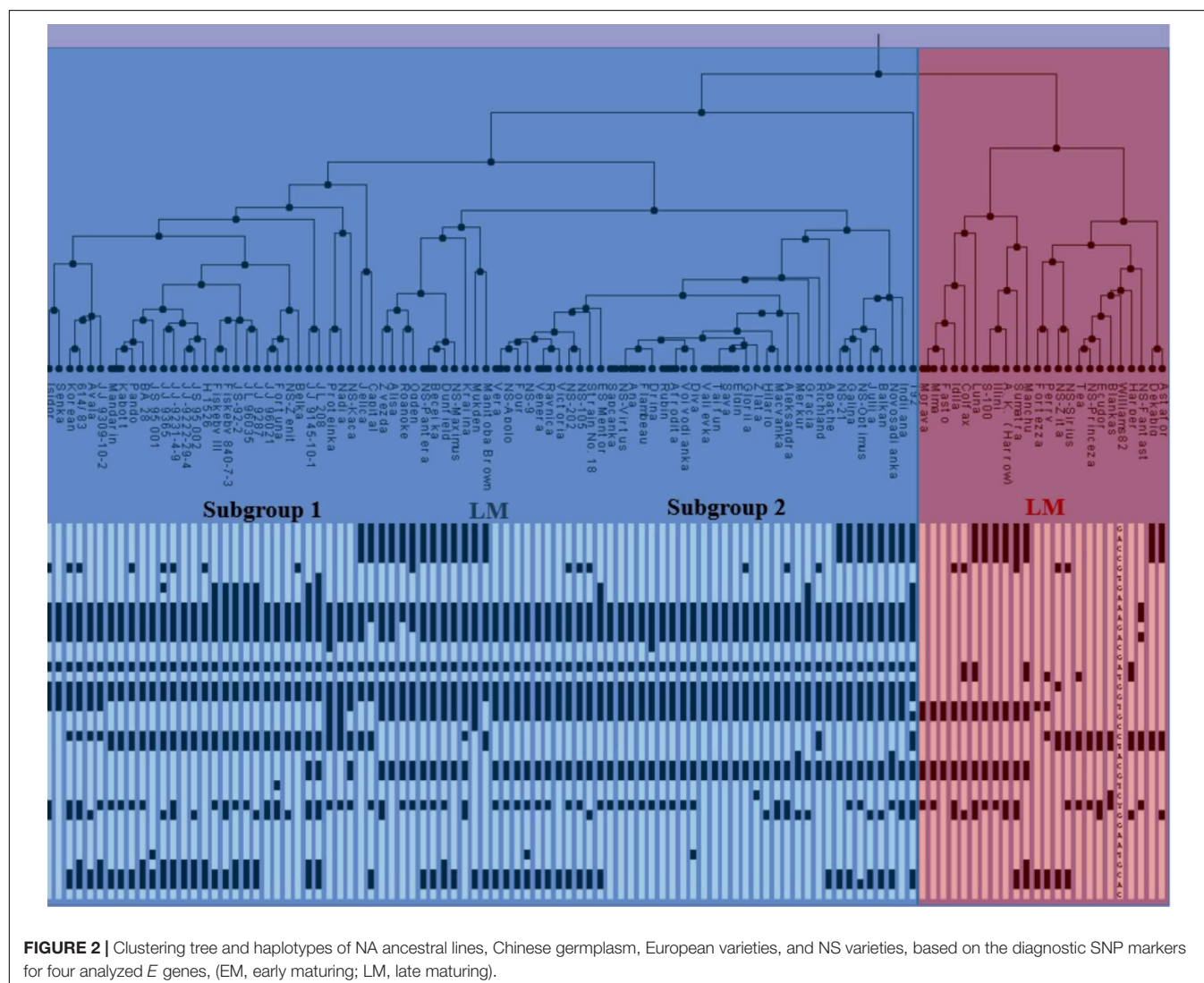


**FIGURE 1 |** Allelic combinations of maturity loci *E1*, *E2*, *E3*, and *E4* over different soybean populations: (A) North-American ancestral lines; (B) Chinese germplasm; (C) European varieties; (D) NS varieties; (E) NS breeding lines.

from China were *E1/e2/E3/E4* followed by *e1-as/e2/e3/E4* (Jia et al., 2014; Jiang et al., 2014), the most abundant haplogroup in Chinese germplasm in our collection was *e1-as/e2/e3-tr/E4* including 57% of lines. The remaining lines mostly contained *e1-nl* allele with *e1-nl/e2/e3-tr/E4* and *e1-nl/e2/e3-tr/e4*, covering 21 and 14% of genotypes, respectively. Of all examined genotypes, only three lines from the NA ancestral and Chinese germplasm had the non-functional *e4* allele, which is in agreement with the fact that *e4* alleles were found in small geographical locations with high latitudes and low temperature (Tsubokura et al., 2013; Langewisch et al., 2014). European varieties from our collection differed from the varieties studied by Kurasch et al. (2017) in their *E* haplotype, as more than 30% of varieties from our collection were *e1-as/E2/E3/E4* and the same number of genotypes had *e1-as/E2/e3-tr/E4*, compared to 1.3 and 12%, respectively, in the study of Kurasch et al. (2017). The haplogroup *e1-as/e2/E3/E4* comprised 17% of lines in European varieties. Similar to the results of Kurasch et al. (2017), NS varieties primarily had *e1-as/e2/E3/E4* haplotype identified

in 63% of genotypes. The same haplotype was found to be predominant in the Kazakh breeding collection (Abugalieva et al., 2016), which was attributed to the frequent exchange of germplasm with partners from former USSR countries. The second largest haplotype in NS varieties is *e1-as/e2/e3-tr/E4* with 11% of genotypes. All other allelic combinations were found in less than 10% of genotypes. For the *E1* locus, allele *e1-as* was detected in 88% of NS varieties classified into MG000-MGIII. Nearly half of the NS breeding lines were predicted to be *e1-as/e2/E3/E4*, with the additional 13% of genotypes predicted as *e1-as/e2/e3-tr/E4* and 13% of *e1-as/E2/e3-tr/E4*.

Novi Sad collection exhibited a high diversity of allelic combinations of *E* genes. The most abundant haplogroup in NS varieties and breeding lines was *e1-as/e2/E3/E4*. The same haplogroup comprised 26% of genotypes in NA ancestral lines which represent more than 85% of the pedigree of NS varieties, indicating that *e1-as/e2/E3/E4* haplotype was favored by artificial selection during soybean breeding under the environmental



**FIGURE 2 |** Clustering tree and haplotypes of NA ancestral lines, Chinese germplasm, European varieties, and NS varieties, based on the diagnostic SNP markers for four analyzed *E* genes, (EM, early maturing; LM, late maturing).



conditions of Central-Eastern Europe. NS varieties and breeding lines had predominantly *e1-as* allele, indicating that they might have been purposely selected for similar maturity and high-yield capacity. This finding is compatible with identified genomic regions affected by positive selection during breeding of NS soybean varieties, where two loci (*Satt357* and *Satt557*) on chromosome 6 were identified as strong positive selection candidates (Tomičić et al., 2015). Similar results were obtained during evaluation of the allelic variants at the *E* genes in 75 European cultivars from five MGs (MG000-MGII), where none of the cultivars carried functional *E1* allele (Kurasch et al., 2017). It was concluded that the *E1* locus plays a major role in early flowering and maturity, and that photoperiod insensitivity at this locus is a probable requirement for soybeans adapted to Central or Northern Europe.

Based on the used SNP markers, two major clusters can be identified and described as early material (EM) consisting of MG000-MGI and late material (LM) consisting mainly of MGII-MGIV genotypes (Figure 2). The first subgroup from EM included Chinese germplasm, NA ancestral lines and NS varieties. Recessive haplotype *e1-as/e2/e3-tr/E4* was the dominant haplotype in this subgroup. The majority of NS varieties belonged to the second subgroup, along with a few NA and European varieties. The most abundant allelic combination in this subcluster was *e1-as/e2/E3/E4*, which represents the major haplogroup within environmental conditions of Central-Eastern Europe. The second group mostly comprised LM genotypes, including NA ancestral lines, European and NS varieties, from MGI to MGIV, with the exception of NS Princeza which belongs to MG0. Most lines from the LM cluster had three or four dominant alleles. The observed grouping of genotypes into late and early maturing clusters indicates that alleles of *E1–E4* genes can be used in the assessment and classification of the assortment according to maturity. However, four genes used in this study are not sufficient to make the precise distinction between different MGs, as there is no unique allelic combination specific for each group. Further work is needed to identify molecular mechanisms behind soybean photoperiod sensitivity, and linking specific allelic combination with a certain MG.

## FUTURE PROSPECTS

Flowering time and maturity are agronomically important traits, which affect soybean adaptation, quality traits, and yield. Complete understanding of their regulation will therefore allow breeding of varieties with optimal flowering or maturity for particular geographic regions. Considering this, we have studied the distribution of different haplogroups of *E* genes in soybean collections of NA ancestral lines, Chinese germplasm, European varieties, NS varieties, and NS breeding lines. Overall, the observed allelic combinations of *E1–E4* genes significantly determined the adaptation of varieties to different geographic regions, although they have different impacts on maturity.

The study was also the first attempt to assess NS soybean varieties and breeding line adaptation based on *E* gene variation, as well as comparatively assess *E* gene variation in NA, Chinese

and European genotypes. The study aims to identify the most prominent *E* allelic combinations in NS genotypes and apply the obtained results in target breeding and introgression of yield and quality traits for environments of Central-Eastern Europe. As *e1-as/e2/E3/E4* was the most common haplotype in the NS population comprising high-yielding varieties grown throughout Central and Eastern Europe, the specific allele combination was proposed as the optimal combination for this environment.

The obtained data provide useful information about the selection of parental genotypes most appropriate for use in local breeding programs for improved soybean productivity. This allows plant breeders to transfer traits more effectively into different MGs, enhance germplasm exploitation, and increase overall efficiency of soybean breeding. Furthermore, the obtained knowledge will significantly contribute to the marker-assisted selection and molecular breeding of soybean, and increase breeding efficiency by enhancing genomic prediction models.

## AUTHOR CONTRIBUTIONS

JM, MĆ, SB-T, and VuĐ planned and designed the experiment. MĆ, KP, VuĐ, and VoĐ performed the research. JM, DM, MĆ, and VuĐ contributed to the interpretation and analysis of results. JM, MĆ, and DM wrote and approved the manuscript.

## FUNDING

This research was part of the project TR31022, which is supported by Ministry of Education, Science and Technological Development of the Republic of Serbia.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01286/full#supplementary-material>

**FIGURE S1** | Diagnostic haplotypes of NA ancestral lines, Chinese germplasm, European varieties and NS varieties, based on selected SNP markers for *E1* gene.

**FIGURE S2** | Diagnostic haplotypes of NA ancestral lines, Chinese germplasm, European varieties and NS varieties, based on selected SNP markers for *E2* gene.

**FIGURE S3** | Diagnostic haplotypes of NA ancestral lines, Chinese germplasm, European varieties and NS varieties, based on selected SNP markers for *E3* gene.

**FIGURE S4** | Diagnostic haplotypes of NA ancestral lines, Chinese germplasm, European varieties and NS varieties, based on selected SNP markers for *E4* gene.

**TABLE S1** | SNP calls located in the genomic regions of *E1*, *E2*, *E3*, and *E4* genes in soybean collections of North-American (NA) ancestral lines, Chinese germplasm, European varieties, and genotypes developed and released at Institute of Field and Vegetable Crops – NS (Novi Sad) varieties, NS breeding lines, high-protein and high-oil genotypes from NS collection.

**TABLE S2** | Allelic variation on the *E1*, *E2*, *E3*, and *E4* maturity genes in soybean collections of North-American (NA) ancestral lines, Chinese germplasm, European varieties, and genotypes developed and released at Institute of Field and Vegetable Crops – NS (Novi Sad) varieties, NS breeding lines, high-protein and high-oil genotypes from NS collection.

## REFERENCES

- Abugalieva, S., Didorenko, S., Anuarbek, S., Volkova, L., Gerasimova, Y., Sidorik, I., et al. (2016). Assessment of soybean flowering and seed maturation time in different latitude regions of Kazakhstan. *PLoS One* 11:e0166894. doi: 10.1371/journal.pone.0166894
- Bandillo, N., Jarquin, D., Song, Q., Nelson, R., Cregan, P., Specht, J., et al. (2015). A population structure and genome-wide association analysis on the USDA soybean germplasm collection. *Plant Genome* 8, 1–13. doi: 10.3835/plantgenome2015.04.0024
- Bernard, R. L. (1971). Two major genes for time of flowering and maturity in soybeans. *Crop Sci.* 11, 242–244. doi: 10.2135/cropsci1971.0011183X001100020022x
- Board, J. E., and Hall, W. (1984). Premature flowering in soybean yield reductions at nonoptimal planting dates as influenced by temperature and photoperiod. *Agron. J.* 76, 700–704. doi: 10.2134/agronj1984.00021962007600040043x
- Bonato, E. R., and Vello, N. A. (1999). E6, a dominant gene conditioning early flowering and maturity in soybeans. *Genet. Mol. Biol.* 22, 229–232. doi: 10.1590/S1415-47571999000200016
- Buzzell, R. I. (1971). Inheritance of a soybean flowering response to fluorescent daylength conditions. *Can. J. Genet. Cytol.* 13, 703–707. doi: 10.1139/g71-100
- Buzzell, R. I., and Voldeng, H. D. (1980). Inheritance of insensitivity to long daylength. *Soyb. Genet. Newsl.* 7, 26–29. doi: 10.1093/jhered/esp113
- Cober, E. R., Molnar, S. J., Charette, M., and Voldeng, H. D. (2010). A new locus for early maturity in soybean. *Crop Sci.* 50, 524–527. doi: 10.2135/cropsci2009.04.0174
- Cober, E. R., Tanner, J. W., and Voldeng, H. D. (1996). Soybean photoperiod-sensitivity loci respond differentially to light quality. *Crop Sci.* 36, 606–610. doi: 10.2135/cropsci1996.0011183X003600030014x
- Cober, E. R., and Voldeng, H. D. (2001). A new soybean maturity and photoperiod-sensitivity locus linked to E1 and T. *Crop Sci.* 41, 698–701. doi: 10.2135/cropsci2001.413698x
- Colson, J., Bouniols, A., and Jones, J. W. (1995). Soybean reproductive development: adapting a model for european varieties. *Agron. J.* 87, 1129–1139. doi: 10.2134/agronj1995.00021962008700060015x
- Davidson, H. R., and Campbell, C. A. (1983). The effect of temperature, moisture and nitrogen on the rate of development of spring wheat as measured by degree days. *Can. J. Plant Sci.* 63, 833–846. doi: 10.4141/cjps83-106
- Dissanayaka, A., Rodriguez, T. O., Di, S., Yan, F., Githiri, S. M., Rodas, F. R., et al. (2016). Quantitative trait locus mapping of soybean maturity gene E5. *Breed. Sci.* 66, 407–415. doi: 10.1270/jsbbs.15160
- Egli, D., TeKrony, D., Heitholt, J., and Rupe, J. (2005). Air temperature during seed filling and soybean seed germination and vigor. *Crop Sci.* 45, 1329–1335. doi: 10.2135/cropsci2004.0029
- Egli, D. B., and Wardlaw, I. F. (1980). Temperature response of seed growth characteristics of soybeans. *Agron. J.* 72, 560–564. doi: 10.2134/agronj1980.00021962007200030036x
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379
- Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., et al. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* 18:161. doi: 10.1186/s13059-017-1289-9
- Garner, W. W., and Allard, H. A. (1930). Photoperiodic response of soybeans in relation to temperature and other environmental factors. *J. Agric. Res.* 41, 719–735.
- Hartwig, E. E. (1973). “Varietal development,” in *Soybeans: Improvement, Production and Uses*, ed. B. E. Caldwell (Madison WI: ASA), 187–207.
- Hodges, T., and French, V. (1985). Soyphen: soybean growth stages modeled from temperature, daylength, and water availability. *Agron. J.* 77, 500–505. doi: 10.2134/agronj1985.00021962007700030031x
- Hrustić, M., and Miladinović, J. (2011). “Importance, origin and expansion of soybean,” in *Soybean*, eds J. Miladinović, M. Hrustić, and M. Vidić (Bečej: Sojaprotein), 11–40.
- Jia, H., Jiang, B., Wu, C., Lu, W., Hou, W., Sun, S., et al. (2014). Maturity group classification and maturity locus genotyping of early-maturing soybean varieties from high-latitude cold regions. *PLoS One* 9:e94139. doi: 10.1371/journal.pone.0094139
- Jiang, B., Nan, H., Gao, Y., Tang, L., Yue, Y., Lu, S., et al. (2014). Allelic combinations of soybean maturity loci E1, E2, E3 and E4 result in diversity of maturity and adaptation to different latitudes. *PLoS One* 9:e106042. doi: 10.1371/journal.pone.0106042
- Jones, P. G., and Laing, D. R. (1978). Simulation of the phenology of soybeans. *Agric. Syst.* 3, 295–311. doi: 10.1371/journal.pone.0124682
- Kilen, T. C., and Hartwig, E. E. (1971). Inheritance of a light-quality sensitive character in soybeans. *Crop Sci.* 11, 559–561. doi: 10.2135/cropsci1971.0011183X001100040030x
- Kong, F., Nan, H., Cao, D., Li, Y., Wu, F., Wang, J., et al. (2014). A new dominant gene E9 conditions early flowering and maturity in soybean. *Crop Sci.* 54, 2529–2535. doi: 10.2135/cropsci2014.03.0228
- Kurasch, A. K., Hahn, V., Leiser, W. L., Vollmann, J., Schori, A., Bétrix, C.-A., et al. (2017). Identification of mega-environments in Europe and effect of allelic variation at maturity E loci on adaptation of European soybean. *Plant Cell Environ.* 40, 765–778. doi: 10.1111/pce.12896
- Langewisch, T., Lenis, J., Jiang, G.-L., Wang, D., Pantalone, V., and Bilyeu, K. (2017). The development and use of a molecular model for soybean maturity groups. *BMC Plant Biol.* 17:91. doi: 10.1186/s12870-017-1040-4
- Langewisch, T., Zhang, H., Vincent, R., Joshi, T., Xu, D., and Bilyeu, K. (2014). Major soybean maturity gene haplotypes revealed by SNPviz analysis of 72 sequenced soybean genomes. *PLoS One* 9:e94150. doi: 10.1371/journal.pone.0094150
- Li, J., Wang, X., Song, W., Huang, X., Zhou, J., Zeng, H., et al. (2017). Genetic variation of maturity groups and four E genes in the Chinese soybean mini core collection. *PLoS One* 12:e0172106. doi: 10.1371/journal.pone.0172106
- Liu, B., Kanazawa, A., Matsumura, H., Takahashi, R., Harada, K., and Jun, A. (2008). Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. *Genetics* 180, 995–1007. doi: 10.1534/genetics.108.092742
- McBlain, B. A., and Bernard, R. L. (1987). A new gene affecting the time of maturity in soybeans. *J. Hered.* 78, 160–162. doi: 10.1371/journal.pone.0172106
- Meckel, L., Egli, D. B., Phillips, R. E., Radcliffe, D., and Leggett, J. E. (1984). Effect of moisture stress on seed growth in soybean. *Agron. J.* 76, 647–650. doi: 10.1371/journal.pone.0172106
- Miladinović, J., and Đorđević, V. (2011). “Soybean morphology and stages of development,” in *Soybean*, eds J. Miladinović, M. Hrustić, and M. Vidić (Bečej: Sojaprotein), 45–68. doi: 10.2134/agronj1984.00021962007600040033x
- Miladinović, J., Kurosaki, H., Burton, J. W., Hrustić, M., and Miladinović, D. (2006). Adaptability of shortseasons soybeans to varying longitudinal regions. *Eur. J. Agron.* 25, 243–249. doi: 10.1016/j.eja.2006.05.007
- Owen, F. V. (1927). Inheritance studies in soybeans. II. Glabrousness, color of pubescence, time of maturity, and linkage relations. *Genetics* 12, 519–529.
- Patil, G., Mian, R., Vuong, T., Pantalone, V., Song, Q., Chen, P., et al. (2017). Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. *Theor. Appl. Genet.* 130, 1975–1991. doi: 10.1007/s00122-017-2955-8
- Piper, E. L., and Boote, K. I. (1999). Temperature and cultivar effects on soybean seed oil and protein concentrations. *J. Am. Oil Chem. Soc.* 76, 1233–1241. doi: 10.1007/s11746-999-0099-y
- Putterill, J., Stockum, C., and Warman, G. (2009). “Photoperiodic flowering in long-day plant *Arabidopsis thaliana*,” in *Photoperiodism: The Biological Calendar*, eds R. J. Nelson, D. L. Denlinger, and D. E. Somers (Oxford: Oxford University Press), 9–37.
- Ray, J. D., Hinson, K., Mankono, E. B., and Malo, F. M. (1995). Genetic control of a long-juvenile trait in soybean. *Crop Sci.* 35, 1001–1006. doi: 10.1038/ng.3819
- Ren, C., Bilyeu, K. D., and Beuselinck, P. (2009). Composition, vigor, and proteome of mature soybean seeds developed under high temperature. *Crop Sci.* 49, 1010–1022. doi: 10.2135/cropsci2008.05.0247
- Samanfar, B., Molnar, S. J., Charette, M., Schoenrock, A., Dehne, F., Golshani, A., et al. (2017). Mapping and identification of a potential candidate gene for a novel maturity locus, E10, in soybean. *Theor. Appl. Genet.* 130, 377–390. doi: 10.1007/s00122-016-2819-7
- Scott, W. O., and Aldrich, S. R. (1983). *Modern soybean production*. Champaign: S & A Publications Inc.

- Seddigh, M., Jolliff, G. D., and Orf, J. H. (1989). Night temperature effects on soybean phenology. *Crop Sci.* 29, 400–406. doi: 10.1371/journal.pone.0124682
- Thomas, J. M. G. (2001). *Impact of Elevated Temperature and Carbon Dioxide on Development and Composition of Soybean Seed*. Doctoral dissertation, University of Florida, Gainesville, FL.
- Tomičić, M., Đorđević, V., Obreht, D., Miladinović, J., Brbaklić, L., Mikić, A., et al. (2015). Tracking footprints of selection associated with soybean adaptation to Central-East Europe environments. *Euphytica* 203, 701–713. doi: 10.1007/s10681-014-1313-z
- Tsubokura, Y., Matsumura, H., Xu, M., Liu, B., Nakajima, H., Anai, T., et al. (2013). Genetic variation in soybean at maturity locus E4 is involved in adaptation to long days at high latitudes. *Agronomy* 3, 117–134. doi: 10.3390/agronomy3010117
- Tsubokura, Y., Watanabe, S., Xia, Z., Kanamori, H., Yamagata, H., Kaga, A., et al. (2014). Natural variation in the genes responsible for maturity loci *E1*, *E2*, *E3* and *E4* in soybean. *Ann. Bot.* 113, 429–441. doi: 10.1093/aob/mct269
- Valliyodan, B., Qiu, D., Patil, G., Zeng, P., Huang, J., Dai, L., et al. (2016). Landscape of genomic diversity and trait discovery in soybean. *Sci. Rep.* 6:23598. doi: 10.1038/srep23598
- Wang, J., McBlain, B. A., Wooley, J. T., Hesketh, J. D., and Bernard, R. L. (1987). A soybean phenology model and associated data base. *Biotronics* 16, 25–38.
- Watanabe, S., Hideshima, R., Xia, Z., Tsubokura, Y., Sato, S., Nakamoto, Y., et al. (2009). Map based cloning of the gene associated with the soybean maturity locus *E3*. *Genetics* 182, 1251–1262. doi: 10.1534/genetics.108.098772
- Watanabe, S., Xia, Z., Hideshima, R., Tsubokura, Y., Sato, S., Yamanaka, N., et al. (2011). A map based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. *Genetics* 188, 395–407. doi: 10.1534/genetics.110.125062
- Wolf, R., Cavins, J., Kleiman, R., and Black, L. (1982). Effect of temperature on soybean seed constituents: oil, protein, moisture, fatty acids, amino acids and sugars. *J. Am. Oil Chem. Soc.* 59, 230–232. doi: 10.1007/BF02582182
- Wolfgang, G., and An, Y.-Q. C. (2017). Genetic separation of southern and northern soybean breeding programs in North America and their associated allelic variation at four maturity loci. *Mol. Breed.* 37:8. doi: 10.1007/s11032-016-0611-7
- Xia, Z., Watanabe, S., Yamada, T., Tsubokura, Y., Nakashima, H., Zhai, H., et al. (2012). Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2155–E2164. doi: 10.1073/pnas.1117982109
- Xu, M., Xu, Z., Liu, B., Kong, F., Tsubokura, Y., Watanabe, S., et al. (2013). Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. *BMC Plant Biol.* 13:91. doi: 10.1186/1471-2229-13-91
- Yue, Y., Liu, N., Jiang, B., Li, M., Wang, H., Jiang, Z., et al. (2017). A single nucleotide deletion in *J* encoding GmELF3 confers long juvenility and is associated with adaption of tropic soybean. *Mol. Plant.* 10, 656–658. doi: 10.1016/j.molp.2016.12.004
- Zhai, H., Lü, S., Wang, Y., Chen, X., Ren, H., Yang, J., et al. (2014). Allelic variations at four major maturity *E* genes and transcriptional abundance of the *E1* gene are associated with flowering time and maturity of soybean cultivars. *PLoS One* 9:e97636. doi: 10.1371/journal.pone.0097636
- Zhao, C., Takeshima, R., Zhu, J., Xu, M., Sato, M., Watanabe, S., et al. (2016). A recessive allele for delayed flowering at the soybean maturity locus *E9* is a leaky allele of FT2a, a FLOWERING LOCUS T ortholog. *BMC Plant Biol.* 16:20. doi: 10.1186/s12870-016-0704-r9

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Miladinović, Čeran, Đorđević, Balešević-Tubić, Petrović, Đukić and Miladinović. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# OliveCan: A Process-Based Model of Development, Growth and Yield of Olive Orchards

Álvaro López-Bernal<sup>1</sup>, Alejandro Morales<sup>2</sup>, Omar García-Tejera<sup>3</sup>, Luca Testi<sup>3\*</sup>, Francisco Orgaz<sup>3</sup>, J. P. De Melo-Abreu<sup>4</sup> and Francisco J. Villalobos<sup>1,3</sup>

<sup>1</sup> Departamento de Agronomía, ETSIAM, Universidad de Córdoba, Córdoba, Spain, <sup>2</sup> Centre for Crop Systems Analysis, Wageningen University & Research, Wageningen, Netherlands, <sup>3</sup> Department of Agronomy, Institute for Sustainable Agriculture, Spanish National Research Council (CSIC), Córdoba, Spain, <sup>4</sup> Departamento de Ciencias e Engenharia de Biosistemas, Instituto Superior de Agronomia, Universidade de Lisboa, Lisbon, Portugal

## OPEN ACCESS

### Edited by:

Mariela Torres,  
Instituto Nacional de Tecnología  
Agropecuaria (INTA), Argentina

### Reviewed by:

Mukhtar Ahmed,  
Pir Mehr Ali Shah Arid Agriculture  
University, Pakistan  
Hartmut Stützel,  
Leibniz University of Hanover,  
Germany  
Peter S. Searles,  
Consejo Nacional de Investigaciones  
Científicas y Técnicas (CONICET),  
Argentina

### \*Correspondence:

Luca Testi  
lucatesti@ias.csic.es

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 04 September 2017

**Accepted:** 23 April 2018

**Published:** 09 May 2018

### Citation:

López-Bernal Á, Morales A,  
García-Tejera O, Testi L, Orgaz F,  
De Melo-Abreu JP and Villalobos FJ  
(2018) OliveCan: A Process-Based  
Model of Development, Growth  
and Yield of Olive Orchards.  
Front. Plant Sci. 9:632.  
doi: 10.3389/fpls.2018.00632

Several simulation models of the olive crop have been formulated so far, but none of them is capable of analyzing the impact of environmental conditions and management practices on water relations, growth and productivity under both well-irrigated and water-limiting irrigation strategies. This paper presents and tests OliveCan, a process-oriented model conceived for those purposes. In short, OliveCan is composed of three main model components simulating the principal elements of the water and carbon balances of olive orchards and the impacts of some management operations. To assess its predictive power, OliveCan was tested against independent data collected in two 3-year field experiments conducted in Córdoba, Spain, each of them applying different irrigation treatments. An acceptable level of agreement was found between measured and simulated values of seasonal evapotranspiration ( $ET$ , range 393 to 1016 mm year<sup>-1</sup>; RMSE of 89 mm year<sup>-1</sup>), daily transpiration ( $E_p$ , range 0.14–3.63 mm d<sup>-1</sup>; RMSE of 0.32 mm d<sup>-1</sup>) and oil yield ( $Y_{oil}$ , range 13–357 g m<sup>-2</sup>; RMSE of 63 g m<sup>-2</sup>). Finally, knowledge gaps identified during the formulation of the model and further testing needs are discussed, highlighting that there is additional room for improving its robustness. It is concluded that OliveCan has a strong potential as a simulation platform for a variety of research applications.

**Keywords:** carbon assimilation, crop model, *Olea europaea* L., SPAC model, water stress, water uptake

## INTRODUCTION

Olive orchards represent the main component of agricultural systems in many semiarid regions with Mediterranean climate, reaching 10.1 Mha worldwide in 2011 (FAOSTAT, 2014). In countries where the cultivation of this tree species is done in extensive areas, olive cropping systems have become of high relevance not only from an economic perspective, but also from an ecological one. Olive orchards have been traditionally cultivated at low planting densities under low-input rainfed conditions. However, the increase in the demand for oil of recognized and consistently high quality in recent years has triggered the development and adoption of farming techniques aimed to improve productivity, such as localized irrigation, fertigation and mechanical pruning and harvesting. As a result, traditional rainfed olive orchards (<200 trees ha<sup>-1</sup>) coexist nowadays with new intensive (250–850 trees ha<sup>-1</sup>) or super-intensive (1200–3000 trees ha<sup>-1</sup>) irrigated

plantations. The rapid changes in olive farming have raised questions on the economic and environmental sustainability of the different olive cropping systems under present and future climate scenarios. Given that an olive orchard is a complex system, its quantitative study via modeling is a crucial step in understanding its behavior in response to climatic and management factors.

To our knowledge, Abdel-Razik (1989) was the first researcher to describe a model for estimating the productivity of olive orchards. The model describes the growth of different organs by simulating radiation interception, photosynthesis, respiration, and applying simple allocation rules, but it does not consider the effects of planting density, canopy structure or pruning, and many of its equations lack a consistent theoretical basis. Villalobos et al. (2006) proposed a simpler approach to estimate biomass production and yield in olive canopies, based on the concept of annual radiation use efficiency and partitioning coefficients, yet this approach does not give insight about the dynamics of the system, its response to climatic variables (besides solar radiation) or the effect of management. More recently, Morales et al. (2016) presented a mechanistic model of olive oil production in the absence of any biotic or abiotic stress, based on a three-dimensional model of canopy photosynthesis and respiration and dynamic distribution of assimilates among organs. However, water stress is the main limiting factor for biomass production in rainfed and deficit-irrigated olive orchards (Moriani et al., 2003; Iniesta et al., 2009).

Simulating the water balance of an irrigated olive orchard is a particularly challenging task as the trees are typically watered by point-source emitters that keep a small fraction of the surface frequently wet while the remaining area remains dry, unless it rains. This fact results in differences between these two soil areas in relation to soil water content, the water fluxes determining the water balance (i.e., runoff, drainage, redistribution along the soil profile, soil evaporation, and root water uptake) and root length density (Fernández et al., 1991). Therefore, traditional modeling approaches based on the use of the average soil water content can lead to large errors, besides giving a poor insight into the system. One alternative consists of using a two-compartment model that solves the water balance separately for each zone of the soil. In this regard, Testi et al. (2006) proposed a model capable of simulating potential transpiration, separately calculating runoff, drainage and soil evaporation from the wet and dry fractions of the soil surface under localized irrigation. The model was developed to determine the potential irrigation needs of olive orchards, so its use is unfortunately limited to unstressed conditions. Lately, García-Tejera et al. (2017a) have formulated a soil-plant-atmosphere-continuum (SPAC) model capable of calculating root water uptake from soils with spatially heterogeneous distributions of water content and root length densities. Such a model also discretizes the soil into different soil zones and layers and, for the canopy, it considers two leaf classes (i.e., sunlit and shaded). Furthermore, the model by García-Tejera et al. (2017a) provides estimates of gross assimilation ( $A$ ), offering an opportunity to link the water and carbon balances of olive trees.

The goal of this study is to present and test a process-oriented model integrating existing knowledge on the growth and development processes of olive orchards and capable to account for the impacts of water stress, management and climate on their productivity, in the absence of nutrient deficiencies, diseases and pests. The model, hereafter named 'OliveCan' -which comes from 'Olive Canopy-', was formulated using the models by Testi et al. (2006); Morales et al. (2016) and García-Tejera et al. (2017a) as starting point.

## MATERIALS AND METHODS

### Model Description

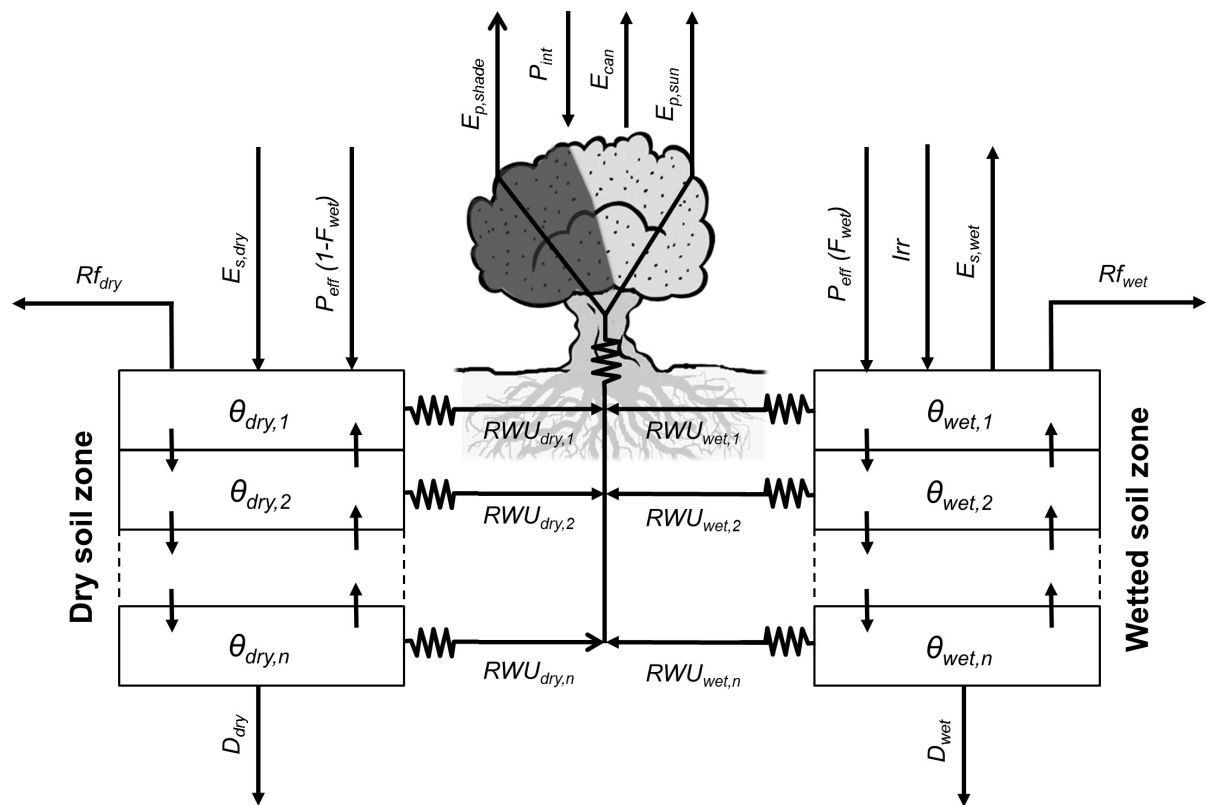
This section provides an overview of the main features and processes within OliveCan. An in-depth description of the model, along with its equations and scientific rationale is given as Supplementary Material. The code of OliveCan was written in Visual Basic 6.0.

OliveCan is subdivided into three main components (Supplementary Figure S1) that are devoted to the computation of the water and carbon balances of the olive orchard and to simulate the impacts of some management operations. The water and carbon balance components are interdependent (i.e., each one needs data provided by the other) and both of them require information on soil traits and weather data.

Although most processes in the model run at daily time steps, others (i.e., root water uptake, photosynthesis, maintenance respiration and chilling accumulation) are computed over the diurnal course and integrated to yield daily values. The number of sub-day periods per day to be considered is customizable through a user-defined parameter ( $N$ ). The meteorological input data required consist of daily values of the following variables: maximum and minimum air temperatures ( $T_{\max}$  and  $T_{\min}$ , respectively), average vapor pressure ( $e_a$ ), solar radiation ( $I_{G,D}$ ) average wind speed ( $U$ ) and precipitation ( $P$ ). For those processes simulated at sub-day intervals, OliveCan incorporates routines that disaggregate the daily values of weather data into theoretical diurnal time series.

### Water Balance Component

This modeling component simulates different physical and physiological processes relevant to olive water use (Figure 1). The model solves separately the water balance for two soil zones representing the dry and wetted (by localized irrigation) surface fractions. This approach enables the model to simulate the spatial heterogeneities in soil water content dynamics associated to the use of point-source emitters. Hence, the model considers that the water supplied by irrigation ( $Irr$ ) is only applied to the wetted soil zone, whereas the dry soil zone is only watered by rainfall. On the other hand, runoff ( $Rf$ ), soil evaporation ( $E_s$ ), root water uptake ( $RWU$ ), and drainage ( $D$ ) represent the water effluxes for both soil zones and are computed independently for each of those. Besides, each soil zone is subdivided into a user-defined number of soil layers ( $n$ ) which are also customizable in thickness. Vertical soil water fluxes between adjacent soil layers are simulated, but no lateral flow between soil zones is considered.



**FIGURE 1 |** Relational diagram of the water balance component, highlighting the processes simulated. Solid arrows represent fluxes of water. In OliveCan, the water balance is solved separately for two soil compartments (i.e., dry and wetted soil zones), as irrigation ( $Irr$ ) is applied to only one of them. For each soil zone, soil water content (i.e.,  $\theta_{dry}$  and  $\theta_{wetted}$ ) are updated every day after calculating the fluxes of effective precipitation ( $P_{eff,dry}$  and  $P_{eff,wetted}$ ), runoff ( $Rf_{dry}$  and  $Rf_{wetted}$ ), drainage ( $D_{dry}$  and  $D_{wetted}$ ), water redistribution between the different soil layers, soil evaporation ( $E_{s,dry}$  and  $E_{s,wetted}$ ) and root water uptake ( $RWU_{dry}$  and  $RWU_{wetted}$ ). Direct evaporation of rain water intercepted by the canopy ( $E_{can}$ ) is also considered.

Daily effective precipitation ( $P_{eff}$ ) is calculated by discounting rainfall interception by the canopy ( $P_{int}$ ) from total daily precipitation ( $P$ ).  $P_{int}$  is calculated using a simplified version of the model of Gómez et al. (2001) and the resulting  $P_{eff}$  is distributed proportionally between the two soil zones as a function of the surface fractions that remain rainfed or are wetted by localized irrigation. With regard to  $P_{int}$ , the canopy is treated as a capacitor capable of storing rain water up to a certain limit determined by canopy dimensions and leaf area index ( $LAI$ ), according to Gómez et al. (2001). The stored water is subsequently lost by direct evaporation, which is simulated based on the Penman–Monteith equation assuming a null canopy resistance. As in Testi et al. (2006), the aerodynamic resistance is deduced from the model proposed by Raupach (1994), parametrized and validated specifically for olive orchards following Verhoef et al. (1997). The direct evaporation from wet foliage prevents tree transpiration ( $E_p$ ), until the intercepted water is totally lost.

Runoff and infiltration are calculated following a Soil Conservation Service curve number methodology that was specifically calibrated and validated for different typologies of olive orchards (Romero et al., 2007). The approach requires information on the canopy ground cover ( $GC$ ) and the soil

hydrological condition ( $SHC$ ) -i.e., an indicative of the capacity of infiltration of the soil when it is wet. The water content at field capacity ( $\theta_{UL}$ ), wilting point ( $\theta_{LL}$ ) and saturation ( $\theta_{sat}$ ) are also needed for the computation of infiltration and all the remaining simulated processes.

Drainage and soil water redistribution processes are simulated by CERES-type sub-models (Jones and Kiniry, 1986), while soil evaporation rates ( $E_s$ ) are estimated using the model of Bonachela et al. (2001). The latter, calculates  $E_s$  with a modified Penman-FAO equation for stage-one evaporation and uses the model of Ritchie (1972) for the soil-limited evaporation stage. For the wetted soil zone, microadvection effects are considered (Bonachela et al., 2001).

The model by García-Tejera et al. (2017a) is used to compute root water uptake ( $RWU$ ) from each layer in the two soil zones, canopy transpiration ( $E_p$ ) and gross assimilation ( $A'$ ). By analogy with the Ohm's law for electric circuits, the model assumes that water transport through the SPAC is driven by differences in water potential and hydraulic resistances. In this regard, three hydraulic resistances are considered: from the soil to the root-soil-interface ( $R_s$ ), from the soil-root interface to the root xylem ( $R_r$ ) and from the root xylem to the canopy ( $R_x$ ).  $R_s$  depends on soil texture, root length density ( $L_v$ ), soil



water content ( $\theta$ ) (Gardner, 1960).  $R_r$  is a function of  $L_v$  and root permeability, the latter being mediated by  $\theta$  (Bristow et al., 1984) and temperature (García-Tejera et al., 2016). Finally,  $R_x$  is calculated from xylem anatomical traits and tree height. In the canopy, two leaf populations are considered (i.e., sunlit and shaded). For each one, gross assimilation ( $A'$ ), stomatal conductance ( $g_s$ ), intercellular  $\text{CO}_2$  concentration ( $C_i$ ) and leaf water potential ( $\Psi_l$ ) are calculated iteratively, considering both the models by Farquhar et al. (1980) and Tuzet et al. (2003). In doing so, the environmental  $\text{CO}_2$  concentration ( $C_a$ ) is explicitly taken into account for calculating both  $A'$  and  $g_s$  on the one hand. On the other, the model requires information on the intercepted photosynthetically active radiation ( $IPAR$ ) as well as the sunlit and shaded fractions of the canopy. These inputs are provided by a simple geometric model of radiation interception which assumes a spheroidal shape for the crown and accounts for the shadowing from neighboring trees. Finally,  $E_p$  is estimated from the imposed evaporation equation assuming that the canopy is coupled to the atmosphere, whereas  $RWU$  is deduced in each layer of each soil zone from the corresponding water potential differences and hydraulic resistances.

### Carbon Balance Component

This modeling component is aimed to simulate the growth and development of the trees and the carbon exchange of the orchard (Figure 2). First, the model calculates the daily pool of assimilates from the rate of  $A'$  determined by the SPAC model. Based on experimental evidence, OliveCan also assumes that such pool can be increased to some extent by reserve remobilization (Bustan et al., 2011) or fruit photosynthesis (Proietti et al., 1999). The available assimilates are allocated to growth and respiration, the latter being segregated into maintenance and growth respiration ( $RESP_M$  and  $RESP_g$ , respectively), of the different organs. In this regard, OliveCan considers six organ types: leaves, shoots (i.e., stems of up to 3 years-old), branches (including the trunk), coarse roots (i.e., roots with secondary growth), fine roots (i.e., absorbing roots) and fruits.

$RESP_M$  is calculated as a function of temperature and biomass, and it is subtracted directly from the pool of assimilates. Whenever maintenance respiration exceeds the pool of assimilates, the deficit is discounted from the reserve pool. The remaining assimilates are distributed among the different organs with partitioning rules being mediated by phenology. The loss of carbon during the synthesis of new biomass was included by calculating a production value ( $PV$ ) (Penning de Vries et al., 1974) for each type of organ according to its biochemical composition.

Two phenological stages are considered for the vegetative organs: (i) a dormant stage characterized by an absence of growth that is induced by chilling accumulation during autumn and (ii) a phase of active growth that starts in late winter, by the time average temperature is above a threshold. In relation to the reproductive growth, the date of flowering is determined with the two-phase model by De Melo-Abreu et al. (2004). Fruit growth is assumed to start after a given amount of thermal time is accumulated from the date of flowering and ceases when either maturity or the harvest date is reached.

During the vegetative rest period and provided that fruits are not present, all the available assimilates after discounting maintenance respiration are allocated to a virtual pool of reserves. Such reserve pool is subsequently used for the growth of vegetative organs and fruits during the growth season. Fruit growth can either be source-limited or sink-limited. In the former case, the associated partitioning coefficient is fixed whereas in the latter, it is calculated as a function of the number of fruits ( $FN$ ), which in turn is modeled as a function of the number of fruits and nodes produced in the previous year. In doing so, the model may be prone to errors in the estimates of productivity and vegetative growth for a given year when performing long runs, but such errors are to be compensated if those model outputs are averaged over biennia. With regard to the vegetative organs, fixed partitioning coefficients are adopted. Whenever fruits are present, the model considers that they become the priority sink of assimilates, thus the vegetative partitioning coefficients are applied after discounting the fruit demand from the daily pool of assimilates. Therefore, partitioning coefficients to vegetative organs are assumed to be independent of tree size, management factors and environmental conditions, as in the model of Morales et al. (2016). As a final remark, inspired by the CERES-type models (Jones and Kiniry, 1986), the growth of fine roots is distributed among the different layers in the two soil zones as a function of the size and water content of each soil compartment.

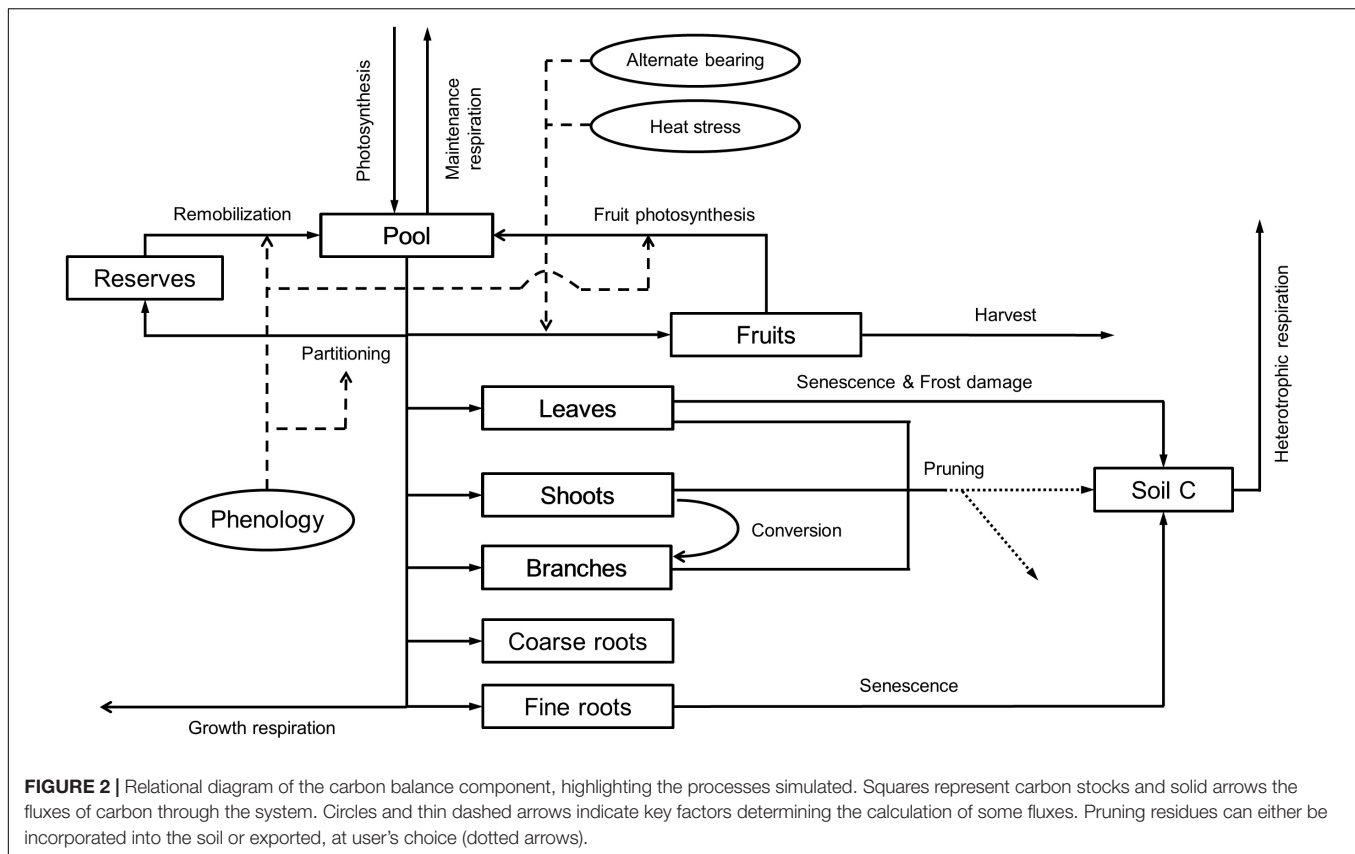
Senescence of leaves and fine roots are simulated using a similar approach to that in the model by Morales et al. (2016). OliveCan takes also into account the conversion of shoots into branches when they exceed 3 years-old. Besides that, the model considers some of the effects of frost events and heat stress. Frost damage is simulated by assuming that a fraction of the standing leaves is defoliated when minimum air temperature falls below a certain temperature threshold. A similar approach is used for simulating the effect of extremely high temperatures during flowering on fruit set: when maximum air temperature exceeds a given threshold, a reduction in the final  $FN$  is triggered.

Variables related to canopy characteristics such as leaf area index ( $LAI$ ) or  $GC$  are updated from the estimates of biomass of leaves assuming that the crowns present a spheroidal shape with constant leaf area density ( $LAD$ ) and ratio of vertical to horizontal canopy radii ( $R_{zx}$ ). Similarly, the biomass of fine roots in each soil compartment is used to compute root length density ( $L_v$ ) by adopting a constant specific root length ( $SRL$ ).

Finally, the soil carbon balance and heterotrophic respiration ( $RESP_H$ ) are computed with an adaptation of the model proposed by Huang et al. (2009) and modified to take into account the effect of soil moisture on the rate of decomposition according to Verstraeten et al. (2006). Then, by considering the different computed fluxes of assimilation and respiration within the orchard, OliveCan provides estimates of the ecosystem respiration ( $RESP_{eco}$ ) and net ecosystem exchange ( $NEE$ ).

### Management Component

Four management operations are considered in OliveCan: tillage, irrigation, harvest and pruning. In the model, tillage operations have an impact on  $CN$  whereas irrigation provides an additional



water input for the wetted soil zone. Irrigation amounts and dates can either be defined explicitly by the users or implicitly calculated through a dedicated routine that, at customizable intervals, applies a fraction of the maximum  $ET$  lost since the last irrigation. Harvesting takes place on a user-defined day of the year and results in the removal of fruits. At harvest, the model provides an estimate of oil yield ( $Y_{oil}$ ) by multiplying the dry biomass of fruits and a fixed coefficient representing the ratio of oil content to dry matter. Finally, pruning is simulated by setting a customizable fraction of  $LAI$  to be removed ( $F_{prune}$ ) and an interval between pruning operations. The model also reduces the biomasses of shoots and branches by the same fraction  $F_{prune}$ . The user should indicate whether pruning residues are incorporated into the soil or exported.

### Initialization Requirements

Apart from the weather dataset and some orchard (e.g., planting density, age, and latitude) and soil (e.g., depth,  $\theta_{UL}$ ,  $\theta_{LL}$ ) basic traits, the user is required to enter the initial values of  $GC$  and  $L_v$  to deduce the biomasses of the different organs following simple criteria (see Supplementary Material). For the computation of  $FN$  in the first season, an estimate of dry yield for the year preceding the start of the simulation is also needed. To initialize the state variables related to phenology, simulations must start at the beginning of a year and the temperature records of the preceding 3 months must be provided. Some simulation settings such as the number of years to simulate and  $N$  must also be provided.

Finally, the user is to indicate the management operations to be implemented and provide values to their parameters.

### Model Parameterization

When available, the values of the different parameters were taken from the literature. Supplementary Table S2 provides a complete list with the parameter values used for the simulations and the source from which they were taken. In short, the parameters of the SPAC model were taken from García-Tejera et al. (2017a,b), who, in turn, gathered most of the parameter values from different sources. Parameters related to phenology were obtained from reports by De Melo-Abreu et al. (2004) and López-Bernal et al. (2014, 2017). The studies by Mariscal et al. (2000) and Pérez-Priego et al. (2014) were used for setting the maintenance respiration and  $PV$  coefficients, respectively. Parameters related to the calculation of fruit number and yield were taken from several sources, including experimental data (see section “Number of Fruits and Alternate Bearing” in Supplementary Material). The coefficient of oil yield to dry fruit matter was taken from experimental data collected in a hedgerow cv. ‘Arbequina’ orchard (López-Bernal et al., 2015). Partitioning coefficients were based on findings by Mariscal et al. (2000); Villalobos et al. (2006) and Scariano et al. (2008). Reports from Barranco et al. (2005) and Koubouris et al. (2009) were used to parametrize the routines modeling the impacts of frost damage and heat stress, respectively. Coefficients modulating fine root growth distribution were directly taken from Jones and Kiniry

(1986). Finally, parameters implied in the soil carbon balance were taken from Verstraeten et al. (2006); Huang et al. (2009) and, to a lesser extent, from other studies.

## Model Testing

Experimental measurements conducted in two mature olive orchards located in the Alameda del Obispo Research Station, Córdoba, Spain (37.8°N, 4.8°W, 110 m) were used for assessing the reliability of OliveCan. The climate in the area is typically Mediterranean, with around 600 mm of average annual rainfall and 1390 mm and average annual  $ET_0$  of 1390 mm (Testi et al., 2004), respectively. The soil for both orchards is classified as a Typic Xerofluvent of sandy loam texture and exceeds 2 m in depth, with field capacity ( $\theta_{UL}$ ) and permanent wilting point ( $\theta_{LL}$ ) water contents of  $0.23 \text{ m}^3 \text{ m}^{-3}$  and  $0.07 \text{ m}^3 \text{ m}^{-3}$ , respectively (Testi et al., 2004). Weather data were collected using a station placed 500 m away from the orchards. Within both orchards, irrigation experiments comprising several irrigation treatments were performed. Each irrigation treatment was simulated separately with OliveCan.

## Experiment I

Extensive information on the orchard characteristics and dataset of Experiment I is provided by Iniesta et al. (2009). In short, the experiment was performed between 2004 and 2006 in a high density cv. 'Arbequina' olive orchard (tree spacing was  $7 \text{ m} \times 3.5 \text{ m}$ , i.e.,  $408 \text{ trees ha}^{-1}$ ) planted in 1997. Irrigation was applied 5 days a week by drip, with seven emitters of  $4 \text{ L h}^{-1}$  per tree. A randomized complete-block design was used with three replications of 12 trees each, and the following irrigation treatments:

- Control irrigation (CON), which applied the required water to match the maximum  $ET$ , discounting rainfall. The maximum  $ET$  was estimated using the model of Orgaz et al. (2006).
- Continuous deficit irrigation (CDI), which applied 25% of the irrigation supplied to CON, distributed throughout the irrigation season.
- Regulated deficit irrigation (RDI), which applied the same seasonal water as CDI, with a midsummer (July 1st to September 10th–15th) deficit period without irrigation.

The measurements (only performed for the central trees of the replicates) used for testing the model were oil yield ( $Y_{oil}$ ,  $\text{g m}^{-2}$ ), seasonal  $ET$  and daily transpiration ( $E_p$ ,  $\text{mm d}^{-1}$ ). With regard to the former, each tree was manually harvested and the fresh yield weighed in the field.  $Y_{oil}$  was subsequently determined from sub-samples of 5 kg of fresh fruits. Cumulative  $ET$  was determined by water balance for the whole 2005 and 2006 seasons by measuring soil water content with a neutron probe (model 503, Campbell Pacific Nuclear Corp, Pacheco, CA, United States). Eight access tubes were installed between two trees per replicate, normal to tree rows. Measurements were taken at several depths (from 0.075 to 2.65 m deep). Finally,  $E_p$  was measured in 2006 with a sap-flow system device developed and assembled in the IAS-CSIC in Córdoba and described by Testi and Villalobos (2009). The system uses the Compensated Heat Pulse (CHP)

method in combination with the Calibrated Average Gradient (CAG) procedure. The probes performed readings every 15 min at 4 depths in the xylem, spaced 10 mm. Six RDI, six CDI and four CON trees were instrumented with two probes per tree, at a height of 30 cm. The outputs of each probe were integrated first along the trunk radius and then around the azimuth angle. Average sap flow records for each treatment were calibrated against the estimates of  $E_p$  deduced from the difference between the measured  $ET$  and soil evaporation in a period of several weeks with no rainfall events during the summer. The model of Orgaz et al. (2006) was used to calculate soil evaporation. The calibrated sap flow data have not been published so far.

Values of  $GC$ ,  $LAD$ , and  $R_{zx}$  required to initialize the model were taken from measurements of tree silhouettes. A record of  $Y_{dry}$  of the year preceding simulations was also considered. Initial  $L_v$  values were taken from records measured by Moriana (2001) for the trees of Experiment II.

## Experiment II

Extensive information on the orchard characteristics and dataset of Experiment II is provided by Moriana et al. (2003). In short, the experiment was performed between 1997 and 1999 in a high density cv. 'Picual' olive orchard (tree spacing was  $6 \text{ m} \times 6 \text{ m}$ , i.e.,  $278 \text{ trees ha}^{-1}$ ) of 18 years of age. Irrigation was applied 5 days a week by drip, with four emitters of  $4 \text{ L h}^{-1}$  per tree. A randomized complete-block design was used with three replications of 16 trees each, and the following irrigation treatments:

- Control irrigation (CON), which applied the required water to match the maximum  $ET$ , based on the fully replenishing soil water extraction from April to October.
- Regulated deficit irrigation (RDI), which applied 75% of the water received by CON (i.e., rainfall plus irrigation) with a midsummer deficit period (15 July to 15 September) without irrigation.
- Continuous deficit irrigation (CDI), which also applied 75% of the water received by CON (i.e., rainfall plus irrigation), but for the whole irrigation season.
- Alternate year irrigation (AYI), which was rainfed in the year of low crop load (1998) and fully irrigated, as CON, during the heavy crop years (1997 and 1999).
- Rainfed (DRY), which received no irrigation during the whole experiment.

The measurements (only performed for the central trees of the replicates) used for the model were  $Y_{oil}$  and seasonal  $ET$ . On the one hand, trees were harvested between December 15th and January 15th for the 3 years. Individual fruit weight of each tree was measured and a subsample of 150 fruits from each tree was used for determining oil content. On the other, cumulative  $ET$  was determined by water balance for each season by measuring soil water content with a neutron probe (model 503, Campbell Pacific Nuclear Corp, Pacheco, CA, United States). Eight access tubes were installed between two trees per replicate in the four irrigation treatments and six tubes were placed in the rainfed treatment. Measurements were taken were performed at several depths (from 0.075 to 2.4 m deep).

Values of  $GC$ ,  $LAD$ , and  $R_{zx}$  required to initialize the model were taken from dedicated measurements. A record of  $Y_{dry}$  of the year preceding simulations was also considered. Initial  $L_v$  values were taken from records measured by Moriana (2001).

### Statistical Analysis

Model performances in reproducing measured data were assessed using mean absolute error (MAE, from 0 to  $+\infty$ , optimum 0), root mean square error (RMSE; from 0 to  $+\infty$ , optimum 0), coefficient of residual mass (CRM, from  $-\infty$  to  $+\infty$ , optimum 0) and modeling efficiency (EF, from  $-\infty$  to 1, optimum 1):

$$MAE = \sum_i^n |S_i - M_i| / n \quad (1)$$

$$RMSE = \sqrt{\sum_i^n (S_i - M_i)^2 / n} \quad (2)$$

$$CRM = 1 - \sum_i^n S_i / \sum_i^n M_i \quad (3)$$

$$EF = \frac{\sum_{i=1}^n (M_i - \bar{M})^2 - \sum_{i=1}^n (S_i - M_i)^2}{\sum_{i=1}^n (M_i - \bar{M})^2} \quad (4)$$

Where  $M_i$  is the  $i$ th measured variable,  $\bar{M}$  is the average value of all measurements,  $S_i$  is the  $i$ th simulated variable and  $n$  is

the number of measured values. In addition, the slope, intercept and coefficient of determination ( $r^2$ ) obtained by regressing the simulated and measured values were also used.

## RESULTS

Measured and simulated values of  $Y_{oil}$  for the two experiments are also presented in **Table 1** in relation to the year and the irrigation treatment. Field measurements showed two common patterns, irrespective of the experiment. On the one hand, there were consistent differences in  $Y_{oil}$  between treatments, with the fully irrigated treatments (i.e., CON) exhibiting higher values than the deficit or rainfed ones. On the other hand, measurements revealed “high”  $Y_{oil}$  in the first and third experimental seasons and “low”  $Y_{oil}$  in the intermediate one. The only exception to this rule occurred in the second biennia of Experiment II for the DRY and RDI treatments, probably as a consequence of the high level of water stress reached in 1999 (Moriana et al., 2003). OliveCan reproduced well both the differences between treatments and the general alternating trend in  $Y_{oil}$  for both experiments. Pooling all the data together, MAE and CRM were close to zero, RMSE was  $63.2 \text{ g m}^{-2}$  and EF was 0.48. The regression analysis yielded an  $r^2$  of 0.51 and a slope of 0.65. Slightly better results in terms of regression parameters, RMSE

**TABLE 1** | Observed (Obs.) and simulated (Sim.) values of seasonal evapotranspiration ( $ET$ ) and oil yield ( $Y_{oil}$ ) in the model tests using the experimental data from Iniesta et al. (2009) (Experiment I) and Moriana et al. (2003) (Experiment II).

Experiment	Treatment	Year	$ET$ (mm)		$Y_{oil}$ (g m <sup>-2</sup> )	
			Observed	Simulated	Observed	Simulated
I	FI	2004	*	798	319	255
		2005	958	862	151	119
		2006	1016	934	317	343
	CDI	2004	*	697	247	241
		2005	621	612	131	115
		2006	709	672	267	243
	RDI	2004	*	702	285	239
		2005	609	617	122	115
		2006	679	641	279	238
II	AYI	1997	822	642	357	254
		1998	576	509	21	91
		1999	839	643	244	191
	CON	1997	814	724	298	358
		1998	778	641	154	121
		1999	899	654	262	235
	DRY	1997	550	573	129	252
		1998	582	505	186	91
		1999	393	519	13	161
	RDI	1997	582	658	268	303
		1998	589	572	173	106
		1999	526	594	171	196
	SDI	1997	649	685	221	302
		1998	636	601	150	106
		1999	562	641	173	201

\*Not measured.



**TABLE 2** | Performances of OliveCan in reproducing the experimental data gathered for the model testing.

Process	Data	MAE	RMSE	CRM	EF	Slope	Intercept	$r^2$
Seasonal $ET$ (mm)	21	36.04	89.37	0.05	0.62	0.59	245.85	0.75
Daily $E_p$ (mm)	1095	0.17	0.32	0.14	0.84	1.01	-0.19	0.90
Annual $Y_{oil}$ (g m <sup>-2</sup> )	24	2.42	63.23	0.01	0.48	0.65	68.57	0.51
Biennial $Y_{oil}$ (g m <sup>-2</sup> )	16	8.81	21.59	0.05	0.64	0.68	52.00	0.70

MAE is mean absolute error, RMSE is root mean square error, CRM is coefficient of residual mass, EF is modeling efficiency and  $r^2$  is coefficient of determination.

and EF were found when analyzing the data grouped in biennia (Table 2).

A good agreement was found between model estimates of seasonal  $ET$  and the measurements by Moriana et al. (2003) and Iniesta et al. (2009) (Table 1). Simulated values were close to experimental data and tracked reasonably well the differences between treatments. MAE and RMSE were 36.04 mm and 89.4 mm, respectively. The statistical analysis revealed no systematic bias, with CRM close to 0, and values of EF of 0.62 (Table 2). Nevertheless, the model was found to underestimate seasonal  $ET$  for the treatments receiving more irrigation, which led to a rather low slope for the linear regression fit (0.59) between simulated and observed values.

Beyond the seasonal scale, OliveCan also provided good estimates of daily  $E_p$ , as evidenced by comparing the model outputs with the calibrated sap flow measurements recorded in Experiment I (Figure 3). The model was found to slightly underestimate  $E_p$  in winter and late autumn, particularly for the CON treatment, which led to a CRM of 0.14. In addition,  $E_p$  was slightly overestimated during the RDI midsummer non-irrigated period. However, simulated and observed values were generally very close in the three treatments. Considering the three treatments together ( $n = 1095$ ), the comparison between simulated and observed  $E_p$  values yielded a MAE of 0.17 mm d<sup>-1</sup>, a RMSE of 0.32 mm d<sup>-1</sup>, a EF of 0.84 and a satisfactory linear regression fit (Table 2).

Figure 4 illustrates the time course of measured and modeled  $E_p$  for the CON and CDI treatments over a typical sunny summer day of Experiment I (July 21st 2006, DOY 202). The four plotted  $E_p$  curves follow a bell shape, the CON ones exhibiting higher values than those of CDI. Apart from that, Figure 4 shows a remarkable lag between the observed and the simulated daily course of  $E_p$ , regardless of the treatment, with the diurnal trends of  $E_p$  being anticipated by the model. Hence, OliveCan tended to overestimate  $E_p$  in the hours following sunrise and to underestimate it around sunset.

## DISCUSSION

### Model Performance

Model tests generally revealed a high level of agreement between simulations and experimental measurements. Given the variety of the simulated treatments and the many assumptions that a model like OliveCan must take, we found the results satisfactory. Notwithstanding that, there were situations in which model estimates departed from observations. For example, some

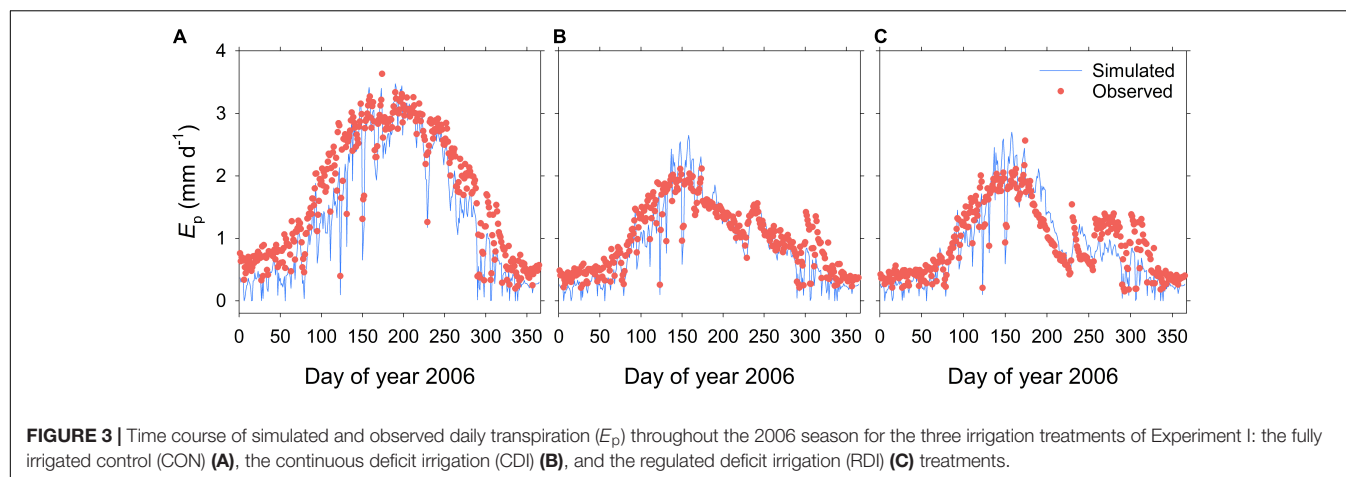
discrepancies were found for some of the simulations of  $Y_{oil}$ ,  $ET$  (Table 1) and  $E_p$  (Figure 3), but, considering that the general trends and differences between treatments were captured by the model, we believe that the results are highly acceptable. Some of the divergences between measured and simulated  $Y_{oil}$  might be attributed to the fact that the approach followed by OliveCan to simulate alternate bearing is limited, as far as the physiological bases of alternate bearing are not completely understood yet (Connor, 2005; Dag et al., 2010). However, biennial comparisons (Table 2) only improved slightly the results. Apart from that, the remarkable lag between the simulated and measured diurnal courses of  $E_p$  (Figure 4) was to be expected: measurements were performed in the trunk with sap flow sensors and OliveCan does not simulate the buffering effect of the water stored in aboveground organs (Cermák et al., 2007). Also, the model assumes that stomatal conductance responds instantaneously to changes in environmental conditions, but the slow dynamics of stomatal opening and closing can cause lags in diurnal transpiration (Violet-Chabrand et al., 2013).

Considering all the simulations together, the maximum simulated oil yield was 358 g m<sup>-2</sup> (Table 1), which is comparable to the maximum values estimated by the model of Morales et al. (2016) and to available experimental data (Villalobos et al., 2006; Pastor et al., 2007). Simulated values of radiation use efficiency for oil production (i.e., the amount of oil produced per unit of intercepted PAR) averaged over biennia ranged between 0.17 and 0.10 g MJ<sup>-1</sup>. These estimates are within the range of variation found by Villalobos et al. (2006) across a wide range of commercial orchards in Southern Spain.

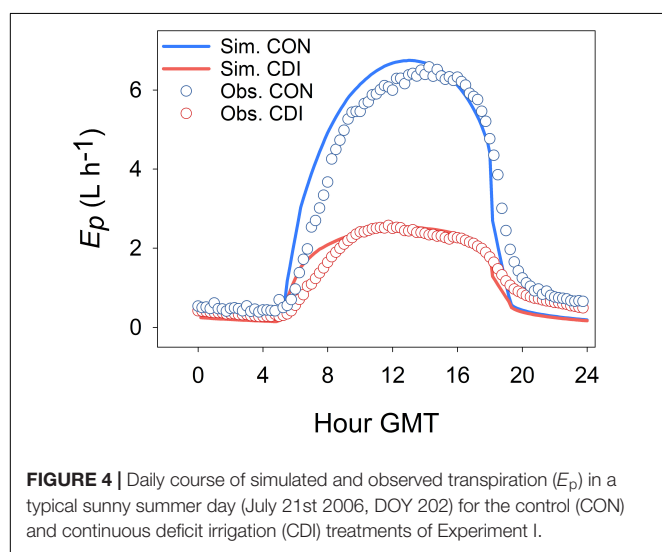
Overall, the results of all the aforementioned comparisons suggest that model performance is fairly satisfactory. However, further testing against experimental data taken from different environmental conditions and orchard characteristics seems highly desirable. This would help to provide additional evidence on the predictive power of OliveCan, or else to identify situations for which model accuracy could be improved through either better calibrations or reformulation of some routines. Apart from that, it should be noted that the reliability of OliveCan for estimating certain output parameters (e.g.,  $NEE$ ,  $RESP_H$ ) has not been tested specifically in the present study, which should also be the focus of future research efforts.

### Model Applicability

Considering its mechanistic approach, the vast quantity of simulated processes and its potential uses, OliveCan represents a momentous step forward in relation to previous olive growth



**FIGURE 3 |** Time course of simulated and observed daily transpiration ( $E_p$ ) throughout the 2006 season for the three irrigation treatments of Experiment I: the fully irrigated control (CON) (A), the continuous deficit irrigation (CDI) (B), and the regulated deficit irrigation (RDI) (C) treatments.



**FIGURE 4 |** Daily course of simulated and observed transpiration ( $E_p$ ) in a typical sunny summer day (July 21st 2006, DOY 202) for the control (CON) and continuous deficit irrigation (CDI) treatments of Experiment I.

simulation models. In this regard, OliveCan enables one to assess the combined effects of management operations and weather over crop performance for different olive orchard and soil typologies both under unstressed and water deficit conditions. Thus, the model shows potential for a broad range of research applications. For instance, OliveCan seems particularly suitable for assessing the performance of olive orchards under future climatic scenarios, as the model explicitly accounts for the multiple effects of reduced rainfall and increased environmental  $\text{CO}_2$  and temperature for the water and carbon balances of the orchard and the development of trees.

Obviously, the comprehensive nature and the wide range of simulated processes come at the expense of both model complexity and high input requirements. The latter is likely to be its main limitation, as far as some of the inputs (e.g., soil depth,  $L_v$  distribution) are not easy to measure in the field. In any case, it is noteworthy to emphasize that OliveCan has not been primarily conceived as a decision support system for farmers, but as a research tool.

## Further Research

During the development of the model, it became apparent that our current understanding of some of the physiological processes to be simulated was limited. For example, timing of vegetative bud break, dynamics of leaf senescence, fruit photosynthesis and the use of reserves are among the phenomena that have received less attention in the literature. Also, OliveCan is missing a sub-model aimed to properly simulate the dynamics of oil accumulation during the fruit growth period. Further research on these and other topics (e.g., alternate bearing) are clearly needed and might result in model improvements through either a more consistent parametrization or the formulation of better equations for simulating such processes.

Further research regarding genetic variability in model parameters is also desirable. With the exception of those related to the simulation of flowering date (De Melo-Abreu et al., 2004) and frost damage (Barranco et al., 2005), all parameters have been taken from past experiments carried out either with only one cultivar each ('Arbequina' being the most frequent) or averaging the results obtained for a few of them. Although the scarce literature does not allow us to disentangle how many of these crop parameters are cultivar-specific, it is clear that exploring their genetic variability might be important for enhancing model reliability. Moreover, the quantification of such cultivar variability may be used for evaluating its impact on tree physiology and productivity under different management, weather or orchard characteristics using OliveCan, which may be useful for breeding purposes.

Finally, future improvements of OliveCan might include additional sub-models for simulating nutrient uptake and the impact of pests and diseases. Apart from that, the model shows potential for being adapted to other tree species, so its interest may not be only restricted to olive researchers.

## CONCLUSION

The model presented here targets the simulation of the interactions between olive trees and their environment through a detailed characterization of the water and carbon balances of

the orchard as affected by weather variables, soil attributes and management operations. The generally high level of agreement found between measured and simulated data evidence the suitability of OliveCan for estimating olive orchard dynamics. These results encourage the application of the model to simulate the growth, carbon exchange and water relations of olive orchards in a wide range of research contexts, including studies on the performance of olive trees under climate change scenarios. The development of OliveCan has also highlighted significant knowledge gaps in relation to some physiological processes and the cultivar specificity of some of the parameters. Further research on these aspects may contribute to improve the reliability of the model.

## AUTHOR CONTRIBUTIONS

All authors played a significant role in the conception and development of the model. FV led out the coding, with contributions from ÁL-B, AM, OG-T, and LT. ÁL-B, LT, and FV gathered the datasets for testing the model. ÁL-B led out the writing with significant contributions from all co-authors.

## REFERENCES

- Abdel-Razik, M. (1989). A model of the productivity of olive trees under optional water and nutrient supply in desert conditions. *Ecol. Modell.* 45, 179–204. doi: 10.1016/0304-3800(89)90081-1
- Barranco, D., Ruiz, N., and Gómez-del Campo, M. (2005). Frost tolerance of eight olive cultivars. *HortScience* 40, 558–560.
- Bonachela, S., Orgaz, F., Villalobos, F. J., and Fereres, E. (2001). Soil evaporation from drip irrigated olive orchards. *Irrig. Sci.* 20, 65–71. doi: 10.1007/s002710000030
- Bristow, K. L., Campbell, G. S., and Calissendorff, C. (1984). The effects of texture on the resistance to water-movement within the rhizosphere. *Soil Sci. Soc. Am. J.* 48, 266–270. doi: 10.2136/sssaj1984.03615995004800020007x
- Bustan, A., Avni, A., Lavee, S., Zipori, I., Yeselson, Y., Schaffer, A., et al. (2011). Role of carbohydrate reserves in yield production of intensively cultivated oil olive (*Olea europaea* L.) trees. *Tree Physiol.* 31, 519–530. doi: 10.1093/treephys/tpq036
- Cermák, J., Kucera, J., Bauerle, W. L., Phillips, N. G., and Hinckley, T. M. (2007). Tree water storage and its diurnal dynamics related to sap flow and changes in stem volume in old-growth Douglas-fir trees. *Tree Physiol.* 27, 181–198. doi: 10.1093/treephys/27.2.181
- Connor, D. J. (2005). Adaptation of olive (*Olea europaea* L.) to water-limited environments. *Aust. J. Agric. Res.* 56, 1181–1189. doi: 10.1071/AR05169
- Dag, A., Bustan, A., Avni, A., Tzipori, I., Lavee, S., and Riov, J. (2010). Timing of fruit removal affects concurrent vegetative growth and subsequent return bloom and yield in olive (*Olea europaea* L.). *Sci. Hort.* 123, 469–472. doi: 10.1016/j.scienta.2009.11.014
- De Melo-Abreu, J. P., Barranco, D., Cordeiro, A. M., Tous, J., Rogado, B. M., and Villalobos, F. J. (2004). Modelling olive flowering date using chilling for dormancy release and thermal time. *Agric. For. Meteorol.* 125, 117–127. doi: 10.1016/j.agrformet.2004.02.009
- FAOSTAT (2014). Available at: <http://faostat.fao.org>
- Farquhar, G. D., von Caemmerer, S., and Berry, J. A. (1980). A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C<sub>3</sub> species. *Planta* 90, 78–90. doi: 10.1007/BF00386231
- Fernández, J. E., Moreno, F., Cabrera, F., Arrue, J. L., and Martín-Aranda, J. (1991). Drip irrigation, soil characteristics and the root activity of olive trees. *Plant Soil* 133, 239–251. doi: 10.1007/BF00009196
- García-Tejera, O., López-Bernal, A., Testi, L., and Villalobos, F. J. (2017a). A soil-atmosphere continuum (SPAC) model for simulating tree transpiration with a

## FUNDING

The research leading to these results has received funding from Ministerio de Economía y Competitividad (Grant Nos. AGL-2010-20766 and AGL2015-69822), from Junta de Andalucía (Grant No. P08-AGR-04202), from the European Community's Seven Framework Programme-FP7 (KBBE.2013.1.4-09) under Grant Agreement No. 613817. 2013–2016 “MODElling vegetation response to EXTREMe Events” (MODEXTREME, [modextreme.org](http://modextreme.org)) and from ERA-NET FACCE SURPLUS (Grant No. 652615, project OLIVE-MIRACLE), the latter co-funded by INIA (PCIN-2015-259). Besides, ÁL-B was funded by a postdoctoral fellowship (‘Juan de la Cierva-Formación 2015’ Programme, FJCI-2015-24109) from Ministerio de Economía y Competitividad.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00632/full#supplementary-material>

- soil multi-compartment solution. *Plant Soil* 412, 215–233. doi: 10.1007/s11104-016-3049-0
- García-Tejera, O., López-Bernal, A., Testi, L., and Villalobos, F. J. (2017b). Analysing the interactions between wetted area and irrigation volume in an olive orchard using a SPAC model with a multi-compartment soil solution. *Irrig. Sci.* 35, 409–423. doi: 10.1007/s00271-017-0549-5
- García-Tejera, O., López-Bernal, A., Villalobos, F. J., Orgaz, F., and Testi, L. (2016). Effect of soil temperature on root resistance: implications for different trees under Mediterranean conditions. *Tree Physiol.* 36, 469–478. doi: 10.1093/treephys/tpv126
- Gardner, W. R. (1960). Dynamic aspects of water availability to plants. *Soil Sci.* 89, 63–73. doi: 10.1097/00010694-196002000-00001
- Gómez, J. A., Giráldez, J. V., and Fereres, E. (2001). Rainfall interception by olive trees in relation to leaf area. *Agric. Water Manag.* 49, 65–76. doi: 10.1016/S0378-3774(00)00116-5
- Huang, Y., Yu, Y., Zhang, W., Sun, W., Liu, S., Jiang, J., et al. (2009). Agro-C: a biogeophysical model for simulating the carbon budget of agroecosystems. *Agric. For. Meteorol.* 95, 203–223. doi: 10.1016/j.agrformet.2008.07.013
- Iniesta, F., Testi, L., Orgaz, F., and Villalobos, F. J. (2009). The effects of regulated and continuous deficit irrigation on the water use, growth and yield of olive trees. *Eur. J. Agron.* 30, 258–265. doi: 10.1016/j.eja.2008.12.004
- Jones, C. S., and Kiniry, J. R. (1986). *CERES-Maize: A Simulation Model of Maize Growth and Development*. College Station: Texas A & M University Press.
- Koubouris, G. C., Metzikakis, I. T., and Vasilakakis, M. D. (2009). Impact of temperature on olive (*Olea europaea* L.) pollen performance in relation to relative humidity and genotype. *Environ. Exp. Bot.* 67, 209–214. doi: 10.1016/j.envexpbot.2009.06.002
- López-Bernal, A., García-Tejera, O., Orgaz, F., Testi, L., and Villalobos, F. J. (2014). “Olive bud dormancy is induced by low temperatures” in *Proceedings of the XIIIth Congress of the European Society for Agronomy*, eds P. Pepó and J. Csajbók (Debrecen: University of Debrecen), 253–254.
- López-Bernal, A., García-Tejera, O., Vega, V. A., Hidalgo, J. C., Testi, L., Orgaz, F., et al. (2015). Using sap flow measurements to estimate net assimilation in olive trees under different irrigation regimes. *Irrig. Sci.* 33, 357–366. doi: 10.1007/s00271-015-0471-7
- López-Bernal, A., Villalobos, F. J., García-Tejera, O., Testi, L., and Orgaz, F. (2017). Do olive vegetative buds undergo a real dormant state in Winter? *Acta Hort.* 1160, 227–230. doi: 10.17660/ActaHortic.2017.1160.33



- Mariscal, M. J., Orgaz, F., and Villalobos, F. J. (2000). Radiation-use efficiency and dry matter partitioning of a young olive (*Olea europaea*) orchard. *Tree Physiol.* 20, 65–72. doi: 10.1093/treephys/20.1.65
- Morales, A., Leffelaar, P. A., Testi, L., Orgaz, F., and Villalobos, F. J. (2016). A dynamic model of potential growth of olive (*Olea europaea* L.). *Eur. J. Agron.* 74, 93–102. doi: 10.1016/j.eja.2015.12.006
- Moriana, A. (2001). *Relaciones hídricas del olivo (Olea europaea L.) bajo riego deficitario*. Ph.D. thesis, University of Córdoba, Córdoba.
- Moriana, A., Orgaz, F., Pastor, M., and Fereres, E. (2003). Yield responses of a mature olive orchard to water deficits. *J. Am. Soc. Hortic. Sci.* 128, 425–431.
- Orgaz, F., Testi, L., Villalobos, F. J., and Fereres, E. (2006). Water requirements of olive orchards-II: determination of crop coefficients for irrigation scheduling. *Irrig. Sci.* 24, 77–84. doi: 10.1007/s00271-005-0012-x
- Pastor, M., García-Vila, M., Soriano, M. A., Vega, V., and Fereres, E. (2007). Productivity of olive orchards in response to tree density. *J. Hortic. Sci. Biotechnol.* 82, 555–562. doi: 10.3389/fpls.2017.01280
- Penning de Vries, F. W. T., Brunsting, A. H. M., and van Laar, A. H. (1974). Products requirements and efficiency of biosynthesis, a quantitative approach. *J. Theor. Biol.* 45, 339–377. doi: 10.1016/0022-5193(74)90119-2
- Pérez-Priego, O., Testi, L., Kowalski, A. S., Villalobos, F. J., and Orgaz, F. (2014). Aboveground respiratory CO<sub>2</sub> effluxes from olive trees (*Olea europaea* L.). *Agrofor. Syst.* 88, 245–255. doi: 10.1007/s10457-014-9672-y
- Proietti, P., Famiani, F., and Tombesi, A. (1999). Gas exchange in olive fruit. *Photosynthetica* 36, 423–432. doi: 10.1016/j.ecoenv.2016.02.024
- Raupach, M. R. (1994). Simplified expressions for vegetation roughness length and zero-plane displacement as functions of canopy height and area index. *Bound. Layer Meteorol.* 71, 211–216. doi: 10.1007/BF00709229
- Ritchie, J. T. (1972). Model for predicting evaporation from a row crop with incomplete cover. *Water Resour. Res.* 8, 1204–1213. doi: 10.1029/WR008i005p01204
- Romero, P., Castro, G., Gómez, J. A., and Fereres, E. (2007). Curve number values for olive orchards under different soil management. *Soil Sci. Soc. Am. J.* 71, 1758–1769. doi: 10.2136/sssaj2007.0034
- Scarano, L., Lo Bianco, R., Di Marco, L., and Policarpo, M. (2008). “Dynamics of dry matter partitioning in young ‘Nocellara del Belice’ olive trees,” in *Proceedings of the Fifth International Symposium on Olive Growing*, Vols 1 and 2, ed. M. T. L. S. F. L. Ozkaya (Leuven: Acta Horticulturae), 397–401. doi: 10.17660/ActaHortic.2008.791.58
- Testi, L., and Villalobos, F. J. (2009). New approach for measuring low sap velocities in trees. *Agric. For. Meteorol.* 149, 730–734. doi: 10.1016/j.agrformet.2008.10.015
- Testi, L., Villalobos, F. J., and Orgaz, F. (2004). Evapotranspiration of a young irrigated olive orchard in southern Spain. *Agric. For. Meteorol.* 121, 1–18. doi: 10.1016/j.agrformet.2003.08.005
- Testi, L., Villalobos, F. J., Orgaz, F., and Fereres, E. (2006). Water requirements of olive orchards: I simulation of daily evapotranspiration for scenario analysis. *Irrig. Sci.* 24, 69–76. doi: 10.1007/s00271-005-0011-y
- Tuzet, A., Perrier, A., and Leuning, R. (2003). A coupled model of stomatal conductance, photosynthesis and transpiration. *Plant Cell Environ.* 26, 1097–1116. doi: 10.1046/j.1365-3040.2003.01035.x
- Verhoef, A., McNaughton, K. G., and Jacobs, A. F. G. (1997). A parameterization of momentum roughness length and displacement height for a wide range of canopy densities. *Hydrol. Earth Syst. Sci.* 1, 81–91. doi: 10.5194/hess-1-81-1997
- Verstraeten, W., Veroustraete, F., and Feyen, J. (2006). On the temperature and water limitation of net ecosystem productivity: implementation in the C-Fix model. *Ecol. Modell.* 199, 4–22. doi: 10.1016/j.ecolmodel.2006.06.008
- Violet-Chabrand, S., Dreyer, E., and Brendel, O. (2013). Performance of a new dynamic model for predicting diurnal time courses of stomatal conductance at the leaf level. *Plant Cell Environ.* 36, 1529–1546. doi: 10.1111/pce.12086
- Villalobos, F. J., Testi, L., Hidalgo, J., Pastor, M., and Orgaz, F. (2006). Modelling potential growth and yield of olive (*Olea europaea* L.) canopies. *Eur. J. Agron.* 24, 296–303. doi: 10.1016/j.eja.2005.10.008

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 López-Bernal, Morales, García-Tejera, Testi, Orgaz, De Melo-Abreu and Villalobos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Association Mapping for Important Agronomic Traits in Safflower (*Carthamus tinctorius* L.) Core Collection Using Microsatellite Markers

Heena Ambreen<sup>†</sup>, Shivendra Kumar<sup>†</sup>, Amar Kumar, Manu Agarwal, Arun Jagannath\* and Shailendra Goel\*

Department of Botany, University of Delhi, New Delhi, India

## OPEN ACCESS

### Edited by:

Johann Vollmann,  
Universität für Bodenkultur Wien,  
Austria

### Reviewed by:

Leonardo Velasco,  
Instituto de Agricultura Sostenible  
(CSIC), Spain  
Ahmad Arzani,  
Isfahan University of Technology, Iran  
Ankica Kondic-Spika,  
Institute of Field and Vegetable Crops,  
Serbia

### \*Correspondence:

Arun Jagannath  
jagannatharun@yahoo.co.in  
Shailendra Goel  
shailendragoel@gmail.com

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 14 October 2017

**Accepted:** 13 March 2018

**Published:** 29 March 2018

### Citation:

Ambreen H, Kumar S, Kumar A,  
Agarwal M, Jagannath A and Goel S  
(2018) Association Mapping for  
Important Agronomic Traits in  
Safflower (*Carthamus tinctorius* L.)  
Core Collection Using Microsatellite  
Markers. *Front. Plant Sci.* 9:402.  
doi: 10.3389/fpls.2018.00402

*Carthamus tinctorius* L. (safflower) is an important oilseed crop producing seed oil rich in unsaturated fatty acids. Scarcity of identified marker-trait associations is a major limitation toward development of successful marker-assisted breeding programs in safflower. In the present study, a safflower panel (CartAP) comprising 124 accessions derived from two core collections was assayed for its suitability for association mapping. Genotyping of CartAP using microsatellite markers revealed significant genetic diversity indicated by Shannon information index ( $H = 0.7537$ ) and Nei's expected heterozygosity ( $I = 0.4432$ ). In Principal Coordinate Analysis, the CartAP accessions were distributed homogeneously in all quadrants indicating their diverse nature. Distance-based Neighbor Joining analysis did not delineate the CartAP accessions in consonance with their geographical origin. Bayesian analysis of population structure of CartAP demonstrated the unstructured nature of the association panel. Kinship analysis at population ( $G_{ij}$ ) and individual level ( $F_{ij}$ ) revealed absence of or weak relatedness between the CartAP accessions. The above parameters established the suitability of CartAP for association mapping. We performed association mapping using phenotypic data for eight traits of agronomic value (*viz.*, seed oil content, oleic acid, linoleic acid, plant height, number of primary branches, number of capitula per plant, 100-seed weight and days to 50% flowering) available for two growing seasons (2011–2012 and 2012–2013) through General Linear Model and Mixed Linear Model. Our study identified ninety-six significant marker-trait associations (MTAs;  $P < 0.05$ ) of which, several MTAs with correlation coefficient ( $R^2$ )  $> 10\%$  were consistently represented in both models and in both seasons for traits *viz.*, oil content, oleic acid content, linoleic acid content and number of primary branches. Several MTAs with high  $R^2$ -values were detected either in a majority or in some environments (models and/or seasons). Many MTAs were also common between traits (*viz.*, oleic/linoleic acid content; plant height/days to 50% flowering; number of primary branches/number of capitula per plant) that showed positive or negative correlation in their phenotypic values. The marker-trait associations identified in this study will facilitate marker-assisted breeding and identification of genetic determinants of trait variability.

**Keywords:** safflower, association mapping, SSR markers, core collections, population structure, kinship analysis

## INTRODUCTION

*Carthamus tinctorius* L., commonly known as “safflower,” contains seed oil with significantly high levels of nutritionally desirable unsaturated fatty acids, which is unique among oilseed crops (Fernandez-Martinez et al., 1993). Safflower is also used for extraction of dyes, several medicinal applications and as a plant factory for production of pharmaceuticals (Dajue and Mündel, 1996; Flider, 2013; Carlsson et al., 2014; Zhou et al., 2014). Additionally, its ability to grow under low moisture and high salinity conditions would give it a competitive edge over other oilseed crops in arid zones (Kaya et al., 2011; Bahrami et al., 2014; Yeilaghi et al., 2015; Ebrahimi et al., 2016). Currently, safflower is cultivated in around 20 countries in a total area of 1,140,002 hectares and production of 948,516 tons (FAOSTAT)<sup>1</sup>. The major producers of safflower are the Russian Federation (286,351 tons), Kazakhstan (167,243 tons), Mexico (121,767 tons), USA (99,830 tons), Turkey (58,000 tons), and India (53,000 tons) accounting for ~71% of total production (FAOSTAT)<sup>1</sup>. In spite of significant fluctuations in acreage under safflower cultivation, India was the highest average producer of the crop during 1994 to 2016. Nevertheless, safflower has not been able to create a niche for itself as a major oilseed crop. The primary factors that deter its cultivation are low yield and oil content, susceptibility to several biotic stresses, presence of spines, and lack of seed dormancy (Nimbkar, 2008). Genetic improvement of safflower is therefore, essential to increase its acceptability and utility as an oilseed crop of global importance.

A primary requisite for crop improvement is the presence of molecular and phenotypic diversity and identification of genes or quantitative trait loci (QTL) governing traits of agronomic importance. Genetic diversity in safflower has been analyzed using several molecular markers viz., Random Amplified Polymorphic DNA (RAPD), Inter Simple Sequence Repeat (ISSR), Amplified Fragment Length Polymorphism (AFLP), Simple Sequence Repeats (SSRs), and Single Nucleotide Polymorphism (SNPs) (Johnson et al., 2007; Yang et al., 2007; Amini et al., 2008; Khan et al., 2009; Sehgal et al., 2009; Chapman et al., 2010; Lee et al., 2014; Pearl and Burke, 2014; Ambreen et al., 2015; Kumar et al., 2015). These studies revealed high genetic differentiation among global safflower accessions and validated few “centers of similarity” of safflower that were identified based on morphological traits (Johnson et al., 2007; Chapman et al., 2010; Pearl and Burke, 2014; Kumar et al., 2015). Studies on development of linkage maps and tagging of agronomic traits using RAPDs, Restriction Fragment Length Polymorphism (RFLP), SSRs and SNP markers have been initiated in safflower (Hamdan et al., 2008, 2012; Mayerhofer et al., 2010; García-Moreno et al., 2011; Pearl et al., 2014). Linkage analysis and/or QTL mapping is a conventional method of identifying genomic regions governing simple and/or complex traits but requires development of bi-parental mapping populations, which is a time consuming process. Moreover, the allelic variation captured in QTL mapping is limited due

to the use of bi-parental crosses and fewer recombination events are tested leading to low mapping resolution (Flint-Garcia et al., 2005). In contrast, association mapping (AM) is a faster and more efficient approach for high-resolution evaluation of complex traits and appears as a promising tool to circumvent limitations of linkage mapping (Yu and Buckler, 2006; Oraguzie and Wilcox, 2007; Abdurakhmonov and Abdurakimov, 2008). AM identifies relationships between traits and genetic polymorphisms in a heterogeneous assemblage of unrelated individuals using naturally occurring recombination events, thus enabling fine-scale mapping of traits. It has emerged as a useful approach for identifying marker-trait associations for several agronomic traits in various crop species (Abdurakhmonov and Abdurakimov, 2008; Zhu et al., 2008; Blair et al., 2009; Yang et al., 2010; Li et al., 2011; Upadhyaya et al., 2012, 2013; Xu et al., 2013; Zhang et al., 2014). Till date, Ebrahimi et al. (2017) is the only study that has explored marker-trait associations in safflower through association mapping using AFLP markers.

The choice of germplasm used for association mapping is critical and should incorporate a wide range of diversity capturing maximum number of historical recombination events (Flint-Garcia et al., 2005; Yan et al., 2011). Core collections have performed well as association mapping panels in various crop species (Blair et al., 2009; Li et al., 2011; Upadhyaya et al., 2013; Soto-Cerda et al., 2014; Zhang et al., 2014) since they are enriched with the maximum possible range of diversity (genetic and phenotypic) existing in the crop germplasm. Another advantage offered by core collections is that these collections often include unrelated individuals (accessions) which decreases the chances of identifying spurious marker-trait associations due to pre-existing population structure. In safflower, evaluation of global germplasm collections identified significant diversity for most of the desirable traits (oil content, fatty acid composition, resistance to biotic and abiotic stresses, and spines) (Ashri, 1975; Dwivedi et al., 2005; Kumar et al., 2016). To facilitate genetic dissection of complex traits, small, operational core collections have been developed in safflower (Johnson et al., 1993; Dwivedi et al., 2005; Kumar et al., 2016). These core collections can be effectively used for association mapping provided they demonstrate presence of high genetic variance (for better mapping resolution) and possess weak population structure and low kinship association among individuals (to avoid spurious associations between molecular markers and functional loci) during association analysis (Pritchard et al., 2000).

The present study describes evaluation of a safflower panel of 124 accessions derived from two core collections (reported earlier from our group; Kumar et al., 2016) for its suitability for association mapping and identification of SSR loci associated with eight traits of agronomic value (plant height, number of primary branches, number of capitula per plant, 100-seed weight, days to 50% flowering, seed oil content, oleic acid content and linoleic acid content) in the crop. The current work will facilitate mining of elite genes or loci for safflower breeding and conservation.

<sup>1</sup> www.fao.org/faostat/en/#data/QD

## MATERIALS AND METHODS

### Plant Material and Phenotypic Evaluation

The present work utilize *C. tinctorius* L. accessions selected from two composite core collections of safflower, CartC1 (57 accessions) and CartC2 (106 accessions; Kumar et al., 2016). We merged accessions from these two core collections, removed redundancy (44 accessions were common to the two core collections) and included five Indian cultivars (Sharda, Manjira, Annigeri, PBNS-12 and TSF-1) assembling a final association panel with a non-redundant set of 124 accessions, which will hereafter be referred to as “CartAP” collection. Seed material of the accessions was obtained from USDA\_ARS, WRPIS, Pullman, WA, USA while the Indian cultivars were procured from National Bureau of Plant Genetic Resources, New Delhi, India and University of Agricultural Sciences, Dharwad, India. Details of the plant material used including their plant identification (PI) number, geographical origin and information regarding their distribution in the two core collections are provided in Supplementary Table 1. Phenotypic evaluation of eight agronomic traits utilized in the present work *viz.*, plant height, number of primary branches, number of capitula per plant, 100-seed weight, days to 50% flowering, seed oil content, oleic and linoleic acid was described by Kumar et al. (2016). The safflower accessions were evaluated in field conditions for two consecutive growing seasons (2011–2012 and 2012–2013) for the studied traits. The climatic variable data for the two growing seasons is provided in Supplementary Table 2.

### Isolation of Genomic DNA and SSR Genotyping

Genomic DNA was extracted from leaf tissues of 10 week-old seedlings of each accession as described by Ambreen et al. (2015). Ninety-three polymorphic SSR markers developed earlier in safflower (Ambreen et al., 2015) were used for genotyping the accessions. A three-primers PCR protocol (Perry, 2004) was followed for genotyping and the reaction mixture included 50 ng of template DNA, 1.5 mM MgCl<sub>2</sub>, 0.2 mM of each dNTPs, 0.05 μM IR700-labeled M13 primer, 0.05 μM M13-tailed forward primer, 0.05 μM reverse primer and 0.75 units of *Taq* DNA polymerase (Biotools, Spain). PCR amplifications were conducted on Veriti thermal cycler (Applied Biosystems, USA) under the following cycling conditions: initial denaturation at 96°C for 5 min followed by 28–30 cycles of 96°C for 45 s, appropriate annealing temperature (Ta°C, ranging from 55 to 65°C) for 30 s, extension at 72°C for 1 min and final extension at 72°C for 7 min. The IR700-labeled amplification products were size fractionated on 6.5% PAGE using 4,300 DNA analyzer system and analyzed following instructions provided in the user manual (LICOR, USA).

### Genetic Analysis

Genetic diversity statistics including number of alleles (Na), major allele frequency (M<sub>AF</sub>), observed heterozygosity (*H*<sub>o</sub>), expected heterozygosity (*H*<sub>e</sub>) and degree of polymorphism generated by each marker (computed as polymorphism information content; PIC) were calculated using PowerMarker version 3.25 (Liu and Muse, 2005). Rare/private allelic richness

per locus (*R*<sub>p</sub>) was determined based on rarefaction approach using HP-Rare v.1.0 (Kalinowski, 2005). Rarefaction allows evaluation of the allelic richness independent of the sample size (Kalinowski, 2004). Hardy Weinberg equilibrium and assessment of genetic diversity through expression of Shannon diversity index (*H*) and Nei's expected gene diversity (*I*) was tested using POPGENE version 1.32 (Yeh et al., 1999). Information about the genetic location of these SSR markers on safflower linkage groups was derived from Bowers et al. (2016) through BLASTN search using standalone Blast+ version 2.2.26 (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.2.26/>). The genomic sequences of 93 SSR loci were used as query for search against the genetic map assembled into twelve linkage groups of safflower by Bowers et al. (2016).

### Distance-Based Genetic Analysis

Genetic relationships between sampled accessions were elucidated through construction of an unrooted Neighbor Joining (NJ) dendrogram based on simple matching coefficient using Darwin version 6.0 (Perrier and Jacquemoud-Collet, 2006). A bootstrap value of 1,000 replicates was used to test the reliability of the NJ dendrogram. Principal Coordinate Analysis (PCoA) was performed using Darwin version 6.0.

### Analysis of Population Structure Using Bayesian Method

The population genetic structure was studied using Bayesian clustering method implemented in STRUCTURE version 2.3.4 (Pritchard et al., 2000). An admixture model and correlated allele frequencies were chosen for estimating proportion of ancestral contribution in each accession. We tested various *K*-values ranging from 1 to 25 with 10 independent replications at each *K*, 100,000 generations burnin period and 200,000 Markov Chain Monte Carlo (MCMC) repetitions. The analysis was performed independent of the geographical origin of accessions. The optimal *K*-value for the dataset was obtained by assessment of the results using STRUCTURE HARVESTOR (Earl, 2012). Accessions with membership proportions (*Q*-value) >80% were considered as pure and part of their corresponding cluster while accessions with membership proportions lesser than 80% were adjudged as admixtures. Calculation of pairwise *F*<sub>ST</sub> and Analysis of molecular variance (AMOVA) between the sub-populations of STRUCTURE were performed using GenAlEx version 6.5 (Peakall and Smouse, 2012) with 1,000 permutations.

### Kinship Coefficient

Relatedness between individuals was estimated through computation of kinship coefficients (*F*<sub>ij</sub>; individual level) following Loiselle et al. (1995). *F*<sub>ij</sub> measures the relative probabilities of identity in descent for random genes between two individuals, *i* and *j*. Kinship coefficient matrix between all pairs of accessions of the association panel was generated using SPAGeDi v 1.5 (Hardy and Vekemans, 2002). All the negative *F*<sub>ij</sub>-values were treated as zero (Hardy and Vekemans, 2002; Yu et al., 2006). The same program was used for evaluation of average kinship coefficient (*G*<sub>ij</sub>; population level) between sub-populations inferred using STRUCTURE.



## Association Mapping

Association mapping was performed for eight quantitative traits *viz.*, seed oil content, oleic acid content, linoleic acid content, 100-seed weight, plant height, number of primary branches, days to 50% flowering and number of capitula per plant for the two seasons independently. Only SSR markers with known positions on linkage groups were used for association mapping. Both General linear model (GLM) and Mixed linear model (MLM) were applied for assessment of marker-trait associations (MTA) in TASSEL v 5.0 (Bradbury et al., 2007) following the user manual. The GLM model includes population structure (Q matrix) but does not include kinship relationships (K matrix) and can often identify false positives in association mapping. To overcome this limitation, the MLM was proposed which incorporates both population structure (Q) and kinship (K) information to decrease spurious associations (Yu et al., 2006). The association between marker and trait was considered significant at a  $P < 0.05$ . The phenotypic variation explained by each marker-trait association was studied through correlation coefficient ( $R^2$ ).

## RESULTS

### SSR-Based Genotyping and Statistical Analysis

Genotyping of 124 safflower accessions of the CartAP collection was performed using 93 polymorphic SSR markers. The SSR markers generated a total of 311 alleles with an average of 3.34 alleles per locus. The number of alleles per locus ranged from 2 to 8 (Supplementary Table 3). The highest number of alleles (8) was detected for loci NGSaf\_265 and NGSaf\_282 (Table 1). The major allele frequency ( $M_{AF}$ ) was high and varied from 0.342 to 0.97 with an average of 0.66 (Supplementary Table 3). The observed heterozygosity ( $H_o$ ) had a low average value of 0.112 and ranged from 0 to 0.96 (Supplementary Table 3). The gene diversity or expected heterozygosity ( $H_e$ ) of studied markers ranged from 0.016 for NGSaf\_117 to 0.76 for NGSaf\_281 with an average of 0.438 (Supplementary Table 3). Information on genetic variability at the 93 SSR loci among accessions of the CartAP collection is summarized in Table 1.

The relative informativeness of these markers was estimated by measurement of polymorphic information content (PIC) for each locus. The PIC-value averaged at 0.38 and ranged from 0.02 for NGSaf\_117 to 0.73 for NGSaf\_281. Nineteen microsatellite loci were highly polymorphic with  $PIC > 0.5$  and ranged from 0.502 to 0.736 (Supplementary Table 3). A major proportion of microsatellite loci (63) were moderately polymorphic with PIC-values between 0.25 to 0.5 while 12 loci generated low polymorphism with  $PIC < 0.25$ . Based on marker statistics, the NGSaf\_281 locus was most informative and had highest utility value among the 93 SSR loci. We were able to assign map positions to 48 out of the 93 SSR loci based on the genetic map of safflower reported by Bowers et al. (2016). SSR locus specific details ( $N_a$ ,  $M_{AF}$ ,  $H_e$ ,  $H_o$ , and PIC) and their map positions are provided in Supplementary Table 3. Mean allelic richness and private allelic richness (number of unique alleles)

**TABLE 1 |** Summary statistics for genetic variability at 93 SSR loci.

MARKER PARAMETERS	
Number of accessions sampled	124
<b>ALLELES</b>	
Total number of alleles	311
Average number of alleles per locus	3.34
Range of alleles	2–8
SSR loci with highest number of alleles <sup>#</sup>	NGSaf_265 and NGSaf_282 (8 alleles)
<b><math>H_o</math></b>	
Average $H_o$	0.112
SSR locus with highest $H_o$ <sup>#</sup>	NGSaf_294 (0.958)
SSR locus with lowest $H_o$ <sup>#</sup>	NGSaf_14, NGSaf_45, NGSaf_63, NGSaf_98, NGSaf_114, NGSaf_117, NGSaf_130, NGSaf_145, NGSaf_151, NGSaf_154, NGSaf_173, NGSaf_248 (0)
<b><math>H_e</math></b>	
Average $H_e$	0.438
SSR locus with highest $H_e$ <sup>#</sup>	NGSaf_281 (0.76)
SSR locus with lowest $H_e$ <sup>#</sup>	NGSaf_117 (0.016)
<b>PIC</b>	
Average PIC-value	0.38
SSR locus with highest PIC-value <sup>#</sup>	NGSaf_281 (0.73)
SSR locus with lowest PIC-value <sup>#</sup>	NGSaf_117 (0.02)

*H<sub>e</sub>*, gene diversity or expected heterozygosity; *H<sub>o</sub>*, observed heterozygosity; *PIC*, polymorphic information content; <sup>#</sup> Values for each parameter are provided in parenthesis.

were measured based on rarefaction approach implemented in HP-rare v 1.0. The mean allelic richness per locus was estimated as 1.287 while private allelic richness was 0.0186. At a significance level of  $P < 0.05$ , all loci except NGSaf\_300 deviated from Hardy-Weinberg equilibrium (HWE) while at  $P < 0.01$ , all loci deviated from HWE (Supplementary Table 3).

### Genetic Diversity

High genetic differentiation was observed among CartAP accessions based on the genetic diversity indices—Shannon information index ( $H$ ; 0.7537) and Nei's expected heterozygosity ( $I$ ; 0.4432). Genetic diversity in the 10 proposed regional gene pools and regions of secondary introduction of safflower (America and Australia) were also evaluated (Table 2). Among the regional gene pools, the Shannon information index varied from 0.4158 to 0.7211 while Nei's expected heterozygosity ranged from 0.2595 to 0.4265. Based on diversity indices, accessions from the Indian subcontinent and American region were the most diverse with an estimated Shannon information index and Nei's expected heterozygosity value of  $H = 0.7028$  and 0.7211 and  $I = 0.4265$  and 0.4227, respectively. Near East ( $H = 0.4158$ ;  $I = 0.2595$ ) and Turkey ( $H = 0.4650$ ;  $I = 0.2981$ ) were genetically least diverse based on our assessment using SSR markers. Within-region segregation assessed using SSR markers revealed the highest number of polymorphic loci among accessions from the Indian subcontinent (91) followed by America (90) and Iran-Afghanistan (85). The least number of polymorphic loci were recorded for accessions from Near East (60) and Turkey (60; Table 2).

## Genetic Relationships and Population Structure

### Neighbor Joining (NJ) Dendrogram and Principal Coordinate Analysis (PCoA)

The genetic relatedness between safflower accessions of CartAP was analyzed by construction of Neighbor-Joining (NJ) dendrogram based on simple matching coefficient (**Figure 1A**). The 124 accessions were grouped into three major clusters (named as NJ I–NJ III) with 7, 48, and 69 accessions, respectively. However, no distinction could be drawn among these groups based on their geographical origin with each of the major cluster exhibiting heterogeneous clustering of accessions (**Figure 1A**). We also obtained low bootstrap support (< 50%) for majority

of the nodes of the NJ dendrogram and only very few internal nodes exhibited high bootstrap value > 50%. All the five Indian cultivars analyzed in the present study clustered together in NJ II indicating their narrow genetic base. Distance-based NJ analysis indicated lack of genetic relationships among these accessions. In Principal Coordinate Analysis (PCoA) based on SSR data, the principal axes 1 and 2 explained 29 and 25% of the total molecular variance, respectively (**Figure 1B**). In agreement with the clustering pattern of NJ dendrogram, aggregation of accessions into discrete groups relative to their geographical origin was not observed in the PCoA plot. The CartAP accessions tend to disperse homogeneously in all the four quadrants indicating the diverse nature of sampled accessions (**Figure 1B**).

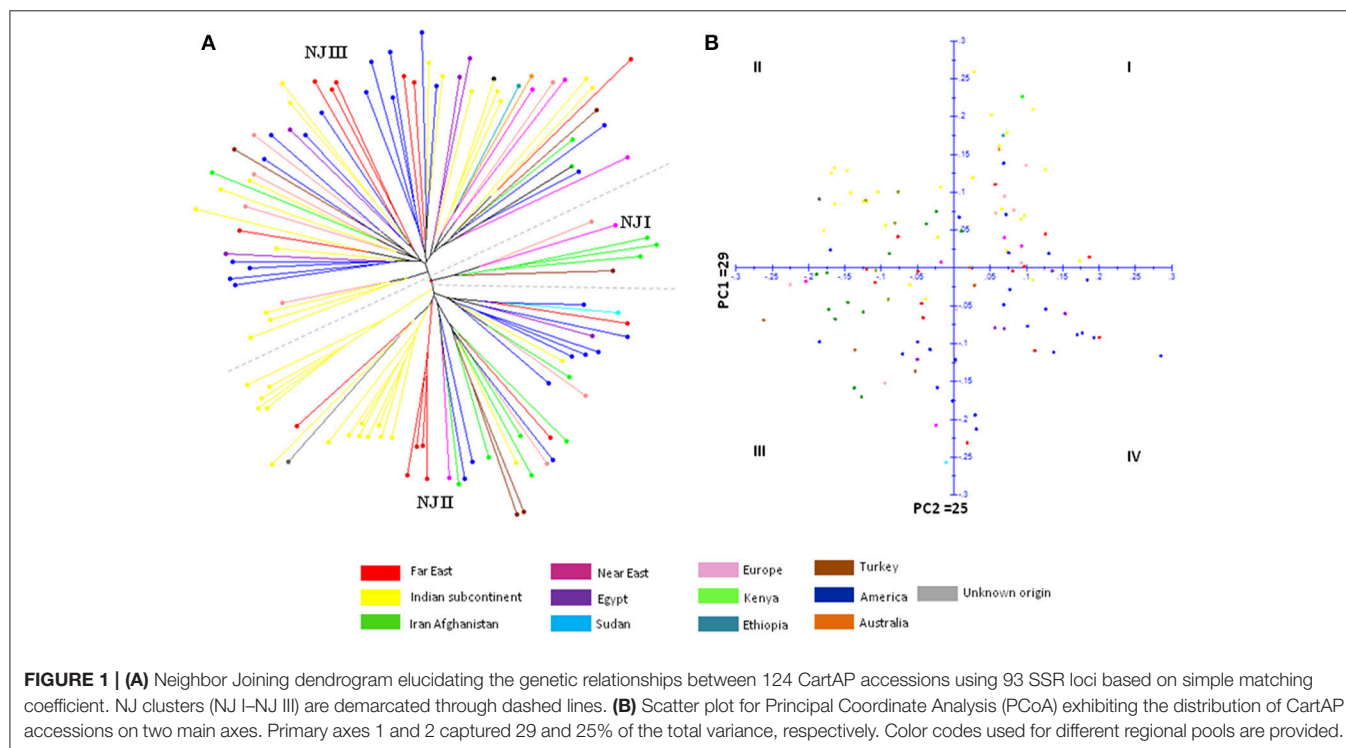
**TABLE 2 |** Genetic diversity indices of CartAP accessions based on their regional gene pool distribution.

Regional gene pool	N	P <sub>i</sub>	PP <sub>i</sub> (%)	H	I
Far East	15	84	90.3	0.6029	0.3782
Indian subcontinent	32	91	97.8	0.7028	0.4265
Iran-Afghanistan	12	85	91.4	0.6016	0.3596
Egypt	5	72	77.4	0.5289	0.3449
Europe	9	84	90.3	0.5966	0.3818
Near East	6	60	64.5	0.4158	0.2595
Turkey	5	60	64.5	0.4650	0.2981
America	30	90	96.8	0.7211	0.4227

N, Number of accessions; P<sub>i</sub>, Number of polymorphic loci; PP<sub>i</sub>, Percent polymorphism; H, Shannon information index; I, Nei's gene diversity.

### Population Structure and Differentiation

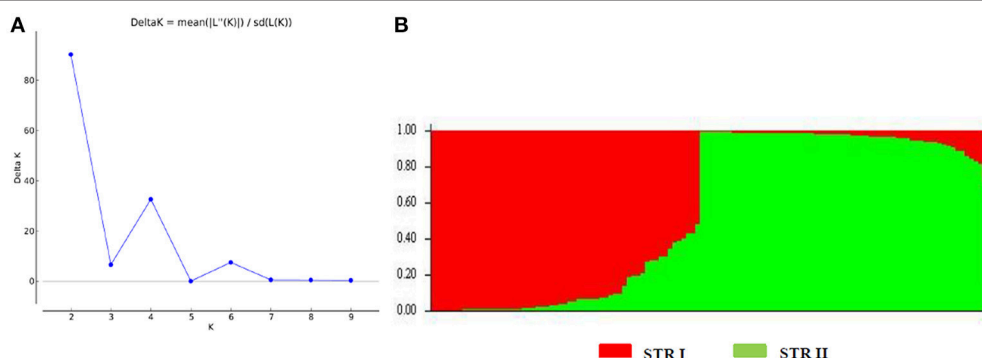
We also analyzed the underlying population structure of CartAP through Bayesian-based approach using STRUCTURE v 2.3.4. Log mean probability and change in log probability ( $\Delta K$ ) were determined using STRUCTURE HARVESTER, which generated the highest peak at  $K = 2$  (**Figure 2A**). Based on membership proportions, safflower accessions with  $\geq 80\%$  of shared ancestry were delineated into two major clusters (STR I and STR II) with 46 and 62 accessions, respectively (**Figure 2B**). Each of these clusters included diverse accessions originating from different regional pools. Sixteen accessions with < 80% of shared ancestry were categorized as admixtures. Additional smaller peaks observed at  $K = 4$  and  $K = 6$  implied presence of sub-groups within the two major groups (**Figure 2A**). Therefore, an independent STRUCTURE run was performed for STR I and



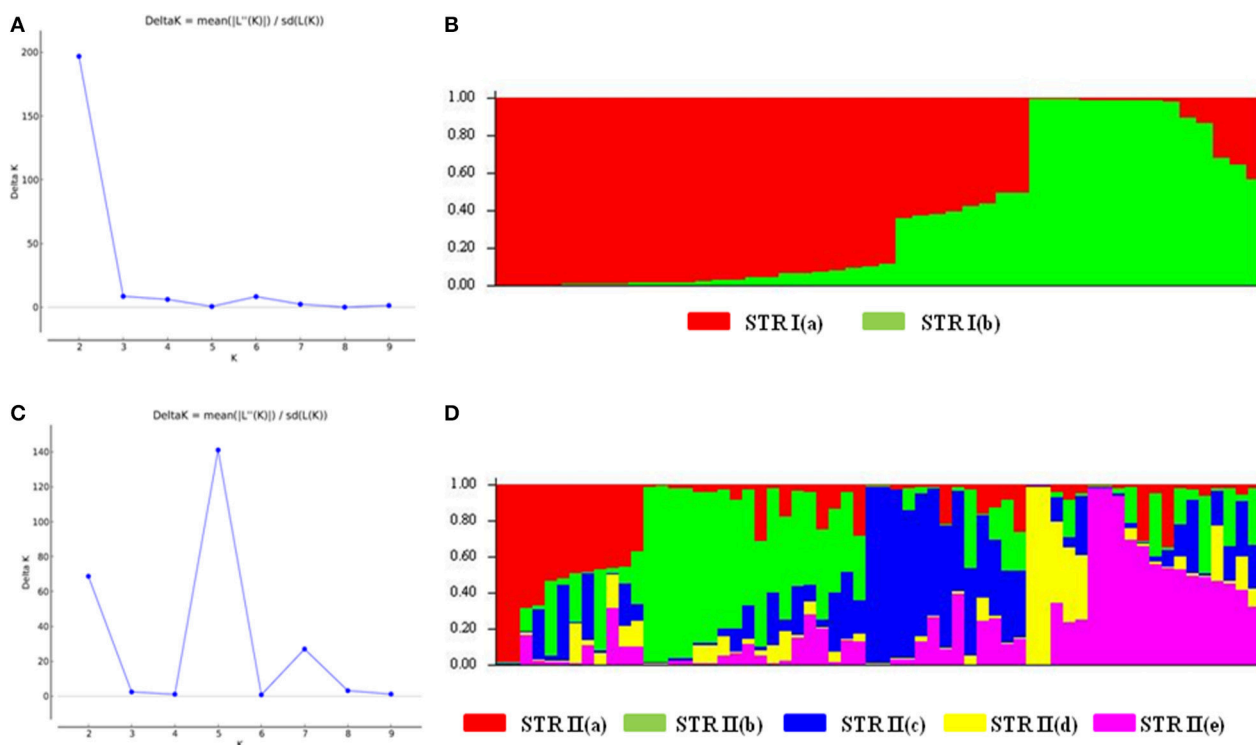
STR II with  $K$ -values ranging from 1 to 10. Sub-clustering of STR I identified the highest peak at  $K = 2$  (**Figure 3A**). The two sub-clusters designated as STR I(a) and STR I(b), contained 32 and 14 accessions, respectively (**Figure 3B**). Sub-clustering of the second major group, STR II yielded highest peak at  $K = 5$  (**Figure 3C**). The five sub-clusters were designated as STR II(a), STR II(b), STR II(c), STR II(d), and STR II(e) and contained 12, 18, 13, 5, and 14 accessions, respectively (**Figure 3D**). Clustering

of accessions in STRUCTURE analysis was heterogeneous and in concordance with the results of NJ and PCoA. The regional distribution of accessions in the seven sub-populations is given in **Table 3**. Among the sub-populations, a substantial number of accessions were admixtures (**Figures 3B,D**).

The expected heterozygosity for the seven sub-clusters varied from 0.1726 for STR II(d) to 0.4352 for STR I(a) (**Table 4**). The genetic divergence among the seven sub-populations was



**FIGURE 2 |** Population structure of CartAP collection inferred using Bayesian-based program STRUCTURE. **(A)** Estimation of hypothetical sub-populations using  $\Delta K$ -values. The number of identified sub-populations were two ( $K = 2$ ). **(B)** Population structure analysis of CartAP accessions at  $K = 2$  based on inferred ancestry (Q matrix). The sub-populations were named STR I and STR II.



**FIGURE 3 |** Hierarchical population structure analysis **(A)** Estimation of sub-population for STR I. Maximum number of sub-populations were inferred at  $K = 2$ . **(B)** Population structure for STR I at  $K = 2$  based on Q matrix. The sub-populations were designated as STR I(a) and STR I(b) **(C)** Estimation of sub-populations for STR II. Maximum number of subpopulations were inferred at  $K = 5$ . **(D)** Population structure for STR II at  $K = 5$  based on Q matrix. The sub-populations were named as STR II(a)–STR II(e). Color codes for each sub-population are provided.

**TABLE 3 |** Distribution of CartAP accessions in different sub-populations derived from STRUCTURE analysis.

Regional gene pools	Number of accessions	STRI		STRII					Admixtures
		STRI(a)	STRI(b)	STRII(a)	STRII(b)	STRII(c)	STRII(d)	STRII(e)	
Far East (FE)	15	3	–	2	5	–	–	1	4
Indian subcontinent (IS)	32	5	12	2	1	8	1	3	4
Iran-Afghanistan (IA)	12	9	–	–	–	–	–	3	1
Near East (NE)	6	2	1	–	–	1	–	2	–
Turkey (TU)	5	3	–	–	–	–	–	1	1
Egypt (EG)	5	1	–	2	2	–	–	–	–
Sudan (SU)	1	1	–	–	–	–	–	–	–
Kenya (KE)	1	–	–	–	–	1	–	–	–
Ethiopia (ET)	1	–	–	–	–	1	–	–	–
Europe (EU)	9	2	–	2	1	–	3	1	–
America (US)	30	6	–	4	9	1	1	3	6
Australia (AUS)	1	–	–	–	–	1	–	–	–
Unknown origin (UN)	1	–	1	–	–	–	–	–	–
Total	124	32	14	12	18	13	5	14	16

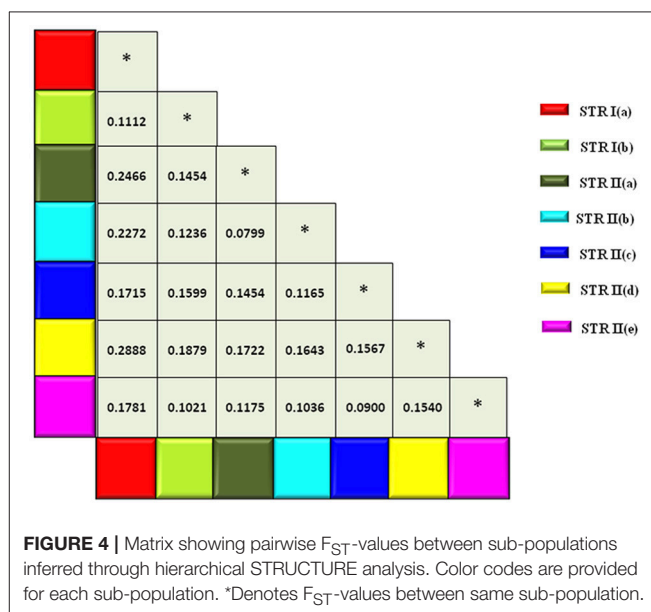
**TABLE 4 |** Expected heterozygosity of the sub-populations derived through STRUCTURE analysis.

Sub-clusters	Expected heterozygosity
STR I	0.469
STR I(a)	0.4352
STR I(b)	0.2838
STR II	0.4217
STR II(a)	0.2230
STR II(b)	0.3475
STR II(c)	0.2669
STR II(d)	0.1726
STR II(e)	0.2845

measured through pair-wise  $F_{ST}$ , which ranged from 0.0799 for STR II(b) and STR II(a) to 0.2888 for STR II(d) and STR I(a) ( $P < 0.001$ ; **Figure 4**). A hierarchical AMOVA analysis was performed among and within regional pools of CartAP accessions, which provided an estimated molecular variance of 7 and 93%, respectively (**Table 5**). Similar evaluation was conducted for seven sub-clusters derived through STRUCTURE analysis of safflower accessions. It provided a molecular variance of 16 and 84% among and within sub-clusters, respectively (**Table 5**). Genetic differentiation among the sub-clusters was highly significant ( $P < 0.0001$ ).

## Kinship Analysis at Individual and Population Level

Based on 93 SSR markers, the co-ancestry or kinship coefficient ( $F_{ij}$ ) between any two safflower accessions of CartAP averaged at a value of 0.043. A major proportion of the pairs of safflower accessions (~54.3%) had zero pairwise kinship


**FIGURE 4 |** Matrix showing pairwise  $F_{ST}$ -values between sub-populations inferred through hierarchical STRUCTURE analysis. Color codes are provided for each sub-population. \*Denotes  $F_{ST}$ -values between same sub-population.

estimates followed by 16.59% pairs having kinship estimates between 0.001 to 0.05, 12.24% pairs between 0.051 to 0.1, 7.8% pairs between 0.11 to 0.15, 4.5% pairs between 0.151 to 0.20 and 2.4% pairs between 0.201 to 0.25 (**Figure 5A**). A minimal fraction of safflower accession pairs (0.022%) exhibited kinship estimate  $>0.25$ . At the population level,  $G_{ij}$ -values were estimated to assess kinship associations between the seven inferred sub-populations derived from STRUCTURE analysis. The mean kinship coefficients ( $G_{ij}$ ) ranged from 0 to 0.0783 (**Figure 5B**) between the sub-populations. Negative values of  $G_{ij}$  were treated as zero for the current analysis. The kinship analysis indicated that most CartAP accessions and the inferred sub-populations exhibited no or weak level



of genetic relatedness at the individual and population level, respectively.

## Association Mapping

The 48 mapped SSR markers were distributed across all the 12 linkage groups of safflower (Supplementary Table 3) with maximum number of SSR loci (9) on linkage group (LG) 6. Identification of markers linked with each of the eight agronomic traits of safflower (seed oil content, oleic acid content, linoleic acid content, 100-seed weight, plant height, number of primary branches, days to 50% flowering and number of capitula per plant) was done through association analysis and performed independently for the two growing seasons. The range and mean values of the eight phenotypic traits evaluated for CartAP accessions are provided in **Table 6**. Association mapping using GLM and MLM models identified 96 significant marker-trait associations ( $P < 0.05$ ) for the eight traits across two

growing seasons (2011–2012 and 2012–2013). Twenty-eight out of 48 SSR loci accounted for these 96 significant associations. A comparatively higher number of significant marker-trait associations were observed through GLM (60) than MLM (36). Out of the 96 significant marker-trait associations (MTAs) detected based on  $P$ -value, 63 associations exhibited correlation value ( $R^2$ )  $\geq 10\%$ . Those MTAs, which were detected in more than one season/model and exhibited  $R^2 \geq 10\%$  in at least one model and/or season were retained for further analysis (**Table 7**). As an exception, we have included the MTA for NGSaf\_279 with number of capitula per plant, since it was consistently detected in all the seasons and models with  $R^2$  ranging from 7.4 to 9.3%.

For the 2011–2012 season, 19 and 14 significant MTAs with  $R^2 \geq 10\%$  were detected using GLM and MLM analysis, respectively (**Table 7**). Twelve MTAs were common between both the models for the studied traits. In GLM, the least number of significant associations ( $R^2 \geq 10\%$ ) were recorded for oleic acid, number of capitula per plant and number of primary branches (1) while the maximum number of MTAs (6) was recorded for plant height. The correlation value ( $R^2$ ) for these MTA ranged from 10 to 24.1%. The MTA between 100-seed weight and NGSaf\_306 displayed the highest  $R^2$  (24.1%; **Table 7**). In MLM analysis, the highest number of significant loci was associated with oleic

**TABLE 5 |** Analysis of molecular variance (AMOVA) between and within regional gene pools and between and within sub-populations derived through hierarchical STRUCTURE analysis.

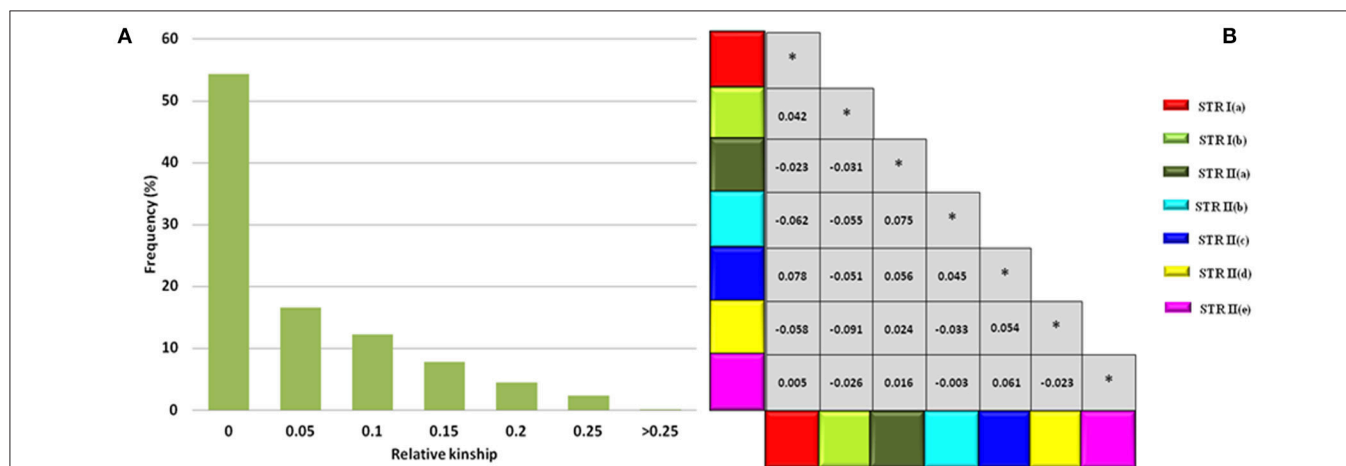
	Est. Var.*	Percent of variance	P-value <sup>a</sup>
<b>REGIONAL GENE POOLS</b>			
Between populations	1.413	7	0.001
Within populations	19.024	93	0.001
Total	20.437	100	
<b>SEVEN SUB-POPULATIONS DERIVED FROM STRUCTURE</b>			
Between populations	3.367	16	0.001
Within populations	17.524	84	0.001
Total	20.891	100	

\*Estimated variance.

<sup>a</sup>With 999 data permutations.

**TABLE 6 |** Mean and range of different phenotypic traits for CartAP accessions.

Phenotypic traits	Mean	Range
Oil content (%)	32	16–50
Oleic acid (%)	27	10–79
Linoleic acid (%)	65	13–87
100-seed weight (gm)	4	1–8
Plant height (cm)	156	94–226
Number of capitula per plant	82	16–203
Number of primary branches	16	6–33
Days to 50% flowering	141	119–160



**FIGURE 5 |** Kinship estimates at individual and population level. **(A)** Frequency distribution for global pairwise kinship estimates ( $F_{ij}$ ) for CartAP accessions. **(B)** Matrix showing pairwise  $G_{ij}$ -values between sub-populations inferred through hierarchical STRUCTURE analysis. Color codes are provided for each sub-population.

\*Denotes  $G_{ij}$ -values values between same sub-population.

**TABLE 7 |** Marker-trait associations identified through General linear model (GLM) and Mixed linear model (MLM) using phenotypic data of season 2011–2012 and 2012–2013.

Traits	SSR marker ID	LG <sup>®</sup>	SEASON 2011–2012				SEASON 2012–2013			
			GLM		MLM		GLM		MLM	
			<i>R</i> <sup>2</sup> (%)	<i>P</i> -value	<i>R</i> <sup>2</sup> (%)	<i>P</i> -value	<i>R</i> <sup>2</sup> (%)	<i>P</i> -value	<i>R</i> <sup>2</sup> (%)	<i>P</i> -value
OIL CONTENT										
	NGSaf_15	3	22.3	1.363E-5	10.4	0.0436	23.4	6.85E-06	16	0.004
	NGSaf_148	7	–	–	–	–	11.5	0.02	–	–
	NGSaf_201	11	12.9	0.011	–	–	–	–	–	–
	NGSaf_255	1	9.7	0.033	–	–	16	0.001	–	–
	NGSaf_300	12	16.7	6.026E-4	16.2	0.016	17.9	2.99E-04	13.8	0.042
OLEIC ACID										
	NGSaf_67	7	11.4	.005	11.4	0.017	11.6	0.004	11.6	0.016
	NGSaf_210	4	–	–	14.3	0.031	–	–	14.6	0.028
	NGSaf_309	5	–	–	34.1	.002	–	–	34	0.002
LINOLEIC ACID										
	NGSaf_67	7	10.9	0.006	10.9	0.021	–	–	16.7	0.039
	NGSaf_83	3	–	–	–	–	12.4	0.014	–	–
	NGSaf_155	10	–	–	–	–	19.1	0.041	–	–
	NGSaf_210	4	14.3	0.013	14.4	0.031	14	0.007	17.9	0.004
	NGSaf_309	5	–	–	34.1	0.002	–	–	–	–
100-SEED WEIGHT										
	NGSaf_101	4	–	–	–	–	8.9	0.006	15.3	0.034
	NGSaf_306	6	14.7	0.031	24.52	0.011	13	0.034	–	–
	NGSaf_309	5	24.1	4.2291E-4	22.9	0.044	15.5	0.022	–	–
PLANT HEIGHT										
	NGSaf_23	6	10	0.009	–	–	–	–	–	–
	NGSaf_83	3	12.5	0.013	–	–	–	–	–	–
	NGSaf_101	4	14.1	0.019	–	–	12.2	0.049	–	–
	NGSaf_156	6	11.8	0.023	–	–	13.7	0.01	13.9	0.033
	NGSaf_173	5	10.6	3.84E-4	4.4	0.027	–	–	–	–
	NGSaf_296	6	14.7	0.007	–	–	12.7	0.021	13.1	0.018
NUMBER OF CAPITULA PER PLANT										
	NGSaf_279	6	7.4	0.037	9.3	0.037	8.1	0.024	8.5	0.051
	NGSaf_309	5	15.4	0.023	25.7	0.023	–	–	–	–
NUMBER OF PRIMARY BRANCHES										
	NGSaf_15	3	–	–	–	–	11.8	0.008	12.4	0.021
	NGSaf_83	3	13.3	0.009	–	–	10.5	0.037	10.5	0.047
	NGSaf_279	6	8.9	0.015	10.8	0.018	8	0.024	8.9	0.043
DAYS TO 50% FLOWERING										
	NGSaf_92	9	–	–	11.4	0.014	7.9	0.026	11.6	0.014
	NGSaf_101	4	12.6	0.037	–	–	–	–	–	–
	NGSaf_201	11	11.9	0.017	11.4	0.014	11.8	0.019	–	–
	NGSaf_255	1	10.7	0.021	–	–	–	–	–	–
	NGSaf_296	6	–	–	–	–	11.1	0.043	–	–

<sup>a</sup>Linkage group; *R*<sup>2</sup>, correlation coefficient.

acid and linoleic acid (3) while the least number (1) of MTA were detected for number of capitula per plant and number of primary branches. MLM analysis could not identify any MTA with  $R^2 \geq 10\%$  for plant height. Among the MTA detected through MLM, the  $R^2$ -values ranged from 10.4 to 34.1%. MTA between oleic/linoleic acid and NGSaf\_309 displayed the highest  $R^2$  (34%).

For the 2012–2013 season, we detected 17 and 13 significant MTAs with  $R^2 \geq 10\%$  for the eight agronomic traits using GLM and MLM analysis, respectively (Table 7). Twelve MTAs were common between both the models for the studied traits. In GLM, the number of significant SSR loci associated with each trait ranged from 1 (for oleic acid) to 4 (for oil content) with  $R^2$ -value ranging from 10.5 to 23.4%. The MTA between oil content and

NGSaf\_15 had the highest  $R^2$ -value of 23.4% (Table 7). In the MLM model, the least number of loci (1) were observed for 100-seed weight and days to 50% flowering while maximum number of loci (3) was found to be linked with oleic acid. These 13 MTAs had  $R^2$ -values ranging from 10.5 to 34.1%. Among these MTAs, we observed the highest  $R^2$ -value of 34% between oleic acid and NGSaf\_309.

A cumulative assessment of MTAs for each trait identified several significant ( $P < 0.05$ ;  $R^2 > 10$ ) associations, which were consistently represented in both models and in both seasons. For example, NGSaf\_15 and NGSaf\_300 for oil content, NGSaf\_67 for oleic acid content, NGSaf\_210 for linoleic acid content, NGSaf\_279 for number of primary branches and number of capitula per plant were detected (Table 7). We also observed MTAs with high correlation values, which were detected either in a majority or only in some environments (models and/or seasons). Many MTAs were also common between traits that showed positive or negative correlation in their phenotypic values (Table 7).

## DISCUSSION

Our earlier work on analysis of a representative global germplasm collection of ~531 safflower accessions identified significant genetic and phenotypic variability in the crop (Kumar et al., 2015, 2016) that could be effectively used to improve traits of agronomic value to enhance crop yield and productivity. Since identification of desirable alleles from large germplasm collections is tedious and time-consuming, core collections encompassing the prevailing genetic and phenotypic diversity of the crop are often developed. In safflower, three studies have reported development of core collections (Johnson et al., 1993; Dwivedi et al., 2005; Kumar et al., 2016). Core collections described by Johnson et al. (1993) and Dwivedi et al. (2005) were based only on morphological and geographical parameters and did not include information on molecular diversity of the crop. They were also comparatively larger with 210 and 570 accessions, respectively. In contrast, two core collections, CartC1 and CartC2, described by Kumar et al. (2016) were constructed based on Maximization (M) strategy using genetic (AFLP), phenotypic and geographical data and were much

smaller (57 and 106 accessions, respectively). Additionally, CartC1 and CartC2 include accessions from eight and ten regional gene pools of safflower (Ashri, 1975), respectively and also have representation from secondary regions of introduction. The present study analyzed the suitability of the above core collections (CartC1 and CartC2) for association mapping through assessment of its genetic diversity and population structure using SSR markers followed by establishment of marker-trait associations for eight agronomically important traits.

## CartAP Collection Is an Effective Panel for Association Mapping

### High Genetic Variation in CartAP Accessions

SSR markers have been extensively used for assessment of association mapping panel and marker-trait associations (Brescaghello and Sorrells, 2006; Agrama et al., 2007; Jun et al., 2008; Yang et al., 2010; Rezaeizad et al., 2011). Being co-dominant markers, they are considered more robust for estimation of population structure and kinship relatedness than dominant markers (Zhu et al., 2008). A large proportion of our tested microsatellite loci (88%) generated moderate to high level of polymorphism, which indicates their efficiency in differentiating *C. tinctorius* L. accessions of CartAP. The mean PIC-value (0.38) obtained in our study using SSR markers on the CartAP collection corroborated well with previous SSR-based studies in safflower (Table 8; Hamdan et al., 2011; Barati and Arzani, 2012; Derakhshan et al., 2014; Lee et al., 2014). The allelic range and average number of alleles per locus observed in the present study were lower than those observed by Chapman et al. (2009; Table 8) probably due to the inclusion of wild species in their study. High genetic variability among CartAP accessions are also illustrated by the diversity indices,  $H = 0.7537$  and  $I = 0.4432$ . This wide genetic diversity is particularly advantageous for association studies as it offers greater allelic diversity with inclusion of rare variants and higher mapping resolution due to representation of historical recombination events from a global population. Thus, the CartAP panel fulfilled the prerequisite of a highly differentiated panel for association mapping (Flint-Garcia et al., 2005).

**TABLE 8 |** Comparative account of SSR marker statistics of the present and previous studies in safflower.

	Present study	(Chapman et al., 2009)	(Hamdan et al., 2011)	(Barati and Arzani, 2012)	(Lee et al., 2014)	(Derakhshan et al., 2014)
Number of accessions sampled	124	27	10	48	100	42
Number of polymorphic SSR markers	93	104	64	42	30	33
Allelic range	2–8	2–15	2–8	2–8	2–7	2–8
Average number of alleles per locus	3.3	6	3.2	3.4	2.8	3.8
Mean expected heterozygosity ( $H_e$ )	0.438	0.54	0.52	0.37	0.386	*
PIC	0.38	0.32	*	0.32	0.325	0.3

\*Not available.

## Weak Population Structure and Low Molecular Relatedness in CartAP Accessions

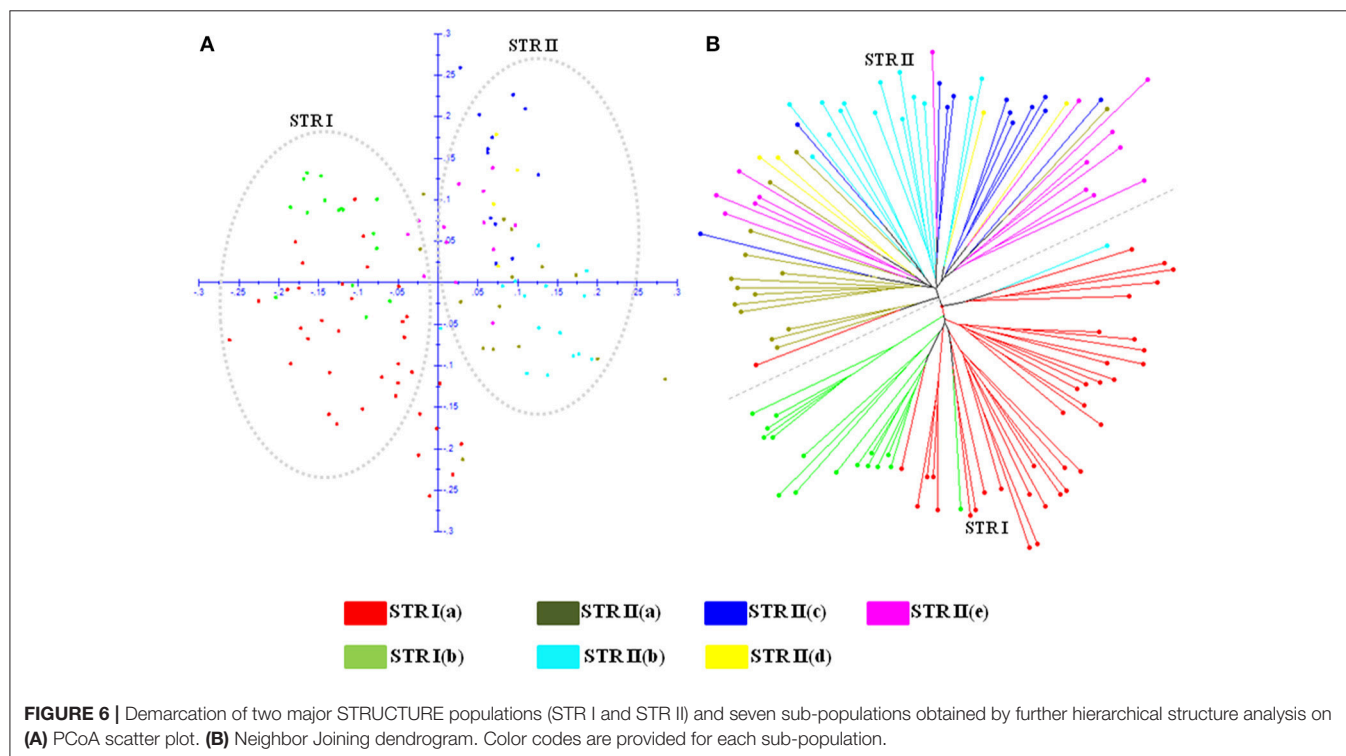
Investigation of the population structure and genetic relatedness between accessions of an association panel is critical to circumvent spurious marker-trait associations (Yu and Buckler, 2006; Zhu et al., 2008; Yang et al., 2010; Nachimuthu et al., 2015). Structure analysis predicts the number of hypothetical sub-populations and identify admixed individuals among the studied panel while clustering analysis display the genetic relationships among accessions. In our study, we observed complete concordance between distance-based NJ and PCoA and Bayesian-based STRUCTURE analysis (Figures 6A,B). NJ clusters, NJ I and NJ II corresponded with STRUCTURE cluster STR I while NJ III coincided with STR II. Neither distance-based NJ analysis nor Bayesian-based STRUCTURE analysis could delimit the accessions on the basis of their geographic origin and/or distribution of studied traits. The lack of geographical structuring among CartAP accessions was expected since *Maximization* strategy was used for development of these core collections (Kumar et al., 2016), which emphasizes on maximization of the allelic diversity (both genetic and phenotypic) with minimum redundancy. For increasing representation of allelic diversity in the core collections, accessions with unique allelic combinations would have been selected and this might have assembled heterogeneous accessions from different regional pools resulting in an unstructured population. Additionally, an extensive number of genetic admixtures were identified with an ancestry share of < 80% in STRUCTURE analysis (Figures 3B,D). The presence of large number of admixed individuals can

be attributed to incomplete lineage sorting during safflower diversification.

The range of *Fst*-value (0.0799–0.2888) suggests low to moderate differentiation between the seven sub-populations derived from hierarchical population structure analysis (Figure 4). At the population level, mean kinship coefficient (*Gij*) values (0–0.075) indicate low genetic similarity between these sub populations of the CartAP collection (Figure 5B). Additionally, high proportion of admixed individuals and the low pairwise kinship or co-ancestry values (*Fij*) at the individual level reflects no or weak relatedness between the safflower accessions (Figure 5A). The weak population structure coupled with low molecular relatedness between the sampled accessions of CartAP would prevent spurious marker-trait associations and confirms suitability of the CartAP panel for association mapping.

## Association Mapping Identifies Significant Marker-Trait Associations

Identification of loci influencing agronomic traits facilitates marker-assisted breeding to increase crop productivity. In safflower, efforts to identify molecular markers linked to traits of agronomic value are limited (Hamdan et al., 2008, 2012; Mayerhofer et al., 2010; García-Moreno et al., 2011; Pearl et al., 2014; Ebrahimi et al., 2017). In the present study, we performed association mapping and identified markers associated with eight traits of agronomic importance (seed oil content, oleic acid content, linoleic acid content, 100-seed weight, plant height, number of primary branches, days to 50% flowering and number of capitula per plant), many of which have been reported to influence crop yield (Patil, 1998). Of these, several





marker-traits associations were found to be stable over the two growing seasons. Since MTAs are influenced by variations in environmental conditions, the strength of these associations needs to be analyzed through multi-location trials.

Low seed oil content is a serious impediment to adoption of safflower as a major oilseed crop globally. Oil content is a quantitative trait controlled by several genomic loci imparting small to moderate genetic effects, which are also influenced by environmental conditions (Hwang et al., 2014). Until now, no attempts have been made to map QTLs for oil content in safflower. Seed oil content in the CartAP accessions ranges from 16 to 50% (Table 6) indicating that the association panel includes diverse accessions which can be used in breeding programs. Our association study identified 11 MTAs correlated with oil content and supported by high correlation coefficient ( $R^2 \geq 10\%$ ; Table 7). Among these, two SSR loci, NGSaf\_15 ( $R^2 = 10.4\text{--}23.4\%$ ) and NGSaf\_300 ( $R^2 = 13.8\text{--}17.9\%$ ) were strongly associated with oil content in both GLM and MLM models and in both seasons (Table 7) and can be used to increase oil content of safflower cultivars.

The fatty acid composition of edible oil determines its utility and market value. Ever since the importance of oleic acid-rich dietary fats was elucidated, the demand for edible oil with high oleic acid content has increased substantially. Safflower is unique among oilseed crops due to the wide range of seed oil compositions available in its global germplasm (Velasco et al., 2005). Two major oil types are found in safflower—one with high linoleic acid and the other rich in oleic acid (Fernandez-Martinez et al., 1993). In addition, several lines with similar levels of linoleic and oleic acid have also been identified (Fernandez-Martinez et al., 1993). Modifying the fatty acid composition of safflower oil is particularly significant from the Indian perspective since majority of Indian cultivars are high in linoleic acid in spite of availability of high oleic lines in the Indian germplasm. Only one major QTL associated with oleic acid has been identified and it mapped on linkage group 3 using a cross between CL-1 (oleic acid: 14–22%) and CL-9 (oleic acid >84%; Hamdan et al., 2012). In addition to the major QTL associated with oleic acid, several modifying genes, which either enhance or repress the expression of major QTLs have been postulated to determine oleic acid content in safflower (Hamdan et al., 2008, 2012). In CartAP accessions, the oleic acid content ranged from 10 to 79% while linoleic acid varied from 13 to 87%. By association mapping, we identified three SSR loci (NGSaf\_67, NGSaf\_210 and NGSaf\_309) linked with oleic acid content in safflower (Table 7). Among these loci, NGSaf\_67 ( $R^2 = 11.4\text{--}11.6\%$ ) was consistently associated with oleic acid content in both seasons and in both models while NGSaf\_210 (displaying a phenotypic variation of ~14%) and NGSaf\_309 (with the highest  $R^2$ -value (~34%) among all the recorded marker-trait associations) were identified to be associated with oleic acid through MLM approach in both the seasons. We identified 10 significant marker-trait associations for linoleic acid content with  $R^2 \geq 10\%$ . The NGSaf\_210 locus ( $R^2 = 14.3\text{--}17.9\%$ ) was associated with linoleic acid in both the models and seasons. In addition, NGSaf\_67 ( $R^2 = 10.9\text{--}16.7\%$ ) and NGSaf\_309 ( $R^2 = 34\%$ ) loci detected for oleic acid were also linked with linoleic acid content. Linoleic acid is synthesized

from oleic acid by activity of fatty acid desaturase 2 enzyme encoded by the *FAD2* gene. A strong negative correlation has been reported between oleic acid and linoleic acid content in safflower (Kumar et al., 2016). Identification of common SSR loci (NGSaf\_67, NGSaf\_210 and NGSaf\_309) for both these fatty acids indicates their strong linkage with the corresponding gene/s of the fatty acid biosynthetic pathway. These MTAs can serve as a useful resource for achieving the desired oil composition in the crop.

Plant height determines plant architecture and also influences crop yield. It is quantitatively inherited and a large number of QTLs associated with plant height have been reported in different crop systems (Wu et al., 2010; Morris et al., 2013; Zanke et al., 2014). In previous studies, safflower germplasm have revealed significant variations for plant height (Yeilaghi et al., 2015; Kumar et al., 2016). However, genetic loci influencing this trait have not yet been identified. The CartAP association panel harbors wide variation for plant height ranging from 94 to 226 cm. Association analysis identified 11 marker-trait associations with a  $R^2 \geq 10\%$ . Two loci, NGSaf\_156 ( $R^2$ -value from 11.8 to 13.9%) and NGSaf\_296 ( $R^2$ -value from 12.7 to 14.7%) were associated with plant height in all environments except in the MLM model of 2011–2012 season. Plant height and flowering time are correlated traits as the onset of reproductive phase marks the termination of apical growth in safflower. Early maturing cultivars are desirable and several accessions with early flowering time were identified in CartAP (Kumar et al., 2016). Association analysis revealed that days to 50% flowering correlated significantly with 8 marker-trait associations which had  $R^2 \geq 10\%$ . The NGSaf\_92 locus ( $R^2 = 7.9\text{--}11.6\%$ ) was associated with days to 50% flowering for all models and seasons except GLM of 2011–2012 season. Interestingly, we detected two markers (NGSaf\_101 and NGSaf\_296), which were associated with both plant height and days to 50% flowering in either one of the models or seasons (Table 7). Plant height and flowering time are traits that are highly influenced by environmental effects. Therefore, seasonal variations between the two growing periods could have led to differences in detection of these MTAs in the studied seasons and models. Nevertheless, these common MTAs provide an initial platform to study the genetic relationships between the two traits.

The number of primary branches also defines plant architecture and influences yield since primary branches produce secondary to quaternary branches, each of which terminates into the characteristic globular head of safflower. Association analysis for number of primary branches detected 6 marker-trait associations, which showed correlation coefficient ( $R^2$ -value)  $\geq 10\%$ . The locus NGSaf\_279 ( $R^2$ -value ranging from 8 to 10.8%) was associated with primary branches using both the models in both growing seasons. The number of capitula in accessions of the CartAP collection ranged from 16 to 203 (Kumar et al., 2016). The number of capitula per plant exhibited 2 significant associations which had  $R^2 \geq 10\%$ . The NGSaf\_279 locus ( $R^2 = 7.4\text{--}9.3\%$ ) was common among all models and for both seasons. A positive correlation was reported by Kumar et al. (2016) between the number of capitula per plant and number of primary branches, which is in congruence with our results on

identification of the SSR locus NGSaf\_279, as a common marker for both the traits.

A mature safflower achene consists of 33–60% hull and 40–67% kernel (Dajue and Mündel, 1996). It has been suggested that reducing hull content can significantly enhance seed oil content in safflower (Dajue and Mündel, 1996). Association analysis of 100-seed weight among CartAP accessions identified 7 significant marker-trait associations with  $R^2 \geq 10\%$ . With the exception of season 2012–2013 MLM approach, two loci NGSaf\_306 ( $R^2 = 13\text{--}24.5\%$ ) and NGSaf\_309 ( $R^2 = 15.5\text{--}24.1\%$ ), were consistently associated with 100-seed weight (Table 7). These loci may play an important role in deciphering the molecular relationship between hull thickness and oil content.

Although safflower is an important oilseed crop with global distribution, there have been limited attempts to identify marker-trait associations for crop improvement. The present study established the suitability of CartAP collection for association mapping due to the absence of population structure and presence of significant genetic and phenotypic diversity. Using association mapping we were able to identify significant marker-trait associations for eight important agronomic traits. Many of these associations were consistently detected over two growing seasons and multiple models. The stability and utility of these marker-trait associations need to be analyzed further under additional environments through multi-location trials. The potential marker-trait associations identified in this study

will be useful in facilitating marker-assisted breeding for crop improvement and identification of candidate genes for trait variability in safflower.

## AUTHOR CONTRIBUTIONS

SG and AJ conceived and designed the experiments; HA and SK conducted the experiments, collected and analyzed the data; HA, SK, AJ, and SG wrote the manuscript; MA, AK, AJ, and SG provided facilities for completion of experiments and reviewed the manuscript.

## ACKNOWLEDGMENTS

This work was supported by the DST-PURSE grant of Department of Science and Technology, Government of India [grant no. Dean(R)/2009/868] provided to the University of Delhi. HA was supported by a research fellowship from University Grants Commission, India. We are thankful for the support of our technical and field staff in completion of the work.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00402/full#supplementary-material>

## REFERENCES

- Abdurakhmonov, I. Y., and Abdurakimov, A. (2008). Application of association mapping to understanding the genetic diversity of plant germplasm resources. *Int. J. Plant Genomics* 2008:574927. doi: 10.1155/2008/574927
- Agrama, H. A., Eizenga, G. C., and Yan, W. (2007). Association mapping of yield and its components in rice cultivars. *Mol. Breed.* 19, 341–356. doi: 10.1007/s11032-006-9066-6
- Ambreen, H., Kumar, S., Variath, M. T., Joshi, G., Bali, S., Agarwal, M., et al. (2015). Development of genomic microsatellite markers in *Carthamus tinctorius* L. (Safflower) using next generation sequencing and assessment of their cross-species transferability and utility for diversity analysis. *PLoS ONE* 10:e0135443. doi: 10.1371/journal.pone.0135443
- Amini, F., Saeidi, G., and Arzani, A. (2008). Study of genetic diversity in safflower genotypes using agro-morphological traits and RAPD markers. *Euphytica* 163, 21–30. doi: 10.1007/s10681-007-9556-6
- Ashri, A. (1975). Evaluation of the germplasm collection of safflower, *Carthamus tinctorius* L. V. Distribution and regional divergence for morphological characters. *Euphytica* 24, 651–659.
- Bahrami, F., Arzani, A., and Karimi, V. (2014). Evaluation of yield-based drought tolerance indices for screening safflower genotypes. *Agron. J.* 106, 1219–1224. doi: 10.2134/agnonj13.0387
- Barati, M., and Arzani, A. (2012). Genetic diversity revealed by EST-SSR markers in cultivated and wild safflower. *Biochem. Syst. Ecol.* 44, 117–123. doi: 10.1016/j.bse.2012.04.013
- Blair, M. W., Díaz, L. M., Buendía, H. F., and Duque, M. C. (2009). Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 119, 955–972. doi: 10.1007/s00122-009-1064-8
- Bowers, J. E., Pearl, S. A., and Burke, J. M. (2016). Genetic mapping of millions of SNPs in safflower (*Carthamus tinctorius* L.) via whole-genome resequencing. *Genes Genom. Genet.* 6, 2203–2211. doi: 10.1534/g3.115.026690
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Breseghello, F., and Sorrells, M. E. (2006). Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Sci.* 46, 1323–1330. doi: 10.2135/cropsci2005.09-0305
- Carlsson, A. S., Zhu, L. H., Andersson, M., and Hofvander, P. (2014). Platform crops amenable to genetic engineering—a requirement for successful production of bio-industrial oils through genetic engineering. *Biocatal. Agric. Biotechnol.* 3, 58–64. doi: 10.1016/j.bcab.2013.12.007
- Chapman, M. A., Hvala, J., Strever, J., and Burke, J. M. (2010). Population genetic analysis of safflower (*Carthamus tinctorius* L.; Asteraceae) reveals a near Eastern origin and five centers of diversity. *Am. J. Bot.* 97, 831–840. doi: 10.3732/ajb.0900137
- Chapman, M. A., Hvala, J., Strever, J., Matvienko, M., Kozik, A., Micheltore, R. W., et al. (2009). Development, polymorphism, and cross-taxon utility of EST-SSR markers from safflower (*Carthamus tinctorius* L.). *Theor. Appl. Genet.* 120, 85–91. doi: 10.1007/s00122-009-1161-8
- Dajue, L., and Mündel, H. H. (1996). *Safflower Carthamus tinctorius L. Promoting the Conservation and Use of Underutilized and Neglected Crops*. 7. Institute of Plant Genetics and Crop Plant Research, Gatersleben. International Plant Genetic Resources Institute, Rome.
- Derakhshan, E., Majidi, M. M., Sharafi, Y., and Mirlohi, A. (2014). Discrimination and genetic diversity of cultivated and wild safflowers (*Carthamus* spp.) using EST-microsatellites markers. *Biochem. Syst. Ecol.* 54, 130–136. doi: 10.1016/j.bse.2014.01.003
- Dwivedi, S. L., Upadhyaya, H. D., and Hegde, D. M. (2005). Development of core collection using geographic information and morphological descriptors in safflower (*Carthamus tinctorius* L.) germplasm. *Genet. Resour. Crop Evol.* 52, 821–830. doi: 10.1007/s10722-003-6111-8

- Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conser. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- Ebrahimi, F., Majidi, M. M., Arzani, A., and Mohammadi-Nejad, G. (2016). Oil and seed yield stability in a worldwide collection of safflower under arid environments of Iran. *Euphytica* 212, 131–144. doi: 10.1007/s10681-016-1779-y
- Ebrahimi, F., Majidi, M. M., Arzani, A., and Mohammadi-Nejad, G. (2017). Association analysis of molecular markers with traits under drought stress in safflower. *Crop Pasture Sci.* 68, 167–175. doi: 10.1071/cp16252
- Fernandez-Martinez, J., Del Rio, M., and De Haro, A. (1993). Survey of safflower (*Carthamus tinctorius* L.) germplasm for variants in fatty acid composition and other seed characters. *Euphytica* 69, 115–122.
- Flider, F. J. (2013). Development and commercialization of GLA safflower oil. *Lipid Tech.* 25, 227–229. doi: 10.1002/lite.201300302
- Flint-Garcia, S. A., Thuillet, A. C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., et al. (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J.* 44, 1054–1064. doi: 10.1111/j.1365-3113X.2005.02591.x
- García-Moreno, M. J., Fernández-Martínez, J. M., Velasco, L., and Pérez-Vich, B. (2011). Molecular tagging and candidate gene analysis of the high gamma-tocopherol trait in safflower (*Carthamus tinctorius* L.). *Mol. Breed.* 28, 367–379. doi: 10.1007/s11032-010-9489-y
- Hamdan, Y. A., García-Moreno, M. J., Fernández-Martínez, J. M., Velasco, L., and Pérez-Vich, B. (2012). Mapping of major and modifying genes for high oleic acid content in safflower. *Mol. Breed.* 30, 1279–1293. doi: 10.1007/s11032-012-9714-y
- Hamdan, Y. A., Velasco, L., and Pérez-Vich, B. (2008). Development of SCAR markers linked to male sterility and very high linoleic acid content in safflower. *Mol. Breed.* 22, 385–393. doi: 10.1007/s11032-008-9183-5
- Hamdan, Y., García-Moreno, M. J., Redondo-Navado, J., Velasco, L., and Pérez-Vich, B. (2011). Development and characterization of genomic microsatellite markers in safflower (*Carthamus tinctorius* L.). *Plant Breed.* 130, 237–241. doi: 10.1111/j.1439-0523.2010.01826.x
- Hardy, O. J., and Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Resour.* 2, 618–620. doi: 10.1046/j.1471-8286.2002.00305.x
- Hwang, E. Y., Song, Q., Jia, G., Specht, J. E., Hyten, D. L., Costa, J., et al. (2014). A genome-wide association study of seed protein and oil content in soybean. *BMC Genom.* 15:1. doi: 10.1186/1471-2164-15-1
- Johnson, R. C., Kisha, T. J., and Evans, M. A. (2007). Characterizing safflower germplasm with AFLP molecular markers. *Crop Sci.* 47, 1728–1736. doi: 10.2135/cropsci2006.12.0757
- Johnson, R. C., Stout, D. M., and Bradley, V. L. (1993). “The US collection: a rich source of safflower germplasm”, in *Proceedings of the Third International Safflower Conference*, eds L. Dajue and H. Yuanzhou (Beijing: Beijing Botanical Garden; Institute of Botany, Chinese Academy of Sciences), 202–208.
- Jun, T. H., Van, K., Kim, M. Y., Lee, S. H., and Walker, D. R. (2008). Association analysis using SSR markers to find QTL for seed protein content in soybean. *Euphytica* 162, 179–191. doi: 10.1007/s10681-007-9491-6
- Kalinowski, S. T. (2004). Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conserv. Genet.* 5, 539–543. doi: 10.1023/B:COGE.0000041021.91777.1a
- Kalinowski, S. T. (2005). HP-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol. Ecol. Notes* 5, 187–189. doi: 10.1111/j.1471-8286.2004.00845.x
- Kaya, M. D., Bayramin, S., Kaya, G., and Uzun, O. (2011). Seed vigor and ion toxicity in safflower (*Carthamus tinctorius* L.) seedlings produced by various seed sizes under NaCl stress. *Arch. Biol. Sci.* 63, 723–729. doi: 10.2298/ABS1103723K
- Khan, M. A., von Witzke-Ehbrecht, S., Maass, B. L., and Becker, H. C. (2009). Relationships among different geographical groups, agromorphology, fatty acid composition and RAPD marker diversity in safflower (*Carthamus tinctorius* L.). *Genet. Resour. Crop Evol.* 56, 19–30. doi: 10.1007/s10722-008-9338-6
- Kumar, S., Ambreen, H., Murali, T. V., Bali, S., Agarwal, M., Kumar, A., et al. (2015). Assessment of genetic diversity and population structure in a global reference collection of 531 accessions of *Carthamus tinctorius* L. (Safflower) using AFLP markers. *Plant Mol. Biol. Rep.* 33, 1299–1313. doi: 10.1007/s11105-014-0828-8
- Kumar, S., Ambreen, H., Variath, M. T., Rao, A. R., Agarwal, M., Kumar, A., et al. (2016). Utilization of molecular, phenotypic, and geographical diversity to develop compact composite core collection in the oilseed crop, safflower (*Carthamus tinctorius* L.) through maximization strategy. *Front. Plant Sci.* 7:1554. doi: 10.3389/fpls.2016.01554
- Lee, G. A., Sung, J. S., Lee, S. Y., Chung, J. W., Yi, J. Y., Kim, Y. G., et al. (2014). Genetic assessment of safflower (*Carthamus tinctorius* L.) collection with microsatellite markers acquired via pyrosequencing method. *Mol. Ecol. Resour.* 14, 69–78. doi: 10.1111/1755-0998.12146
- Li, Z. M., Ding, J. Q., Wang, R. X., Chen, J. F., Sun, X. D., Chen, W., et al. (2011). A new QTL for resistance to *Fusarium* ear rot in maize. *J. Appl. Genet.* 52, 403–406. doi: 10.1007/s13353-011-0054-0
- Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282
- Loisele, B. A., Sork, V. L., Nason, J., and Graham, C. (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am. J. Bot.* 1, 1420–1425. doi: 10.1002/j.1537-2197.1995.tb12679.x
- Mayerhofer, R., Archibald, C., Bowles, V., and Good, A. G. (2010). Development of molecular markers and linkage maps for the *Carthamus* species, *C. tinctorius* and *C. oxyacanthus*. *Genome* 53, 266–276. doi: 10.1139/g10-002
- Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., et al. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. U.S.A.* 110, 453–458. doi: 10.1073/pnas.1215985110
- Nachimuthu, V. V., Muthurajan, R., Duraialaguraja, S., Sivakami, R., Pandian, B. A., Ponniah, G., et al. (2015). Analysis of population structure and genetic diversity in rice germplasm using SSR markers: an initiative towards association mapping of agronomic traits in *Oryza Sativa*. *Rice* 8:30. doi: 10.1186/s12284-015-0062-5
- Nimbkar, N. (2008). “Issues in safflower production in India”, in *Safflower: Unexploited Potential and World Adaptability, Proceedings of the Seventh International Safflower Conference*, eds S.E. Knights and T.D. Potter (Wagga Wagga, NSW).
- Oraguzie, N. C., and Wilcox, P. L. (2007). “An overview of association mapping”, in *Association Mapping in Plants*, eds N. C. Oraguzie, E. H. A. Rikkerink, and S. E. Gardiner (New York, NY: Springer), 1–9.
- Patil, H. S. (1998). Genetic variability, association and path analysis in safflower. *Indian. J. Agric. Res.* 32, 46–50.
- Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460
- Pearl, S. A., Bowers, J. E., Reyes-Chin-Wo, S., Michelmores, R. W., and Burke, J. M. (2014). Genetic analysis of safflower domestication. *BMC Plant Biol.* 14:43. doi: 10.1186/1471-2229-14-43
- Pearl, S. A., and Burke, J. M. (2014). Genetic diversity in *Carthamus tinctorius* (Asteraceae: safflower), an underutilized oilseed crop. *Am. J. Bot.* 101, 1640–1650. doi: 10.3732/ajb.1400079
- Perrier, X., and Jacquemoud-collet, J. P. (2006). *DARwin Software, version 5.0.158*. Montpellier: Department Systems Biologiques (BIOS), CIRAD.
- Perry, D. J. (2004). Identification of Canadian durum wheat varieties using a single PCR. *Theor. Appl. Genet.* 109, 55–61. doi: 10.1007/s00122-004-1597-9
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Rezaei, Z., Wittkop, B., Snowdon, R., Hasan, M., Mohammadi, V., Zali, A., et al. (2011). Identification of QTLs for phenolic compounds in oilseed rape (*Brassica napus* L.) by association mapping using SSR markers. *Euphytica* 177, 335–342. doi: 10.1007/s10681-010-0231-y
- Sehgal, D., Rajpal, V. R., Raina, S. N., Sasanuma, T., and Sasakuma, T. (2009). Assaying polymorphism at DNA level for genetic diversity diagnostics of the safflower (*Carthamus tinctorius* L.) world germplasm resources. *Genetica* 135, 457–470. doi: 10.1007/s10709-008-9292-4
- Soto-Cerda, B. J., Duguid, S., Booker, H., Rowland, G., Diederichsen, A., and Cloutier, S. (2014). Genomic regions underlying agronomic traits in linseed (*Linum usitatissimum* L.) as revealed by association mapping. *J. Integr. Plant Biol.* 56, 75–87. doi: 10.1111/jipb.12118

- Upadhyaya, H. D., Wang, Y. H., Gowda, C. L. L., and Sharma, S. (2013). Association mapping of maturity and plant height using SNP markers with the sorghum mini core collection. *Theor. Appl. Genet.* 126, 2003–2015. doi: 10.1007/s00122-013-2113-x
- Upadhyaya, H. D., Wang, Y. H., Sharma, S., Singh, S., and Hasenstein, K. H. (2012). SSR markers linked to kernel weight and tiller number in sorghum identified by association mapping. *Euphytica* 187, 401–410. doi: 10.1007/s10681-012-0726-9
- Velasco, L., Pérez-Vich, B., Hamdan, Y., and Fernández-Martínez, J. M. (2005). “Genetic study of several seed oil quality traits in safflower,” in *Safflower: A Unique Crop for Oil Spices and Health Consequently, a Better Life for you, Proceedings of the Sixth International Safflower Conference*, eds E. Esendal, J. W. Bergman, N. Kandemir, R. C. Johnson, and A. Corleto (Istanbul), 2–79.
- Wu, X., Wang, Z., Chang, X., and Jing, R. (2010). Genetic dissection of the developmental behaviours of plant height in wheat under diverse water regimes. *J. Exp. Bot.* 61, 2923–2937. doi: 10.1093/jxb/erq117
- Xu, J., Ranc, N., Muñoz, S., Rolland, S., Bouchet, J. P., Desplat, N., et al. (2013). Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species. *Theor. Appl. Genet.* 126, 567–581. doi: 10.1007/s00122-012-2002-8
- Yan, J., Warburton, M., and Crouch, J. (2011). Association mapping for enhancing maize (L.) genetic improvement. *Crop Sci.* 51, 433–449. doi: 10.2135/cropsci2010.04.0233
- Yang, X., Yan, J., Shah, T., Warburton, M. L., Li, Q., Li, L., et al. (2010). Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. *Theor. Appl. Genet.* 121, 417–431. doi: 10.1007/s00122-010-1320-y
- Yang, Y. X., Wu, W., Zheng, Y. L., Chen, L., Liu, R. J., and Huang, C. Y. (2007). Genetic diversity and relationships among safflower (*Carthamus tinctorius* L.) analyzed by inter-simple sequence repeats (ISSRs). *Genet. Resour. Crop Evol.* 54, 1043–1051. doi: 10.1007/s10722-006-9192-3
- Yeh, F. C., Yang, R. C., Boyle, T., Ye, Z. H., and Mao, J. X. (1999). *POPGENE, Version 1.32: The User Friendly Software for Population Genetic Analysis*. Edmonton, AB: Molecular Biology and Biotechnology Centre, University of Alberta.
- Yeilaghi, H., Arzani, A., and Ghaderian, M. (2015). Evaluating the contribution of ionic and agronomic components toward salinity tolerance in safflower. *Agron. J.* 107, 2205–2212. doi: 10.2134/agronj15.0202
- Yu, J., and Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17, 155–160. doi: 10.1016/j.copbio.2006.02.003
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702
- Zanke, C. D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., et al. (2014). Whole genome association mapping of plant height in winter wheat (*Triticum aestivum* L.). *PLoS ONE* 9:e113287. doi: 10.1371/journal.pone.0113287
- Zhang, P., Liu, X., Tong, H., Lu, Y., and Li, J. (2014). Association mapping for important agronomic traits in core collection of rice (*Oryza sativa* L.) with SSR markers. *PLoS ONE* 9:e111508. doi: 10.1371/journal.pone.0111508
- Zhou, X., Tang, L., Xu, Y., Zhou, G., and Wang, Z. (2014). Towards a better understanding of medicinal uses of *Carthamus tinctorius* L. In traditional Chinese medicine: a phytochemical and pharmacological review. *J. Ethnopharmacol.* 151, 27–43. doi: 10.1016/j.jep.2013.10.050
- Zhu, C., Gore, M., Buckler, E. S., and Yu, J. (2008). Status and prospects of association mapping in plants. *Plant Genome* 1, 5–20. doi: 10.3835/plantgenome2008.02.0089

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Ambreen, Kumar, Kumar, Agarwal, Jagannath and Goel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Genetics, Host Range, and Molecular and Pathogenic Characterization of *Verticillium dahliae* From Sunflower Reveal Two Differentiated Groups in Europe

Alberto Martín-Sanz<sup>1</sup>, Sandra Rueda<sup>1</sup>, Ana B. García-Carneros<sup>2</sup>, Sara González-Fernández<sup>2</sup>, Pedro Miranda-Fuentes<sup>2</sup>, Sandra Castuera-Santacruz<sup>2</sup> and Leire Molinero-Ruiz<sup>2\*</sup>

<sup>1</sup> Pioneer Hi-Bred International, Inc., La Rinconada, Spain, <sup>2</sup> Department of Crop Protection, Institute for Sustainable Agriculture, Spanish National Research Council, Córdoba, Spain

## OPEN ACCESS

### Edited by:

Thomas Miedaner,  
University of Hohenheim, Germany

### Reviewed by:

Thomas John Gulya,  
Agricultural Research Service (USDA),  
United States  
Matthew Denton-Giles,  
Curtin University, Australia

### \*Correspondence:

Leire Molinero-Ruiz  
leire.molinero@csic.es

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 05 December 2017

**Accepted:** 19 February 2018

**Published:** 09 March 2018

### Citation:

Martín-Sanz A, Rueda S, García-Carneros AB, González-Fernández S, Miranda-Fuentes P, Castuera-Santacruz S and Molinero-Ruiz L (2018) Genetics, Host Range, and Molecular and Pathogenic Characterization of *Verticillium dahliae* From Sunflower Reveal Two Differentiated Groups in Europe. *Front. Plant Sci.* 9:288. doi: 10.3389/fpls.2018.00288

Verticillium wilt and leaf mottle of sunflower, caused by the fungus *Verticillium dahliae* (Vd) has become a major constraint to sunflower oil production in temperate European countries. Information about Vd from sunflower is very scarce despite genetics, molecular traits and pathogenic abilities of fungal strains affecting many other crops being widely known. Understanding and characterizing the diversity of Vd populations in those countries where sunflowers are frequent and severely affected by the fungus are essential for efficient breeding for resistance. In this study, we have analyzed genetic, molecular and pathogenic traits of Vd isolates affecting sunflower in European countries. When their genetics was investigated, almost all the isolates from France, Italy, Spain, Argentina, and Ukraine were assigned to vegetative compatibility group (VCG) 2B. In Bulgaria, Turkey, Romania, and Ukraine, some isolates were assigned to VCG6, but some others could not be assigned to any VCG. Genotyping markers used for Vd affecting crops other than sunflower showed that all the isolates were molecularly identified as race 2 and that markers of defoliating (D) and non-defoliating (ND) pathotypes distinguished two well-differentiated clusters, one (E) grouping those isolates from Eastern Europe and the other (W) all those from the Western Europe and Argentina. All the isolates in cluster W were VCG2B, while the isolates in cluster E belonged to an unknown VCG or to VCG6. When the host range was investigated in the greenhouse, the fungus was highly pathogenic to artichoke, showing the importance of farming alternatives in the management of Verticillium attacks. Sunflower genotypes were inoculated with a selection of isolates in two experiments. Two groups were identified, one including the isolates from Western Europe, Argentina, and Ukraine, and the other including isolates from Bulgaria, Romania, and Turkey. Three pathogenic races were differentiated: V1, V2-EE (Eastern Europe) and V2-WE (Western Europe). Similarly, three differentials are proposed for race identification: HA 458 (universal susceptible),

HA 89 (resistant to V2-EE, susceptible to V2-WE) and INRA2603 (susceptible to V2-EE, resistant to V2-WE). The diversity found in *Vd* affecting sunflower must be taken into account in the search for resistance to the pathogen for European environments of sunflower production.

**Keywords:** control strategies, crop rotation, genetic resistance, molecular markers, races of *V. dahliae*, pathotypes of *V. dahliae*, soilborne fungus

## INTRODUCTION

Verticillium wilt and leaf mottle (VWLM), caused by the fungus *Verticillium dahliae* Kleb. (*Vd*), has traditionally been a major disease of sunflower in Argentina and the United States (Gulya et al., 1997; Radi and Gulya, 2007; Galella et al., 2012) as well as in temperate European countries (Harveson and Markell, 2016). However, disease incidence in France, Italy, Spain, and countries around the Black Sea has dramatically increased in recent years and in some regions, like southern France, it is becoming a major constraint to sunflower oil production (Harveson and Markell, 2016; Debaeke et al., 2017). *Verticillium dahliae* is a soilborne ascomycete with a wide range of host crops. Besides sunflower, it causes important yield losses in artichoke (*Cynara cardunculus* L. var. *scolymus*), cauliflower (*Brassica oleracea* var. *botrytis* L.), cotton (*Gossypium hirsutum* L.), eggplant (*Solanum melongena* L.), lettuce (*Lactuca sativa* L.), olive tree (*Olea europaea* L.), potato (*Solanum tuberosum* L.), tobacco (*Nicotiana tabacum* L.), and tomato (*Solanum lycopersicum* L.), among others (Pegg and Brady, 2002). In Spain, *Verticillium* constitutes an important constraint for the production of cotton, artichoke, and, particularly, of olive tree (Bejarano-Alcázar et al., 1996; Berbegal et al., 2010; Jiménez-Díaz et al., 2011; López-Escudero and Mercado-Blanco, 2011). Also in Spain, VWLM outbreaks have repeatedly been observed in the last few years in sunflower fields of Cadiz province (García-Ruiz et al., 2014), where it is grown in alternation with other crops, particularly cotton and/or tomato. Some fields have in fact even been turned into olive tree groves. Host specialization occurs in *Vd*, meaning that isolates from a given host may be pathogenic on other hosts but, generally, they are more virulent (symptoms are more severe) on the hosts from which they were obtained. In some *Vd* isolates, host specialization is more pronounced (Bhat and Subbarao, 1999; Douhan and Johnson, 2001). In areas where sunflower is grown in alternation with other crop species, determination of the host range specificity of *Vd* affecting sunflowers is important for the correct management of the whole cropping system.

Clonality in *Vd* is described by means of vegetative compatibility, which refers to the genetically controlled ability of individual fungal strains to undergo hyphal anastomosis and form stable heterokaryons. Vegetatively compatible isolates are placed in the same vegetative compatibility group (VCG). In spite of there being many studies on the genetic diversity of *Vd* from artichoke (Jiménez-Díaz et al., 2006), cotton (Dervis et al., 2008; Korolev et al., 2008), eggplant (Dervis et al., 2009b), olive tree (Navas-Cortés et al., 2009; Dervis et al., 2010), potato and mint (Dung et al., 2013) or sugarbeet (Strausbaugh et al., 2016) among other crops, the

genetic characterization of *Vd* isolates from sunflower has been scarcely addressed. Isolates of *Vd* affecting sunflowers in Canada showed weak reactions with testers from VCG4A and 4B; one isolate was identified as VCG3 and another one was compatible with all VCG groups except VCG2A (El-Bebany et al., 2013). In previous works by our research group, isolates from Argentina and Spain were ascribed to VCG2B (García-Carneros et al., 2014). Since *Vd* reproduces asexually, isolates in the same VCG could be genetically distinct populations with a similarity in a number of physiological, ecological, pathogenic, and host-range traits (Jiménez-Díaz et al., 2006; Collado-Romero et al., 2008; Korolev et al., 2009). Thus, genetic diversity of *Vd* isolates from sunflower could be intimately associated with disease occurrence and severity as a consequence of particular interactions *Vd* isolate – sunflower genotype.

In some crops, such as cotton or olive tree, isolates of *Vd* causing infection are pathogenically characterized by assignment to defoliating (D) or non-defoliating (ND) pathotypes, which are identified on the basis of their capacity to cause, or not, the complete fall of green leaves (Rodríguez-Jurado et al., 1993; Bejarano-Alcázar et al., 1996). Molecular analyses using SCAR markers differentiated a genetically homogeneous group of D isolates belonging to VCG1A (Mercado-Blanco et al., 2002). In contrast, these markers showed a high molecular diversity of ND pathotypes belonging to 2A, 2B, and 4B VCGs (Pérez-Artés et al., 2000; Mercado-Blanco et al., 2001, 2002, 2003). Preliminary results from our research group showed that the molecular pattern of *Vd* isolates infecting sunflower in Argentina and Spain matched that of the ND pathotype of artichoke and/or cotton, pointing to the closeness between ND isolates affecting these three crops and, therefore, suggesting that any of them could serve as a carrier and source of inoculum for *Verticillium* outbreaks (García-Carneros et al., 2014).

While isolates of *Vd* infecting crops like cotton and olive tree are assigned to D or ND pathotypes, races of *Vd* pathogenic to tomato, lettuce, and sunflower are distinguished depending on the genes of resistance that they overcome. Race 1 and race 2 have been described for isolates of *Vd* pathogenic to tomato (Alexander, 1962) and lettuce (Vallad et al., 2006; Hayes et al., 2007). Moreover, the races of *Vd* on tomato and lettuce have been reported as being correlated (Maruthachalam et al., 2010). Sequence similarity between the resistance gene of lettuce (*Vr1*) and that of tomato (*Ve1*) suggests that they share similar race 1-specific genes for resistance to *Vd* (Hayes et al., 2011). Also, race 1 is characterized by the presence of the effector gene *Ave1*, conferring avirulence to lettuce or tomato that carry the resistance genes *Vr1* or *Ve1*, respectively

(de Jonge et al., 2012). Conversely, race 2 *Vd* isolates lack *Ave1* and they are, therefore, potentially virulent on plants carrying resistance to race 1 (de Jonge et al., 2012; Short et al., 2014). Race 1 seems to have arisen once by horizontal gene transfer and is genetically much less diverse than race 2 (de Jonge et al., 2012; Jiménez-Díaz et al., 2017). Race 2 occurs worldwide and it causes disease on cultivars from a range of crops for which effective resistance has not been reported (Maruthachalam et al., 2010; Short et al., 2014; Sandoya et al., 2017). Regarding *Vd* on sunflower, the first race (NA-1) was detected in the United States, and it was controlled by the resistance in HA 89, which was associated to a single major gene (Fick and Zimmer, 1974; Gulya et al., 1997). New races overcoming this resistance, and apparently different to each other, have been reported later on the basis of phenotypic characterization: one (NA-Vd2) in the United States (Gulya, 2007), four in Argentina (Bertero de Romano and Vázquez, 1982; Galella et al., 2004; Clemente et al., 2017) and one in Spain (García-Ruiz et al., 2014). No comparative studies of these proposed races of *Vd* affecting sunflower have so far been conducted, and neither are any relationships with races 1 and 2 on tomato and lettuce known.

From a practical point of view, genetic resistance has been the most effective method for controlling VWLM in sunflower for nearly 50 years. Initial sources of resistance were identified in Canada in the 1950s (Putt, 1958). The inheritance of resistance in some inbred lines was found to be qualitative or of a complete dominance and designated as  $V_1$  (Putt, 1964). The same type of resistance was found 10 years later in certain inbred lines from the USDA collection, such as HA 89 (Fick and Zimmer, 1974), which is a recurrent parent in the development of resistant hybrids, particularly in public sunflower breeding programs. The new races of *Vd* are not controlled by the resistance in HA 89 (see above). Instead, some of them seem to be controlled by the resistance of some entries of the USDA sunflower collection, such as PI507901 (Radi and Gulya, 2007) or the inbred lines HA 300 and HA 371 (Gulya et al., 1997). Moreover, the inheritance of resistance to race NA-Vd2 appears to be recessive or additive in some lines, and the breeding alternative of pyramiding quantitative resistance is being explored in Argentina (Galella et al., 2012). Frequent outbreaks of VWLM in sunflower-growing countries suggest that the resistance in commercial hybrids was overcome by the pathogen and this makes it urgent to identify plant material which could serve as a donor of resistance against the current races of the fungus worldwide.

This work was conceived from a holistic perspective since a bewildering amount of scientific information is available for *Vd* affecting many crops but *Vd* being pathogenic on sunflower is largely unknown. Here we describe the population structure of the *Vd* affecting sunflowers in countries of Europe where VWLM recurrently threatens oil production: Bulgaria, France, Italy, Romania, Spain, Ukraine, and Turkey. Genetic, molecular and pathogenic traits of the fungal collection were studied and, because of its epidemiological significance, we also addressed to what extent *Vd* from sunflower can be pathogenic on other crops.

## MATERIALS AND METHODS

### Isolates of *Verticillium dahliae* From Sunflower

All the isolates of the fungus were recovered from affected sunflowers that were collected between 2009 and 2016 in Argentina, Bulgaria, France, Italy, Romania, Spain, Turkey, and Ukraine. Because of the importance *Verticillium* wilt has in sunflowers in Argentina and although we did not have a set of isolates representing the race diversity of *Vd* in that country, one isolate from Argentina was included as representative. The plants showed interveinal chlorosis and yellowing, as well as wilt symptoms (see **Supplementary Figure S1A**). Their reference and information on the year of collection and geographical location of the samples are presented in **Table 1**. Cross sections of the stem base and petiole tissues of all the plants were analyzed. Each section was divided into two–six pieces that were surface-disinfested for 3 min by immersion in 10% household bleach (40 g of active chlorine per liter), rinsed in deionized water for 3 min and air dried using a vertical laminar flow cabinet. Segments 2 to 4 mm long of sunflower tissue were aseptically transferred to petri plates containing potato dextrose agar (PDA). Plates were incubated at 25°C for 72 h in darkness. Colonies were morphologically confirmed by observation under the stereoscope. Only one colony among all those recovered from the same field was selected for further studies, and a minimum of two monoconidial cultures were obtained by the following procedure. Each isolated colony was transferred to PDA and incubated in the laboratory at 25°C. After 8–10 days, the plates were flooded with 5 ml of sterile deionized water each and swirled gently. The conidial suspension was filtered through two layers of sterile gauze. Five serial 1:10 dilutions were prepared from the initial suspension and, from each of them, a small volume was streaked onto Water Agar (WA) medium following a zigzag distribution. Plates were incubated at 28°C for 24 h in darkness. Germinating conidia were then identified under the stereoscope and individually transferred to PDA. The final colonies were confirmed as being *Vd* based on morphological characters, and labeled as monoconidial isolates. Monoconidial and original isolates were stored in PDA as part of the fungal collection of the Laboratory of Field Crop Diseases at the Institute for Sustainable Agriculture, Córdoba, Spain.

### Genetic Diversity: Determination of Vegetative Compatibility Groups

The VCG of 38 monoconidial isolates, including those previously characterized by our research group (García-Carneros et al., 2014), were determined by generation and characterization of nitrate non-utilizing (*nit*) mutants of each of them and determination of vegetative compatibility. *Nit* mutants were generated on Water Agar Chlorate medium as colonies presenting a faint growth on Czapek-Dox Agar (CDA) with no aerial mycelium (Korolev and Katan, 1997) and phenotyped on CDA amended with hypoxanthine as described by Correll et al. (1987). Complementation tests were done by pairing *nit* mutants of the isolates with the complementary mutants of the

**TABLE 1** | Isolates of *Verticillium dahliae* used in this work, listed by year of collection, geographic origin and host, and genetic characterization by means of assignment to Vegetative Compatibility Group (VCG).

Isolate*	Year of collection	Geographical location (Country, Province)	Host	VCG
VdS0109	2009	Spain, Andalusia	Sunflower	2B
VdS0209	2009	Spain, Andalusia	Sunflower	2B
VdS0112*	2012	Argentina, Balcarce	Sunflower	2B
VdS0212	2012	Spain, Andalusia	Sunflower	2B
VdS0312	2012	Spain, Andalusia	Sunflower	2B
VdS0113** <sup>a</sup>	2013	Spain, Cádiz	Sunflower	2B
VdS0213	2013	Spain, Cádiz	Sunflower	2B
VdO0913	2013	Spain, Jaen, Sant. Puerto	Olive tree	1A
VdO1113	2013	Spain, Jaen, Cast. Locubin	Olive tree	2A
VdS0114	2014	Turkey, Thrace	Sunflower	–
VdS0214*	2014	Turkey, Thrace	Sunflower	6
VdS0314	2014	Turkey, Thrace	Sunflower	–
VdS0414	2014	Turkey, Thrace	Sunflower	–
VdS0514*	2014	Turkey, Thrace	Sunflower	6
VdS0614	2014	Turkey, Thrace	Sunflower	–
VdS0714*	2014	Bulgaria, Krumovo	Sunflower	6
VdS0814	2014	Bulgaria, Krumovo	Sunflower	–
VdS0914*	2014	Romania, Valu lui Traian	Sunflower	–
VdS1014	2014	Romania, Valu lui Traian	Sunflower	–
VdS1114	2014	Romania, Valu lui Traian	Sunflower	–
VdS1314	2014	Romania, Movila Braila	Sunflower	–
VdS1414*	2014	France, Montgiscard	Sunflower	2B
VdS1514*	2014	France, Montbeilhan, Marignac	Sunflower	–
VdS1614*	2014	France, Vacquiès, Saint-Aignan	Sunflower	2B
VdS1714*	2014	France, Vacquiès, Saint-Aignan	Sunflower	2B
VdS0115	2015	Spain, Cádiz	Sunflower	2B
VdS0215	2015	Spain, Cádiz	Sunflower	2B
VdS0116*	2016	Bulgaria, Dobrich	Sunflower	6
VdS0216*	2016	France, Montech	Sunflower	2B
VdS0316*	2016	Italy, Macerata	Sunflower	2B
VdS0416*	2016	Italy, Pisa	Sunflower	2B
VdS0516*	2016	Romania, Cuocorova	Sunflower	–
VdS0616*	2016	Romania, Slava Rusă	Sunflower	–
VdS0716*	2016	Spain, Seville	Sunflower	2B
VdS0816*	2016	Turkey, Edirne	Sunflower	–
VdS0916*	2016	Turkey, Kastamonu, Ahmetbey	Sunflower	6
VdS1016*	2016	Ukraine, Manvelivka, Dnepropet	Sunflower	2B
VdS1116*	2016	Ukraine, Velyki Sorochyntsi, Polta	Sunflower	6

<sup>a</sup>Isolate identified as a new race of *Vd* overcoming *V<sub>1</sub>* gene in HA89 by García-Ruiz et al. (2014).

\*Isolates included in the pathogenic characterization with sunflower genotypes.

international OARDC (The Ohio State University, Wooster, OH, United States) reference testers, Israeli *nit* testers and testers of Washington State University (Pullman WA, United States): T9 isolate (VCG1A), Ep8M and Ep52 isolates (VCG2A), Cot200 and Cot254 isolates (VCG2B), 70-21 isolate (VCG3), 131M isolate (VCG4A), Pt15M isolate (VCG4B), and MT isolate (VCG6). Pairings were done following the methodology of Collado-Romero et al. (2006). Mycelial plugs of *nit* testers and *nit* mutants of test (unknown) isolates were placed 1.5 cm apart on CDA in petri plates and incubated at 25°C in the dark. Plates were scored for prototrophic growth every 7 days and until 28–35 days of incubation. Positive complementation was indicated by the

formation of a dense, aerial growth where mycelia from the tester and the test *nit* mutant had met and formed a prototrophic heterokaryon (**Supplementary Figure S2**). The test *nit* mutant was thus considered vegetatively compatible with the tester strain and was assigned to its VCG.

## Molecular Characterization

The molecular characterization of all the isolates was performed using diagnostic primers of race 1 (Usami et al., 2007; de Jonge et al., 2012) and race 2 (Short et al., 2014), as well as with the diagnostic markers of D and ND pathotypes described for *Vd* infecting olive tree and artichoke (Carder et al., 1994;



Mercado-Blanco et al., 2001, 2002, 2003; Collado-Romero et al., 2009). Isolates of the D pathotype of *Vd* are pathogenic on sunflower among other crop species (Jiménez-Díaz et al., 2017), but previous findings by our research group showed that those affecting sunflower are molecularly similar to ND isolates of the fungus that are pathogenic to artichoke or cotton (García-Carneros et al., 2014). Besides, races of *Vd* from sunflower have been determined on the basis of the reaction of particular host genotypes (or differentials) carrying genes of resistance from different sources, but racial characterization by means of molecular markers for race has not been addressed so far. Total genomic DNA from each isolate was purified using the i-Genomic Plant DNA Extraction Minikit (Intron Biotechnology, Sangdaewon-Dong, South Korea) according to the manufacturer's instructions. Quality and concentration of DNA samples were determined using a Qubit™ 3.0 Fluorometer (Invitrogen™, Carlsbad, CA, United States). Finally, DNA samples were adjusted to a final concentration of 10 ng/μL and stored at −20°C until required.

The primer pairs used for the diagnosis of D and ND pathotypes were: DB19/DB22, DB19/espdef01, INTD2f/INTD2r, INTND2f/INTND2r, INTNDf/INTNDR and INTND2f/INTND3r. Optimized PCR assays were carried out in a final volume of 25 μL containing 0.4 μM each primer, 800 μM dNTPs, 2.5 μL 10x PCR buffer (800 mM tris-HCl, pH 8.3–8.4 at 25°C, 0.2% Tween20 wt/V), 0.75 U Taq-DNA Polymerase (Dominion MBL, Córdoba, Spain), 1.5 mM (DB19/DB22 primers) or 2 mM (rest of primers) MgCl<sub>2</sub>. Amplification conditions were as follows: 4 min denaturation at 94°C; followed by 35 cycles of 1 min denaturation at 94°C, 1 min of annealing at 54°C (DB19/DB22), 62°C (DB19/espdef01), 64°C (INTD2f/INTD2r, INTND2f/INTND2r, INTNDf/INTNDR), 60°C (INTND2f/INTND3r), and 1 min of extension at 72°C; and a final extension step of 6 min at 72°C. Determination of races 1 and/or 2 was conducted using diagnostic primer pairs Tr1/Tr2, VdAve1F/VdAve1R and VdR2F/VdR2R. Optimized PCR assays were carried out in a final volume of 25 μL containing 10 μM each primer, 400 μM dNTPs, 2.5 μL 10x PCR buffer (800 mM tris-HCl, pH 8.3–8.4 at 25°C, 0.2% Tween 20 wt/V), 0.75 U Taq-DNA Polymerase (Dominion MBL, Córdoba, Spain), 3 mM MgCl<sub>2</sub>. The following profiles were set for amplifications: 2 min initial denaturation at 94°C; 35 cycles of 1 min denaturation at 94°C, 1 min annealing at 64°C (Tr1/Tr2 and VdR2F/VdR2R) or 62°C (VdAve1F/VdAve1R), and 1 min of extension at 72°C; and a final extension step of 10 min at 72°C.

All reactions were done in a T1 Thermocycler (Whatman Biometra, Göttingen, Germany). Amplification products were separated by horizontal electrophoresis in 1.5 or 2% agarose gels containing 0.05 μl/ml GoldView Nucleic Acid Stain (SBS Genetech, Beijing, China) and visualized over a UV light source. A 100- to 2,000-bp or 100- to 1,000-bp ladder (Dominion MBL, Córdoba, Spain) was included in the electrophoresis.

A binary matrix based on presence (1) or absence (0) of PCR product was generated. Cluster analysis using the unweighted paired group method with arithmetic averages

(UPGMA) algorithm and Jaccard's similarity coefficient (Jaccard, 1908) were used to classify the isolates and determine genetic similarities among them. Analyses were performed with InfoStat Software® v. 2010 (Di Rienzo et al., 2010).

## Host Range: Pathogenicity of Isolates of *Verticillium dahliae* From Sunflower on Herbaceous Crop Species

An experiment was conducted under greenhouse conditions. Six crop species were inoculated with three isolates of *Vd* from sunflower (VdS0112, VdS0113, and VdS0213) and two from olive tree (one from the D pathotype, VdO0913, and another from the ND pathotype, VdO1113, both of them belonging to the fungal collection of the Laboratory of Field Crop Diseases at the Institute for Sustainable Agriculture, Córdoba, Spain).

Seeds of genetically susceptible artichoke ('Talpiot'), cotton ('Avangard'), eggplant ('Cristal'), lettuce ('Maravilla de verano'), tomato ('Manacor') and sunflower (HA 89 hybrid) were surface-sterilized by surface-sterilized by immersing them in 10% sodium hypochlorite for 10 min, then thoroughly rinsed in deionised water and incubated in the dark at saturation humidity in a germinator at 26 ± 2°C until radicles were 2–5 mm long. Seedlings were then transplanted into vermiculite and incubated in the greenhouse at 15–25°C and photoperiod of 14 h light per day for 1 month. Fertilization was applied weekly using a commercial solution (COMPO Universal Fertilizer) following manufacturer's recommendation. Then, six plants (replications) of each crop species were uprooted and inoculated by root immersion in conidial suspensions of the *Vd* isolates (10<sup>6</sup> conidia/ml). Roots of control plants were immersed in deionized water. The experiment was carried out in a completely randomized 6 × 6 factorial design. Plants were incubated for 8 weeks in the greenhouse under the same conditions previously described. Severity of symptoms (SS) in each plant was assessed weekly as a percentage of the foliar tissue affected. Sequential SS values were used to calculate the area under the disease progress curve (AUDPC) by trapezoidal integration method (Campbell and Madden, 1990). The experiment was performed twice and, as no significant differences between the two replicates were found for AUDPC (McIntosh, 1983), data were pooled and analysed using analysis of variance (ANOVA). Mean values of AUDPC were compared using Fisher's protected least significant difference (LSD) tests ( $P = 0.05$ ). Statistical analyses of data were performed using STATISTIX 10.0 software (Analytical Software, Tallahassee, FL, United States).

## Pathogenic Characterization of *Verticillium dahliae* in Sunflower Genotypes

The pathogenic characterization of *Vd* from sunflower was conducted by means of two phenotyping experiments carried out under greenhouse conditions (20–25°C day and 15–20°C night, with 14 h light). Both experiments were replicated and similar results were obtained. In the first experiment, seven genotypes of sunflower, inbred lines and commercial

hybrids with different responses to *Verticillium* wilt according to previous unpublished data (from field and greenhouse experiments), were independently inoculated with 21 isolates (Table 1). Inbred lines were Pioneer 1 and the public lines HA 458, HA 89, and INRA2603. The hybrids included were Pioneer 2, Pioneer 3, and Pioneer 4. Four-week-old plants grown as previously described (experiment of host range in previous subheading) were uprooted and inoculated by immersing the roots in a suspension of  $10^6$  conidia per ml for 30 min. Roots of the control plants were immersed in water. Inoculated plants were individually transplanted to 0.75 l pots filled with peat:sand (2:1). Four replications (pots) were used for each genotype and *Vd* isolate. Plants were incubated for 25 days and, at the end of the experiment, VWLM was assessed by means of a Disease Index (DI) that was calculated as:  $DI = AN \times SS$ , where AN is the percent of affected nodes and SS represents the severity of symptoms according to a 0–5 scale based on chlorosis and necrosis of leaves proposed by Alkher et al. (2009) (0 = no chlorosis or necrosis, 1 = visible chlorosis with <1% necrosis, 2 = up to 40% chlorosis and 1–20% necrosis, 3 = up to 65% chlorosis and 20–35% necrosis, 4 = 100% chlorosis and 35–70% necrosis and 5 = 100% chlorosis and 70–100% necrosis). This scale was used because it represents, with only one value, both the area of the plant with symptoms and the severity of those symptoms. The experiment was performed twice and, as no significant difference between the two replicates was found for DI (McIntosh, 1983), data were pooled and VWLM assessed using ANOVA. Hierarchical cluster analysis with UPGMA algorithm and Euclidean distance was made to classify the isolates into different pathotypes. A principal component analysis (PCA) was also carried out to visualize the distribution of the variability found. Statistical, clustering and PCA analyses were performed with InfoStat Software® v. 2010.

A second phenotyping experiment was established in order to confirm the existence of pathogenic variants and to identify differentials of races of *Vd*. This experiment was performed with inbred lines and a selection of those isolates representing the diversity found after UPGMA analysis in the first experiment. Thus, 12 isolates of *Vd* were inoculated into HA 458, HA 89, INRA2603, and Pioneer 1. The experiment was conducted like experiment 1, with slight modifications. Three-week-old sunflowers were grown and inoculated as described. Aiming at obtaining similar disease data to those occurring under field conditions, sunflowers were transplanted into 3.5 l pots for 6 weeks. At the end of the experiment, VWLM in the plants was evaluated using the DI explained above. The experiment was conducted twice and the data pooled since no significant difference between the two replicates was found for DI (McIntosh, 1983). The DI results were analyzed using ANOVA and, when significant effects were obtained, Fisher's protected LSD tests ( $P = 0.05$ ) were used for comparisons of genotypes, *Vd* isolates, and their interaction. Statistical analyses of data were performed using InfoStat Software® v. 2010.

## RESULTS

### Genetic Diversity: Determination of Vegetative Compatibility Groups

Of the 36 isolates of *Vd* from sunflower characterized to VCG, 17 were assigned to VCG2B and 6 were assigned to VCG6. The remaining isolates could not be assigned to any VCG. The two isolates from olive tree belonged to VCG1A (VdO0913 isolate, D pathotype), and to VCG2A (VdO1113 isolate, ND pathotype). Surprisingly, VCGs of *Vd* from sunflower were related to their geographical origin. All the isolates from Argentina, France, Italy, and Spain were assigned to VCG2B, while isolates from Bulgaria, Turkey, Romania, and Ukraine were assigned only to VCG6. In this latter group, 13 of the 19 isolates failed to form stable heterokaryons with any of the *nit* mutant testers (Table 1).

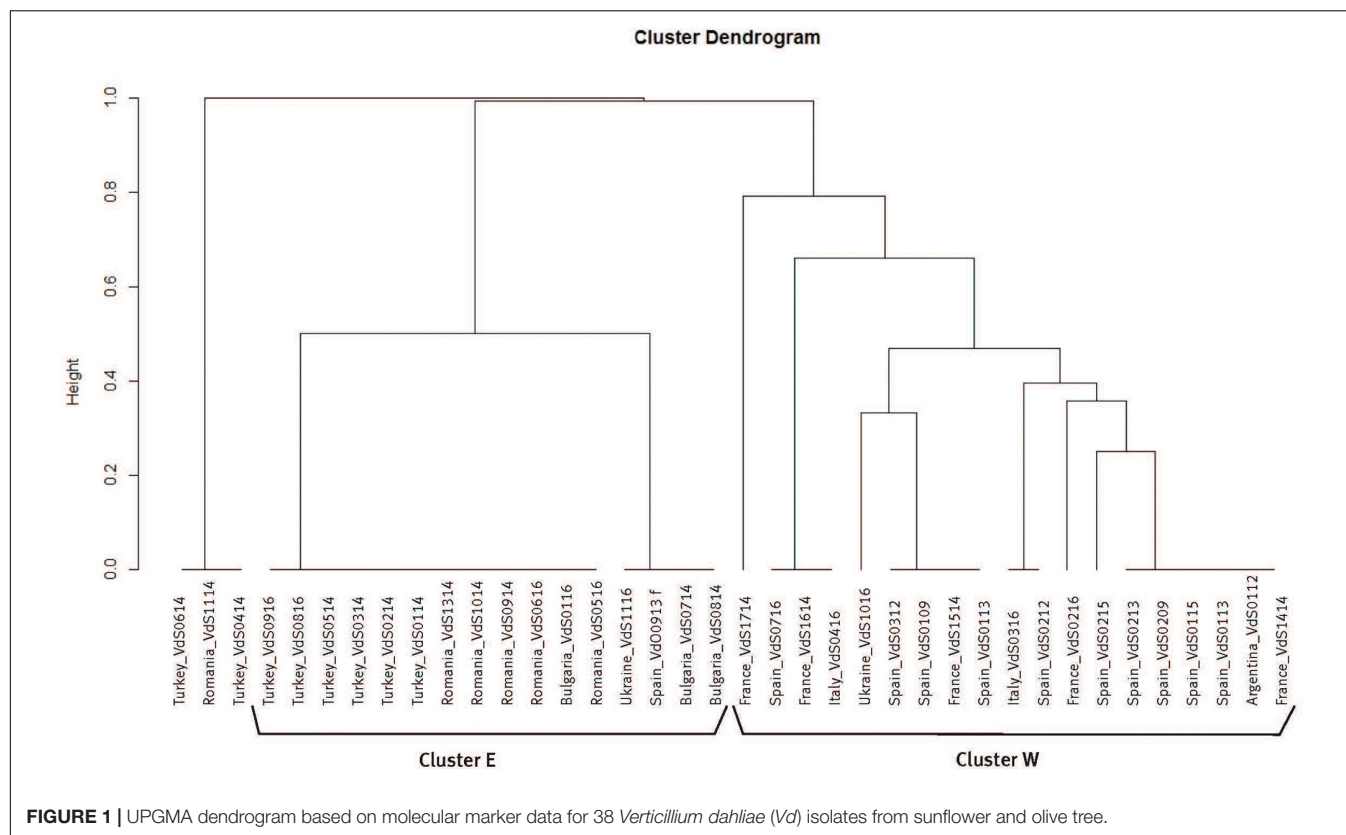
### Molecular Characterization

With regard to molecular characterization, all the isolates amplified, as expected, the 543- or 526-bp marker specific to *Vd* (DB19/DB22 primers).

When amplified using primer pairs INTD2f/INTD2r (462-bp marker) and DB19/espdef01 (334-bp marker), all the isolates from Argentina and Western Europe showed the 462 (–), 334 (–) pattern. When isolates of *Vd* from Eastern European countries were amplified using the same pairs of primers, they had the following patterns: 462 (+), 334 (+); 462 (–), 334 (+); and 462 (–), 334 (–). This same group of isolates had a single molecular pattern after amplifications with INTNDf/INTNdr (1,163-bp marker), INTND2f/INTND2r (824-bp marker), and INTND2f/INTND3r (688-bp marker): 1,163 (–), 824 (–), 688 (–). Conversely, PCR assays of *Vd* isolates from Argentina and Western European countries using these three pairs of primers amplified either the three markers, only the 688-bp marker, or any combination of two out of the three of them (Supplementary Table S1).

When amplified with race-specific primers, all 38 isolates yielded 256-bp amplicons with VdR2F/VdR2R and failed to amplify with Tr1/Tr2 and VdAve1F/VdAve1R (Supplementary Table S1). Since no polymorphisms were detected in our *Vd* isolates when using race-specific primers, these data were omitted for the molecular analysis.

The dendrogram resulting from the UPGMA analysis of the molecular data set for pathogenic characterization distinguished three well-differentiated clusters among the 38 isolates of *Vd* (Figure 1). The first cluster (Cluster E) grouped the 16 isolates collected in countries from Eastern Europe as well as the isolate VdO0913 from olive tree collected in Spain, which shared about a 50% similarity. Moreover, all the isolates grouped in cluster E belonged to an unknown VCG or to VCG6. All isolates of *Vd* from sunflower of Argentina and Western Europe countries, as well as one isolate from Ukraine and the isolate VdO0113 from olive tree of Spain, shared a 21% similarity and were grouped in a second cluster (cluster W) irrespective of their country of origin. Interestingly, all the isolates in cluster W were assigned to VCG2B. Finally, two isolates from Turkey (VdS0614 and



VdS0414) and one from Romania (VdS1114) were genetically very distant from the rest of the isolates.

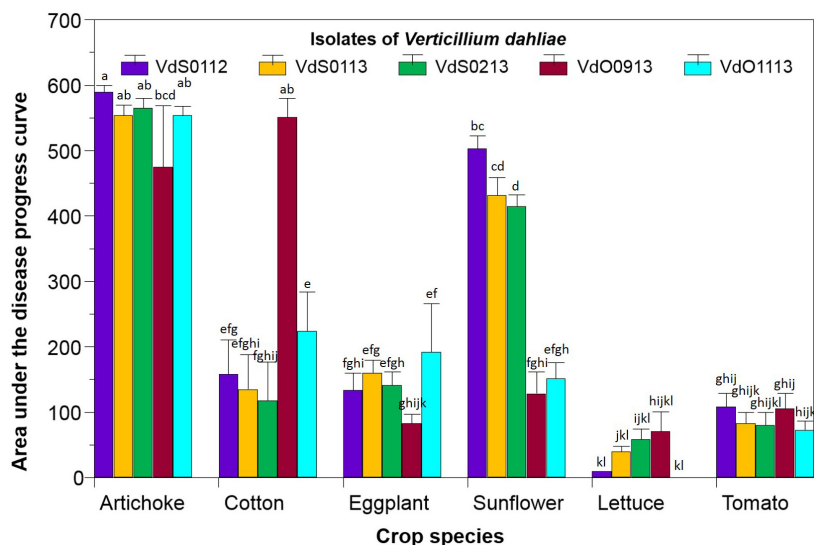
## Host Range: Pathogenicity of Isolates of *Verticillium dahliae* From Sunflower in Herbaceous Crop Species

The disease by *Vd*, expressed as the AUDPC, caused by the three isolates from sunflower and the two from olive tree on artichoke, cotton, eggplant, sunflower, lettuce, and tomato is depicted in **Figure 2**. Data for non-inoculated control plants are not shown since they were zero. Statistical analyses showed that both main factors –crop species and isolate of *Vd*– had a significant impact ( $P < 0.0001$ ) on the disease, but also a significance ( $P < 0.0001$ ) of the crop species  $\times$  isolate of *Vd* interaction was obtained, indicating that the AUDPC of each crop species was influenced by the particular *Vd* isolate infecting it. Overall, the susceptibility of the crop species studied varied from low in the case of lettuce and tomato (36 and 90 AUDPC across isolates, respectively), to high and very high in that of sunflower (326 AUDPC across isolates) and artichoke (548 AUDPC across isolates), respectively (**Figure 2**). Likewise, the pathogenic ability of *Vd* isolates was dependent on the crop species that they were infecting. None of them were pathogenic to lettuce or tomato, since disease values did not significantly differ from those of the non-inoculated controls. On the contrary, all the isolates were highly pathogenic to artichoke, with AUDPC values ranging from 590 for 112 to 475 for 1333 (**Figure 2**).

The most interesting reactions to *Vd* were those of cotton and sunflower. Cotton was highly susceptible to the isolate VdO0913 (551 AUDPC), and was moderately susceptible to the rest of the isolates (average 159 AUDPC across them). Sunflower was highly susceptible to those isolates recovered from sunflower samples (450 AUDPC averaged across VdS0112, VdS0113, and VdS0213) but not to those from olive tree (140 AUDPC averaged across them). Finally, eggplant displayed moderate susceptibility to VdO1113 isolate (192 AUDPC in comparison to 129 AUDPC averaged across the remaining four isolates).

## Pathogenic Characterization: Identification of Sunflower Genotypes as Differentials of Races of *Verticillium dahliae* From Sunflower

The VdS1116 isolate was pathogenic to HA 458 and Pioneer 3 genotypes but it did not cause any symptom in the rest of the genotypes. All the isolates except VdS1116 induced symptoms in those two genotypes but also in at least one of the others. The UPGMA dendrogram generated with the phenotypic data of experiment 1 shows a first approach to the pathogenic diversity of *Vd* isolates (**Figure 3**). Individual phenotypic information for each isolate is presented in **Supplementary Table S2**. The UPGMA dendrogram shows two big clusters, one including isolates from Western Europe (France, Spain, and Italy), the isolate from Argentina and the two isolates from Ukraine. The other big cluster grouped isolates from Eastern Europe (Turkey,



**FIGURE 2 |** Reaction, expressed as the area under the disease progress curve, of six crop species upon inoculation with three isolates of *Verticillium dahliae* from sunflower (VdS0112, VdS0113, and VdS0213) and two from olive tree (VdO0913 and VdO1113). Bars with the same lower case letters are not significantly different according to the least significant difference test ( $P = 0.05$ , critical least significant difference value = 86.80).

Bulgaria, and Romania). This clustering reflects the reactions of the sunflower genotypes: in the Eastern Europe group, Pioneer 4, HA 89 and Pioneer 1 were the most resistant genotypes, while in the case of isolates from Western Europe, Pioneer 1, Pioneer 2, and INRA2603 were the most resistant ones. In both groups, HA 458 and Pioneer 3 were the most susceptible genotypes. The PCA biplot represented in **Supplementary Figure S3** shows these relationships.

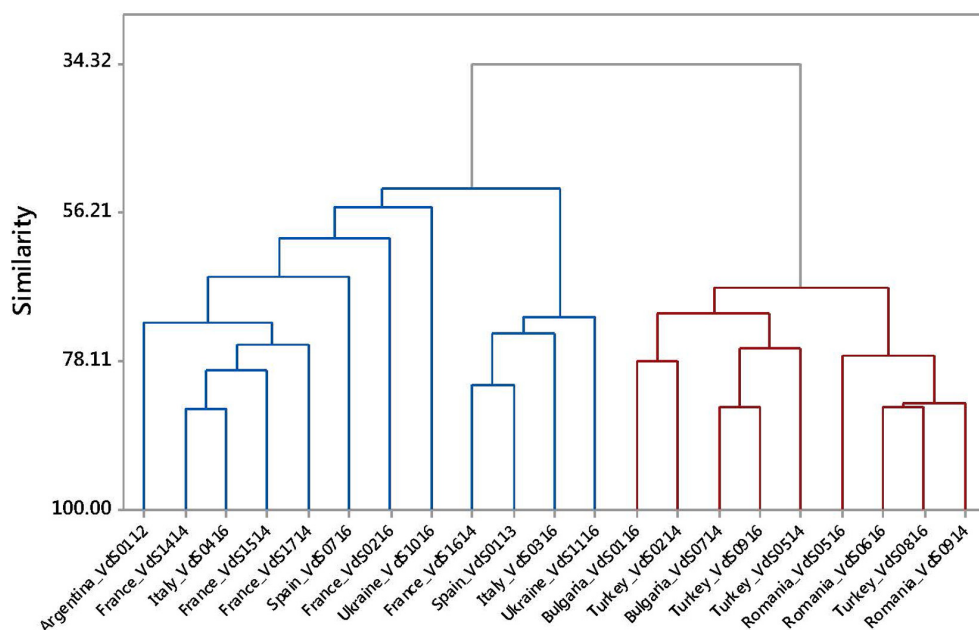
**Figure 4** shows the results of experiment 2 in which four sunflower inbred lines (HA 458, HA 89, INRA2603, and Pioneer 1) and 12 isolates representing the diversity observed in experiment 1 were included. Characteristic and differential reactions of sunflower genotypes in experiment 1 were observed again in this experiment. Thus, INRA2603 was resistant to isolates from Western Europe (including those from Argentina and Ukraine) while it was susceptible to isolates from Eastern Europe, and the opposite situation was observed for HA 89 genotype. Pioneer 1 was resistant to both groups of isolates and HA 458 was the one most susceptible to all of them. There were significant differences for DI between genotypes, isolates of *Vd* and their interaction ( $P < 0.0001$  for all). In general, genotype  $\times$  isolate combinations resulting in values of DI between 0 and 100 did not significantly differ from those of the non-inoculated plants so that this criterion of  $<100$  was used to determine resistance interactions. Pioneer 1 line was resistant to all isolates (31 DI averaged across isolates) and HA 458 presented DI values of 200 or higher with all the isolates. The INRA2603 and HA 89 responses depended on the isolates, showing resistance reactions for some of them and susceptible ones for others. As in experiment 1, the VdS1116 isolate was pathogenic only to HA 458 (230 DI). Isolates from Turkey, Bulgaria, and Romania caused higher disease responses than 100 in INRA2603 (144 DI averaged across isolates). This same inbred line was resistant to the rest

of the isolates. By contrast, HA 89 was resistant to VdS0714, VdS0516, VdS0616, VdS0816, and VdS0916 (Turkey, Bulgaria, and Romania) (68 DI averaged across isolates) and susceptible to the remaining seven *Vd* isolates (195 DI averaged across them). **Supplementary Figure S1B** shows the symptoms in HA 458, HA 89, INRA2603, and Pioneer 1 inbred lines after inoculation with isolate VdS0316 which corresponds to the typical profile observed for Western Europe *Vd* isolates.

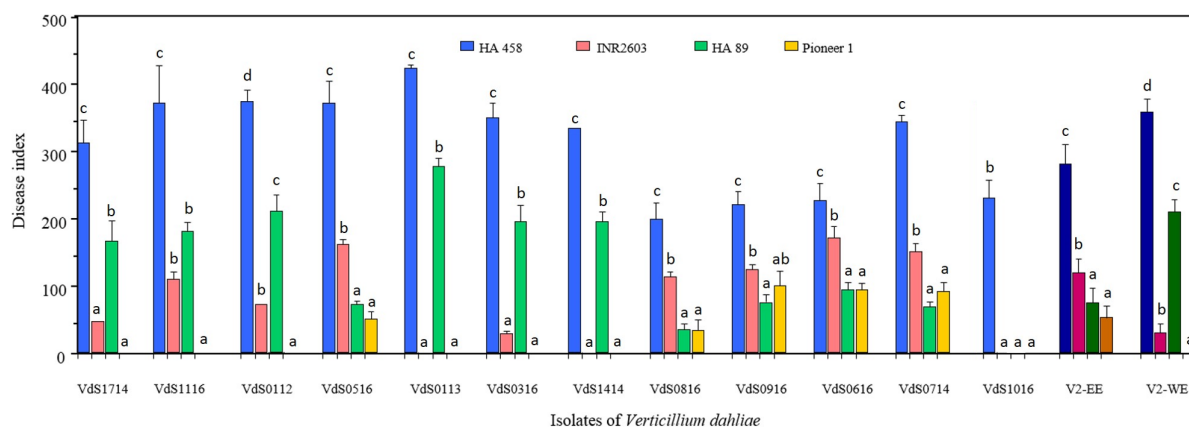
## DISCUSSION

In this study, a low genetic diversity of *Vd* from sunflower was found, with only VCG2B and VCG6 being identified among the isolates. One important finding has been the identification of sunflower as the second host, after pepper (Bhat et al., 2003), in which *Vd* is assigned to VCG6. Moreover, isolates belonging to VCG6 were restricted to Eastern Europe, where we also found a high proportion of *Vd* isolates not assigned to any VCG. This could be related to a genetic diversity of *Vd* that is unidentifiable with the available VCG testers. In contrast, VCG2B was the only group identified for all the *Vd* isolates from Western Europe and Argentina in agreement with previous reports from our group (García-Carneros et al., 2014). The VCG2B has been identified for isolates from *Vd* pathogenic to other herbaceous crops species, such as mint (Douhan and Johnson, 2001), spinach (Iglesias-Garcia et al., 2013), cotton (Göre et al., 2014), watermelon (Dervis et al., 2009a), or eggplant (Dervis et al., 2009b) among others. Our host range results are consistent with host adaptation (Jiménez-Díaz et al., 2017) in *Vd* from sunflower, which was clearly pathogenic to sunflower but not to the other crop species, with the exception of artichoke. Host adaptation means that isolates may be pathogenic on multiple





**FIGURE 3 |** UPGMA dendrogram based on Disease Index values for seven sunflower genotypes inoculated with 21 isolates of *Verticillium dahliae* from sunflower. Blue and red colors are used to indicate Clusters W and E, respectively.



**FIGURE 4 |** Verticillium wilt and leaf mottle in four sunflower genotypes caused by 12 *Verticillium dahliae* isolates from different geographical origins and expressed by a disease index calculated on the basis of the percent of affected nodes and the severity of symptoms in the plants (see section "Materials and Methods" for details). The last two groups of darker bars represent values averaged for isolates from the east (V2-EE) and the west (V2-WE) of Europe. Bars with the same lower case letters are not significantly different according to the least significant difference test ( $P = 0.05$ ).

hosts but are usually more virulent on some hosts, typically, but not exclusively, on those from which they were recovered (Bhat and Subbarao, 1999; Douhan and Johnson, 2001; Jiménez-Díaz et al., 2006). Furthermore, the finding that *Vd* from sunflower is pathogenic and highly virulent on artichoke is in agreement with the results of its genetic characterization. All three isolates from sunflower included in the host range study were identified as VCG2B, a frequent VCG in *Vd* from artichoke (Bergal et al., 2010). Additionally, it was not unexpected to find that isolates of *Vd* from olive tree affected cotton, since cross pathogenicity of the fungus in both crops has been

reported (López-Escudero and Mercado-Blanco, 2011). However, according to our results about the molecular and pathogenic diversity of *Vd* in Europe, our conclusions on host range might not be applicable to all isolates of *Vd* in Europe. Host range of *Vd* from sunflower would be better precised if isolates of the fungus from Eastern Europe and VCG different to VCG2B were considered. From the phytopathological point of view, root tissues and plant debris of any crop species infected by *Vd* strains from sunflower can serve as carriers and sources of inoculum. Studies on cross pathogenicity in *Vd* belonging to VCG6 and infecting sunflower and pepper, as well as that of *Vd* belonging

to VCG2B and pathogenic to herbaceous hosts such as sunflower, mint, spinach, watermelon, or eggplant are needed to better understand the concern that these crops as farming alternatives can raise for possible severe outbreaks or increased severities of Verticillium wilt.

Molecular markers revealed a haplotype diversity that suggests a clear divergence between *Vd* from the east and west of Europe. An important finding of this study is that molecular differences in *Vd* from sunflower were mostly related to ND and D pathotypes, since all the isolates were race 2. Since *Ave1* was not amplified from any of the haplotypes of *Vd* from sunflower nor from either of the two from olive tree, all of them lack this gene (de Jonge et al., 2012). Hu et al. (2015) found that ND and D isolates of *Vd* from cotton correlated with races 1 and 2. In our research ND and D pathotypes of *Vd* from sunflower were identified irrespective of race. Geographical differences were found instead: the ND pathotype was identified for haplotypes of *Vd* from Western Europe (cluster W) and the D pathotype for haplotypes from Eastern Europe (cluster E). Another interesting outcome of our study is the unexpectedly strong agreement between haplotype clustering and genetic characterization, with only VCG6 identified in cluster E (east of Europe) and VCG2B the only genetic group in haplotypes from cluster W (west of Europe). Although race identification using molecular biological methods is more useful than time-consuming inoculation experiments, the molecular identification of those pathogenic differences in *Vd* from sunflower that are clearly distinguished on the basis of phenotypic data is still not possible. Differential pathogenicity within race 2 has also recently been reported by Usami et al. (2017) for *Vd* from tomato.

Little is known about the pathogenic diversity of *Vd* in sunflower. Understanding this diversity of *Vd* populations in Europe and determining the pathogenic races that are present in the area is an essential and determining requirement for efficient resistance breeding. In this study we found that the new race overcoming the *V<sub>1</sub>* gene into HA 89 (VdS0113 isolate) (García-Ruiz et al., 2014) can be effectively controlled by the resistance in the public line INRA2603. Moreover, INRA2603 frequently presented reverse reactions to the *Vd* isolates to those of HA 89. Overall, those *Vd* isolates effectively controlled by INRA2603 were not controlled by HA 89. This was the case of *Vd* isolates from Argentina, France, Italy, and Spain. On the contrary, INRA2603 was susceptible to isolates from Bulgaria, Turkey, and Romania to which HA 89 was resistant. These results suggest that the nature of the resistance of INRA2603 to *Vd*, and, probably, its associated resistance mechanism/s, is different to that in HA 89. On the other hand, we propose to name these pathogenic races of *Vd* as: (a) V2-EE ("Verticillium race 2 East Europe," pathogenic on INRA2603 but not on HA 89), (b) V2-WE ("Verticillium race 2 West Europe," pathogenic on HA 89 but not on INRA2603), and (c) V1 (the race controlled by both HA 89 and INRA2603). Another finding of our study is that race V2-WE of *Vd* is not only present in Spain, but also in Argentina (VdS0112), France (VdS1414 and VdS1714), Italy (VdS0316), and Ukraine (VdS1016). Whether or not race V2-WE has the same pathogenic abilities as isolates of *Vd* overcoming *V<sub>1</sub>* (HA 89) in Argentina (Bertero de Romano and Vázquez, 1982; Galella

et al., 2004) and/or in the United States (Gulya, 2007) remains unknown.

The most useful differentials for race characterization of plant pathogens -including those of sunflower- are public inbred lines, since their genetic background is known and they can easily be exchanged between research groups (Molinero-Ruiz et al., 2015). The presence of four pathogenic races has recently been reported in Argentina (Clemente et al., 2017) on the basis of the use of a set of differentials that is not public. Since INRA2603 and HA 89 are public lines and differentially resistant and/or susceptible to *Vd* in Europe, we propose that they should be used as differentials for pathogenic races of *Vd*. Thus, the set for identification of pathogenic races of *Vd* would be: HA 458 (universal susceptible), HA 89 and INRA2603.

## CONCLUSION

The current study constitutes the first research work focused on the characterization of *Vd* on sunflower in Europe. Its findings provide new insights into *Vd* populations affecting sunflower, a preliminary description of three genotypes to establish a universal set of race differentials like, for example, those in downy mildew – sunflower (Tourvieille de Labrouhe et al., 2000), and have fundamental implications for resistance breeding. First, we found that the *Vd* isolates from sunflower lack the *Ave1* gene and are molecularly distinguished into two different groups: Western Europe and Eastern Europe, their differences being associated with ND and D pathotypes, respectively. Even genetic differences were found between both groups, VCG2B being described in *Vd* from the west of Europe and VCG6 being assigned only to isolates from the east of Europe. With respect to pathogenic characterization of *Vd* from sunflower, and in addition to race V1, races V2-EE and V2-WE were determined according to the sources of resistance that they overcome (HA 89 and INRA2603 inbred lines). Secondly, any search for resistance to *Vd* for European environments of sunflower production should take this diversity into account in order to find donors with a broad resistance that can be effective to both V2-EE and V2-WE races. Otherwise, this pathogenic variability must be properly managed through the development of hybrids with resistance to specific geographical areas (Western and Eastern Europe). This research constitutes a milestone in analyzing the diversity of *Vd* in countries of Europe where sunflowers are grown. Collaborations between public and private sectors similar to that of this work should be advisable in other areas where Verticillium poses a threat to this oil crop.

## AUTHOR CONTRIBUTIONS

AM-S and LM-R analyzed the data, interpreted the results, conceived and designed the experiments, and contributed materials, equipment, and analysis tools. AM-S, SR, AG-C, SG-F, PM-F, and SC-S conducted the experiments. AM-S, AG-C, PM-F, SC-S, and LM-R wrote the manuscript. All authors reviewed the manuscript and approved the final version.

## FUNDING

Financial support for this research was partially provided by the Spanish Ministry of Economy, Industry and Competitiveness (AGL2010-17909 and AGL2016-80483-R grants) and the European Regional Development Fund (ERDF). *Nit* testers were kindly provided by Prof. R. Jiménez-Díaz (University of Córdoba, Córdoba, Spain) and by Dr. D. A. Johnson and Mr. D. L. Wheeler (Washington State University, Pullman, WA, United States). VdO0913 and VdO1113 isolates of *V. dahliae* from olive tree were kindly provided by Ms. M. Herrera (Laboratorio de Producción y Sanidad Vegetal de Jaén, AGAPA, Consejería de Agricultura, Pesca y Desarrollo Rural, Junta de Andalucía, Spain). Seeds of artichoke, cotton, eggplant, lettuce, and tomato were provided by Semillas Fitó (Barcelona, Spain).

## ACKNOWLEDGMENTS

We sincerely thank Mr. D. M. Martínez-Rosales (Dow Dupont) for his support in data analysis.

## REFERENCES

- Alexander, L. J. (1962). Susceptibility of certain Verticillium-resistant tomato varieties to an Ohio isolate of the pathogen. *Phytopathology* 52, 998–1000.
- Alkher, H., El Hadrami, A., Rashid, K. Y., Adam, L. R., and Daayf, F. (2009). Cross-Pathogenicity of *Verticillium dahliae* between potato and sunflower. *Eur. J. Plant Pathol.* 124, 505–519. doi: 10.1007/s10658-009-9437-z
- Bejarano-Alcázar, J., Blanco-López, M. A., Melero-Vara, J. M., and Jiménez-Díaz, R. M. (1996). Etiology, importance and distribution of Verticillium wilt of cotton in southern Spain. *Plant Dis.* 80, 1233–1238. doi: 10.1094/PD-80-1233
- Berbegal, M., Ortega, A., Jiménez-Gasco, M. M., Olivares-García, C., Jiménez-Díaz, R. M., and Armengol, J. (2010). Genetic diversity and host range of *Verticillium dahliae* isolates from artichoke and other vegetable crops in Spain. *Plant Dis.* 94, 396–404. doi: 10.1094/PDIS-94-4-0396
- Bertero de Romano, A. B., and Vázquez, A. (1982). “A new race of *Verticillium dahliae* Kleb,” in *Proceedings of the 10th International Sunflower Conference*, Toowoomba, QLD, 177–178.
- Bhat, R., Smith, R. F., Koike, S. T., Wu, B. M., and Subbarao, K. V. (2003). Characterization of *Verticillium dahliae* isolates and wilt epidemics of pepper. *Plant Dis.* 87, 789–797. doi: 10.1094/PDIS.2003.87.7.789
- Bhat, R. G., and Subbarao, K. V. (1999). Host range specificity in *Verticillium dahliae*. *Phytopathology* 89, 1218–1225. doi: 10.1094/PHYTO.1999.89.12.1218
- Campbell, C. L., and Madden, L. V. (1990). *Introduction to Plant Disease Epidemiology*. New York, NY: John Wiley & Sons.
- Carder, J. H., Morton, A., Tabrett, A. M., and Barbara, D. J. (1994). “Detection and differentiation by PCR of subspecific groups within two *Verticillium* species causing vascular wilts in herbaceous hosts,” in *Modern Assays for Plant Pathogenic Fungi*, eds A. Schots, F. M. Dewey and R. Oliver (Oxford: CAB International), 91–97.
- Clemente, G. E., Bazzalo, M. E., and Escande, A. R. (2017). New variants of *Verticillium dahliae* causing sunflower leaf mottle and wilt in Argentina. *J. Plant Pathol.* 99, 445–451. doi: 10.4454/jpp.v99i2.3875
- Collado-Romero, M., Berbegal, M., Jiménez-Díaz, R. M., Armengol, J., and Mercado-Blanco, J. (2009). A PCR-based “molecular tool box” for in planta differential detection of *Verticillium dahliae* vegetative compatibility groups infecting artichoke. *Plant Pathol.* 58, 515–526. doi: 10.1111/j.1365-3059.2008.01981.x
- Collado-Romero, M., Mercado-Blanco, J., Olivares-García, C., and Jiménez-Díaz, R. M. (2008). Phylogenetic analysis of *Verticillium dahliae* vegetative

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00288/full#supplementary-material>

**FIGURE S1 | (A)** Severe infection by *Verticillium dahliae* (Vd) in sunflower. **(B)** Symptoms produced by Vd isolate VdS0316 in the four sunflower genotypes used for race characterization.

**FIGURE S2 |** Example of the formation of a prototrophic heterokaryon.

**FIGURE S3 |** Principal coordinates analysis of the seven sunflower genotypes used in the pathogenic characterization and the 21 *Verticillium dahliae* isolates used in this study.

**TABLE S1 |** Molecular characterization of isolates of *Verticillium dahliae* (Vd) from sunflower using markers diagnostic of defoliating (D) and not defoliating (ND) pathotypes and of races 1 and 2 of the fungal species.

**TABLE S2 |** Phenotypic data of experiment 1 of the pathogenic characterization of 21 *Verticillium dahliae* isolates inoculated on seven sunflower genotypes.

- compatibility groups. *Phytopathology* 98, 1019–1028. doi: 10.1094/PHYTO-98-9-1019
- Collado-Romero, M., Mercado-Blanco, J., Olivares-García, C., Valverde-Corredor, A., and Jiménez-Díaz, R. M. (2006). Molecular variability within and among *Verticillium dahliae* vegetative compatibility groups determined by fluorescent AFLP and PCR markers. *Phytopathology* 96, 485–495. doi: 10.1094/PHYTO-96-0485
- Correll, J. C., Klittich, C. J. R., and Leslie, J. F. (1987). Nitrate nonutilizing mutants of *Fusarium oxysporum* and their use in vegetative compatibility tests. *Phytopathology* 77, 1640–1646. doi: 10.1094/Phyto-77-1640
- de Jonge, R., van Esse, H. P., Maruthachalam, K., Bolton, M. D., Santhanam, P., Saber, M. K., et al. (2012). Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5110–5115. doi: 10.1073/pnas.1119623109
- Debaeke, P., Bedoussac, L., Bonnet, C., Bret-Mestries, E., Seassau, C., Gavaland, A., et al. (2017). Sunflower crop: environmental-friendly and agroecological. *OCOL* 24:D304. doi: 10.1051/oc/2017020
- Dervis, S., Kurt, S., Soylu, S., Erten, L., Soylu, E. M., Yildiz, M., et al. (2008). Vegetative compatibility groups of *Verticillium dahliae* from cotton in the Southeastern Anatolia Region of Turkey. *Phytoparasitica* 36, 74–83. doi: 10.1007/BF02980750
- Dervis, S., Mercado-Blanco, J., Erten, L., Valverde-Corredor, A., and Pérez-Artés, E. (2010). Verticillium wilt of olive in Turkey: a survey on disease importance, pathogen diversity and susceptibility of relevant olive cultivars. *Eur. J. Plant Pathol.* 127, 287–301. doi: 10.1007/s10658-010-9595-z
- Dervis, S., Yetisir, H., Tok, F. M., Kurt, S., and Karaca, F. (2009a). Vegetative compatibility groups and pathogenicity of *Verticillium dahliae* isolates from watermelon in Turkey. *Afr. J. Agric. Res.* 4, 1268–1275. doi: 10.1046/j.1365-3059.1998.00273.x
- Dervis, S., Yetisir, H., Yildirim, H., Tok, F. M., Kurt, S., and Karaca, F. (2009b). Genetic and pathogenic characterization of *Verticillium dahliae* isolates from eggplant in Turkey. *Phytoparasitica* 37, 467–476. doi: 10.1007/s12600-009-0061-4
- Di Rienzo, J. A., Casanoves, F., Balzarini, M. G., Gonzalez, L., Tablada, M., and Robledo, C. W. (2010). *InfoStat (Version 2010)*. Córdoba: Unidad Nacional de Córdoba.
- Douhan, L. I., and Johnson, D. A. (2001). Vegetative compatibility and pathogenicity of *Verticillium dahliae* from spearmint and peppermint. *Plant Dis.* 85, 297–302. doi: 10.1094/PDIS.2001.85.3.297

- Dung, J. K. S., Peever, T. L., and Johnson, D. A. (2013). *Verticillium dahliae* populations from mint and potato are genetically divergent with predominant haplotypes. *Phytopathology* 103, 445–459. doi: 10.1094/PHYTO-06-12-0133-R
- El-Bebany, A. F., Alkher, H., Lorne, R. A., and Daayf, F. (2013). Vegetative compatibility of *Verticillium dahliae* isolates from potato and sunflower using nitrate non-utilizing (nit) mutants and PCR-based approaches. *Can. J. Plant Pathol.* 35, 1–9. doi: 10.1080/07060661.2012.702128
- Fick, G. N., and Zimmer, D. E. (1974). Monogenic resistance to *Verticillium* wilt in sunflowers. *Crop Sci.* 14:895. doi: 10.2135/cropsci1974.0011183X001400060037x
- Galella, M. T., Bazzalo, M. E., and León, A. (2004). “Compared pathogenicity of *Verticillium dahliae* isolates from Argentina and the USA,” in *Proceedings of the 16th International Sunflower Conference*, Fargo, ND, 177–180.
- Galella, M. T., Bazzalo, M. E., Morata, M., Cimmino, C., Kaspar, M., Grondona, M., et al. (2012). “Pyramiding QTLs for *Verticillium dahliae* resistance,” in *Proceedings of the 18th International Sunflower Conference*, Mar del Plata, 219–224.
- García-Carneros, A. B., García-Ruiz, R., and Molinero-Ruiz, L. (2014). Genetic and molecular approach to *Verticillium dahliae* infecting sunflower. *Helia* 37, 205–214. doi: 10.1515/helia-2014-0014
- García-Ruiz, R., García-Carneros, A. B., and Molinero-Ruiz, L. (2014). A new race of *Verticillium dahliae* causing leaf mottle of sunflower in Europe. *Plant Dis.* 98:1435. doi: 10.1094/PDIS-04-14-0360-PDN
- Göre, M. E., Erdoğan, O., Caner, O. K., Aydın, M. H., and Berk, S. (2014). VCG diversity and virulence of *Verticillium dahliae* from commercially available cotton seed lots in Turkey. *Eur. J. Plant Pathol.* 140, 689–699. doi: 10.1007/s10658-014-0500-z
- Gulya, T. (2007). New strain of *Verticillium dahliae* in North America. *Helia* 30, 115–120. doi: 10.2298/HEL0747115G
- Gulya, T. J., Rashid, K. Y., and Marisevic, S. M. (1997). “Sunflower diseases,” in *Sunflower Technology and Production*, ed. A. A. Schneiter (Madison, WI: ASA), 263–380.
- Harveson, R. M., and Markell, S. G. (2016). “Verticillium wilt,” in *Compendium of Sunflower Diseases*, eds R. M. Harveson, S. G. Markell, C. C. Block and T. J. Gulya (St. Paul, MN: The American Phytopathological Society), 59–61.
- Hayes, R. J., McHale, L. K., Vallad, G. E., Truco, M. J., Michelmore, R. W., Klosterman, S. J., et al. (2011). The inheritance of resistance to *Verticillium* wilt caused by race 1 isolates of *Verticillium dahliae* in the lettuce cultivar La Brillante. *Theor. Appl. Genet.* 123, 509–517. doi: 10.1007/s00122-011-1603-y
- Hayes, R. J., Vallad, G. E., Qin, Q. M., Grube, R. C., and Subbarao, K. V. (2007). Variation for resistance to *Verticillium* wilt in lettuce (*Lactuca sativa* L.). *Plant Dis.* 91, 439–445. doi: 10.1094/PDIS-91-4-0439
- Hu, X. P., Gurung, S., Short, D. P. G., Sandoya, G. V., Shang, W. J., Hayes, R. J., et al. (2015). Nondefoliating and defoliating strains from cotton correlate with races 1 and 2 of *Verticillium dahliae*. *Plant Dis.* 99, 1713–1720. doi: 10.1094/PDIS-03-15-0261-RE
- Iglesias-García, A. M., Villarroel-Zeballos, M. I., Feng, C., du Toit, L. J., and Correll, J. C. (2013). Pathogenicity, virulence, and vegetative compatibility grouping of *Verticillium* isolates from spinach seed. *Plant Dis.* 97, 1457–1469. doi: 10.1094/PDIS-01-13-0016-RE
- Jaccard, P. (1908). Nouvelles recherches Sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.* 44, 223–270.
- Jiménez-Díaz, R. M., Mercado-Blanco, J., Olivares-García, C., Collado-Romero, M., Bejarano-Alcázar, J., Rodríguez-Jurado, D., et al. (2006). Genetic and virulence diversity in *Verticillium dahliae* populations infecting artichoke in eastern-central Spain. *Phytopathology* 96, 288–298. doi: 10.1094/PHYTO-96-0288
- Jiménez-Díaz, R. M., Olivares-García, C., Landa, B. B., Jiménez-Gasco, M. M., and Navas-Cortés, J. A. (2011). Region-wide analysis of genetic diversity in *Verticillium dahliae* populations infecting olive in southern Spain and agricultural factors influencing the distribution and prevalence of vegetative compatibility groups and pathotypes. *Phytopathology* 101, 304–315. doi: 10.1094/PHYTO-07-10-0176
- Jiménez-Díaz, R. M., Olivares-García, C., Trapero-Casas, J. L., Jiménez-Gasco, M. M., Navas-Cortés, J. A., Landa, B. B., et al. (2017). Variation of pathotypes and races and their correlations with clonal lineages in *Verticillium dahliae*. *Plant Pathol.* 66, 651–666. doi: 10.1111/ppa.12611
- Korolev, N., and Katan, T. (1997). Improved medium for selecting nitrate nonutilizing (nit) mutants of *Verticillium dahliae*. *Phytopathology* 87, 1067–1070. doi: 10.1094/PHYTO.1997.87.10.1067
- Korolev, N., Katan, T., and Katan, J. (2009). “Physiological races and vegetative compatibility groups among *Verticillium dahliae* isolates from tomato in Israel,” in *Proceedings of the 2nd International Symposium on Tomato Diseases*, Kusadasi, 57–64. doi: 10.17660/ActaHortic.2009.808.7
- Korolev, N., Pérez-Artés, E., Mercado-Blanco, J., Bejarano-Alcázar, J., Rodríguez-Jurado, D., Jiménez-Díaz, R., et al. (2008). Vegetative compatibility of cotton defoliating *Verticillium dahliae* in Israel and its pathogenicity to various hosts. *Eur. J. Plant Pathol.* 122, 603–617. doi: 10.1007/s10658-008-9330-1
- López-Escudero, F. J., and Mercado-Blanco, J. (2011). *Verticillium* wilt of olive: a case study to implement an integrated strategy to control a soil-borne pathogen. *Plant Soil* 344, 1–50. doi: 10.1007/s11104-010-0629-2
- Maruthachalam, K., Atallah, Z. K., Vallad, G. E., Klosterman, S. J., Hayes, R. J., Davis, R. M., et al. (2010). Molecular variation among isolates of *Verticillium dahliae* and polymerase chain reaction based differentiation of races. *Phytopathology* 100, 1222–1230. doi: 10.1094/PHYTO-04-10-0122
- McIntosh, M. S. (1983). Analysis of combined experiments. *Agron. J.* 75, 153–155. doi: 10.2134/agronj1983.00021962007500010041x
- Mercado-Blanco, J., Rodríguez-Jurado, D., Parrilla-Araujo, S., and Jiménez-Díaz, R. M. (2003). Simultaneous detection of the defoliating and nondefoliating *Verticillium dahliae* pathotypes in infected olive plants by duplex, nested polymerase chain reaction. *Plant Dis.* 87, 1487–1494. doi: 10.1094/PDIS.2003.87.12.1487
- Mercado-Blanco, J., Rodríguez-Jurado, D., Pérez-Artés, E., and Jiménez-Díaz, R. M. (2001). Detection of the nondefoliating pathotype of *Verticillium dahliae* in infected olive plants by nested PCR. *Plant Pathol.* 50, 609–619. doi: 10.1046/j.1365-3059.2001.00601.x
- Mercado-Blanco, J., Rodríguez-Jurado, D., Pérez-Artés, E., and Jiménez-Díaz, R. M. (2002). Detection of the defoliating pathotype of *Verticillium dahliae* in infected olive plants by nested-PCR. *Eur. J. Plant Pathol.* 108, 1–13. doi: 10.1023/A:1013994827836
- Molinero-Ruiz, L., Delavault, P., Pérez-Vich, B., Pacureanu-Joita, M., Bulos, M., Altieri, E., et al. (2015). History of the race structure of *Orobanche cumana* and the breeding of sunflower for resistance to the parasitic weed: a review. *Span. J. Agric. Res.* 13:e10R01. doi: 10.5424/sjar/2015134-8080
- Navas-Cortés, J. A., Olivares, C., Trapero-Casas, J. L., Landa, B. B., Jiménez-Gasco, M. M., and Jiménez-Díaz, R. M. (2009). “The influence of agronomic factors on prevalence and distribution of *Verticillium dahliae* vegetative compatibility groups and pathotypes infecting olive in Andalusia, southern Spain,” in *Proceedings of the Abstracts Book 10th International Verticillium Symposium*, de Corfu, 79.
- Pegg, G. F., and Brady, B. L. (2002). *Verticillium Wilts*. Wallingford: CAB International. doi: 10.1079/9780851995298.0000
- Pérez-Artés, E., García-Pedrajas, M. D., Bejarano-Alcázar, J., and Jiménez-Díaz, R. M. (2000). Differentiation of cotton-defoliating and nondefoliating pathotypes of *Verticillium dahliae* by RAPD and specific PCR analyses. *Eur. J. Plant Pathol.* 106, 507–517. doi: 10.1094/PHYTO-98-2-0167
- Putt, E. D. (1958). Note on resistance of sunflowers to leaf mottle disease. *Can. J. Plant Sci.* 38, 274–276. doi: 10.4141/cjps58-044
- Putt, E. D. (1964). Breeding behavior of resistance to leaf mottle disease or *Verticillium* in sunflowers. *Crop Sci.* 4, 177–179. doi: 10.2135/cropsci1964.0011183X000400020016x
- Radi, S. A., and Gulya, T. J. (2007). “Sources of resistance to a new strain of *Verticillium dahliae* on sunflower in North America-2006,” in *Proceedings of the 29th Sunflower Research Workshop*, Bismarck, ND, 10–11.
- Rodríguez-Jurado, D., Blanco-López, M. A., Rapoport, H., and Jiménez-Díaz, R. M. (1993). Present status of *Verticillium* wilt of olive in Andalusia (Southern Spain). *Bull. OEPP EPPO Bull.* 23, 513–516. doi: 10.1111/j.1365-2338.1993.tb01362.x
- Sandoya, G. V., Gurung, S., Short, D. P., Subbarao, K. V., Michelmore, R. W., and Hayes, R. J. (2017). Genetics of resistance in lettuce to races 1 and 2 of *Verticillium dahliae* from different host species. *Euphytica* 213:20. doi: 10.1007/s10681-016-1813-0
- Short, D. P. G., Gurung, S., Maruthachalam, K., Atallah, Z. K., and Subbarao, K. V. (2014). *Verticillium dahliae* race 2-specific PCR reveals a high frequency of



- race 2 strains in commercial spinach seed lots and delineates race structure. *Phytopathology* 104, 779–785. doi: 10.1094/PHYTO-09-13-0253-R
- Strausbaugh, C. A., Eujayl, I. A., and Martin, F. N. (2016). Pathogenicity, vegetative compatibility and genetic diversity of *Verticillium dahliae* isolates from sugar beet. *Can. J. Plant Pathol.* 38, 492–505. doi: 10.1080/07060661.2016.1260639
- Tourvieille de Labrouhe, D., Gulya, T. J., Masirevic, S., Penaud, A., Rashid, K. Y., and Virányi, F. (2000). “New nomenclature of races of *Plasmopara halstedii* sunflower downy mildew,” in *Proceedings of the 15th International Sunflower Conference*, (Toulouse: International Sunflower Association), 61–66.
- Usami, T., Ishigaki, S., Takashina, H., Matsubara, Y., and Amemiya, Y. (2007). Cloning of DNA fragments specific to the pathotype and race of *Verticillium dahliae*. *J. Gen. Plant Pathol.* 73, 89–95. doi: 10.1007/s10327-006-0334-4
- Usami, T., Momma, N., Kikuchi, S., Watanabe, H., Hayashi, A., Mizukawa, M., et al. (2017). Race 2 of *Verticillium dahliae* infecting tomato in Japan can be split into two races with differential pathogenicity on resistant rootstocks. *Plant Pathol.* 66, 230–238. doi: 10.1111/ppa.12576
- Vallad, G. E., Qin, Q. M., Grube, R., Hayes, R. J., and Subbarao, K. V. (2006). Characterization of race-specific interactions among isolates of *Verticillium dahliae* pathogenic on lettuce. *Phytopathology* 96, 1380–1387. doi: 10.1094/PHYTO-96-1380
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Martin-Sanz, Rueda, García-Carneros, González-Fernández, Miranda-Fuentes, Castuera-Santacruz and Molinero-Ruiz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Fruit Phenolic Profiling: A New Selection Criterion in Olive Breeding Programs

Ana G. Pérez<sup>1\*</sup>, Lorenzo León<sup>2</sup>, Carlos Sanz<sup>1</sup> and Raúl de la Rosa<sup>2</sup>

<sup>1</sup> Department of Biochemistry and Molecular Biology of Plant Products, Instituto de la Grasa, CSIC, Seville, Spain, <sup>2</sup> Instituto Andaluz de Investigación y Formación Agraria, Pesquera, Alimentaria y de la Producción Ecológica (IFAPA), Centro Alameda del Obispo, Córdoba, Spain

## OPEN ACCESS

### Edited by:

Mariela Torres,  
Instituto Nacional de Tecnología  
Agropecuaria (INTA), Argentina

### Reviewed by:

Agnese Taticchi,  
University of Perugia, Italy  
Primo Proietti,  
University of Perugia, Italy  
Arnon Dag,  
Agricultural Research Organization  
(Israel), Israel

### \*Correspondence:

Ana G. Pérez  
agracia@cica.es

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 11 October 2017

**Accepted:** 12 February 2018

**Published:** 27 February 2018

### Citation:

Pérez AG, León L, Sanz C and de la  
Rosa R (2018) Fruit Phenolic Profiling:  
A New Selection Criterion in Olive  
Breeding Programs.  
Front. Plant Sci. 9:241.  
doi: 10.3389/fpls.2018.00241

Olive growing is mainly based on traditional varieties selected by the growers across the centuries. The few attempts so far reported to obtain new varieties by systematic breeding have been mainly focused on improving the olive adaptation to different growing systems, the productivity and the oil content. However, the improvement of oil quality has rarely been considered as selection criterion and only in the latter stages of the breeding programs. Due to their health promoting and organoleptic properties, phenolic compounds are one of the most important quality markers for Virgin olive oil (VOO) although they are not commonly used as quality traits in olive breeding programs. This is mainly due to the difficulties for evaluating oil phenolic composition in large number of samples and the limited knowledge on the genetic and environmental factors that may influence phenolic composition. In the present work, we propose a high throughput methodology to include the phenolic composition as a selection criterion in olive breeding programs. For that purpose, the phenolic profile has been determined in fruits and oils of several breeding selections and two varieties (“Picual” and “Arbequina”) used as control. The effect of three different environments, typical for olive growing in Andalusia, Southern Spain, was also evaluated. A high genetic effect was observed on both fruit and oil phenolic profile. In particular, the breeding selection UCI2-68 showed an optimum phenolic profile, which sums up to a good agronomic performance previously reported. A high correlation was found between fruit and oil total phenolic content as well as some individual phenols from the two different matrices. The environmental effect on phenolic compounds was also significant in both fruit and oil, although the low genotype × environment interaction allowed similar ranking of genotypes on the different environments. In summary, the high genotypic variance and the simplified procedure of the proposed methodology for fruit phenol evaluation seems to be convenient for breeding programs aiming at obtaining new cultivars with improved phenolic profile.

**Keywords:** *Olea europaea*, olive breeding, virgin olive oil, phenolic compounds, genotype, genotype × environment interaction

## INTRODUCTION

Virgin olive oil (VOO) is a key food within the Mediterranean diet whose daily intake has well-known benefits for human health (Estruch et al., 2013). Olive oil has traditionally been produced and consumed in the Mediterranean countries. Thus, Europe is responsible for 78% of the olive oil world production, and Spain is the largest producer with an average production of 1.3 million tons over the past 7 years, being “Picual” and “Arbequina” cultivars the two most important Spanish varieties in terms of oil production (FAOSTAT, 2016<sup>1</sup>). The olive possesses a substantial genetic diversity (Belaj et al., 2012) with more than 1,200 olive varieties cataloged (Bartolini et al., 2005). However, most of those cultivars are traditional as few attempts have been made to produce new varieties by systematic breeding (Bellini et al., 2002; Rallo et al., 2008; Lavee, 2013). This can be due to the limited knowledge on the real variability of olive germplasm for many of the most important agronomic and oil quality traits. Also, the long juvenile phase, high heterozygosity and scarce information on trait heritability have been important limiting factors that have negatively affected olive breeding (de la Rosa et al., 2016).

In the past few years, olive growing and olive oil production had shown an exponential increase in non-Mediterranean countries (FAOSTAT, 2016<sup>1</sup>). The emergence of these new olive producing areas with very different edaphoclimatic conditions respect to those of the Mediterranean countries and the necessary adaptation to intensive production systems and mechanical harvesting have driven the demand for new olive varieties. All these factors have significantly boosted the development of new and more ambitious olive breeding programs (Lavee, 2013). Thus, the objectives of most recent breeding programs are not only agronomic. In these sense, two different marketing strategies need to be fulfilled (i) producing standard quality extra-virgin olive oil at lower prices, and (ii) offering consumers a variety of extra-virgin olive oils with high quality standards and different sensory profiles. The first approach is usually related to super-intensive cultivation and highly mechanized harvesting methods, while the second is associated to preserving olive tree biodiversity and traditional methods as part of the extraordinary food traditions associated with the Mediterranean diet (Ilarioni and Proietti, 2014).

The increasing importance of the nutritional quality of olive oil for consumers and markets has led to olive-breeding programs with new nutritional targets (El Riachy et al., 2012; Rugini and De Pace, 2016). In this sense, although VOO contains a number of minor compounds with interesting biological activities, it is generally accepted that the phenolic compounds are the oil components most directly associated with its health related properties (Servili et al., 2014; Bernardini and Visioli, 2017). In addition to their nutritional properties, the phenolic compounds of VOO also have important organoleptic implications since they are the main contributors to bitter and pungent sensory descriptors. The secoiridoids compounds, containing in their molecules the phenolic alcohol tyrosol (*p*-HPEA) or its hydroxyl derivative hydroxytyrosol (3,4-DHPEA), are the most

abundant class of phenolics in all olive products. Thus, the main phenolic glucosides present in the olive fruit are the secoiridoids oleuropein, ligstroside and demethyloleuropein, and their hydrolytic derivatives, the dialdehydic forms of decarboxymethyloleuropein and decarboxymethylligstroside aglucones (3,4-DHPEA-EDA and *p*-HPEA-EDA, respectively) and the aldehydic forms of oleuropein and ligstroside aglucones (3,4-DHPEA-EA and *p*-HPEA-EA, respectively) are the main phenolic components in most olive oils (Montedoro et al., 2002; Pérez et al., 2014). Many studies reporting the ability of VOO phenolics to reduce chronic inflammation and oxidative damage relate these beneficial effects of VOO with the level of 3,4-DHPEA in plasma (Mateos et al., 2011; Bernardini and Visioli, 2017). This scientific evidence has led the European Union to approve a health claim on olive oil polyphenols which may be applied only for oils containing at least 250 mg/kg of hydroxytyrosol and its derivatives (European Commission, 2012).

The metabolism of phenolic compounds in the olive tree is very complex, and is modulated by genetic (Talhaoui et al., 2016) and environmental factors (Romero and Motilva, 2010) that determine the final phenolic composition of olive fruits. In a similar way, the influence of agricultural practices such as limited irrigation (Cirilli et al., 2017), optimization of pruning to increase light availability (Proietti et al., 2012) or selection of optimum harvest date (Famiani et al., 2000) have also been described. The phenolic glycosides present in the olive fruit are later transformed and modified by endogenous hydrolytic and oxidative enzymes that are activated during the oil extraction process. In this way, the phenolic profile of VOO is directly related to the phenolic content of the olive fruit (Gómez-Rico et al., 2008) and the activity of hydrolytic and oxidative enzymes during milling and malaxation (Romero-Segura et al., 2011). Any breeding program that aims to improve the phenolic composition of the VOO should consider an experimental design that allows the evaluation of all these factors. Likewise, it is important to have accurate analytical tools to find out the influence of each specific factor on the final phenolic composition of VOO. Olive breeding programs focusing on oil quality have additional limitations to those previously mentioned (de la Rosa et al., 2016) due to the very large number of genotypes but very little oil production achieved in the early stages of breeding. For this reason, it is vital to have reliable analytical methodologies to predict the composition of the oil from the analysis of the fruits. In this sense, recent studies have described significant correlations between the composition of olive fruits and oils for components such as fatty acids, sterols, tocopherols, or squalene (Velasco et al., 2014; de la Rosa et al., 2016). To the best of our knowledge, fruit phenolic profiling has not previously been used as selection criteria in olive breeding. Besides, there are no previous reports on olive comparative trials devoted to study the interaction of genetic and environmental factors that may influence fruit phenolic composition which in turn determines the phenolic profile of VOO.

The objective of this work is to describe the use of new analytical tools that allow predicting the phenolic composition of the oils from the phenolic profiling of the olive fruits, without

<sup>1</sup><http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor>

the previous step of oil extraction, and therefore facilitating the scrutiny of large seedling populations. The predictive method developed has been used to select a new olive breeding selection producing VOO with an optimum phenolic composition.

## MATERIALS AND METHODS

### Plant Material

The selection process on the olive breeding program of Cordoba, (Spain) is divided in three steps, seedling stage, intermediate selection and final comparative trials (León et al., 2015). In each step, the number of genotypes is reduced and the number of replications per genotype increases (de la Rosa et al., 2016). The final comparative trials are planted in different environments with a reduced number of selections according to the potential adaptability to each edaphoclimatic conditions.

In the present work, phenolic evaluation was carried out in three comparative trials planted in typical olive growing areas of Andalusia, Sothern Spain: Córdoba, Morón, and Ubeda. Ubeda has lower winter temperatures and rainfall than the other two locations, while Cordoba has the highest rainfall (Table 1). Soils mainly differ on the clay percentage (42.0% in Moron, 31.9% in Ubeda and 22.3% in Cordoba). Two breeding selections (UCI-2-68 and UCI-5-65) were planted in the selected locations together with their parents (“Picual” and “Arbequina”), following an unbalanced design. These two selections showed high productivity, oil content and oleic acid percentage in previous breeding selection stages (León et al., 2004a,b, 2007, 2011). Besides, four more breeding selections were also included in the Ubeda trial (UCI-12-85, UCI-12-104, UCI-19-79, and UCI-19-60) for having good productivity and high oil content (unpublished data). All the breeding selections come from crosses performed in 1992 to 1997, between “Picual” and “Arbequina.” All the trials have a randomized complete design with 3–4 blocks and 3–4 trees per elementary plot. Trees were

planted in 2011 at  $6 \times 7$  m distance and standard fertilization and irrigation practices were carried out. Irrigation supply ( $1,500 \text{ m}^3$  per year and h) by in-line drippers was used to avoid water stress of plants.

Fruit samples of 2 kg were harvested per each genotype and elementary plot in 3 blocks per each of the three trials (Córdoba, Morón, and Ubeda). Breeding selections and parents were collected in a common date, in mid-November 2016, typical for olive harvesting in southern Spain, when most fruits were at turning color (de la Rosa et al., 2013). Random subsamples were taken for both direct fruit analysis and oil extraction.

### Chemicals

Reagents for extraction and other measurements were supplied by Sigma-Aldrich (St. Louis, MO). Oleuropein, verbascoside, luteolin-7glucoside, apigenin-7glucoside, rutin, apigenin, luteolin, tyrosol, hydroxytyrosol, vanillic acid, vainillin, p-coumaric acid, and ferulic acid were obtained from Sigma Chemical (St. Louis, MO, USA) and Extrasynthese (Genay, France). Non-commercial phenolic standards such as ligstroside, hydroxytyrosol-1-glucoside, or the main secoiridoids derivatives were obtained from olive leaves, fruits and oils using a high performance liquid chromatography (HPLC) preparative system.

### Olive Oil Extraction

Olive oil was extracted using an Abencor analyzer (Comercial Abengoa, S.A., Seville, Spain) that simulates the industrial process of VOO production on a laboratory scale (Martínez et al., 1975). Processing parameters have been precisely described in a previous study (Pérez et al., 2014).

### Extraction and Analysis of Fruit and VOO Phenolic Compounds

Fruit phenolic compounds were extracted according to a previously developed protocol (García-Rodríguez et al., 2011).

**TABLE 1 |** Mean temperatures (maximum, minimum, and average) and monthly rainfall during 2016 in the three locations studied.

	Cordoba	Moron	Ubeda	Cordoba	Moron	Ubeda	Cordoba	Moron	Ubeda	Cordoba	Moron	Ubeda
	Avg temp (°C)			Max temp (°C)			Min temp (°C)			Rainfall (mm)		
January	10.6	11.7	9.2	15.7	16.2	16.1	6.4	7.6	4.0	59.6	67.4	44.6
February	10.9	11.3	8.5	16.4	16.3	14.6	5.6	6.8	3.4	42.6	55.8	49.8
March	11.4	11.8	9.1	19.0	18.4	16.6	4.5	5.9	2.3	30.2	37.6	20.6
April	15.7	15.4	14.1	22.0	21.5	21.6	10.0	9.9	6.9	115.4	79.8	77.8
May	18.6	18.7	17.6	25.1	25.1	26.0	12.6	12.8	9.7	92.4	94.0	56.4
June	24.9	24.7	24.5	33.1	32.7	34.3	16.3	16.5	13.4	0.2	0.0	0.0
July	29.1	28.3	28.8	37.3	36.8	38.7	20.5	20.4	18.1	0.4	1.0	6.4
August	28.5	28.4	27.8	36.9	36.4	37.6	20.2	20.7	17.6	0.2	0.8	2.4
September	24.9	24.7	22.9	33.0	32.9	33.4	16.8	17.1	13.3	3.0	6.4	4.2
October	19.6	20.1	17.7	26.5	26.9	27.0	14.2	14.4	10.6	84.0	87.2	25.0
November	12.5	13.0	10.3	18.4	18.6	18.1	8.1	8.3	4.7	142.2	93.8	90.6
December	10.4	11.6	8.6	16.5	17.1	16.6	6.2	6.7	3.6	45.0	38.2	27.0
Mean temp.	18.1	18.3	16.7	25.0	24.9	25.1	11.8	12.3	9.0			
Total rainfall										615.2	562.0	404.8



Longitudinal pieces of mesocarp tissue were cut from 20 olive fruits and kept at 4°C for 24 h in DMSO (6 ml/g of fruit), containing syringic acid (24 mg/ml) as internal standard. The extracts were filtered through a 0.45 µm mesh nylon and kept at –20°C until HPLC analysis.

VOO phenolics were isolated by solid phase extraction (SPE) on a diol-bonded phase cartridge (Supelco, Bellefonte, PA) following a previously described procedure (Mateos et al., 2001). 0.5 ml of a methanol solution containing *p*-hydroxyphenyl-acetic acid and *o*-coumaric acid as internal standards was added to each oil sample (2.5 g) before the extraction.

Phenolic extracts from fruits and oils were analyzed by HPLC on a Beckman Coulter liquid chromatography system equipped with a System Gold 168 detector, a solvent module 126, an autosampler module 508 and a Waters column heater module following a previously described methodology (Pérez et al., 2014). A Superspher RP 18 column (4.6 mm i.d. × 250 mm, particle size 4 µm; Dr Maisch GmbH, Germany) at flow rate 1 mL min<sup>–1</sup> and a temperature of 35°C was used for all the analyses. A total of 15 phenolic compounds were analyzed in fruit phenolic extracts: hydroxytyrosol-4-glucoside, hydroxytyrosol-1-glucoside, demethyloleuropein, verbascoside, luteolin-7-glucoside, demethyligstroside, rutin, oleuropein, comselogside, ligstroside, luteolin 3,4-DHPEA-EA, apigenin, and *p*-HPEA-EA. The last four compounds were also analyzed in VOO extracts in which 12 other phenolic compounds were also detected: hydroxytyrosol, tyrosol, vanillic acid, vainillin, *p*-coumaric acid, hydroxytyrosol acetate, 3,4-DHPEA-DEA, *p*-HPEA-DEA, pinorelinol, acetoxypinorelinol, and ferulic acid. The quantification of flavones and ferulic acid was done at 335 nm while the rest of phenolic components were quantitated at 280 nm. Response factors were calculated for each phenolic compound as described previously (García-Rodríguez et al., 2011).

The tentative identification of compounds by their UV-vis spectra was confirmed by HPLC/ESI-qTOF-HRMS. The liquid chromatograph system was Dionex Ultimate 3000 RS U-HPLC liquid chromatograph system (Thermo Fisher Scientific, Waltham, MA, USA) equipped with a similar Superspher RP 18 column but with formic acid (1%) instead of phosphoric acid (0.5%) in the mobile phase. A split post-column of 0.4 mL/min was introduced directly on the mass spectrometer electrospray ion source. The HPLC/ESI-qTOF operated for mass analysis using a microTOF-QII High Resolution Time-of-Flight mass spectrometer (UHRTOF) with qQ-TOF geometry (Bruker Daltonics, Bremen, Germany) equipped with an electrospray ionization (ESI) interface. Mass spectra were acquired in MS fullscan mode and data were processed using TargetAnalysis 1.2 software (Bruker Daltonics, Bremen, Germany).

## Statistical Analysis

Fruit samples (2 kg) were harvested per each cultivar and elementary plot in 3 blocks per each trial and random subsamples were taken for both direct fruit analysis and oil extraction. All the data were statistically evaluated using STATISTICA 5.0 (Statsoft Inc., Tulsa, OK, USA). Descriptive statistics and

variability plots were obtained for the whole dataset of genotypes and environments. Correlations among phenols or group of phenols were analyzed for the whole dataset using Pearson's correlations (at  $p \leq 0.05$ ;  $p \leq 0.01$ ;  $p \leq 0.001$ ). A subset of data (4 genotypes and 3 environments) was used to evaluate the relative contribution of genetic and environmental factors on the phenolic variability by means of ANOVA and separation of the means was obtained at  $p \leq 0.05$  by least significance differences (LSD). Principal component analysis (PCA) was used to evaluate the levels of association among the phenolic compounds from the cultivars and advanced breeding selections under study. PCA was applied to the same subset of data used for ANOVA (4 genotypes and 3 environments) and to a second subset of data containing all the genotypes in a single environment.

## RESULTS

### Identification and Quantitation of the Main Phenolic Components of Olive Fruits and Oils from New Advanced Breeding Selections

The analysis of the fruit phenolic extracts allowed identifying a great variability in the profile of phenolic glycosides of the eight olive genotypes and three environments under study. Demethyloleuropein, was the most abundant phenolic compound with a mean content of 8,787.4 µg/g fruit and a range of variability from 26.1 to 23,937.9 µg/g fruit. This was followed by oleuropein with a significantly lower mean value (4,966.2 µg/g) and a lower range (175.2–16,542.2 µg/g). The mean contents of verbascoside, ligstroside and luteolin-7-glucoside were considerably lower than those of demethyloleuropein and oleuropein: 825.2, 497.2, and 391.06 µg/g respectively.

Similarly, a great variability was also observed in terms of phenolic components of the oils. Tyrosol (*p*-HPEA) and hydroxytyrosol (3,4-DHPEA) derived compounds were the most important class of phenolic compounds found in all VOOs. Among them 3,4-DHPEA-EDA was the most abundant phenolic component in the studied oils with a mean content of 234.1 µg/g of oil in the range 27.1–576.81 µg/g. The tyrosol derivative *p*-HPEA-EDA was the second most abundant component (mean value of 148.2 µg/g and range 23.3–444.11 µg/g) followed by another hydroxytyrosol derivative 3,4 DHPEA-EA (mean value 85.3 and range 14.2–410.1 µg/g) while lower contents were found for *p*-HPEA-EA (mean value 15.87 µg/g) and 3,4-DHPEA-acetate (mean value 11.06 µg/g). Significant amounts of lignans (acetoxypinorelinol and pinorelinol), flavones (luteolin and apigenin), and phenolic acids (cinnamic acid, *p*-coumaric acid, vanillic acid, and ferulic acid) were also found in the oils obtained from the eight genotypes and three environments analyzed. The highest variability ranges found among them, although less important from a quantitative point of view than those previously mentioned for tyrosol derived compounds, were those of acetoxypinorelinol which possesses promising anticancer activity (mean value 23.7 and range 10.2–74.33,4 µg/g) (Menéndez et al., 2008) and luteolin (mean value 7.32 µg/g).

## Genetic and Environmental Effects on the Phenolic Composition of Olive Fruit and Oil

The well-known influence of edaphoclimatic parameters on the quality of VOO makes extremely important to include different environments in the evaluation of advanced olive breeding selections (León et al., 2016). In the present work, the phenolic evaluation was carried out in three comparative trials (Morón, Córdoba, and Ubeda), representing different edaphoclimatic areas in Southern Spain. The variability specifically induced by both factors, genotype and environment, on the phenolic profile of fruits and oils was analyzed. **Figure 1**, graphically shows the variability observed in terms of the total phenolic content in the fruits, and in the content of the three main secoiridoid glycosides. Thus, among the genotypes grown in different environments, the highest phenolic content was always found in fruits grown in the environment Ubeda, while the lowest content was generally associated to fruits from Córdoba. However, the range of this environmental variability was different for each genotype. **Figure 2** graphically shows the variability plots of the main phenolic components found in VOO. The influence of the environmental factor on the oils was in good agreement with that observed in the fruits. As previously reported for fruit phenolic, in those genotypes grown in the three different environments, higher phenolic contents were always associated to oils from the environment Ubeda. The highest phenolic contents were found in oils obtained from breeding selections UCI 12-85 and UCI 12-104 grown in Ubeda. Besides their extremely high phenolic content, these two genotypes also possess very high levels of p-HPEA-EDA (320 and 450  $\mu\text{g/g}$  oil, respectively), which is closely related to the pungency of VOO (Andrewes et al., 2003).

In order to evaluate the relative contribution of genetic and environmental factors on the variability observed in the phenolic content of fruits and oils a subset of data from the genotypes “Arbequina,” “Picual,” UCI 5-65, and UCI 2-68, was subjected to analysis of variance. Minor phenolic components were excluded and only 15 phenolic variables were included in the analysis. Four variables were selected in fruits (total phenolics; demethyloleuropein, oleuropein and ligstroside) and 11 variables were selected in VOOs (total phenolics; 3,4 DHPEA-EDA; p-HPEA-EDA; 3,4 DHPEA-EA; HPEA-EA, 3,4 DHPEA-acetate, pinoresinol, acetoxypinoresinol, luteolin, apigenin, and the sum of the main phenolic acids (cinnamic acid, p-coumaric acid, vanillic acid, and ferulic acid). The influence of each factor was estimated by the percent of the total variance (**Figure 3**). Analysis of variance shows a significant effect ( $p \leq 0.05$ ) of genotype for all the phenolic compounds analyzed in fruits and VOOs. The environment was the major contributor to the variance of the total phenolic content of the fruits (66.9%), fruit secoiridoids (59.7%), oleuropein content (46.8%), and to a lesser extent to the variance of the total phenolic content of the oils (29.99%). On the contrary, the genotype  $\times$  environment effect was only significant ( $p \leq 0.05$ ) for demethyloleuropein and 3,4 DHPEA-EDA. To analyze this information in detail, the mean values of the 15 selected variables in the four genotypes and three environments analyzed were compared (**Table 2**). Significant differences were found for key phenolic components among the

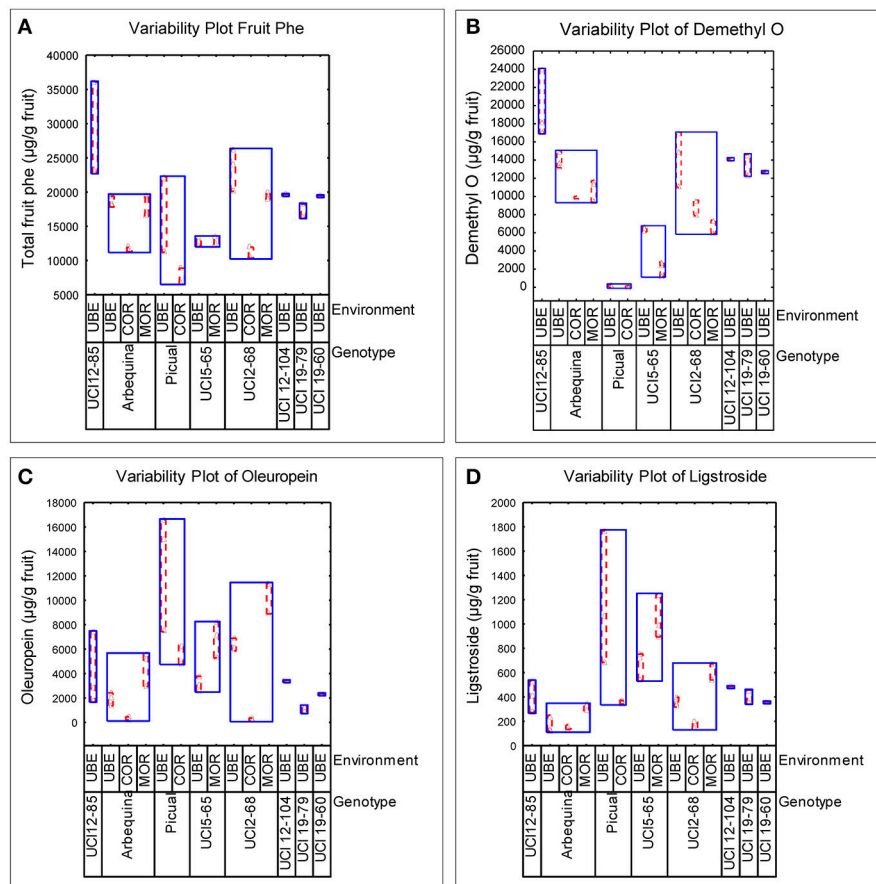
four genotypes. UCI 2-68 displayed the highest phenolic content in both, fruit and VOO. The content of demethyloleuropein, the most abundant phenolic glycoside found in the olive fruits analyzed was significantly higher in UCI 2-68 and “Arbequina,” with moderate levels found in UCI 5-65 and very low amount detected in “Picual” fruits. The genotype UCI 2-68 also had the highest contents of 3,4 DHPEA-EDA, p-HPEA-EDA, 3,4 DHPEA-acetate, and pinoresinol.

## Relationship between Fruit and VOO Phenolic Components

To investigate the possibility to predict the phenolic composition of the VOO from the analysis of the phenolic profiles of the fruits Pearson's correlation coefficients were computed using the data obtained from all the fruits and oils analyzed in this study (**Table 3**). Significant positive correlation was found between total fruit phenolics and total VOO phenolic contents ( $r = 0.685$ ) and even a slightly higher coefficient was calculated for fruit secoiridoids compounds and the total phenolic content of the oil ( $r = 0.735$ ). The highest correlation found between individual fruit phenolic compounds and VOO phenolic content was found for demethyloleuropein ( $r = 0.650$ ). Curiously, oleuropein content correlated poorly to total fruit phenolics ( $r = 0.289$ ) and no correlation at all was found between oleuropein content and the total phenolic content of the oil ( $r = -0.057$ ). Data shown in **Table 3** also provides relevant information on the relationship between the different phenolic components of VOO. Thus, 3,4-DHPEA-EDA was highly correlated to total AOV phenolics ( $r = 0.850$ ) followed by p-HPEA-EDA ( $r = 0.759$ ) while significantly lower correlation coefficients were found for the secoiridoids with monoaldehyde structure, 3,4-DHPEA-EA and p-HPEA-EA ( $r = 0.217$  and  $r = 0.144$ , respectively).

## Phenolic Profiling as a Selection Criterion in Olive Breeding

PCA analysis was first applied to the same subset of data used for the variance analysis (four genotypes and three environments). The first and second principal components described 60% of the total variability (PC1 35.66 and PC2 23.65%). PC1 was strongly linked to demethyloleuropein ( $r = 0.90$ ), 3,4-DHPEA-EDA ( $r = 0.82$ ), p-HPEA-EDA ( $r = 0.78$ ) while negatively correlated to oleuropein ( $r = -0.67$ ), 3,4-DHPEA-EA ( $r = -0.71$ ), p-HPEA-EA ( $r = -0.67$ ). PC2 was positively correlated to total VOO phenolic content ( $r = 0.72$ ), pinoresinol ( $r = 0.633$ ) and to a lesser extent to acetoxypinoresinol ( $r = 0.46$ ) and negatively correlated to luteolin and apigenin content ( $r = -0.57$  and  $r = -0.78$ , respectively). The PCA bi-plot (**Figure 4**) shows the contribution of the PCA analysis to sample profiling for genotype and environment. The distribution of phenolic profiles in the PCA bi-plot suggests a greater influence of genotype vs. environment. In this sense, the distinction of different environments within each genotype is not evident in the four genotypes studied. Thus, the profiles of UCI 5-65 and UCI 2-68 grown in different environments were only partially segregated in the scatter-plot. On the contrary, when PCA analysis was applied to all the breeding selections grown in the environment Ubeda



**FIGURE 1 |** Variability plots of the main phenolic compounds (µg/g fruit pulp) analyzed in fruits from different genotypes and environments. **(A)** Total fruit phenolics. **(B)** Demethyloleuropein. **(C)** Oleuropein. **(D)** Ligstroside.

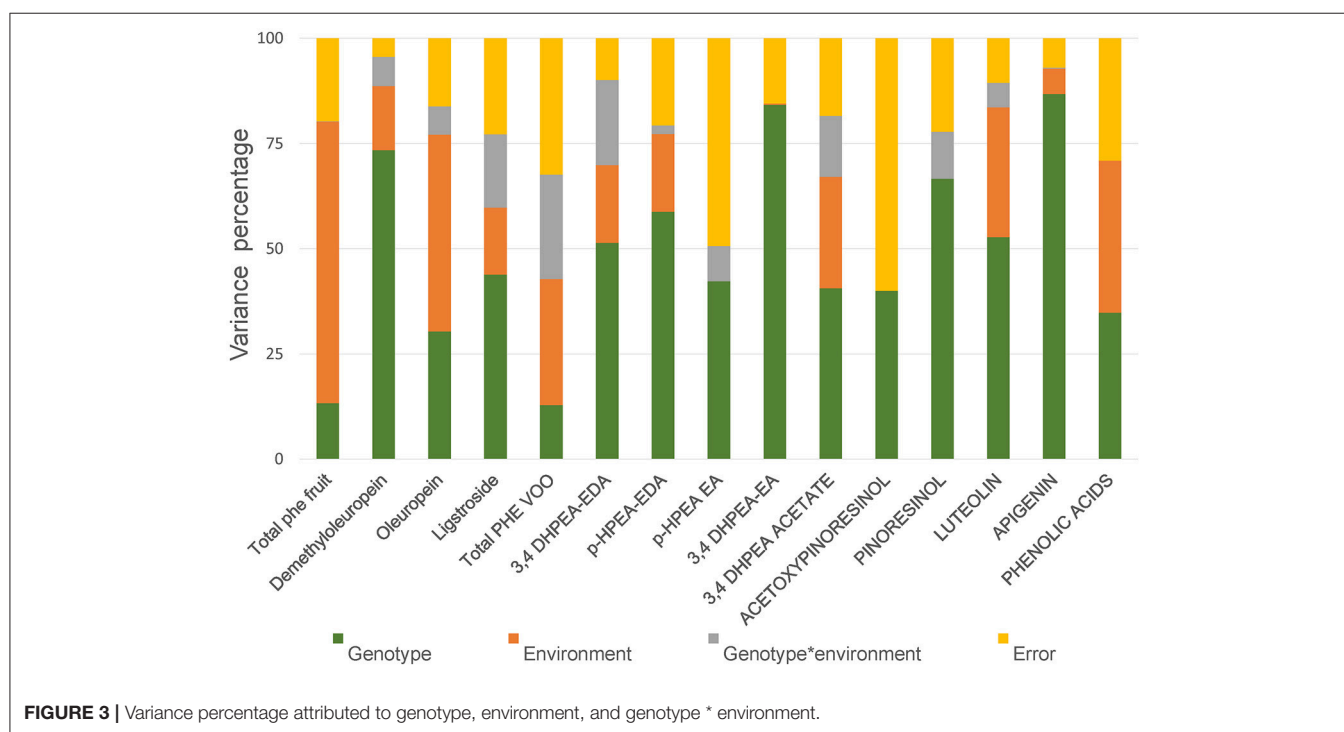
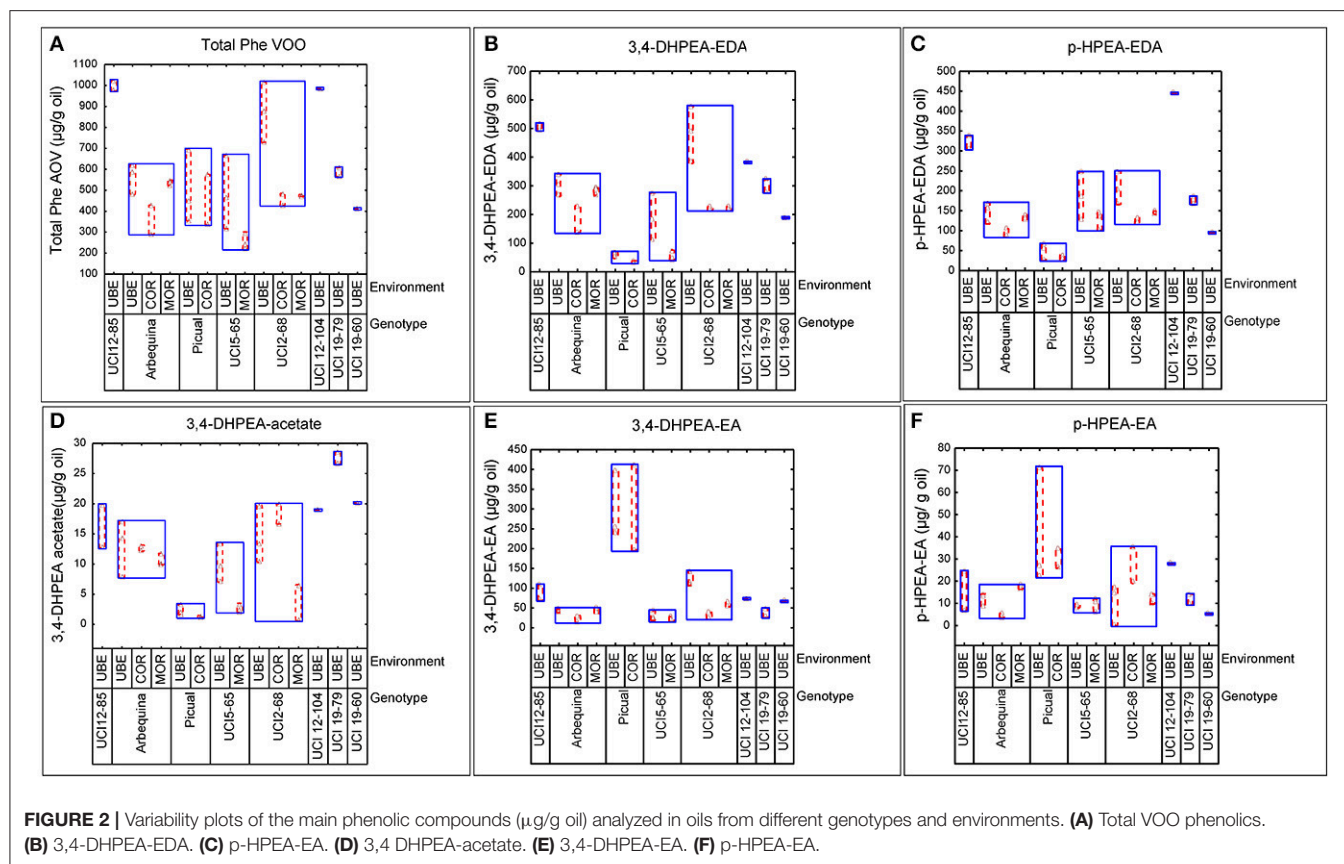
the eight genotypes were clearly separated in the scatter plot (Figure 5).

## DISCUSSION

### Phenolic Composition of New Advanced Olive Breeding Selections. Influence of Genetic and Environmental Factors

The major aim of the olive breeding program of Cordoba is to select new olive cultivars that, together with their good agronomic characteristics, are able to produce oils with an optimum nutritional and organoleptic quality. The genotypes included in this study are advanced selections from this program which in previous breeding selection stages have shown good productivity and high oil content (León et al., 2004a,b, 2007, 2011). The phenolic profiles of fruits and oils of these new breeding selections together with their parents (“Picual” and “Arbequina”) showed a great variability in the three environments studied (Figures 1, 2). The greatest variability in the phenolic profile of the fruits was associated to demethyloleuropein, oleuropein and verbascoside. The mean

value found for demethyloleuropein (8787.4 µg/g) is lower than the mean content analyzed in the “Arbequina” fruits grown in the three environments selected in this study shown in Table 1 (11,771.5 µg/g) but quite similar to other values recently reported for “Arbequina” fruits also grown in south Spain (Romero et al., 2017). However, the significant content of demethyloleuropein found in all the breeding selections analyzed it is quite remarkable considering the fact that this phenolic glycoside is only present in a very short number of traditional olive cultivars (Gómez-Rico et al., 2008). The mean content of oleuropein found among the advanced breeding selections (Figure 1) was lower than the mean value found for “Picual” fruits in the environments selected in this study (Table 2) and also lower than those recently reported for Picual fruits grown in seven orchards representatives of southeast Spain (Romero et al., 2017). The high content of verbascoside found in the breeding selections analyzed might also be of interest from a nutritional point of view given the biological properties and clinical potential described for this compound (Alipieva et al., 2014). However, the content of this glucoside is not so relevant in relation to VOO phenolic composition since verbascoside, due to its chemical structure, is not hydrolysable by olive β-glucosidase so that no significant hydrolytic derivatives of





**TABLE 2 |** Phenolic components ( $\mu\text{g g}^{-1}$ ) of fruits (lowercase) and oils (uppercase letters) with regards to genotype and environment.

	Total phe fruit	Demethyloleuropein	Oleuropein	Ligstroside	Total PHE VOO	3,4 DHPEA- EDA	p-HPEA- EDA	p-HPEA- EA	3,4 DHPEA-EA	3,4 DHPEA ACETATE	ACETOXYPINORESINOL	PINORESINOL	LUTEOLIN	APIGENIN	PHENOLIC ACIDS
<b>GENOTYPES</b>															
Arbequina	16387.3 ab	11771.5 a	2537.4 c	220.4 b	502.3 b	265.4 b	128.2 b	12.0 b	37.7 b	12.0 a	29.1 ab	2.6 c	8.2 b	2.9 b	1.4 b
Picual	14111.4 bc	132.6 c	10029.0 a	847.3 a	482.2 b	49.7 d	42.7 c	36.1 a	298.5 a	1.9 b	38.7 a	3.8 b	6.1 c	2.0 c	2.0 a
UCI2-68	18900.5 a	10527.0 a	5676.7 b	373.7 b	640.6 a	333.6 a	166.0 a	15.7 b	77.9 b	12.5 a	19.6 b	4.5 a	3.9 d	0.8 d	1.3 b
UCI6-65	12706.0 c	4300.5 b	4998.5 b	858.2 a	368.4 b	121.2 c	156.6 ab	9.2 b	25.5 b	6.4 b	15.9 b	2.7 c	11.1 a	5.3 a	1.1 b
<b>ENVIRONMENTS</b>															
cordoba	10170.5 b	6228.9 b	2050.1 b	225.1 b	423.9 b	146.2 b	84.2 b	20.7	118.6 a	10.7 a	30.4	3.9 a	3.2 b	1.2 b	1.1 b
Moron	16220.9 a	6534.3 b	6830.4 a	663.2 a	414.9 b	181.2 b	133.3 a	13.0	39.2 b	5.8 b	20.5	2.9 b	7.9 a	3.5 a	1.5 a
Ubeda	18225.4 a	8723.2 a	6102.4 a	592.3 a	605.1 a	258.9 a	147.4 a	17.8	122.3 a	10.0 a	26.1	3.3 b	8.9 a	3.0 a	1.5 a

Different letters indicate LSD between genotypes and environments ANOVA ( $P \leq 0.05$ ).

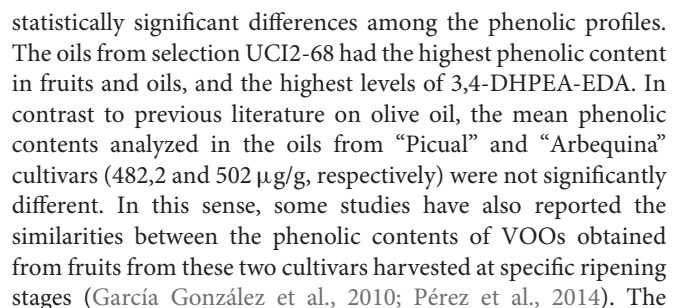
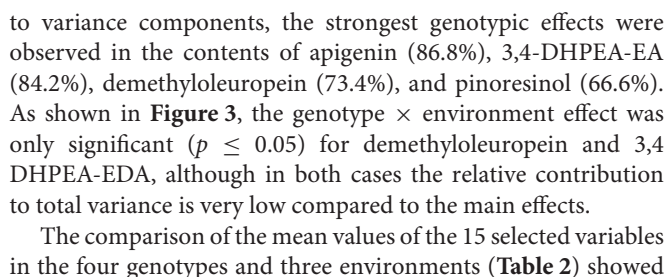
this glucoside are found in VOO (Romero-Segura et al., 2012). The two most relevant phenolic compounds found in the oils of the eight genotypes were 3,4 DHPEA-EDA and p-HPEA-EDA. The variability range of 3,4 DHPEA-EDA was quite similar to that previously analyzed in 136 olive seedlings from a single cross Picual  $\times$  Arbequina (Pérez et al., 2014) and significantly higher than those found among the crosses “Arbequina”  $\times$  “Arbosana” and “Sikitita”  $\times$  “Arbosana” (El Riachy et al., 2012). The well-known biological activity of 3,4 DHPEA-EDA (Grasso et al., 2007; Bernardini and Visioli, 2017) and its high content suggest a key contribution of this compound to the antioxidant activity of the oils. Similarly, the high content of p-HPEA-EDA found among the analyzed breeding selections may also have important quality implications, both nutritional and organoleptic, due to its relation to oil pungency and to its anti-inflammatory properties (Lucas et al., 2011).

One of the main goals of this study was to estimate the relative influence of genetic and environmental factors on the phenolic composition of olive breeding selections. The variability plots obtained for the main phenolic components of the fruits and oils (Figures 1, 2) and the comparison of means (Table 2) show the specific contribution of environments and genotypes to the phenolic variability found in this study. Rainfall and/or irrigation regime of the olive tree it is probably the most studied environmental factor influencing the composition of VOOs, (Romero and Motilva, 2010). In this sense, given that the water applied by irrigation was similar in the three environments, the major differences in water availability corresponds to rainfall (Table 1), that was higher in Cordoba respect to the other two locations, with the lowest value associated to Ubeda. The overall higher phenolic content found in fruits and oils from Ubeda could be related with the lower rainfall of this location and similarly, the lower phenolic content of genotypes grown in Cordoba could be explained by its highest rainfall. The increase of phenolic content in VOO with reduced water availability has been well documented for olive (Marra et al., 2016; Cirilli et al., 2017) although this relation has not always been clearly observed (Pierantozzi et al., 2014). However, taking into account that the three environments differs on many aspects (soil type, temperature regime, rainfall, etc.), differences on phenolic composition among the three environments could not be directly attributed to a single environmental factor. In relation to the contribution of the genotype, the variability pattern of UCI 2-68 in total VOO phenolics is in good agreement with that observed for demethyloleuropein content (Figure 1B), which is the precursor of 3,4-DHPEA-EDA the most abundant phenolic compound analyzed in the oils of UCI 2-68, (Figure 2B). Similarly, the highest variability of “Picual” oils for the 3,4 DHPEA-EA and p-HPAE-EA contents (Figures 2E,F) correlates with that previously mentioned for oleuropein and ligstroside contents in the same fruits (Figures 1C,D). The major contribution of the genotype to the phenolic variability was clearly demonstrated after calculating the percent of the total variance (Figure 3). Thus, while a significant effect of the genotype was found for all the phenolic compounds, the environment was only a major contributor to the variance of the total phenolic content of the fruits (66.9%). According

**TABLE 3 |** Pearson's correlation coefficients among the main phenolic compounds found in fruits (lowercase letters) and oils (uppercase letters) from all genotypes and environments.

	Total phe fruit	Fruit secoiridoids	Demethyl O	Oleuropein	Ligstroside	Verbascoside	TOTAL PHE VOO	AOV SECOIRIDIDS	3,4-DHPEA	p-HPEA	3,4-DHPEA AC	3,4-DHPEA-EDA	p-HPEA-EDA	ACETOXY-P	PINORESINOL	3,4-DHPEA-EA	p-HPEA-EA	CINNAMIC ACID	LUTEOLIN	APIGENIN
Total phe fruit	1																			
Fruit secoiridoids	***0.991	1																		
Demethyl O	***0.594	***0.643	1																	
Oleuropein	0.289	0.241 **	– 0.584	1																
Ligstroside	0.046	0.013 **	– 0.644	***0.765	1															
Verbascoside	0.092	– 0.027	– 0.318	*0.397	0.327	1														
TOTAL PHE VOO	***0.685	***0.735	***0.650	– 0.057	– 0.197	– 0.291	1													
VOO SECOIRIDIDS	***0.688	***0.740	***0.653	– 0.071	– 0.149	– 0.244	***0.999	1												
3,4-DHPEA	– 0.079	– 0.120	– 0.584	*0.585	**0.586	0.310	– 0.232	– 0.082	1											
p-HPEA	0.102	0.082	– 0.234	0.304	*0.501	0.122	0.129	0.272	***0.674	1										
3,4-DHPEA AC	0.261	0.288	***0.758	***	– 0.668	**	0.532	– 0.256	0.413	0.406 **	– 0.565	– 0.206	1							
3,4-DHPEA-EDA	***0.695	***0.753	***0.900	– 0.343	*– 0.492	*– 0.388	***0.850	***0.858	– 0.421	0.077	***0.612	1								
p-HPEA-EDA	***0.478	***0.534	***0.663	– 0.298	– 0.175 *	– 0.479	***0.759	***0.770	– 0.183	0.307	*0.477	***0.756	1							
ACETOXY-P	0.026	0.020	– 0.235	0.335	0.234	0.215	0.122	0.089	– 0.033	– 0.016	– 0.258	– 0.155	– 0.317	1						
PINORESINOL	*0.361	**0.355	*0.347	– 0.103	– 0.121	– 0.154	***0.630	***0.638	0.027	*0.381	0.279	*0.441	***0.694	– 0.095	1					
3,4-DHPEA-EA	0.060	0.037	*– 0.469	***0.611	*0.387	*0.361	0.217	0.203	0.280	0.198	*– 0.437	– 0.213	– 0.343	***0.669	0.117	1				
p-HPEA-EA	– 0.046	– 0.056	*– 0.388	*0.480	**0.435	0.179	0.144	0.124	0.155	0.142	– 0.257	– 0.249	– 0.157	***0.698	0.126	***0.721	1			
CINNAMIC ACID	*0.371	0.326	– 0.125	***0.609	*0.357	*0.492	0.161	0.153	0.321	0.261	– 0.215	– 0.007	– 0.213	0.301	– 0.009	***0.557	0.316	1		
LUTEOLIN	– 0.003	– 0.038	0.045	– 0.096	0.163	– 0.083	– 0.022	– 0.033	– 0.018	0.042	0.093	0.033	0.112	– 0.229	*0.398	– 0.201	0.265	0.028	1	
APIGENIN	– 0.230	– 0.272	– 0.225	– 0.008	*0.362	0.028	– 0.280	– 0.288	0.161	0.188	– 0.256	– 0.259	– 0.034	– 0.163	*0.379	– 0.257	0.257	0.035	**0.776	1

Marked correlations are significant at \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ .



comparison of the mean values determined in each environment indicates that, with the exception of 3,4-DHPEA-EDA and Pinoresinol, significantly lowest phenolic contents were always found in the fruits and oils from Cordoba. On the contrary, although the values found in fruits and oils from Ubeda were usually higher than in the other two environments, for most traits the differences between Ubeda and Moron were not statistically different.

According to the data obtained, the phenolic profile seems to be dependent on the genotype with only quantitative, but no-qualitative, differences associated to the environmental factor. Although it is clear that variations among years might be added to that observed between environments, considering that the climatic conditions (maximum, minimum, and average temperature and rainfall) of 2016 were in the average range of the last 10 years, the results could be extendable to other seasons.

## Relationship between Fruit and VOO Phenolic Components

One of the main objectives of this study was to investigate the suitability of fruit phenolic profiling to predict the phenolic composition of VOO from olive breeding selections. The hydrolysis of secoiridoid glycosides seems to be the key step in the biosynthesis of phenolic components during the extraction of VOO (Obied et al., 2008). In previous studies we have demonstrated that during the milling of the olive fruits, cell integrity is disrupted and phenolic glycosides are transformed to their corresponding aglycones (3,4-DHPEA-EDA, p-HPEA-EDA, 3,4-DHPEA-EA, p-HPEA-EA, luteolin or apigenin) by a highly specific olive  $\beta$ -glucosidase (Romero-Segura et al., 2012). The secoiridoid derivatives formed may be further hydrolyzed to simple phenolic compounds such as 3,4 DHPEA or p-HPEA. Other compounds such as the lignans, pinoresinol and 1-acetoxypinoresinol, not detected in the olive pulp, are presumably formed in the olive seed and only liberated and transferred to the oil after olive stone crushing (Klen et al., 2015). However, a number of studies have reported that only a minimal amount of the phenolic compounds formed during the milling of the olive fruits and the subsequent malaxation of olive pastes are transferred to the oil. This transfer being cultivar dependent (Talhaoui et al., 2016) and significantly different for each class of compounds, with the highest transfer rate corresponding to secoiridoids compounds, followed by flavonoids and simple phenols. The Pearson's correlation coefficients computed with all the phenolic components analyzed in fruits and oils could provide useful information on this issue (Table 3). The correlation between the total phenolic content of fruits and oils was ( $r = 0.685$ ) and slightly higher between total secoiridoids from fruits and oils ( $r = 0.740$ ). Higher correlations have been reported for fatty acids ( $r = 0.98$ ), tocopherols ( $r = 0.96$ ), and other compounds analyzed in olive fruit and VOO (Velasco et al., 2014). However, those compounds are already present in the olive fruit tissue while the biosynthesis of oil phenolic components may involve a number of complex biochemical reactions (Obied et al., 2008; Klen et al., 2015) which are also affected by the oxidative degradation catalyzed by olive polyphenol oxidase

and peroxidase (García-Rodríguez et al., 2011). The significant positive correlation found between demethyloleuropein and the VOO phenolic content ( $r = 0.650$ ) contrasts with the low value found for oleuropein ( $r = -0.057$ ). However, the very different correlation coefficients found for both secoiridoid glycosides are in good concordance with the higher mean value found for demethyloleuropein compared to that of oleuropein in the fruits of the eight genotypes analyzed (Figure 1). Similarly, the highest Pearson's coefficients between individual fruit and oil phenolic components were obtained for demethyloleuropein and 3,4-DHPEA-EDA ( $r = 0.900$ ), which in turn is the most abundant secoiridoid compound in the oils analyzed in this study (Figure 2). This data also corroborates previous findings on the high specificity of the olive  $\beta$ -glucosidase which forms 3,4-DHPEA-EDA as the unique product of demethyloleuropein hydrolysis while 3,4-DHPEA-EA is the main, but not the unique product found after oleuropein hydrolysis (Romero-Segura et al., 2012). In this sense, the correlation coefficient found for 3,4-DHPEA-EA and oleuropein was significantly lower ( $r = 0.611$ ) than that mentioned for 3,4-DHPEA-EDA and demethyloleuropein. Positive correlation coefficients were also found between demethyloleuropein and p-HPEA-EDA ( $r = 0.663$ ) and demethyloleuropein and 3,4-DHPEA acetate ( $r = 0.758$ ) although no conclusive data has been obtained so far on their biosynthesis. In a similar way, ligstroside content positively correlated to p-HPEA-EA ( $r = 0.435$ ) which suggests a similar biosynthetic pathway for the latter compound to that described for 3,4-DHPEA-EA (Romero-Segura et al., 2012). The also high correlation coefficient found between ligstroside and oleuropein ( $r = 0.765$ ) points to a common biosynthesis for both glycosides not fully demonstrated yet (Obied et al., 2008).

Data shown in Table 3 also reveal significant differences between the major phenolic components of the VOO, with very high correlation values between 3,4-DHPEA-EDA ( $r = 0.850$ ) and p-HPEA-EDA ( $r = 0.759$ ) and the total phenolic content of the oils but very low coefficients for 3,4-DHPEA-EA and p-HPEA-EA. Similar correlations were reported by El Riachy et al. (2012) for these monoaldehydic compounds in the analysis of two different segregating populations although non-significant correlation was found for the major secoiridoid component of VOO, 3,4-DHPEA-EDA. An interesting relation was also found also between 3,4-DHPEA-EDA and the non-secoiridoid 3,4-DHPEA acetate ( $r = 0.612$ ). This correlation value, and that previously mentioned between demethyloleuropein and 3,4-DHPEA-acetate, seem to support the formation of 3,4-DHPEA-acetate not by simple hydroxytyrosol acetylation but through a more complex biochemical pathway involving the cleavage of the aglycone formed after demethyloleuropein deglycosilation. The elucidation of 3,4-DHPEA-acetate biosynthesis could have great interest from a biotechnological point of view given the enhanced bioavailability of this hydroxytyrosol derivative (Mateos et al., 2011). Pinoresinol was the non-secoiridoid compound which best correlated to AOV phenolic content (0.630) while a very low correlation coefficient was calculated for acetoxypinoresinol ( $r = 0.122$ ).

Among all the information obtained from the Pearson correlation coefficients computed, it is important to emphasize



that the significant correlations found between specific phenolic compounds, or groups of phenolic compounds, in the olive fruit, and the phenolic content of AOV may have an important predictive value.

## Phenolic Profiling as a Selection Criterion in Olive Breeding

The first and second principal components from the PCA applied to the data set used for analysis of variance described 60% of the total variability found in the four genotypes and three environments (**Figure 4**). This value is quite similar to that found in previous studies on the phenolic variability of olive progenies (El Riachy et al., 2012; Pérez et al., 2014). The distribution of phenolic profiles in the PCA bi-plot seems to confirm the conclusion raised after the analysis of variance, since most of the variability shown in **Figure 4** corresponds to genotype rather than to environment. In this sense, while the four genotypes are clearly segregated, the distinction among the different environments within each genotype is not always possible. Thus, the profiles of UCI 5-65 and UCI 2-68 grown in different environments were only partially segregated in the scatter-plot.

The genotype Picual, located in the second quadrant of the plot (**Figure 4B**), is clearly separated from the other three genotypes. According to the vector distribution (**Figure 4A**) oleuropein, 3,4 DHPEA-EA, p-HPEA-EA, and acetoxypinoresinol are related, and located in the second quadrant in which the genotype Picual is also included (**Figure 4B**). The other three genotypes are mainly located in the right part of the scatter plot, mostly in the first and fourth quadrants. According to the scatter plot the phenolic profiles of the two analyzed breeding selections (UCI 5-65 and UCI 2-68) seem to be closer to “Arbequina” than to “Picual” cultivar. In this sense, UCI 2-68, located in the first quadrant has a phenolic composition quite similar to that of “Arbequina” cultivar but a higher phenolic content. On the contrary, selection UCI 5-65 exhibits also a similar phenolic pattern, with higher luteolin and apigenin levels that increase its potential antioxidant properties (Rice-Evans et al., 1995), but with a significantly lower total phenolic content. As shown in **Figure 4B**, the overall higher phenolic content of fruits and oils from Ubeda are confirmed by the upper position of 5-65-U and 2-68-U in their respective groups. On the contrary, the influence of environmental factors seems to be less important in “Arbequina,” located in the central area of the plot and in “Picual” cultivar located along the second quadrant.

The effect of environmental factors on the phenolic profile of new breeding selections may provide very relevant information within an olive breeding program. In this sense, the total phenolic content of oils from UCI 5-65 grown in Ubeda is 482.2 µg/g oil, but this content is significantly lower in other environments (**Figure 2A**). Thus, the very low phenolic content found in the oils obtained from fruits grown in Moron (253.0 µg/g oil) could negatively affect the nutritional quality of these oils and could exclude them from the European health claim (European Commission, 2012). Taking into account the minimum values established by the EFSA for VOO phenolic health claim (250 µg/g oil), and the additional

benefits of a medium-high phenolic content for the organoleptic properties and the stability of VOO, the breeding selection UCI 2-68 is clearly superior to UCI 5-65 in terms of phenolic composition.

When PCA analysis was applied to breeding selections grown in the same environment genotypes are clearly segregated (**Figure 5**). The selection UCI 12-85, with the highest oil phenolic content (999 µg/g oil) and the selection UCI 12-104 are located in the upper part of the first quadrant, above the selection UCI 2-68. The high phenolic content of the breeding selection UCI 12-85 may have potential in terms of nutritional quality. However, it is important to point out that the very high levels of secoiridoid compounds in its oil may also have a negative impact in its organoleptic properties. In this sense, the VOO from this selection has the highest contents of p-HPEA-EDA (321 µg/g oil) which is highly related to the pungency of VOO (Andrewes et al., 2003) which greatly affects consumer acceptance. According to their respective positions in the scatter plot the breeding selections UCI 19-60 and UCI 19-19 may be categorized as medium and low phenolic content genotypes. The breeding selection UCI 19-79 had slightly higher phenolic content than UCI 5-65 but lower than “Arbequina” cultivar. The localization of this selection in the center of the fourth quadrant (**Figure 5B**) matches that of luteolin and 3,4-DHPEA acetate in the vector distribution of the variables (**Figure 5A**). These two compounds are the most significant phenolic components in the VOO of UCI 19-79, which possesses the highest content of luteolin and 3,4-DHPEA acetate among all the genotypes analyzed (14.4 and 27.6 µg/g oil, respectively).

## CONCLUSION

The high correlation found between fruit and oil phenolic components content, as well as the high genotypic variance for them, indicate that the analysis of fruit phenolic compounds, without the previous step of oil extraction, is an useful tool for olive breeding which could facilitate the selection of olive genotypes with potential interest in terms of oil phenolic composition. In this sense, fruit phenolic profiling could be used as selection criterion in the early stages of olive breeding programs to avoid the selection of genotypes whose oils will never reach an optimum phenolic content (European Commission, 2012). Moreover, the low genotype × environment interaction on phenolic composition, leading to a similar ranking of genotypes on the different environments, could also facilitate the evaluation of new selections from breeding works. The analytical methodology reported in this study allowed the identification of the new selection UCI2-68, characterized by an optimum phenolic profile, together to a good agronomic performance previously reported (León et al., 2004a,b, 2007, 2011), which represents a new olive cultivar producing superior quality oil.

## AUTHOR CONTRIBUTIONS

RdlR and LL: Conceived and designed the breeding experiments; AP and CS: Designed and performed the analytical studies;

AP: Wrote the manuscript. All authors discussed and commented the manuscript.

## FUNDING

This work was partly funded by European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 645595, by Research Project AVA201601.2 from IFAPA (partially

funded by European Regional Development fund) and by Programa Nacional de Recursos y Tecnologías Agroalimentarias financed by the Spanish Government, project AGL2015-67652.

## ACKNOWLEDGMENTS

We thank Juan Luis Tortosa (Instituto de la Grasa, CSIC) for technical assistance.

## REFERENCES

- Alipieva, K., Korkina, L., Orhan, I. E., and Georgiev, M. I. (2014). Verbascoside a review of its occurrence, biosynthesis and pharmacological significance. *Biotechnol. Adv.* 32, 1065–1076. doi: 10.1016/j.biotechadv.2014.07.001
- Andrewes, P., Busch, J. L., de Joode, T., Groenewegen, A., and Alexandre, H. (2003). Sensory properties of virgin olive oil polyphenols: identification of deacetoxyglistrosideglycon as a key contributor to pungency. *J. Agric. Food Chem.* 51, 1415–1420. doi: 10.1021/jf026042j
- Bartolini, G., Prevost, G., Messeri, C., and Carignani, C. (2005). *Olive Germplasm: Cultivars and World-Wide Collections*. Rome: FAO/Plant Production and Protection. Available online at: www.oleadb.it.
- Belaj, A., Dominguez-Garcia, M. D. C., Gustavo Atienza, S., Martin Urdiroz, N., De La Rosa, R., Satovic, Z., et al. (2012). Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DARs, SSRs, SNPs) and agronomic traits. *Tree Genet. Genomes* 8, 365–378. doi: 10.1007/s11295-011-0447-6
- Bellini, E., Giordani, E., and Rosati, A. (2002). Genetic improvement of olive from clonal selection to cross-breeding programs. *Adv Hort. Sci.* 22, 73–86.
- Bernardini, E., and Visioli, F. (2017). High quality, good health: the case for olive oil. *Eur. J. Lipid Sci. Technol.* 119:1500505. doi: 10.1002/ejlt.201500505
- Cirilli, M., Caruso, G., Gennai, C., Urbani, S., Frioni, E., Ruzzi, M., et al. (2017). The role of polyphenoloxidase, peroxidase, and  $\beta$ -glucosidase in phenolics accumulation in *Olea europaea* L. Fruits under different water regimes. *Front. Plant Sci.* 8:717. doi: 10.3389/fpls.2017.00717
- de la Rosa, R., Arias Calderón, R., Velasco, L., and León, L. (2016). Early selection for oil quality components in olive breeding progenies. *Eur. J. Lipid Sci. Technol.* 118, 1160–1167. doi: 10.1002/ejlt.201500425
- de la Rosa, R., Talhaoui, N., Rouis, H., Velasco, L., and León, L. (2013). Fruit characteristics and fatty acid composition in advanced olive breeding selections along the ripening period. *Food Res. Int.* 54, 1890–1896. doi: 10.1016/j.foodres.2013.08.039
- El Riachy, M., Priego-Capote, F., León, L., Luque de Castro, M. D., and Rallo, L. (2012). Virgin olive oil phenolic profile and variability in progenies from olive crosses. *J. Sci. Food Agric.* 92, 2524–2533. doi: 10.1002/jsfa.5662
- Estruch, R., Ros, E., Salas-Salvado, J., Covas, M. I., Corella, D., Arós, F., et al. (2013). Primary prevention of cardiovascular disease with a Mediterranean diet. *N. Engl. J. Med.* 368, 1279–1290. doi: 10.1056/NEJMoa1200303
- European Commission (2012). *Establishing a List of Permitted Health Claims Made on Foods, Other than Those Referring to the Reduction of Disease Risk and to Children's Development and Health*. Regulation EU No 432/2012. European Commission.
- Famiani, F., Proietti, P., Farinelli, D., and Tombesi, A. (2000). Oil quality in relation to olive ripening. *Acta Hort.* 586, 671–674. doi: 10.17660/ActaHortic.2002.586.143
- García González, D., Tena, N., and Aparicio, R. (2010). Quality characterization of the New Virgin Olive Oil Var. Sikitita by phenols and volatile compounds. *J. Agric. Food Chem.* 58, 8357–8364. doi: 10.1021/jf101316d
- García-Rodríguez, R., Romero-Segura, C., Sanz, C., Sánchez-Ortiz, A., and Pérez, A. G. (2011). Role of polyphenol oxidase and peroxidase in shaping the phenolic profile of virgin olive oil. *Food Res. Int.* 44, 629–635. doi: 10.1016/j.foodres.2010.12.023
- Gómez-Rico, A., Fregapane, G., and Salvador, M. D. (2008). Effect of cultivar and ripening on minor components in Spanish olive fruits and their corresponding virgin olive oils. *Food Res. Int.* 41, 433–440. doi: 10.1016/j.foodres.2008.02.003
- Grasso, S., Siracusa, L., Spatafora, C., Renis, M., and Tringali, C. (2007). Hydroxytyrosol lipophilic analogues: enzymatic synthesis, radical scavenging activity and DNA oxidative damage protection. *Bioorganic. Chem.* 35, 137–152. doi: 10.1016/j.bioorg.2006.09.003
- Ilarióni, L., and Proietti, P. (2014). "Olive tree cultivars," in *The Extra-Virgin Olive Oil Handbook*, ed C. Peri (West Sussex: John Wiley & Sons, Ltd.), 59–67.
- Klen, T., Wondra, A. G., Vrhovšek, U., Sivilotti, P., and Vodopivec, B. M. (2015). Olive fruit phenols transfer, transformation, and partition trail during laboratory-scale olive oil processing. *J. Agric. Food Chem.* 63, 4570–4579. doi: 10.1021/jf506353z
- Lavee, S. (2013). Evaluation of the need and present potential of olive breeding indicating the nature of the available genetic resources involved. *Sci. Hort* 161, 333–339. doi: 10.1016/j.scienta.2013.07.002
- León, L., Arias-Calderon, R., de la Rosa, R., Khadari, B., and Costes, E. (2016). Optimal spatial and temporal replications for reducing environmental variation for oil content components and fruit morphology traits in olive breeding. *Euphytica* 207, 675–684. doi: 10.1007/s10681-015-1569-y
- León, L., Beltrán, G., Aguilera, M. P., Rallo, L., Barranco, D., and de la Rosa, R. (2011). Oil composition of advanced selections from an olive breeding program. *Eur. J. Lipid Sci. Technol.* 113, 870–875. doi: 10.1002/ejlt.201000535
- León, L., De La Rosa, R., Barranco, D., Rallo, L., León, L., De La Rosa, R., et al. (2007). Breeding for early bearing in olive. *HortScience* 42, 499–502.
- León, L., Martin, L. M., and Rallo, L. (2004a). Repeatability and minimum selection time for fatty acid composition in olive progenies. *HortScience* 39, 477–480.
- León, L., Rallo, L., Del Rio, C., and Martin, L. M. (2004b). Variability and early selection on the seedling stage for agronomic traits in progenies from olive crosses. *Plant Breed.* 123, 73–78. doi: 10.1046/j.0179-9541.2003.00920.x
- León, L., Velasco, L., and de la Rosa, R. (2015). Initial selection steps in olive breeding programs. *Euphytica* 201, 453–462. doi: 10.1007/s10681-014-1232-z
- Lucas, L., Russell, A., and Keast, R. (2011). Molecular mechanisms of inflammation. Anti-inflammatory benefits of virgin olive oil and the phenolic compound oleocanthal. *Curr. Pharm. Des.* 17:7542768. doi: 10.2174/138161211795428911
- Marra, F. P., Marino, G., Marchese, A., and Caruso, T. (2016). Effects of different irrigation regimes on a super-high-density olive grove cv. "Arbequina" vegetative growth, productivity and polyphenol content of the oil. *Irrig. Sci.* 34, 313–325. doi: 10.1007/s00271-016-0505-9
- Martínez, J. M., Muñoz, E., Alba, J., and Lanzón, A. (1975). Report about the use of the 'Abencor' analyser. *Grasas Aceit.* 26, 379–385.
- Mateos, R., Espartero, J. L., Trujillo, M., Rios, J. J., León Camacho, M., Alcudia, F., et al. (2001). Determination of phenols, flavones, and lignans in virgin olive oils by solid-phase extraction and high-performance liquid chromatography with diode array ultraviolet detection. *J. Agric. Food Chem.* 49, 2185–2192. doi: 10.1021/jf0013205
- Mateos, R., Pereira-Caro, G., Saha, S., Cert, R., Redondo-Horcajo, M., Bravo, L., et al. (2011). Acetylation of hydroxytyrosol enhances its transport across differentiated Caco-2 cell monolayers. *Food Chem.* 125, 865–872. doi: 10.1016/j.foodchem.2010.09.054

- Menéndez, J. A., Vázquez-Martin, A., García-Villalba, R., Carrasco-Pancorbo, A., Oliveras-Ferraro, C., Fernandez-Gutierrez, A., et al. (2008). Anti-HER2 (erbB-2) oncogene effects of phenolic compounds directly isolated from commercial Extra-Virgin Olive Oil (EVOO). *BMC Cancer* 8:377. doi: 10.1186/1471-2407-8-377
- Montedoro, G., Baldioli, M., Selvaggini, R., Begliomini, A. L., and Taticchi, A. (2002). Relationship between phenolic compounds of olive fruit and olive oil: importance of the endogenous enzymes. *Acta Hort.* 586, 551–556. doi: 10.17660/ActaHortic.2002.586.115
- Obied, H. K., Prenzler, P. D., Ryan, D., Servili, M., Taticchi, A., Esposto, S., et al. (2008). Biosynthesis and biotransformations of phenol-conjugated oleosidic secoiridoids from *Olea europaea* L. *Nat. Prod. Rep.* 25, 1167–1179. doi: 10.1039/b719736e
- Pérez, A. G., León, L., Pascual, M., Romero-Segura, C., Sánchez-Ortiz, A., de la Rosa, R., et al. (2014). Variability of virgin olive oil phenolic compounds in a segregating progeny from a single cross in *Olea europaea* L. and sensory and nutritional quality implications. *PLoS ONE* 9:e92898. doi: 10.1371/journal.pone.0092898
- Pierantozzi, P., Torres, M., Lavee, S., and Maestri, D. (2014). Vegetative and reproductive responses, oil yield and composition from olive trees (*Olea europaea*) under contrasting water availability during the dry winter-spring period in central Argentina. *Ann. Appl. Biol.* 164, 116–127. doi: 10.1111/aab.12086
- Proietti, P., Nasini, L., Famiani, F., Guelfi, P., and Standardi, A. (2012). Influence of light availability on fruit and oil characteristics in *Olea europaea* L. *Acta Hort.* 949, 243–250. doi: 10.17660/ActaHortic.2012.949.35
- Rallo, L., Barranco, D., De la Rosa, R., and León, L. (2008). “Chiquitita” olive. *HortScience* 43, 529–533.
- Rice-Evans, C. A., Miller, N. J., Bolwell, P. G., Broamley, P. M., and Pridham, J. B. (1995). The relative antioxidant activities of plant derived polyphenolic flavonoids. *Free Radic. Res.* 22, 375–383. doi: 10.3109/10715769509145649
- Romero, C., Medina, E., Mateo, M. A., and Brenes, M. (2017). Quantification of bioactive compounds in Picual and Arbequina olive leaves and fruits. *J. Sci. Food Agric.* 97, 1725–1732. doi: 10.1002/jsfa.7920
- Romero, M. P., and Motilva, M. J. (2010). “Effect of climatic conditions on quality of virgin olive oil,” in *Olive and Olive Oil in Health and Disease Prevention*, eds V. R. Preedy and R. R. Watson (San Diego, CA: Elsevier Inc), 43–50.
- Romero-Segura, C., García-Rodríguez, R., Sánchez Ortiz, A., Sanz, C., and Pérez, A. G. (2012). The role of olive beta-glucosidase in shaping the phenolic profile of virgin olive oil. *Food Res. Int.* 45, 191–196. doi: 10.1016/j.foodres.2011.10.024
- Romero-Segura, C., García-Rodríguez, R., Sanz, C., and Pérez, A. G. (2011). Virgin olive phenolic profile as a result of the anabolic and catabolic enzymes status in the olive fruit. *Acta Hort.* 924, 379–384. doi: 10.17660/ActaHortic.2011.924.47
- Rugini, E., and De Pace, C. (2016). “Olive breeding with classical and modern approaches,” in *The Olive Tree Genome. Compendium of Plant Genomes*, eds E. Rugini, L. Baldoni, R. Muleo, and L. Sebastiani (Cham: Springer), 163–193.
- Servili, M., Sordini, B., Esposto, S., Urbani, S., Veneziani, G., di Maio, I., et al. (2014). Biological activities of phenolic compounds of extra virgin olive oil. *Antioxidants* 3, 1–23. doi: 10.3390/antiox3010001
- Talhaoui, N., Gómez-Caravaca, A. M., León, L., De la Rosa, R., Fernández-Gutiérrez, A., and Segura-Carretero, A. (2016). From olive fruits to olive oil: phenolic compound transfer in six different olive cultivars grown under the same agronomical conditions. *Int. J. Mol. Sci.* 17:337. doi: 10.3390/ijms17030337
- Velasco, L., Fernandez-Cuesta, A., De la Rosa, R., Ruiz-Mendez, M. V., and León, L. (2014). Selection for some olive oil quality components through the analysis of fruit flesh. *J. Am. Oil Chem. Soc.* 91, 1731–1736. doi: 10.1007/s11746-014-2523-1

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Pérez, León, Sanz and de la Rosa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Using Wild Olives in Breeding Programs: Implications on Oil Quality Composition

Lorenzo León<sup>1\*</sup>, Raúl de la Rosa<sup>1</sup>, Leonardo Velasco<sup>2</sup> and Angelina Belaj<sup>1</sup>

<sup>1</sup> IFAPA Centro Alameda del Obispo, Córdoba, Spain, <sup>2</sup> Instituto de Agricultura Sostenible – Consejo Superior de Investigaciones Científicas, Córdoba, Spain

## OPEN ACCESS

### Edited by:

Mariela Torres,  
Instituto Nacional de Tecnología  
Agropecuaria (INTA), Argentina

### Reviewed by:

Rosario Muleo,  
Università degli Studi della Tuscia, Italy  
Daniela Farinelli,  
University of Perugia, Italy

### \*Correspondence:

Lorenzo León  
lorenzo.leon@juntadeandalucia.es

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 06 October 2017

**Accepted:** 09 February 2018

**Published:** 27 February 2018

### Citation:

León L, de la Rosa R, Velasco L and  
Belaj A (2018) Using Wild Olives  
in Breeding Programs: Implications on  
Oil Quality Composition.  
Front. Plant Sci. 9:232.  
doi: 10.3389/fpls.2018.00232

A wide genetic diversity has been reported for wild olives, which could be particularly interesting for the introgression of some agronomic traits and resistance to biotic and abiotic stresses in breeding programs. However, the introgression of some beneficial wild traits may be paralleled by negative effects on some other important agronomic and quality traits. From the quality point of view, virgin olive oil (VOO) from olive cultivars is highly appreciated for its fatty acid composition (high monounsaturated oleic acid content) and the presence of several minor components. However, the composition of VOO from wild origin and its comparison with VOO from olive cultivars has been scarcely studied. In this work, the variability for fruit characters (fruit weight and oil content, OC), fatty acid composition, and minor quality components (squalene, sterols and tocopherols content and composition) was studied in a set of plant materials involving three different origins: wild genotypes ( $n = 32$ ), cultivars ( $n = 62$ ) and genotypes belonging to cultivar  $\times$  wild progenies ( $n = 62$ ). As expected, values for fruit size and OC in wild olives were lower than those obtained in cultivated materials, with intermediate values for cultivar  $\times$  wild progenies. Wild olives showed a remarkably higher C16:0 percentage and tocopherol content in comparison to the cultivars. Contrarily, lower C18:1 percentage, squalene and sterol content were found in the wild genotypes, while no clear differences were found among the different plant materials regarding composition of the tocopherol and phytosterol fractions. Some common highly significant correlations among components of the same chemical family were found in all groups of plant materials. However, some other correlations were specific for one of the groups. The results of the study suggested that the use of wild germplasm in olive breeding programs will not have a negative impact on fatty acid composition, tocopherol content, and tocopherol and phytosterol profiles provided that selection for these compounds is conducted from early generations. Important traits such as tocopherol content could be even improved by using wild parents.

**Keywords:** breeding, fatty acid composition, minor components, *Olea europaea*, oleasters

## INTRODUCTION

The use of novel genetic diversity in plant breeding is considered of paramount importance to obtain new cultivars adapted to high productive, resilient, and sustainable growing systems. Crop wild relatives represent potential new sources of genetic diversity so that global conservation priorities for crop wild genetic resources are encouraged (Castañeda-Álvarez et al., 2016).



According to recent revisions, the species to which cultivated olive belongs, *Olea europaea*, includes six sub-species based on morphology and geographical distribution (Green, 2002). Among them, the subsp. *europaea*, which can be found throughout the whole Mediterranean basin, is represented by two botanical varieties: cultivated olive (var. *europaea*) and wild olive (var. *sylvestris*). A wide genetic diversity has been reported for wild olives, even higher than the observed for cultivated ones (Baldoni et al., 2006; Breton et al., 2006; Belaj et al., 2010; Erre et al., 2010; Boucheffa et al., 2016). This could be particularly interesting for the introgression of some agronomic characters in breeding programs, such as biotic and abiotic stresses resistance, as already reported for many crops (Hajjar and Hodgkin, 2007). For instance, resistance to verticillium wilt scarcely found in current cultivars, has been observed in wild germplasm (Colella et al., 2008; Arias-Calderón et al., 2015), which suggest its potential use as parents in olive breeding programs. Progenies involving wild parents have also showed shorter juvenile period and more abundant flowering than progenies from cultivated parents, which can represent important advantages for olive breeding (Klepo et al., 2013, 2014). However, it is not fully understood whether the introgression of some beneficial characters could be accompanied by a parallel negative effect regarding some other important agronomic and quality characters.

From the quality point of view, virgin olive oil (VOO) is highly appreciated due to its fatty acid (high monounsaturated oleic acid content) and minor compounds composition. All these components are responsible of well-known healthy properties of VOO (Pérez-Jiménez et al., 2007). High variability for most olive oil quality components has been reported in progenies from breeding programs (Sánchez de Medina et al., 2014, 2015; De la Rosa et al., 2016). However, compositional quality of VOOs from wild origin has been scarcely studied. Some works comparing oil composition between cultivars and wild olives indicate overlapping results between the two groups (Hannachi et al., 2008; Dabbou et al., 2011; Boucheffa et al., 2014). While it seems that fruit size and oil content (OC) are determinant traits to discriminate between wild and cultivated forms (Hannachi et al., 2008; Belaj et al., 2011). Wild olives yielding high quality oils were suggested to be commercially useful (Baccouri et al., 2008; Dabbou et al., 2011), although a possible feral origin of the genotypes reported to have high OC cannot be excluded (Baccouri et al., 2008).

Up to recently, most of the olive breeding programs were based in intra-specific crosses between cultivars of well-known merit (De la Rosa et al., 2013), representing the selection work performed in Australia a rare case of the use of wild olive material for olive breeding. As a result, based on oil yield and quality, some interesting and well adapted genotypes to the conditions of Australia were selected, although feral but not genuine wild olives were used (Sedgley, 2004). In this sense, in the framework of the olive breeding program of Córdoba Spain, many efforts have been dedicated, in the last decade, to the collection, *ex situ* conservation and evaluation of wild genotypes from different origins (Belaj et al., 2007, 2011) as well as initial characterization of some olive

progenies involving wild genotypes as parents (Klepo et al., 2013, 2014).

In this work, the variability for oil quality components including fatty acid composition and minor components such as tocopherols, phytosterols, and squalene was studied in a set of plant materials involving wild genotypes, cultivars, and crosses between them. The main objective of the research was to study the usefulness and implications for oil quality of using wild genetic resources in olive breeding. Additionally, we intended to gain some insights about the genetic determinism for these characters.

## MATERIALS AND METHODS

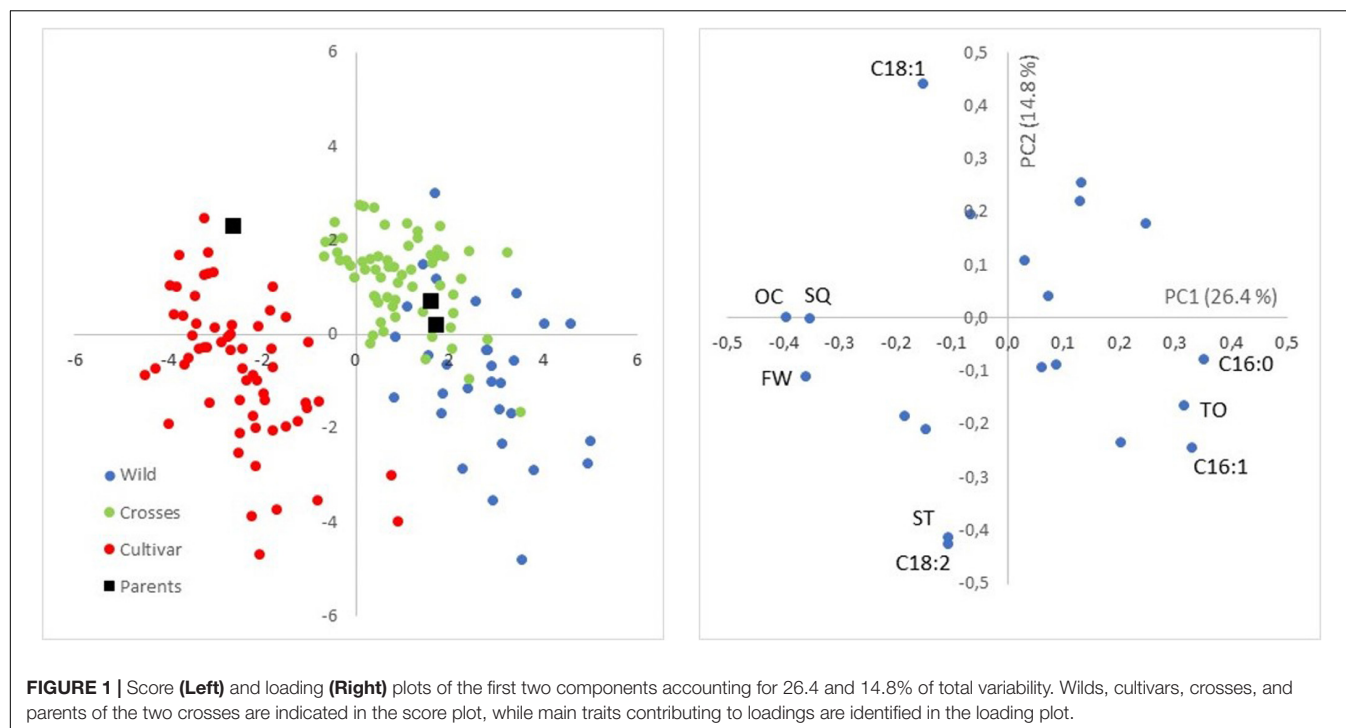
### Plant Materials

The olive germplasm under study include genotypes from three different origins: wild genotypes ( $n = 32$ ), cultivars ( $n = 62$ ) and genotypes belonging to cultivar  $\times$  wild progenies ( $n = 62$ ). All the genotypes were grown in the same conditions at the IFAPA Centro Alameda del Obispo, Córdoba (Spain). Maintained in an *ex situ* wild olive collection (De la Rosa et al., 2014), the wild genotypes came from prospecting surveys in Andalusia and Balearic Islands (21 and 9, respectively), and the rest (2) belonged to subs *Olea guanchica* from Canary Islands. The cultivated plant material under study comprised 62 olive cultivars from 13 different countries maintained at the World Olive Germplasm Collection (Supplementary Table 1). In both cases, each genotype is represented by two trees which have been evaluated for the traits under study during two harvesting seasons. In addition, 62 genotypes from crosses between the cultivar 'Picual' and two wild trees (W1 and W2), including 44 and 18 genotypes, respectively, were also evaluated. In this case, each seedling is represented by a single tree. Most of the plant materials were evaluated in two harvest seasons. However, average values per genotype were used in all analysis as genotype effect represented the main source of variation accounting for a major proportion of sums of squares for the evaluated traits, compared to harvest season effect (Supplementary Table 2).

### Traits Evaluated

Fruit samples of around 0.5 kg were randomly collected for each plant in a common date (mid-November), typical for olive harvesting. Previous works suggest that it seems more efficient to compare genotypes in a common date rather than to harvest the olive samples in a fixed ripening index, which is also quite difficult to achieve for large number of genotypes (De la Rosa et al., 2013; Bodoira et al., 2016). From each sample, three subsamples of around 25 g were randomly selected to produce dried samples sizes suitable for NMR sample holder. Fruit fresh weight (FFW) was measured and, after drying in a forced-air oven at 105°C for 42 h to ensure dehydration, OC was determined using an NMR fat analyzer (Minispec MQone, Bruker Optik GmbH, Ettlingen, Germany), and expressed as a percentage on dry weight basis.

Twenty additional fruits were randomly chosen from each sample, stored at  $-80^{\circ}\text{C}$  and then lyophilized. After lyophilization, the stones were removed and the flesh was milled in a laboratory ball mill. The samples were then stored at



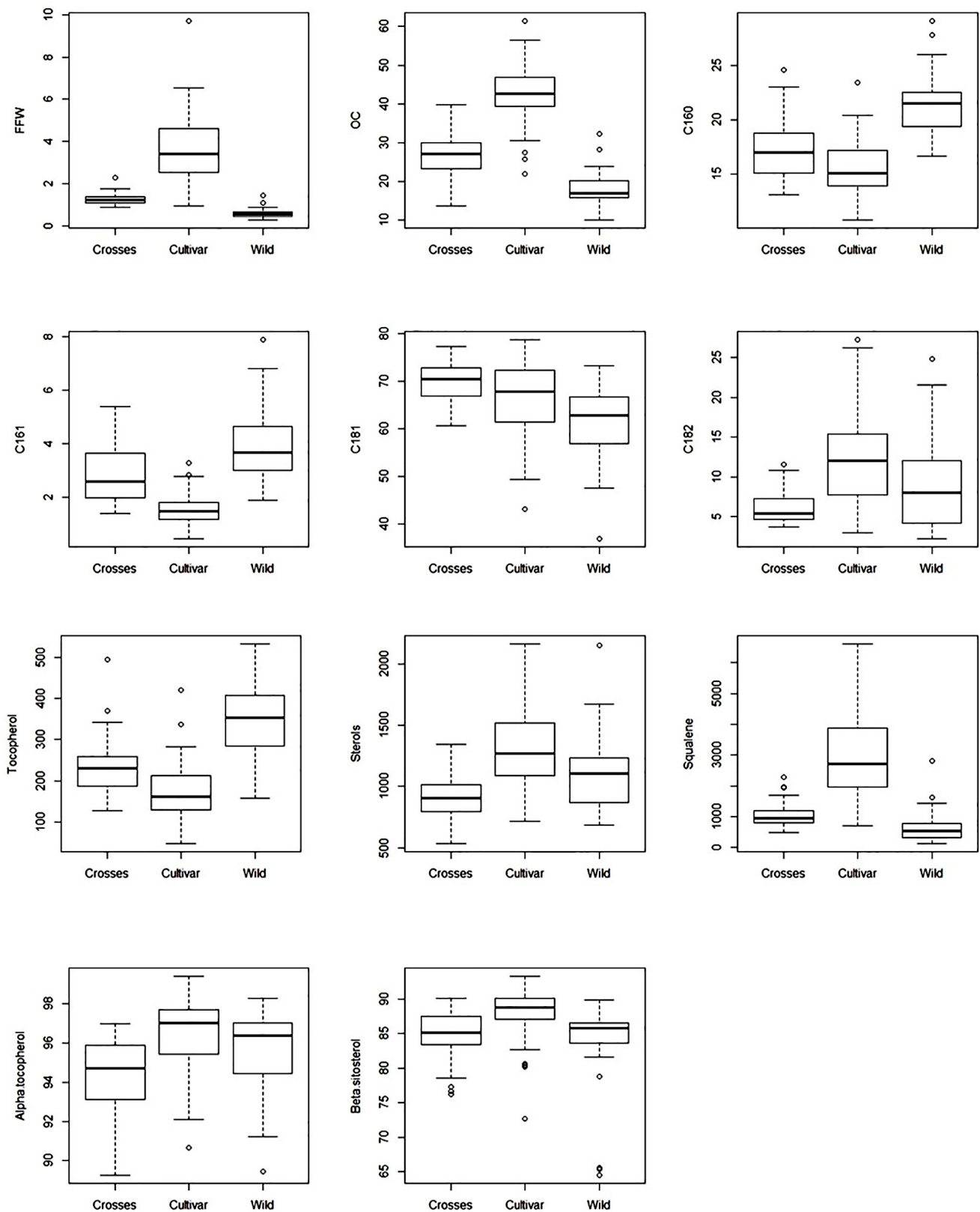
–20°C till analysis, usually within 48–72 h. All the analyses were performed in duplicate following previous procedures developed in our breeding program for direct analysis of fruit flesh (Velasco et al., 2014). In short, fatty acid composition was analyzed by simultaneous oil extraction and fatty acid methylation followed by gas-liquid chromatography (GLC) on a Perkin Elmer Clarus 600 GC (Perkin Elmer Inc., Waltham, MA, United States) equipped with a BPX70 30 m × 0.25 mm internal diameter × 0.25 μm film thickness capillary column (SGE Analytical Science Pty Ltd., Ringwood, VIC, Australia). Fatty acids were named according to C:D nomenclature, number of carbon atoms:number of double bonds in the

chain. Tocopherol extraction, separation by high-performance liquid chromatography (HPLC), and quantification was done on around 100 mg of lyophilized olive flesh using a fluorescence detector (Waters 474) at 295 nm excitation and 330 nm emission and iso-octane/tert-butylmethylether (94:6) as eluent at an isocratic flow rate of 0.8 ml/min. Quantitative determination of tocopherols was done by using rac-5,7-dimethyltolcol (Matreya LLC, Pleasant Gap, PA, United States) as internal standard and total tocopherol content (TO) was calculated as the sum of α, β, γ and δ-tocopherol contents, expressed as mg kg<sup>–1</sup> lyophilized fruit flesh. Finally, sterols (ST) and squalene (SQ) contents in lyophilized olive flesh samples were analyzed by

**TABLE 1 |** Descriptive statistics for fruit traits and main oil quality components evaluated in the three groups of samples (wilds, cultivars, and crosses).

Trait	Wild (n = 32)			Crosses (n = 62)			Cultivars (n = 62)		
	Mean		CV (%)	Mean		CV (%)	Mean		CV (%)
FFW (g)	0.60	c*	42.33	1.24	b	19.68	3.66	a	44.39
OC (%)	18.03	c	25.72	26.74	b	20.83	42.57	a	17.49
C16:0 (%)	21.34	a	13.49	17.09	b	15.28	15.44	c	15.35
C16:1 (%)	3.91	a	36.42	2.83	b	38.28	1.55	c	37.92
C18:1 (%)	61.57	c	13.19	69.87	a	6.01	66.95	b	11.72
C18:2 (%)	9.03	b	60.86	6.06	c	33.50	12.40	a	46.91
Squalene (mg kg <sup>–1</sup> )	674.3	b	80.57	1048.8	b	35.05	3009.6	a	46.72
Tocopherol (mg kg <sup>–1</sup> )	347.1	a	28.61	233.4	b	25.34	172.0	c	38.90
α-tocopherol (%)	95.52	a	2.19	94.36	b	2.01	96.47	a	1.98
Sterols (mg kg <sup>–1</sup> )	1113.7	b	28.20	916.2	c	19.43	1310.2	a	25.58
β-sitosterol (%)	83.63		7.73	84.79		3.96	87.96		3.99

\*For each trait different letters indicate significant differences between groups at  $P < 0.05$ .



**FIGURE 2 |** Box and whisker plots for fruit traits and main oil quality components evaluated in the three groups of samples (wilds, cultivars, and crosses).

**TABLE 2 |** Number of genotypes for each group exceeding thresholds values for the main fatty acids according to IOC trade standards (International Olive Council [IOC], 2016).

	C160		C181		C182	
	<7.5	>20.0	<55.0	>83.0	<2.5	>21.0
Wilds (32)	–	24	6	–	1	2
Cultivars (62)	–	3	5	–	–	5
Parents:						
Picual	13.93		77.33		3.52	
W1	17.64		69.08		6.13	
W2	21.39		62.25		6.30	
Crosses (62)	–	9	–	–	–	–

Values of the parents for each fatty acid are indicated.

GLC of the unsaponifiable fraction following silylation, using a Perkin Elmer Clarus 600 Gas Chromatograph equipped with a ZB-5 capillary column (id = 0.25 mm, length = 30 m, film thickness = 0.10  $\mu$ m; Phenomenex, Torrance, CA, United States). Total sterol content was expressed as mg kg<sup>-1</sup> lyophilized fruit flesh.

## Data Analysis

Principal components analysis (PCA) was used to investigate the variability between and within the different groups of samples evaluated and the relationship among traits, and analysis of variance and box and whiskers plots were performed for the most important variables selected from PCA to analyze differences between groups. Pearson correlation and

linear regression were used to test the relations among the traits measured. Unscrambler (CAMO A/S, Trondheim, Norway) and Statistix (Analytical Software, Tallahassee, FL, United States) software were used for the statistical analysis.

## RESULTS AND DISCUSSION

PCA was used for a preliminary exploratory analysis of variability between and within groups of samples (**Figure 1**). The score biplot of PC1 (26.4% of total variation) and PC2 (14.8%) showed clear separation of the three groups of samples (wilds, cultivars, and crosses), although a wide variability was also observed within each of the three groups. PC1 was positively correlated mainly with C16:0, C16:1, and TC, and negatively with FFW, OC, and SQ content. PC2 was associated positively mainly with C18:1 and negatively with C18:2 and ST. Main separation between groups was obtained through PC1, i.e., higher values for C16:0, C16:1, and TC are expected for wilds, and higher values for FW, OC, and SQ content for cultivars, with intermediate values in progenies from crosses in accordance to the relative position of the parents in the score plot.

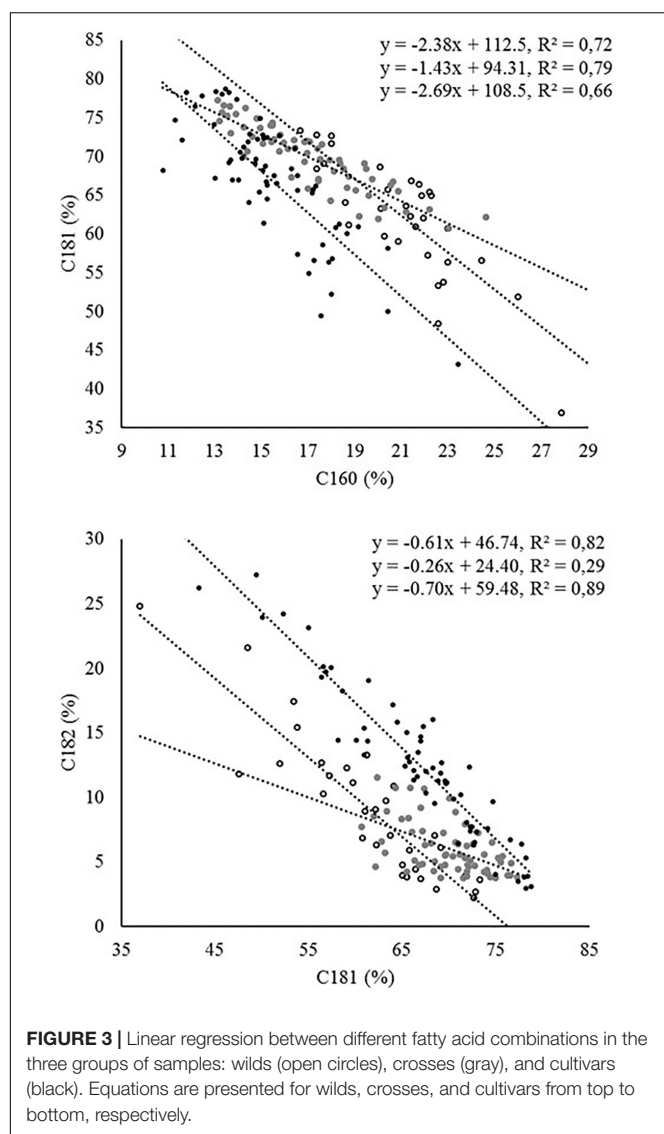
The general results inferred from PCA could be expanded to original data, with significant differences in average values between groups of samples for the most important variables selected from PCA (**Table 1**). A wide and similar variability was observed in all three groups but clear differences in range of variation could be inferred from box and whiskers plots (**Figure 2**). High variability for most of the components has

**TABLE 3 |** Correlations among fruit traits and main oil quality components evaluated in the three groups of samples: wilds, cultivars, and crosses from top to bottom.

	C16:0	C16:1	C18:1	C18:2	TO	SQ	ST
C16:0	x	0.72	−0.85	0.55	0.14	−0.18	0.03
	x	0.68	−0.81	0.60	0.34	−0.31	0.52
	x	0.90	−0.89	0.12	0.13	−0.31	0.06
C16:1	< 0.0001	x	−0.48	0.12	0.02	−0.32	0.18
	< 0.0001	x	−0.35	0.10	0.29	−0.24	0.43
	< 0.0001	x	−0.81	0.09	0.24	−0.40	0.13
C18:1	< 0.0001	0.0055	x	−0.90	0.02	0.24	0.04
	< 0.0001	0.0048	x	−0.95	−0.29	0.41	−0.61
	< 0.0001	< 0.0001	x	−0.54	−0.13	0.40	−0.11
C18:2	0.0011	0.4986	< 0.0001	x	−0.14	−0.20	−0.13
	< 0.0001	0.4476	< 0.0001	x	0.23	−0.42	0.54
	0.3656	0.4913	< 0.0001	x	−0.05	−0.27	0.04
TO	0.4591	0.9109	0.9289	0.4326	x	0.21	0.32
	0.0064	0.0223	0.0206	0.0773	x	−0.49	0.23
	0.3211	0.0626	0.3009	0.6992	x	0.03	0.53
SQ	0.3108	0.0778	0.1882	0.2744	0.2444	x	−0.05
	0.0152	0.0650	0.0011	0.0007	< 0.0001	x	−0.34
	0.0135	0.0012	0.0015	0.0327	0.8108	x	0.21
ST	0.8687	0.3144	0.8383	0.4843	0.0742	0.8012	x
	< 0.0001	0.0006	< 0.0001	< 0.0001	0.0716	0.0066	x
	0.6523	0.3134	0.3818	0.7762	< 0.0001	0.0977	x

Pearson correlation coefficients and signification levels are indicated above and below the diagonal, respectively.





been reported in other cultivars and wilds collections (Aparicio and Luna, 2002; Baccouri et al., 2008; Hannachi et al., 2013; Beltrán et al., 2015; Kyçyk et al., 2016), as well as progenies from breeding programs (De la Rosa et al., 2016). From the commercial point of view, this wide variability could impose some restrictions according to international trade standards (International Olive Council [IOC], 2016) (Table 2). For instance, C16:0 values for most of the wild genotypes would be higher than established (20% maximum). Some of them could also affront problems due to current limits for C18:1 and C18:2. It should be noted, however, that the same problem is currently faced for traditional cultivars with some of them exceeding thresholds values for the main fatty acids set up by the International Olive Council (International Olive Council [IOC], 2016). Interestingly, genotypes from crosses exceeded maximum permitted values for C16:0, probably due to the high content of wild parents, but they complied with the other thresholds. Similar results have been reported in previous evaluations. For instance, Baccouri

et al. (2011) found that only 85 out of 150 studied wild olives showed an oil fatty acid composition within IOC trade standards.

As expected from PCA and previous results, wild olives were characterized by low average values of fruit size and OC, significantly lower than cultivated materials (Hannachi et al., 2008, 2009; Belaj et al., 2011), while intermediate ones were obtained in progenies from crosses. For all groups, C18:1 was clearly the most abundant fatty acid, followed by C16:0 and C18:2 (Table 1). Wilds showed the highest mean content for C16:0 and C16:1, cultivars the opposite results, and progenies from crosses intermediate values. Not so clear results were obtained for C18:1 and C18:2, where progenies from crosses showed the highest and lowest mean values, respectively. However, it should be noted that high oleic acid 'Picual' cultivar was used as parent for these crosses, which may have affected these results due to the high heritability reported for this character (De la Rosa et al., 2016). Similar average composition has been previously reported in the evaluation of wild olives from different origins, such as Tunisia (Baccouri et al., 2011), Turkey (Matthäus et al., 2014), and Pakistan (Anwar et al., 2013). In contrast with our results, much lower values of C16:0 were found in wilds olives from Algeria (Boucheffa et al., 2014) and Tunisia (Dabbou et al., 2011). Besides, no differentiation in the distribution of fatty acid composition was observed between wild and cultivated olives in Tunisia (Hannachi et al., 2008; Dabbou et al., 2011). In addition, extreme values have been reported for subsp. *cuspidata* specimens from Kenya, with lower C18:1 content (up to 44.3%) and subsequently higher C18:2 content (up to 33.3%) than those identified in the present research for wilds. Such contrasting results found in wild materials may be probably attributed to the different environmental factors linked to the geographical position where the genotypes were evaluated *in situ* (Hannachi et al., 2009), to the representativeness of the oleasters included in the studies (Dabbou et al., 2011) as well as to a possible feral origin of them (Hannachi et al., 2008; Dabbou et al., 2011).

Significant differences between groups were also obtained for total amounts of minor components, with cultivars showing the highest total values for SQ and ST and the lowest values for TO. Wilds showed the lowest average value for SQ and the highest average value for TO. High TO content has been previously reported in samples from wild olives (Baccouri et al., 2008; Boucheffa et al., 2014). The lowest average value for ST was found in the progenies from crosses. Again, low ST content of 'Picual' parent could have affected these results due to the high heritability also reported for this character (De la Rosa et al., 2016). No or only minimum differences were obtained regarding minor components composition, being  $\alpha$ -tocopherol and  $\beta$ -sitosterol the predominant TO and ST forms, respectively, in all groups. The ST profile in wild olives is in agreement with the results obtained by Hannachi et al. (2013), although the authors did not quantify total ST content. To the best of our knowledge, no information on total SQ content in wild olive germplasm is found in the literature.

Pearson's correlation coefficients among individual components reflected relationships previously inferred from

PCA analysis (Table 3). Some highly significant values among components of the same chemical family were found in all groups of plant materials, for instance between  $\alpha$ -tocopherol and  $\gamma$ -tocopherol or between  $\delta$ 5-avenasterol and  $\beta$ -sitosterol (data not showed), as previously reported for other materials in our breeding program (De la Rosa et al., 2016). However, no significant correlations were found among components of different chemical families.

In relation to the fatty acid profile, the most significant correlations were found between C16:0 and C16:1, which were positively correlated, and between C18:1 and both C16:0 and C18:2, in both cases negatively correlated. The evaluation of linear regressions between C18:1 and both C16:0 and C18:2 in the three groups of trees revealed marked differences in the coefficients of determination and/or the slope of the regression line (Figure 3). Thus, in the regression between C16:0 and C18:1, crosses showed a lower slope compared to cultivars and wilds. A similar situation was observed in the regression between C18:1 and C18:2 (Figure 3). In the latter case, the coefficient of determination was considerably lower in the crosses than in the parents and wilds, attributable to the existence of lower variability for both traits in the crosses than in the cultivated and wild parents (Figure 3 and Table 1). The fatty acid composition of olive oil depends on both genetic and environmental factors, with genotypic variance having been reported as the main contributor to total variance in studies involving high genetic variability (León et al., 2008; Rondanini et al., 2011), as it is the case for the present study. The occurrence of strong negative correlation between oleic acid and both palmitic and linoleic acid has been reported previously in olive, which has been attributed to the relationship of these fatty acids in the biosynthetic pathway (Dabbou et al., 2012). Biosynthesis of C16 and C18 fatty acids in olive is located in plastids, where a series of enzymatic reactions that elongate an acyl chain bound to an acyl carrier protein (ACP) by the stepwise addition of two-carbon molecules are catalyzed by the fatty acid synthase complex. The last elongation step produces C18:0-ACP from C16:0-ACP. Then, C18:0-ACP can be desaturated by the C18:0-ACP desaturase to produce C18:1-ACP. The fatty acids C16:0, C18:0, and C18:1 can be exported from the plastid after the hydrolysis of the acyl-ACP by acyl-ACP thioesterases. Outside the plastid, C18:1 can be desaturated to C18:2 by the microsomal enzyme oleoyl-phosphatidylcholine desaturase (FAD2) (Somerville et al., 2000). This biosynthetic pathway is complex due to the existence

of several rate-limiting reactions controlling carbon flux through the pathway, which determines that the levels of one fatty acid may alter the activity of other enzymes within the pathway and subsequently the levels of other fatty acids (Salas et al., 2014).

In summary, the results of the study seem to indicate that the use of wild germplasm in olive breeding programs could not have a negative impact on fatty acid composition, tocopherol content, and tocopherol and phytosterol profiles provided that selection for these compounds is conducted from early generations. Important traits such as TO content can be even improved by using wild parents. Conversely, our results indicated a putative negative impact on both total phytosterol and squalene contents, although this needs to be confirmed with studies involving larger populations.

## AUTHOR CONTRIBUTIONS

All authors conceived and designed the experiment. AB was in charge of plant materials selection and samples collection. LL, RdLR, and LV performed fruit traits and oil quality analysis. LL prepared the first draft of the manuscript, all authors critically reviewed the manuscript prior to submission, read and approved the final version of the manuscript and agreed to be accountable for accuracy, integrity, and appropriateness of the manuscript.

## FUNDING

CICE project P11-AGR-7301, IFAPA projects AVA201601.2 and PR.PEI.IDF201601.2 partially funded by European Regional Development Fund, and European Union's Horizon 2020 RISE project BeFore funded under the Marie Skłodowska-Curie grant agreement No. 645595. The conservation and management of olive cultivars at WOGB IFAPA Córdoba has been financially supported by INIA Projects (RFP 2012-00005; RFP 2013-00005).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00232/full#supplementary-material>

## REFERENCES

- Anwar, P., Bendini, A., Gulfranz, M., Qureshi, R., Valli, E., Di Lecce, G., et al. (2013). Characterization of olive oils obtained from wild olive trees (*Olea ferruginea* Royle) in Pakistan. *Food Res. Int.* 54, 1965–1971. doi: 10.1016/j.foodres.2013.09.029
- Aparicio, R., and Luna, G. (2002). Characterisation of monovarietal virgin olive oils. *Eur. J. Lipid Sci. Technol.* 104, 614–627. doi: 10.1002/1438-9312(200210)104:9/10<614::AID-EJLT614>3.0.CO;2-L
- Arias-Calderón, R., Rodríguez-Jurado, D., León, L., Bejarano-Alcazar, J., De la Rosa, R., and Belaj, A. (2015). Pre-breeding for resistance to *Verticillium* wilt in olive: fishing in the wild relative gene pool. *Crop Prot.* 75, 25–33. doi: 10.1016/j.cropro.2015.05.006
- Baccouri, B., Guerfel, M., Zarrouk, W., Taamalli, W., Daoud, D., and Zarrouk, M. (2011). Wild olive (*Olea europaea* L.) selection for quality oil production. *J. Food Biochem.* 350, 161–176. doi: 10.1111/j.1745-4514.2010.00373.x
- Baccouri, B., Zarrouk, W., Baccouri, O., Guerfel, O., Nouairi, I., Krichene, D., et al. (2008). Composition, quality and oxidative stability of virgin olive oils from some selected wild olives (*Olea europaea* L. subsp. *Oleaster*). *Grasas y Aceites* 59, 346–351. doi: 10.3989/gya.2008.v59.i4.528

- Baldoni, L., Tosti, N., Ricciolini, C., Belaj, A., Arcioni, S., Pannelli, G., et al. (2006). Genetic structure of wild and cultivated olives in the central mediterranean basin. *Ann. Bot.* 98, 935–942. doi: 10.1093/aob/mcl178
- Belaj, A., León, L., Satovic, Z., and De la Rosa, R. (2011). Variability of wild olives (*Olea europaea* subsp. *europaea* var. *sylvestris*) analysed by agromorphological traits and SSR markers. *Sci. Hortic.* 129, 561–569. doi: 10.1016/j.scienta.2011.04.025
- Belaj, A., Munoz-Diez, C., Baldoni, L., Porceddu, A., Barranco, D., and Satovic, Z. (2007). Genetic diversity and population structure of wild olives from the north-western Mediterranean assessed by SSR markers. *Ann. Bot.* 100, 449–458. doi: 10.1093/aob/mcm132
- Belaj, A., Muñoz-Diez, C., Baldoni, L., Satovic, Z., and Barranco, D. (2010). Genetic diversity and relationships of wild and cultivated olives at regional level in Spain. *Sci. Hortic.* 124, 323–330. doi: 10.1016/j.scienta.2010.01.010
- Beltrán, G., Bucheli, M. E., Aguilera, M. P., Belaj, A., and Jimenez, A. (2015). Squalene in virgin olive oil: screening of variability in olive cultivars. *Eur. J. Lipid Sci. Technol.* 118, 1250–1253. doi: 10.1002/ejlt.201500295
- Bodoira, R., Torres, M., Pierantozzi, P., Aguete, F., Taticchi, A., Servili, M., and Maestri, D. (2016). Dynamics of fatty acids, tocopherols and phenolic compounds biogenesis during olive (*Olea europaea* L.) fruit ontogeny. *J. Am. Oil Chem. Soc.* 93, 1289–1299. doi: 10.1007/s11746-016-2877-7
- Boucheffa, S., Miazzi, M. M., di Rienzo, V., Mangini, G., Fanelli, V., Tamendjari, A., et al. (2016). The coexistence of oleaster and traditional varieties affects genetic diversity and population structure in Algerian olive (*Olea europaea*) germplasm. *Genet. Resour. Crop Evol.* 64, 379–390. doi: 10.1007/s10722-016-0365-4
- Boucheffa, S., Tamendjari, A., Rovellini, P., and Venturini, S. (2014). Composition and antioxidant activity of some algerian wild extra virgin olive oils. *Riv. Ital. Sostanze Gr.* 91, 177–185.
- Breton, C., Tersac, M., and Bervillé, A. (2006). Genetic diversity and gene flow between the wild olive (oleaster, *Olea europaea* L.) and the olive: several Plio-Pleistocene refuge zones in the Mediterranean basin suggested by simple sequence repeats analysis. *J. Biogeogr.* 33, 1916–1928. doi: 10.1111/j.1365-2699.2006.01544.x
- Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., et al. (2016). Global conservation priorities for crop wild relatives. *Nat. Plants* 2:16022. doi: 10.1038/nplants.2016.22
- Colella, C., Miacola, C., Amenduni, M., D'Amico, M., Bubici, G., and Cirulli, M. (2008). Sources of verticillium wilt resistance in wild olive germplasm from the Mediterranean region. *Plant Pathol.* 57, 533–539. doi: 10.1111/j.1365-3059.2007.01785.x
- Dabbou, S., Chaieb, I., Rjiba, I., Issaoui, M., Echbili, A., Nakbi, A., et al. (2012). Multivariate data analysis of fatty acid content in the classification of olive oils developed through controlled crossbreeding. *J. Am. Oil Chem. Soc.* 89, 667–674. doi: 10.1007/s11746-011-1946-1
- Dabbou, S., Dabbou, S., Selvaggini, R., Urbani, S., Taticchi, A., Servili, M., et al. (2011). Comparison of the chemical composition and the organoleptic profile of virgin olive oil from two wild and two cultivated Tunisian *Olea europaea*. *Chem. Biodivers.* 8, 189–202. doi: 10.1002/cbdv.201000086
- De la Rosa, R., Arias-Calderón, R., Velasco, L., and León, L. (2016). Early selection for oil quality components in olive breeding progenies. *Eur. J. Lipid Sci. Technol.* 118, 1160–1167. doi: 10.1002/ejlt.201500425
- De la Rosa, R., Klepo, T., Arias-Calderón, R., Toumi, A., Domínguez-García, M. C., León, L., Belaj, A., Mariotti, R., et al. (2014). Current status of conservation, evaluation and usefulness of wild olive germplasm. *Acta Hortic.* 1057, 515–520. doi: 10.17660/ActaHortic.2014.1057.65
- De la Rosa, R., Talhaoui, N., Rouis, H., Velasco, L., and León, L. (2013). Fruit characteristics and fatty acid composition in advanced olive breeding selections along the ripening period. *Food Res. Int.* 54, 1890–1896. doi: 10.1016/j.foodres.2013.08.039
- Erre, P., Chessa, I., Munoz-Diez, C., Belaj, A., Rallo, L., and Trujillo, I. (2010). Genetic diversity and relationships between wild and cultivated olives (*Olea europaea* L.) in Sardinia as assessed by SSR markers. *Genet. Resour. Crop Evol.* 279, 41–54. doi: 10.1007/s10722-009-9449-8
- Green, P. S. (2002). A revision of *Olea* L. (Oleaceae). *Kew Bull.* 57, 91–140. doi: 10.2307/4110824
- Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* 156, 1–13. doi: 10.1007/s10681-007-9363-0
- Hannachi, H., Breton, C., Msallem, M., El-Hadj, S. B., El-Gazzah, M., and Bervillé, A. (2008). Differences between native and introduced olive cultivars as revealed by morphology of drupes, oil composition and SSR polymorphisms: a case study in Tunisia. *Sci. Hortic.* 116, 280–290. doi: 10.1016/j.scienta.2008.01.004
- Hannachi, H., Nasri, N., Elfalleh, W., Tlili, N., Ferchichi, A., and Msallem, M. (2013). Fatty Acids, sterols, polyphenols, and chlorophylls of olive oils obtained from Tunisian wild olive trees (*Olea europaea* L. var. *sylvestris*). *Int. J. Food Prop.* 16, 1271–1283. doi: 10.1080/10942912.2011.584201
- Hannachi, H., Sommerlatte, H., Breton, C., Msallem, M., El-Gazzah, M., El-Hadj, S. B., et al. (2009). Oleaster (var. *sylvestris*) and subsp. *cuspidata* are suitable genetic resources for improvement of the olive (*Olea europaea* subsp. *europaea* var. *europaea*). *Genet. Resour. Crop Evol.* 56, 393–403. doi: 10.1007/s10722-008-9374-2
- International Olive Council [IOC] (2016). *Trade Standard Applying to Olive Oil and Olive Pomace Oils. COI/T.15/NC No 3/Rev. 11*. Available at: <http://www.internationaloliveoil.org/>
- Klepo, T., De la Rosa, R., Satovic, Z., León, L., and Belaj, A. (2013). Utility of wild germplasm in olive breeding. *Sci. Hortic.* 152, 92–101. doi: 10.1016/j.scienta.2012.12.010
- Klepo, T., Toumi, A., de la Rosa, R., León, L., and Belaj, A. (2014). Agronomic evaluation of seedlings from crosses between the main Spanish olive cultivar 'Picual' and two wild olive trees. *J. Hortic. Sci. Biotechnol.* 89, 508–512. doi: 10.1080/14620316.2014.11513113
- Kyčyk, O., Aguilera, M. P., Gaforio, J. J., Jiménez, A., and Beltrán, G. (2016). Sterol composition of virgin olive oil of forty-three olive cultivars from the world collection olive germplasm bank of cordoba. *J. Sci. Food Agric.* 96, 4143–4150. doi: 10.1002/jsfa.7616
- León, L., De la Rosa, R., Gracia, A., Barranco, D., and Rallo, L. (2008). Fatty acid composition of advanced olive selections obtained by crossbreeding. *J. Sci. Food Agric.* 88, 1921–1926. doi: 10.1002/jsfa.3296
- Matthäus, B., Özcan, M. M., and Juhaime, F. A. L. (2014). Some physico-chemical properties and composition in wild olive (*Olea europaea* L. subsp. *oleaster*) fruit and oil. *Riv. Ital. Sostanze Gr.* 91, 255–259.
- Pérez-Jiménez, F., Ruano, J., Perez-Martinez, P., Lopez-Segura, F., and Lopez-Miranda, J. (2007). The influence of olive oil on human health: not a question of fat alone. *Mol. Nutr. Food Res.* 51, 1199–1208. doi: 10.1002/mnfr.200600273
- Rondanini, D. P., Castro, D. N., Searles, P. S., and Rousseaux, M. C. (2011). Fatty acid profiles of varietal virgin olive oils (*Olea europaea* L.) from mature orchards in warm arid valleys of Northwestern Argentina (La Rioja). *Grasas y Aceites* 62, 399–409. doi: 10.3989/gya.125110
- Salas, J. J., Martínez-Force, E., Harwood, J. L., Venegas-Calderón, M., Aznar-Moreno, J. A., Moreno-Pérez, A. J., et al. (2014). Biochemistry of high stearic sunflower, a new source of saturated fats. *Prog. Lipid Res.* 55, 30–42. doi: 10.1016/j.plipres.2014.05.001
- Sánchez de Medina, V., Calderón-Santiago, M., El Riachy, M., Priego-Capote, F., and Luque de Castro, M. D. (2014). High-resolution mass spectrometry to evaluate the influence of cross-breeding segregating populations on the phenolic profile of virgin olive oils. *J. Sci. Food Agric.* 94, 3100–3109. doi: 10.1002/jsfa.6653
- Sánchez de Medina, V., El Riachy, M., Priego-Capote, F., and Luque de Castro, M. D. (2015). Composition of fatty acids in virgin olive oils from cross breeding segregating populations by gas chromatography separation with flame ionization detection. *J. Sci. Food Agric.* 95, 2892–2900. doi: 10.1002/jsfa.7030

- Sedgley, M. (2004). *Wild Olive Selection for quality oil production*. Canberra, ACT: Rural Industries Research and Development Corporation.
- Somerville, C., Browse, J., Jaworski, J. G., and Ohlrogge, J. B. (2000). "Lipids," in *Biochemistry and molecular biology of plants*, eds B. B. Buchanan, W. Gruissem, and R. L. Jones (Rockville, MD: American Society of Plant Physiologists), 456–527.
- Velasco, L., Fernandez-Cuesta, A., De la Rosa, R., Victoria Ruiz-Mendez, M., and León, L. (2014). Selection for some olive oil quality components through the analysis of fruit flesh. *J. Am. Oil Chem. Soc.* 91, 1731–1736. doi: 10.1007/s11746-014-2523-1

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 León, de la Rosa, Velasco and Belaj. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Divergent N Deficiency-Dependent Senescence and Transcriptome Response in Developmentally Old and Young *Brassica napus* Leaves

Vajiheh Safavi-Rizi<sup>1</sup>, Jürgen Franzaring<sup>2</sup>, Andreas Fangmeier<sup>2</sup> and Reinhard Kunze<sup>1\*</sup>

<sup>1</sup> Institute of Biology, Dahlem Centre of Plant Sciences, Free University Berlin, Berlin, Germany, <sup>2</sup> Institute of Landscape and Plant Ecology, University of Hohenheim, Stuttgart, Germany

## OPEN ACCESS

### Edited by:

Dragana Miladinović,  
Institute of Field and Vegetable Crops,  
Serbia

### Reviewed by:

Frédéric Marsolais,  
Agriculture and Agri-Food Canada  
(AAFC), Canada  
Astrid Wingler,  
University College Cork, Ireland

### \*Correspondence:

Reinhard Kunze  
reinhard.kunze@fu-berlin.de

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 02 October 2017

**Accepted:** 10 January 2018

**Published:** 01 February 2018

### Citation:

Safavi-Rizi V, Franzaring J,  
Fangmeier A and Kunze R (2018)  
Divergent N Deficiency-Dependent  
Senescence and Transcriptome  
Response in Developmentally Old  
and Young *Brassica napus* Leaves.  
Front. Plant Sci. 9:48.  
doi: 10.3389/fpls.2018.00048

In the spring oilseed rape (OSR) cultivar ‘Mozart’ grown under optimal N supply ( $N_O$ ) or mild N deficiency ( $N_L$ ) the transcriptome changes associated with progressing age until early senescence in developmentally old lower canopy leaves (leaf #4) and younger higher canopy leaves (leaf #8) were investigated. Twelve weeks old  $N_O$  and  $N_L$  plants appeared phenotypically and transcriptomically identical, but thereafter distinct nutrition-dependent differences in gene expression patterns in lower and upper canopy leaves emerged. In  $N_O$  leaves #4 of 14-week-old compared to 13-week-old plants, ~600 genes were up- or downregulated, whereas in  $N_L$  leaves #4 ~3000 genes were up- or downregulated. In contrast, in 15-week-old compared to 13-week-old upper canopy leaves #8 more genes were up- or downregulated in optimally N-supplied plants (~2000 genes) than in N-depleted plants (~750 genes). This opposing effect of N depletion on gene regulation was even more prominent among photosynthesis-related genes (PSGs). Between week 13 and 14 in leaves #4, 99 of 110 PSGs were downregulated in  $N_L$  plants, but none in  $N_O$  plants. In contrast, from weeks 13 to 16 in leaves #8 of  $N_L$  plants only 11 PSGs were downregulated in comparison to 66 PSGs in  $N_O$  plants. Different effects of N depletion in lower versus upper canopy leaves were also apparent in upregulation of autophagy genes and NAC transcription factors. More than half of the regulated NAC and WRKY transcription factor, autophagy and protease genes were specifically regulated in  $N_L$  leaves #4 or  $N_O$  leaves #8 and thus may contribute to differences in senescence and nutrient mobilization in these leaves. We suggest that in N-deficient plants the upper leaves retain their N resources longer than in amply fertilized plants and remobilize them only after shedding of the lower leaves.

**Keywords:** autophagy, *Brassica napus*, leaf senescence, N remobilization, N-deficiency, oilseed rape, transcriptome, transcription factor

## INTRODUCTION

In the past three decades the worldwide oilseed rape acreage has expanded nearly threefold to 36 million ha and the production has increased even fivefold to 73 million tons in 2013 (Food and Agriculture Organization of the United Nations)<sup>1</sup>. In winter oilseed rape production fertilization

<sup>1</sup> <http://faostat3.fao.org>

with up to 200 kg nitrogen (N) ha<sup>-1</sup> year<sup>-1</sup> is common practice. Although oilseed rape (OSR) has a high uptake capacity for inorganic N, its nitrogen use efficiency (NUE; for definitions see Masclaux-Daubresse et al., 2010; Xu et al., 2012) is low. Only 50–60% of the applied N is recovered in the plants and at the time of harvest 80% of the total plant N is localized in the seeds (Schjoerring et al., 1995; Jensen et al., 1997; Malagoli et al., 2005a; Rathke et al., 2006). Accordingly, winter OSR production has a high N balance surplus that often exceeds the limit of 60 kg ha<sup>-1</sup> year<sup>-1</sup> that is effective since 2009 in Germany (Düngeverordnung)<sup>2</sup> and the European Union (Nitrates Directive<sup>3</sup>). To meet these requirements without compromising seed yield, the development of cultivars with improved NUE at reduced fertilizer input is an important agricultural goal in OSR breeding.

Two factors determining the NUE are the N-uptake ability of the plants and the N-remobilization efficiency from old, senescing leaves during pod development and seed ripening. N-uptake increases in young plants approximately until flowering, but stagnates or even decreases during pod ripening and contributes only a minor fraction of the N in the seeds (Schjoerring et al., 1995; Rossato et al., 2001; Malagoli et al., 2005a; Gombert et al., 2010). Indeed, the majority of N required for seed filling and pod ripening is mobilized from senescing leaves and stems (Malagoli et al., 2005a; Gombert et al., 2010). Although the N-efficiency of winter OSR can also be enhanced by breeding cultivars with enhanced N-uptake ability (Schulte auf'm Erley et al., 2007), strengthening the N remobilization activity of leaves during the vegetative phase is a promising approach for improving the NUE of oilseed rape (Gironde et al., 2015). In a simulation model of N partitioning, Malagoli et al. (2005b) came to the conclusion that by optimizing N remobilization from leaves at lower nodes and N retranslocation from vegetative to reproductive tissues, OSR yield could be increased by 15%.

Yet, the shed leaves from lower nodes still have a high N content of up to 3.5% whereas leaves from upper nodes contain at the time of abscission only 1% residual N (Malagoli et al., 2005a). What limits N remobilization from early senescing leaves? Phloem loading of amino acids from degraded leaf proteins appears not to be the limiting step (Tilsner et al., 2005). In many winter OSR cultivars the onset of senescence and abscission of lower node leaves occurs already during the vegetative stages before the development of pods and seeds. This lack of sink organs supposedly leads to a low N remobilization rate from early leaves (Schjoerring et al., 1995; Rossato et al., 2001; Noquet et al., 2004; Malagoli et al., 2005a). Accordingly, winter cultivars with a delayed leaf senescence phenotype ('functional stay-green'; reviewed in Thomas and Ougham, 2014) tend to have a higher N-efficiency (Schulte auf'm Erley et al., 2007; Gregersen et al., 2013; Koeslin-Findeklee et al., 2015a,b).

The onset of leaf senescence is regulated by multiple, endogenous and environmental factors, among them N deficiency (Gregory, 1937; Mei and Thimann, 1984; Masclaux-Daubresse et al., 2007; Bieker and Zentgraf, 2013;

Koeslin-Findeklee et al., 2015b). However, the developmental response to N deficiency and timing of senescence initiation are not uniform throughout the plant body. During development of winter OSR plants senescence progresses sequentially from the bottom toward the top and the sink leaves in young plants later turn into source leaves during pod ripening (reviewed by Avice and Etienne, 2014). N-deprivation triggers earlier onset of senescence in older leaves, whereas in young leaves at higher nodes senescence is delayed (Etienne et al., 2007; Desclos et al., 2008). Thus, the spatially and temporally concerted modulation of senescence initiation is a promising target to improve the N-efficiency of OSR, but it requires a deeper understanding of the metabolic and transcriptional changes associated with leaf senescence initiation and progression in different parts of the plant. Koeslin-Findeklee et al. (2015b) identified in a study of transcriptomic changes following senescence induction by N-depletion in leaves from a lower node of two *B. napus* winter cultivars differing in their stay-green properties and N-efficiency, a large number of cultivar-specifically regulated, senescence-associated genes, but they did not address leaf-rank specific expression differences.

In this study we report that in the doubled haploid OSR spring cultivar 'Mozart' senescence progression and the effect of N-limitation are similar as in winter OSR cultivars and we present a genome-wide developmental transcription analysis of plants grown under standard or reduced N-supply. The developmental transcription changes in lower and upper canopy leaves of plants grown under low N-fertilization indicated that in old (source) leaves senescence was initiated earlier and this onset was accompanied by extensive transcriptional reprogramming. In contrast, in young (sink) leaves at a node below the inflorescence, transcriptional reprogramming was delayed in N-depleted plants. We identified transcription regulator, autophagy and protease genes that were specifically regulated in N-depleted lower canopy leaves or in upper leaves under ample N supply, and genes that were expressed senescence-associated in oilseed rape, but not in *Arabidopsis*. We hypothesize that some of these genes may have OSR-specific functions in N-remobilization during N-deficiency induced leaf senescence and contribute to differences in senescence execution and nutrient mobilization in upper and lower canopy leaves.

## MATERIALS AND METHODS

### Plant Material and Growth Conditions

Oilseed rape spring cultivar *Brassica napus* cv. 'Mozart' plants (BSA Nr. RAS 502, supplied by Norddeutsche Pflanzenzucht Hans-Georg Lembke KG – NPZ, Hohenlieth, Germany) were cultivated in solid medium in growth chambers that simulated the daylight length and average daily temperature profile between 1991 and 2005 in South-West Germany from March 15th (day 0: sowing) onward (Supplementary Table 1). Light intensity (photon flux density) during daylight phases was approximately 1000 μmol m<sup>-2</sup> s<sup>-1</sup>. The average CO<sub>2</sub> concentration during illumination was 396 ppm which

<sup>2</sup>[https://www.gesetze-im-internet.de/d\\_v\\_2017/](https://www.gesetze-im-internet.de/d_v_2017/)

<sup>3</sup>[http://ec.europa.eu/environment/water/water-nitrates/index\\_en.html](http://ec.europa.eu/environment/water/water-nitrates/index_en.html)

approximates ambient atmospheric conditions. During the dark phase the CO<sub>2</sub> concentration increased by approximately 100 ppm. A more detailed description of nursing, growth and physiological parameters of the plants analyzed in this study are presented in Franzaring et al. (2011). Leaf disks from early developing leaf #4 (at 78, 85, 92, and 99 days after sowing, DAS) and leaf #8 (at 92 and 106 DAS) were collected from plants grown at optimal (N<sub>O</sub>) or low (N<sub>L</sub>) N supply. For optimal N nutrition, NH<sub>4</sub>NO<sub>3</sub> was supplied in three equal gifts to each pot at germination (0 DAS; extended BBCH-scale stage GS0; Meier, 2001), 72 DAS (GS35) and 79 DAS (GS59) at an equivalent of 150 kg N ha<sup>-1</sup> t. For N<sub>L</sub> plants fertilizer gifts were reduced by half (75 kg N ha<sup>-1</sup> t). For each leaf sample three biological replicates from different plants were collected. Before harvesting, relative chlorophyll levels of the leaves were determined using a Konica Minolta SPAD-502 chlorophyll meter. For each leaf, SPAD values from two positions were measured and averaged.

## RNA Isolation

After freezing and grinding the samples in liquid nitrogen, total RNA was isolated by a hot phenol method as described (Drechsler et al., 2015). Total RNA was purified further using the RNeasy Mini Kit (Qiagen, Hilden, Germany). RNA quality was monitored on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States).

## *Brassica napus* Custom Microarray Design and Functional Annotation

The *Brassica napus* custom microarray was designed and processed as described in Koeslin-Findeklee et al. (2015b). Briefly, after a probe-preselection strategy established by ImaGenes GmbH (Berlin, Germany; now Source BioScience)<sup>4</sup> (Weltmeier et al., 2011) 60,955 probes representing 59,577 targets (EST clusters termed in this paper *B. napus* 'unigenes') were selected for the production of microarrays in the Agilent 8 × 60k format. Microarray design (GPL19044) and expression data (Series entry GSE97653) are deposited in the NCBI Gene Expression Omnibus (GEO) repository. To assign putative functions to the 59,577 *B. napus* 'unigenes,' they were locally BLASTed (RRID:SCR\_004870) against the TAIR10 *Arabidopsis thaliana* cDNA collection (Lamesch et al., 2012) using the BioEdit alignment editor<sup>5</sup> (RRID:SCR\_007361). Putative functions were attributed to *B. napus* unigenes based on the annotation of the most homologous *Arabidopsis thaliana* genes with a BLAST *E*-value ≤ 10<sup>-6</sup>. When multiple *B. napus* unigenes had the same *Arabidopsis* homolog, the unigene with the lowest *E*-value that is significantly regulated in any one sample was selected for further analysis.

## Processing and Bioinformatic Analysis of Microarray Data

Microarray expression data readouts were generated by the Agilent Feature Extraction software. The raw data files were

processed, normalized and analyzed with the Bioconductor package LIMMA<sup>6</sup> (RRID:SCR\_006442; Smyth, 2004). The *read.maimages* function was used to load the data into an RGList object. Background subtraction and quantile normalization was performed followed by statistical analysis (*moderated t-test*). The average of replicated spots was calculated using the *avereps* function. A design matrix was built for the *linear modeling* function and the intensity values were applied as *lmFit* function. Contrast matrices representing comparisons between different harvest time points and N treatments were created and applied to modeled data for computing the statistical significance. Regulated unigenes (≥3-fold expression change and Benjamini-Hochberg-corrected *p*-value *P*<sub>adj</sub> < 0.05) were clustered by their temporal expression profiles with the Short Time-series Expression Miner (STEM) software, RRID:SCR\_005016, using default settings (Ernst and Bar-Joseph, 2006). Grouping of unigenes into functional categories was performed with the BAR Classification SuperViewer Tool w/Bootstrap<sup>7</sup> (RRID:SCR\_006748) using MapMan (RRID:SCR\_003543) categories as annotation source (Provart and Zhu, 2003; Provart et al., 2003). Enriched GO-terms were identified with the DAVID Bioinformatics Resources 6.8 in the GOTERM\_BP\_DIRECT term compilation using default settings<sup>8</sup> (RRID:SCR\_003033; Huang et al., 2009a,b). Heat maps of differentially regulated genes were created using MultiExperiment Viewer (MeV)<sup>9</sup> (RRID:SCR\_001915; Saeed et al., 2006). *Arabidopsis thaliana* transcription factors were compiled from the AGRIS AtTFDB<sup>10</sup> (RRID:SCR\_006928; Yilmaz et al., 2011), autophagy-related genes from the Autophagy database<sup>11</sup> (RRID:SCR\_002671; Homma et al., 2011), and peptidases from the MEROPS database<sup>12</sup> (RRID:SCR\_002671; Rawlings et al., 2016).

## qPCR Primer Design and Assay for *B. napus* 'Unigenes'

Primers for quantitative real-time PCR (qPCR) were calculated by QuantPrime<sup>13</sup> (Arvidsson et al., 2008) after importing the *B. napus* unigene assemblies (Supplementary Table 13). One μg DNase I-digested total RNA was used for cDNA synthesis using SuperScript III Reverse Transcriptase (ThermoFisher Scientific). qPCR reactions were performed in 5 μl total volume including 2.5 μl Power SYBR Green Master Mix (ThermoFisher Scientific), 0.5 μM forward and reverse primers and 0.5 μl cDNA. *UP1* and *UBC9* were used as reference genes (Chen et al., 2010). The thermal profile used for all qPCRs was: 2 min 50°C > 10 min 95°C > (15 s 95°C > 1 min 60°C)<sub>40x</sub>. Data were analyzed by the 2<sup>-ΔΔCt</sup> method (Schmittgen and Livak, 2008).

<sup>6</sup>www.bioconductor.org/packages/release/bioc/html/limma.html

<sup>7</sup>http://bar.utoronto.ca/ntools/cgi-bin/ntools\_classification\_superviewer.cgi

<sup>8</sup>https://david.ncifcrf.gov

<sup>9</sup>www.tm4.org/

<sup>10</sup>http://arabidopsis.med.ohio-state.edu

<sup>11</sup>www.tanpaku.org/autophagy/index.html

<sup>12</sup>http://merops.sanger.ac.uk

<sup>13</sup>http://quantprime.mpimp-golm.mpg.de

<sup>4</sup>www.sourcebioscience.com

<sup>5</sup>www.mbio.ncsu.edu/BioEdit/bioedit.html

## RESULTS

### The Transcriptome Response to Reduced N Supply Differs in Early and Late Oilseed Rape Leaves

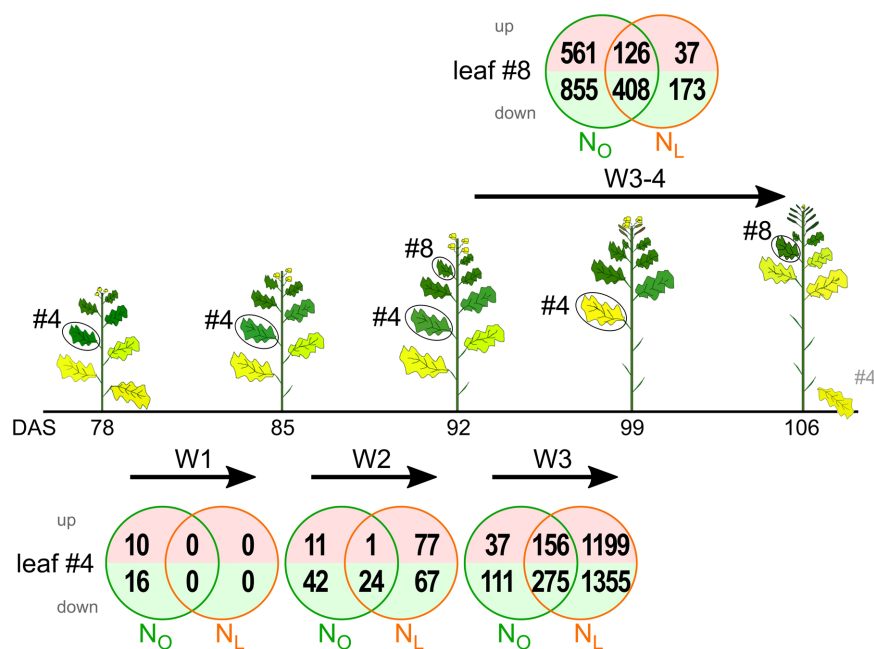
The aim of this study was to investigate if in spring oilseed rape (OSR) a mild N deficiency can be detected at the transcriptomic level, if the transcriptome response differs in developmentally older (source) leaves at a lower node and younger (sink) leaves at a higher node, and if the developmental response to N-deficiency resembles that in winter OSR cultivars. *Brassica napus* cv. 'Mozart' plants were raised under controlled conditions in a growth chamber under optimal N supply ( $N_O$ ) or N supply reduced by 50% ( $N_L$ ). Morphological, physiological and performance data of the same plants we investigated in this study were previously reported by Franzaring et al. (2011). The  $N_L$  conditions caused only subtle developmental and growth phenotypes (Supplementary Figure 1), flowering started on average only 2 days later than in  $N_O$  plants (Figure 3 in Franzaring et al., 2011), but seed yield was reduced (Table 1 in Franzaring et al., 2011). The early developing leaf #4 and the late developing leaf #8, located at the base of a flower developing side shoot, were harvested as representatives of old (source) leaves and young (sink) leaves, respectively. Leaf #4 was harvested at four different time points during development (78, 85, 92, and 99 days after sowing, DAS). Leaf #8 was harvested at the two time points 92 DAS and 106 DAS. Under both N treatments, at 92 DAS leaves #4 were still alive and attached to the stem, whereas at

106 DAS on most plants they were dead and shed (Figure 1 and Supplementary Figure 1).

Transcription analysis was performed using a *B. napus* custom microarray representing 59,577 'unigenes' (Koeslin-Findeklee et al., 2015b). For 54,095 (91%) of the *B. napus* 'unigenes' 19,185 homologs were identified in *Arabidopsis thaliana* (Supplementary Table 2). For only 5,522 of these *Arabidopsis* genes one single *B. napus* unigene is represented on the microarray, whereas for 71% more than one *B. napus* unigene exist (Supplementary Table 3). These unigenes include representatives in each of the 66 biological categories in the MapMan metabolic pathway visualizer (Thimm et al., 2004). In 41 (sub-)categories more than 70% of the corresponding genes are represented by a homologous *B. napus* unigene (Supplementary Table 4).

Under optimal N supply in leaf #4 the number of *B. napus* unigenes up- or downregulated relative to the previous harvest time point with significant ( $P_{adj} < 0.05$ ) and  $\geq 3$ -fold expression changes progressively increased from week 1 (26 genes) to week 2 (78 genes) to week 3 (579 genes) of the observation period (Figure 1 and Supplementary Table 5). The same trend, but with a much steeper increment, was observed in plants that were grown under reduced N fertilization. In the first week the transcriptome did not change at all, whereas 1 week later 169 regulated genes appeared and in week 3 the number of regulated genes jumped to 2,985. In both growth conditions and all time intervals the downregulated genes outnumbered the upregulated genes.

In the upper canopy leaf #8, 1,950 genes were up- or downregulated in  $N_O$  plants and 744 in  $N_L$  plants. Thus, the



**FIGURE 1 |** Transcriptome changes in oilseed rape leaf 4 (#4) and leaf 8 (#8) during development under optimal N supply ( $N_O$ ) and low N supply ( $N_L$ ). Leaves #4 and #8 were harvested between 78 and 106 days after sowing (DAS) as indicated below the sketches of the oilseed rape plants. The Venn diagrams display the numbers of upregulated and downregulated genes highlighted in light red and lime green, respectively. The examined developmental intervals are termed week 1 (W1), week 2 (W2), week 3 (W3) and weeks 3 + 4 (W3-4). Depicted are genes with significant,  $\geq 3$ -fold expression changes ( $P_{adj} < 0.05$ ,  $n = 3$ ).



relation of regulated genes in  $N_O$  and  $N_L$  plants was inverse compared to leaf #4. Also in leaf #8 the downregulated genes outnumbered the upregulated genes under both N fertilization regimes.

## Reduced N Supply Correlates with Differential Senescence Progression in Lower and Upper Canopy Leaves

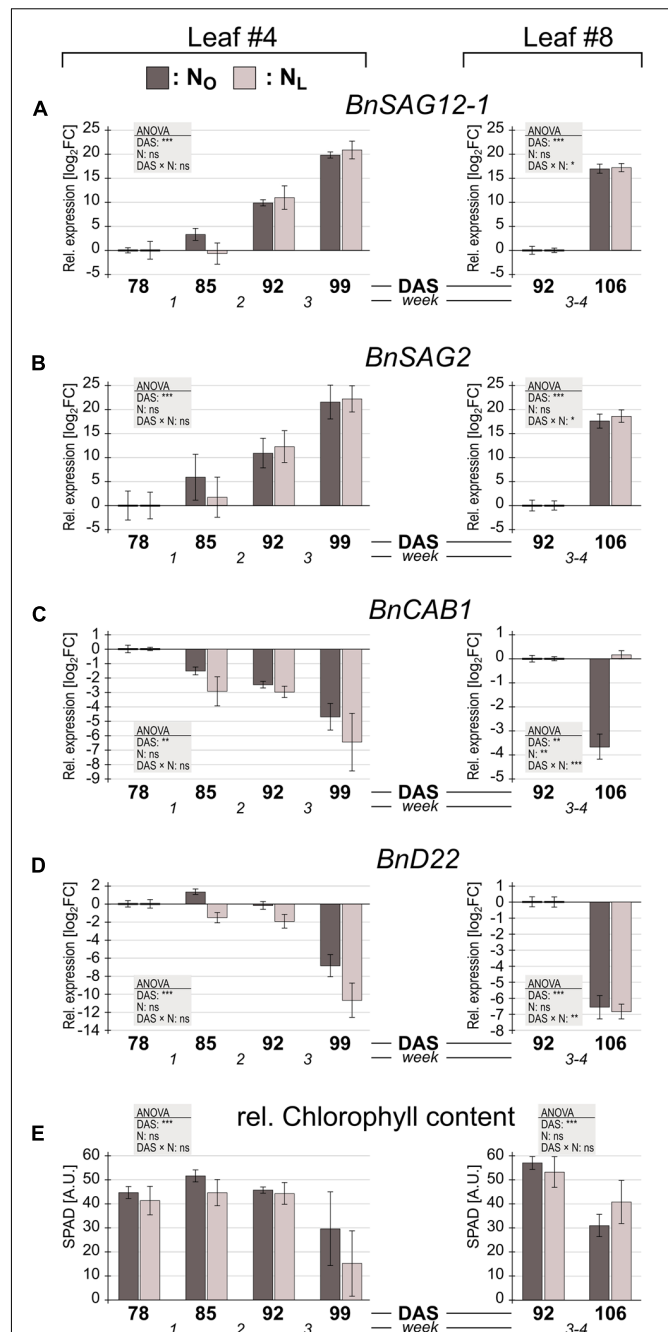
The observed massive increase in gene regulation might indicate the onset of leaf senescence, which is known to be accompanied by transcriptome reorganization (Buchanan-Wollaston et al., 2005; van der Graaff et al., 2006; Koeslin-Findeklee et al., 2015b). We therefore tracked senescence initiation by measuring chlorophyll content and expression of chlorophyll A/B binding protein gene *BnCAB1*, *Brassica napus* drought 22 kD protein gene *BnD22* and the senescence associated genes *BnSAG12-1* and *BnSAG2* by qPCR.

In leaves #4 of  $N_O$  plants, *BnSAG12-1* (Figure 2A) and *BnSAG2* (Figure 2B) expression by trend increased already in week 1 and continued to increase throughout weeks 2 and 3. In  $N_L$  leaves, upregulation of these two genes started only in week 2. Under both N treatments *BnCAB1* transcription appeared to decline in week 1 (Figure 2C), but the expression change was not significant (Supplementary Table 6). *BnD22*, whose expression level was approximately 3.5-fold higher under low N-conditions at the beginning of the observation period than under optimal N supply (Supplementary Table 5) as has also been reported by Desclos et al. (2008), displayed no significant expression change in weeks 1 and 2 (Supplementary Table 6), but a rapid decline in week 3 (Figure 2D). Expression of both SAGs and *CAB*, but not *BnD22*, indicated upcoming senescence one to 2 weeks before also a decline in chlorophyll was measurable (Figure 2E).

In leaf #8, the equal upregulation of *BnSAG12-1* and *BnSAG2* and downregulation of *BnD22* in  $N_O$  and in  $N_L$  leaves indicates that senescence has started during weeks 3–4. The regulation of *BnCAB1* was strikingly different in  $N_O$  and  $N_L$  leaves #8. Under low N supply *BnCAB1* expression was maintained, whereas under optimal N supply its transcription declined.

Neither in the lower or the upper leaves the differences between  $N_L$  and  $N_O$  plants in senescence marker gene expression and chlorophyll content were significant, whereas the expression pattern of *BnCAB1* is very different in  $N_L$  and  $N_O$  leaves #8. We therefore compared the expression levels of the 110 OSR homologs of *A. thaliana* photosynthesis-related genes (PSG) on the microarray (Supplementary Table 7). In the optimally N-supplied leaves #4, none of these genes were significantly up- or downregulated between 78 DAS and 99 DAS (except one downregulated gene in week 2). In striking contrast, in leaves #4 grown under reduced N fertilization, 94% of the PSGs were downregulated (4 PSGs in week 2 and another 99 PSGs in week 3). In the upper canopy leaves #8 the pattern was opposite: in  $N_O$  leaves 66 PSGs were downregulated compared to 11 downregulated PSGs in  $N_L$  leaves.

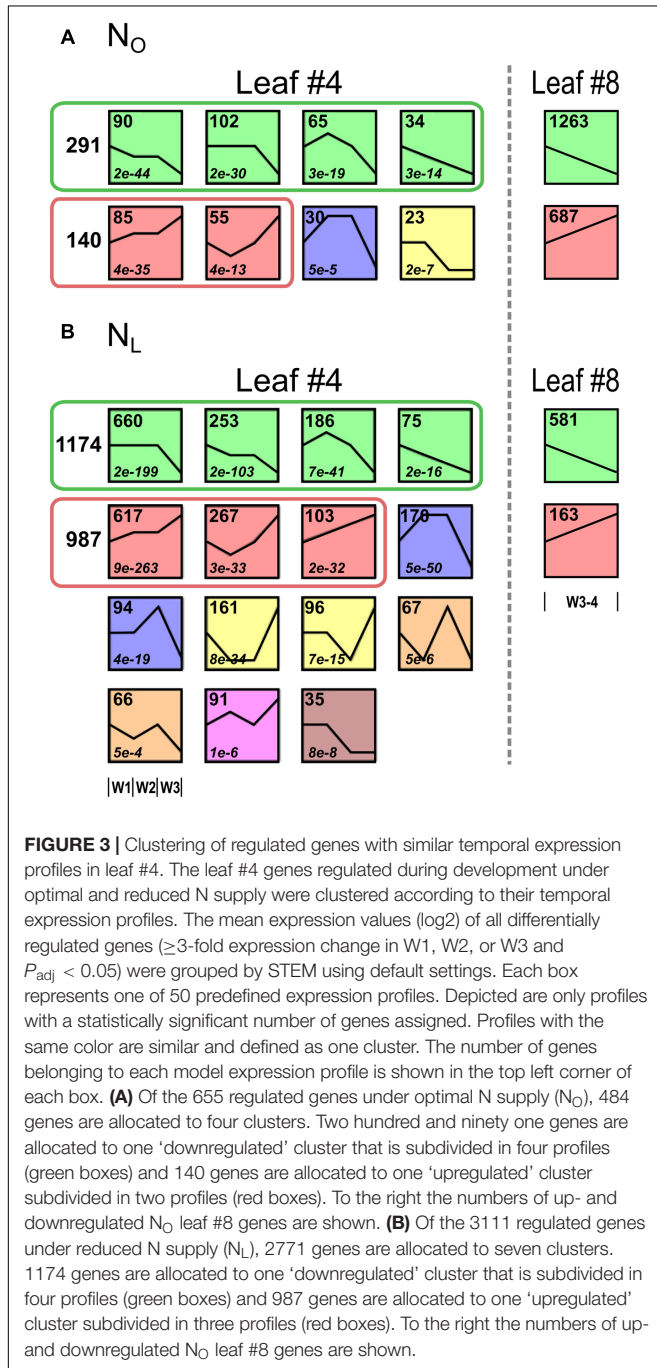
In summary, these data and the higher total number of up- and downregulated genes in  $N_L$  leaves #4 (Figure 1) indicate that a mild N deficiency leads not to a significant earlier initiation, but



**FIGURE 2 |** Senescence marker gene expression and chlorophyll content in *Brassica napus* leaves of plants grown under optimal or reduced N supply. *B. napus* leaves #4 and #8 of plants grown under optimal N supply ( $N_O$ ; dark gray columns) or under reduced N supply ( $N_L$ ; light gray columns) were harvested at the indicated days after sowing (DAS). Relative expression levels were determined by qPCR. For leaf #4, expression changes relative to the level at 78 DAS, and for leaf #8, expression changes relative to the level at 92 DAS are shown of (A) senescence associated gene *BnSAG12-1*, (B) senescence associated gene *BnSAG2*, (C) chlorophyll a/b binding protein gene *BnCAB1* and (D) *B. napus* gene *BnD22*. (E) Relative chlorophyll contents at each harvest time point are shown as SPAD values. Error bars indicate the standard error of the means ( $n = 3$ ). Significant expression differences by N treatment, over time and by interaction of the two parameters (Continued)

**FIGURE 2 | Continued**

was calculated by two way ANOVA (\* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ). Subsequently a Tukey's HSD *post hoc* test was done to identify significant differences between all different harvest time points (Supplementary Table 6).



to a more rapid progression of senescence once it has started. In contrast, in the upper canopy leaf #8 mild N deficiency causes a delay in senescence progression. Thus, in spring OSR the effect of N-deprivation on senescence in older and younger leaves is similar as in winter OSR (Etienne et al., 2007; Desclos et al., 2008).

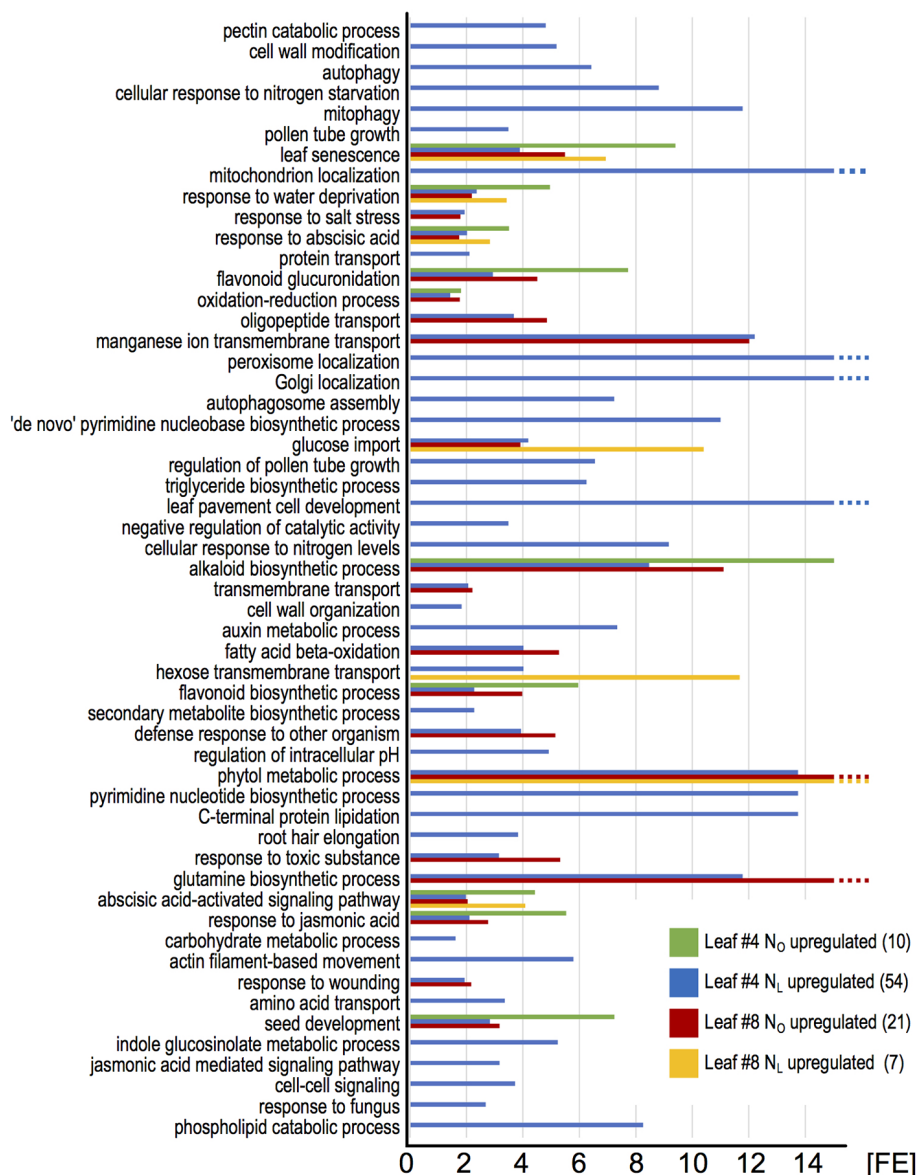
## N Fertilization-Dependent Gene Expression

To identify genes with similar expression change profiles during development under optimal and reduced N supply in the lower canopy leaf #4, the regulated genes were clustered by their expression profiles and assigned to 50 predefined model temporal expression profiles (Supplementary Table 8). Of the 665 genes up- or downregulated in  $N_0$  leaves #4, four clusters with eight profiles had a statistically significant number of genes assigned (Figure 3A). Overall, 291 of the genes (44%) are allocated to a cluster of four downregulated profiles and 140 genes (21%) fall into a cluster of two upregulated profiles. In leaves #4 of  $N_L$  plants, 2941 genes (94% of the 3111 regulated genes) are assigned to 15 model temporal expression profiles with a statistically significant number of genes (Figure 3B). Of these, 1174 genes (38%) are allotted to a cluster of four downregulated profiles and 987 genes (32%) to a cluster of three upregulated profiles, respectively. A conspicuous contrast between the transcriptomes of the developmentally younger upper canopy leaves #8 compared to the older leaves #4 is that in leaves #8 a higher number of genes is regulated during the 2 weeks observation interval in plants grown under optimal than under reduced N supply (Figure 1).

## Functional Classification of N Fertilization-Dependently Regulated Genes

To identify the most highly regulated N deficiency-responsive and senescence-associated pathways in leaves #4 and #8, we performed a Gene Ontology term enrichment analysis with the up- and downregulated genes in the two leaf #4 STEM clusters and in leaf #8 (Figure 3). To visualize considerably regulated biological processes in N-deficient leaves #4, which show the most progressed senescence symptoms, all significantly enriched ( $P < 0.05$ ) GO terms for the up- and downregulated genes are displayed in Figures 4, 5, respectively, and listed in Supplementary Table 9. The most significantly upregulated processes in  $N_L$  leaves #4 include cell wall weakening, cellular response to N starvation, intracellular bulk degradation of cytoplasmic components like chloroplasts and mitochondria (autophagy, mitophagy) and general leaf senescence activities.  $N_L$  leaves #4 shared almost 40% of upregulated GO terms with  $N_0$  leaves #8, but only 19 and 13% with  $N_0$  leaves #4 and  $N_L$  leaves #8, respectively. Of the 73 significantly downregulated GO terms in  $N_L$  leaves #4, approximately one quarter are associated with photosynthesis and related pathways, and another ~20% encompass biosynthesis of chlorophyll, amino acids, fatty acids, glucose, amylopectin, alkanes, and plastoquinone. In optimally N-supplied leaves #8, 40 of these GO terms (55%) were also downregulated, but only ~30% in  $N_0$  leaves #4 and in N-deficient leaves #8. In agreement with photosynthetic gene and senescence marker gene expression levels (Figure 2 and Supplementary Table 7), the GO term analysis suggests that senescence was most advanced in  $N_L$  leaves #4 at 99 DAS, followed by  $N_0$  leaves #8 at 106 DAS,  $N_L$  leaves #8 at 106 DAS and  $N_0$  leaves #4 at 99 DAS. Accordingly, N-deficiency induced in

## upregulated GO terms for biological processes



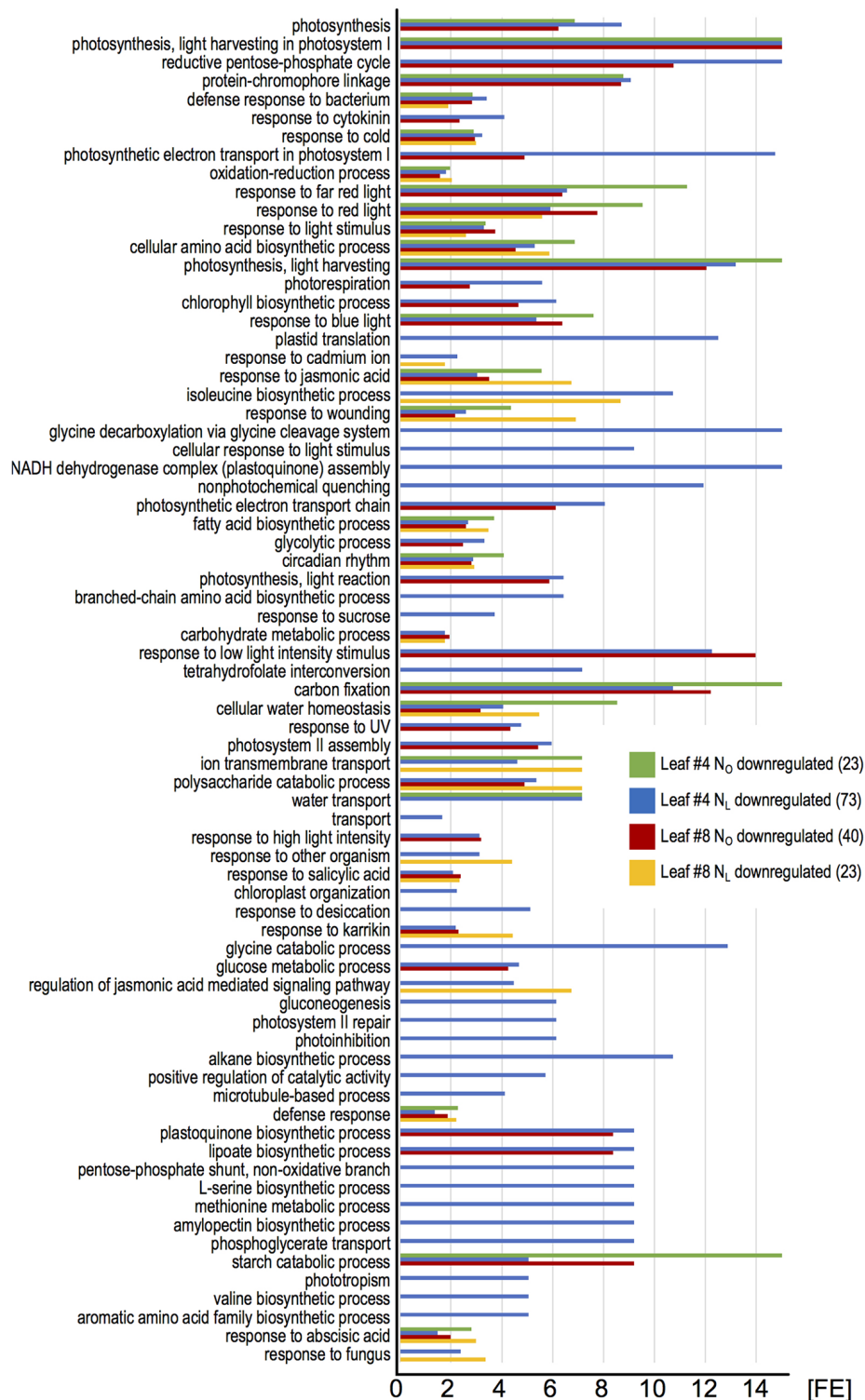
**FIGURE 4 |** Enriched Gene Ontology terms among upregulated leaf #4 genes. The upregulated genes in all samples (red pictograms in **Figure 2**) were analyzed for enriched Gene Ontology terms using the DAVID Bioinformatics Resources 6.8. The blue bars represent all significantly enriched GO\_BP\_DIRECT terms in N<sub>L</sub> leaves #4 ( $P < 0.05$ ). Those terms that were also enriched in other samples are shown as green, red, and yellow bars. The lack of a bar indicates that the fraction of upregulated genes in a GO term is not significantly higher than the overall fraction of upregulated genes. The numbers in brackets denote the count of enriched GO terms.

developmentally old leaves of the spring OSR cultivar 'Mozart' the accelerated progression of senescence and remobilization of nutrients, whereas in developmentally younger leaves in the upper canopy it led to a delay in senescence progression.

In an independent approach to identify up- or downregulated biological pathways, the temporally regulated genes in N<sub>0</sub> and N<sub>L</sub> leaves #4 were grouped by their putative biological functions according to the MapMan classification (Thimm et al., 2004), and categories enriched or depleted for regulated genes were

identified. Although the total number of regulated genes was five times higher in N<sub>L</sub> compared to N<sub>0</sub> plants, the majority of functional gene categories showed no significant differences in the fractions of regulated genes (Supplementary Figure 2). This is consistent with the weak phenotypic differences between N<sub>0</sub> and N<sub>L</sub> plants (Supplementary Figure 1). However, major differences are apparent in the categories tetrapyrrole synthesis and photosynthesis, oxidative pentose phosphate pathway (OPP) and C1-metabolism. In leaves #4 of N<sub>L</sub> plants, half of the genes

## downregulated GO terms for biological processes



**FIGURE 5 |** Enriched Gene Ontology terms among downregulated leaf #4 genes. The downregulated genes in all samples (green pictograms in **Figure 2**) were analyzed for enriched Gene Ontology terms using the DAVID Bioinformatics Resources 6.8. The blue bars represent all significantly enriched GO\_BP\_DIRECT terms in  $N_L$  leaves #4 ( $P < 0.05$ ). Those terms that were also enriched in other samples are shown as green, red, and yellow bars. The lack of a bar indicates that the fraction of downregulated genes in a GO term is not significantly higher than the overall fraction of downregulated genes. The numbers in brackets denote the count of enriched GO terms.



associated with photosynthesis (97 of 206 genes in this category) and 25% of the genes involved in chlorophyll biosynthesis (12 of 48 genes in the tetrapyrrole category) were downregulated, whereas in  $N_O$  plants only 5% of the photosynthesis and no chlorophyll biosynthesis genes were downregulated. Thus, the gene expression data reflect the reduced chlorophyll content in  $N_L$  plants (Figure 2). Also downregulated in  $N_L$ , but not in  $N_O$  plants, were the oxidative pentose phosphate pathway, which generates reductants required for various biosynthetic processes, including fatty acid synthesis and inorganic N and S assimilation (Kruger and von Schaewen, 2003; Bussell et al., 2013) and the one-carbon (C1) metabolism pathway. This pathway is also connected to the S-assimilation pathway by supplying C1 units for the synthesis of S-methylmethionine (SMM), which is transported from source leaves via the phloem to sink organs (Hanson and Roje, 2001). The MapMan classification of up- and downregulated leaf #8 genes revealed overall less differences in significantly regulated categories between  $N_O$  and  $N_L$  plants compared to the older leaves #4 (Supplementary Figure 3). However, noticeable differences are apparent in the categories tetrapyrrole synthesis and photosynthesis. Opposite to leaves #4, in leaves #8 in both categories a large fraction of the genes was downregulated in  $N_O$ , but not in  $N_L$  plants, suggesting that, although the chlorophyll content had not yet much declined between 92 DAS and 106 DAS, senescence was more advanced in leaves #8 of plants grown under optimal N supply than in plants grown under reduced N fertilization.

## Senescence- and N Deficiency-Associated Transcription Factor Genes

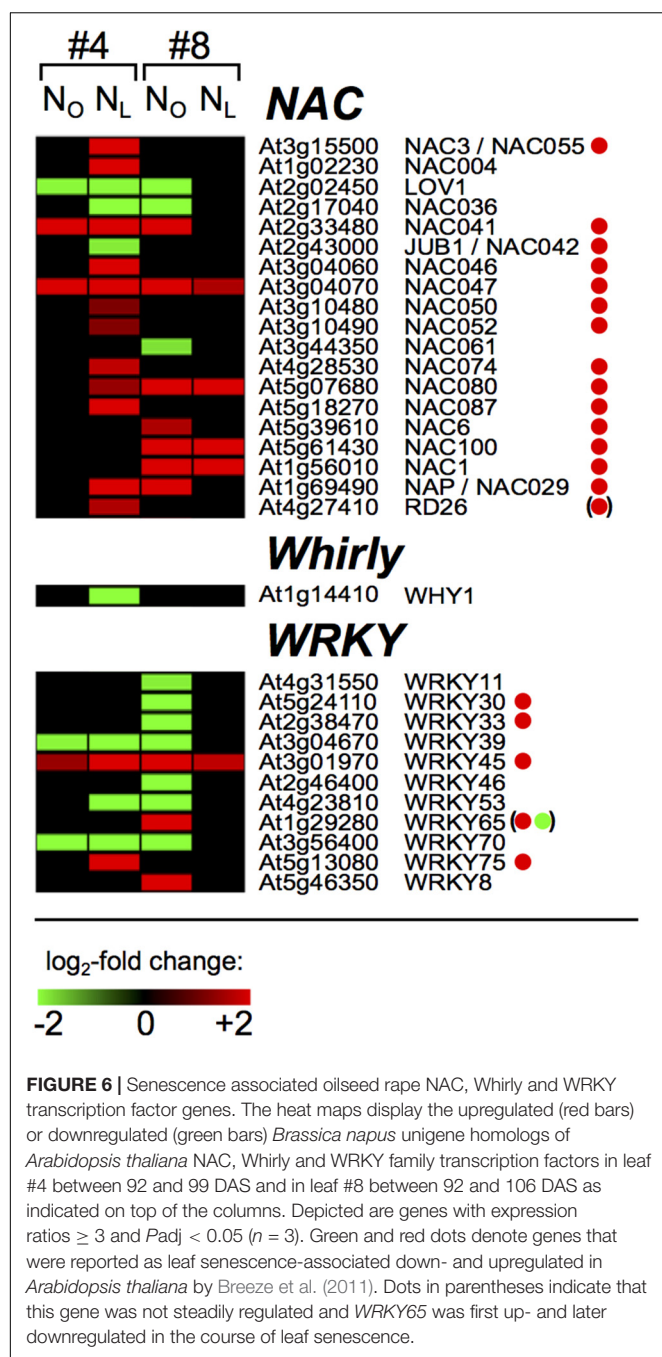
A major process during leaf senescence is remobilization of N and other nutritional degradation products from source to sink organs. The initiation and progression of senescence is orchestrated by transcription factors (TFs) and thus the identification of senescence-associated TFs that are responsive to N deficiency conditions is crucial for understanding the parameters that determine the N-efficiency of oilseed rape. We therefore investigated the range of *B. napus* homologs of *Arabidopsis* TFs that were differentially regulated upon N-deprivation in source leaves #4 and sink leaves #8. In total, 271 regulated OSR homologs of *Arabidopsis* TFs were found in 37 of the 51 *Arabidopsis* TF families (Yilmaz et al., 2011), and in most families more genes were down- than upregulated (Supplementary Table 10). Under both N-regimes, almost all TF genes in leaf #4 were regulated exclusively in week 3, only nine genes showed regulation during weeks 1 or 2 (Supplementary Table 11). Analogous to the frequencies of total regulated genes (Figure 1), between 92 and 99 DAS in N-deficient leaves #4, 3.6-fold more putative TF genes were transcriptionally regulated (78 genes up and 104 genes down) than in optimally N-supplied leaves #4 (22 genes up, 29 genes down). In leaf #8, between 92 and 106 DAS threefold more TF genes were regulated in  $N_O$  (48 genes up, 92 genes down) than in  $N_L$  plants (15 genes up, 21 genes down). None of the TF genes were oppositely regulated in  $N_O$  and  $N_L$  leaves #4 or leaves #8, or in  $N_L$  leaf #4 and  $N_O$  leaf #8.

Twelve genes were upregulated exclusively in the senescing  $N_L$  #4 and  $N_O$  #8 leaves, among them the *NAP/NAC029* homolog, and thus are leaf rank-independent senescence-associated TFs. Twenty-four TF genes were downregulated solely in the  $N_L$  #4 and  $N_O$  #8 leaves, among them the *WRKY53* homolog, suggesting that they are controlling pathways that are downregulated during senescence. The WRKY, Whirly and NAC families deserve special attention, because members of these families were reported to play key roles in controlling leaf senescence and plastid stability in *Arabidopsis*. In contrast to most other TF families, the NAC genes were predominantly upregulated in senescing leaves #4 and #8 (Figure 6 and Supplementary Table 11). With the only exception of *JUB1*, in senescing *Arabidopsis* leaves the corresponding genes are also upregulated (Breeze et al., 2011). Eleven of the 19 regulated OSR NAC genes were specifically regulated in  $N_L$  leaves #4 or  $N_O$  leaves #8, which indicates differences in the regulation of downstream processes in the two canopy levels. Other than the NAC factors, most of the regulated WRKY genes were downregulated and it appears that in this TF family more transcriptional reprogramming occurred in  $N_O$  leaves #8 than in  $N_L$  leaves #4 (Figure 6).

## N-Deficiency Associated Expression of Protein Degradation Genes

The plant-specific developmental process of leaf senescence safeguards the coordinate degradation of proteins, lipids and nucleic acids and remobilization of the resulting low molecular weight nutrients from the senescing leaves to sink organs. Chloroplasts are the most important resource for nitrogen remobilized from senescing source leaves, and autophagy is a crucial process for degradation of chloroplasts during senescence and in response to starvation (Ishida et al., 2014; Michaeli et al., 2014; Izumi et al., 2017). Autophagy mutant plants suffer from premature senescence accompanied by accelerated cell death (reviewed in Minina et al., 2014). Senescence is also accompanied by the activation of various peptidases. To identify the senescence-associated OSR homologs of *Arabidopsis* autophagy genes in leaves #4 and #8 we matched them against the autophagy database (Homma et al., 2011).

We identified 28 OSR homologs of *Arabidopsis* autophagy(-related) genes that showed  $\geq 3$ -fold changes in transcription levels (Figure 7A and Supplementary Table 12). In N-deficient leaves #4, 19 autophagy gene homologs were upregulated, among them ten ATG core genes that are essential for autophagosome formation (reviewed in Michaeli et al., 2016; Have et al., 2017). Seven of these autophagy core genes and seven autophagy-related genes are also upregulated in senescing *Arabidopsis* leaves (van der Graaff et al., 2006; Breeze et al., 2011). Remarkably, although also in  $N_O$  leaves #4 senescence initiated during week 3 of the observation period, as was indicated by marker gene expression (Figure 2) and enrichment of the GO term 'leaf senescence' (Figure 4), except for *ATG4a* (see below) none of the autophagy core genes were regulated yet in week 3. Also in N-deficient leaves #8, where senescence was delayed, no activation of the autophagy genes was observed. In  $N_O$  leaves #8, *ATG7*, *ATG8a* and six autophagy-related genes



were upregulated, however, the intensification of the autophagy pathway was clearly lower than in N-deficient leaves #4. Only two genes were regulated in both leaves #4 and #8 and independent of the N supply. *ATG4a*, which encodes a cysteine protease involved in the ATG8 ubiquitination-like pathway and is linked to autophagosome formation, was downregulated in all four samples. In *Arabidopsis*, *ATG4a* is transcriptionally induced by sudden N-depletion and carbon-starvation (Yoshimoto et al., 2004; Rose et al., 2006), but it has not been reported in the context of senescence yet. *PLDP2* was upregulated in all four samples. This gene is also in *Arabidopsis* upregulated during senescence

and was reported to regulate vesicle trafficking and to play a role in Pi-starvation (Breeze et al., 2011). The downregulation of the salicylic-acid responsive *PR1* gene (Ward et al., 1991) in all samples except N<sub>0</sub> leaves #4 is consistent with the corresponding downregulation of the GO term 'response to salicylic acid' in these samples.

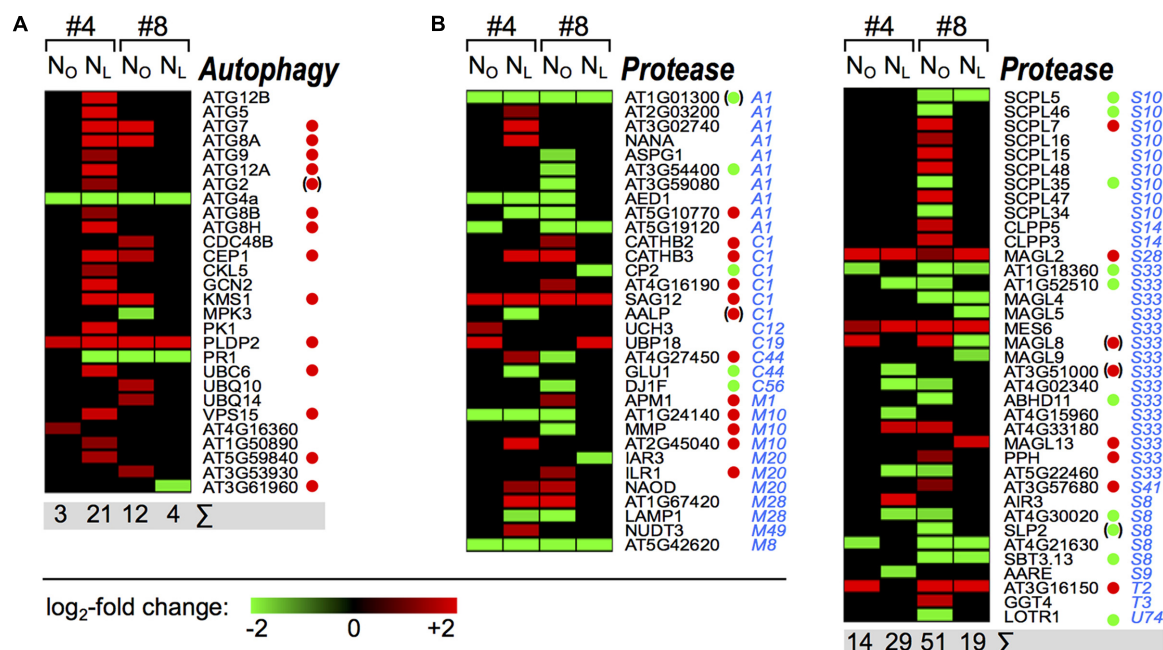
In addition to autophagy-related proteins, transcriptional and proteomic studies identified in various plant species a large number of senescence-associated, mostly upregulated peptidases from diverse families (reviewed in Roberts et al., 2012). Yet, in winter OSR, only few senescence-associated proteases and protease inhibitors were reported (Etienne et al., 2007; Desclos et al., 2009). We identified in spring OSR 'Mozart' overall 69 up- or downregulated OSR homologs of all *A. thaliana* peptidases listed in the MEROPS peptidase database (Rawlings et al., 2016) (Figure 7B). For 35 of these, Breeze et al. (2011) observed leaf senescence-associated regulation of the corresponding *Arabidopsis* genes; 30 of them were regulated in the same direction. Other than with the autophagy-related proteins, regulation of protease genes was heterogeneous and more genes were down- than upregulated (37 vs. 32). Remarkably, in contrast to all other gene classes, the highest number of regulated protease genes occurred in N<sub>0</sub> leaves #8. It is tempting to speculate that in spite of the onset of senescence under ample N supply in leaves #8 (Figure 4), dismantling of chloroplasts and degradation of chlorophyll is still pending (Figure 2) and therefore autophagy is not massively upregulated yet (Figure 7A). However, at that stage leaves #8 likely act already as source leaves and provide nutrients for pod development and seed filling (Supplementary Figure 1, Franzaring et al., 2011). The prominent regulation of many proteases in these leaves may be associated with an elevated nutrient export activity.

## DISCUSSION

### Transcriptome Reprogramming in N Supply-Dependent Senescence of Lower and Upper Node *B. napus* Leaves

Since the divergence of the ancestral Brassicaceae into the *Arabidopsis* and *Brassica* lineages ~17 million years ago (Cheung et al., 2009), genome triplication, allopolyploidization of the *B. napus* parental *B. rapa* and *B. oleracea* genomes, and gene loss events occurred, with the consequence that the modern OSR genome contains zero to more than six orthologs of any *Arabidopsis* gene (Rana et al., 2004). This complicates the identification of orthology relationships between *Arabidopsis* and *B. napus* genes and prevents the distinction between multiple related *B. napus* genes when using *Arabidopsis* microarrays. We therefore used a microarray with 60 nt-probes based predominantly on three EST libraries from *B. napus*, *B. rapa*, and *B. oleracea* and a smaller number of other publically available ESTs (Trick et al., 2009).

In previous studies of the transcriptome response to N starvation in *Arabidopsis thaliana* (Wang et al., 2003; Scheible et al., 2004; Balazadeh et al., 2014) and winter oilseed



**FIGURE 7 |** Senescence associated oilseed rape autophagy and protease genes. The heat maps display the upregulated (red bars) or downregulated (green bars) *Brassica napus* unigenes homologs of *Arabidopsis thaliana* (A) autophagy genes and (B) peptidase genes in leaf #4 between 92 and 99 DAS and in leaf #8 between 92 and 106 DAS as indicated on top of the columns. Peptidase family affiliations (MEROPS Peptidase Database) are indicated in blue letters. Depicted are genes with expression ratios  $\geq 3$  and  $P_{adj} < 0.05$  ( $n = 3$ ). Green and red dots denote genes that were reported as leaf senescence-associated down- and upregulated in *Arabidopsis thaliana* by Breeze et al. (2011). Dots in parentheses indicate that this gene was not steadily regulated in the course of leaf senescence.

rape (Koeslin-Findeklee et al., 2015b), plants were grown hydroponically in low N medium and transcription analysis was performed after nitrate re-addition. This treatment may invoke a rapid and temporal plant response to nutrient shock (Wang et al., 2001; Peng et al., 2007) and thus may not fully reflect the plant adaptive responses to long-term low N conditions. Here, we compared the OSR transcriptome in plants of a spring cultivar grown under optimal or low N fertilization in solid medium under seasonal climate simulating conditions, which is more similar to field conditions (Franzaring et al., 2012).

In the developmentally early winter OSR leaves, senescence typically begins during flowering but before the seed filling stage, and the leaves are shed before the developing reproductive organs have reached their maximal sink strength. This is considered as one reason for the relatively inefficient N remobilization and high residual N content in the fallen leaves that, moreover, increases with N fertilization (Schjoerring et al., 1995; Hocking et al., 1997; Rossato et al., 2001; Noquet et al., 2004; Malagoli et al., 2005a). In cauline leaves in the upper canopy, senescence initiates later and the N content of fallen leaves is lower, indicating a more efficient N remobilization from these leaves driven by the higher sink strength of the developing pods during seed filling (Malagoli et al., 2005a; Etienne et al., 2007). In this study we aimed to determine if the chronology of senescence initiation in different canopy levels is similar in a spring OSR cultivar and how differences between early and late leaves are reflected in their transcriptomes.

Taking the expression changes of the senescence marker genes *BnSAG12-1*, *BnSAG2* and *BnCAB1* as indicators, under optimal as well as low N supply the first signs of senescence initiation appeared in the lower canopy leaf #4 already in week 1 of the observation period. In this early senescence phase the chlorophyll content is not a useful indicator for senescence or N deprivation (Figure 2), as had also been observed by Gombert et al. (2006). In the following 2 weeks senescence progressed under both N regimes, but more rapidly in the N deficient plants as is indicated by the massive downregulation of PSGs. In the upper canopy leaf #8 the *BnSAG12-1*, *BnSAG2* and *BnD22* expression changes did not show a difference in the senescence status of  $N_0$  and  $N_L$  leaves. However, downregulation of *BnCAB* and many PSGs indicated that N deficiency led to a delay of senescence progression in younger leaves. This conclusion was corroborated by the extent of transcriptome reprogramming and the affected metabolic processes. In leaves #4 essentially no change in gene regulation in either  $N_0$  or  $N_L$  leaves was observed in week 1. One week later at an overall low level already twice as many genes were regulated in  $N_L$  compared to  $N_0$  leaves, and in week 3 in N-deficient leaves the number of regulated genes increased another 15-times to almost 3,000 regulated genes, whereas under ample N supply less than 600 genes were regulated. The effect of N deprivation was opposite in the upper canopy leaves #8, where 2.5-times more genes were regulated in  $N_0$  plants. The functional classification of regulated genes revealed that senescence-associated transcriptome reprogramming in spring oilseed rape cv. 'Mozart' comprises largely the same biological



processes as in *Arabidopsis thaliana* (Buchanan-Wollaston et al., 2005; van der Graaff et al., 2006; Breeze et al., 2011).

## Divergent Regulation of Transcription Factors in Senescing Young and Old Leaves

The age-dependent expression of thousands of senescence-associated genes is orchestrated by transcription factors, many of which are themselves transcriptionally regulated during senescence. Several of these TFs are also induced by various biotic or abiotic stresses, indicating that senescence is an integrated response of plants to endogenous developmental signals and environmental cues (Woo et al., 2013). In *Arabidopsis*, transcriptomic analyses revealed the enrichment of upregulated TF genes of the NAC, WRKY, AP2/EREBP, MYB, C2H2 zinc-finger, bZIP, and GRAS families during leaf senescence (Buchanan-Wollaston et al., 2005; van der Graaff et al., 2006; Balazadeh et al., 2008, 2010; Breeze et al., 2011). We observed also in spring OSR a senescence-associated transcriptional reorganization in all these TF families. The comparison of the expression changes of the 112 senescence-associated TF genes that were identified in *Arabidopsis* by Breeze et al. (2011) and here in OSR reveals largely congruent transcription increases or decreases (Supplementary Table 11). Interestingly though, we note in OSR that 107 TF genes were only in N<sub>L</sub> leaf #4 and 65 genes only in N<sub>O</sub> leaves #8 more than threefold up- or downregulated, which indicates distinct regulation of individual senescence processes and nutrient remobilization in upper and lower canopy leaves. Also noticeable is a virtually perfect congruence of TF gene regulation in certain families between OSR and *Arabidopsis* (NAC, C2C2, C2H2) and more divergent regulation in others (AP2-EREBP, C3H, homeobox).

The expression profiles of well characterized key regulators of senescence in OSR and *Arabidopsis* attest that, in spite of their very different sporophyte architectures, the regulatory network controlling senescence is similar in these two Brassicaceae. For example, in both species the positive regulators of chlorophyll degradation *NAC046* and *NAC055* and the senescence promoting *NAP/NAC029* factor are upregulated (Figure 6) (Guo and Gan, 2006; Hickman et al., 2013; Oda-Yamamizo et al., 2016). On the other hand, the negative regulator of senescence *WRKY70* (Ülker et al., 2007; Zentgraf et al., 2010; Besseau et al., 2012) and the early induced *WRKY53* factor, which interacts with other senescence regulators (Hinderhofer and Zentgraf, 2001; Miao et al., 2004; Zentgraf et al., 2010), were downregulated. However, in a few cases also divergent regulation of senescence-controlling factors in *Arabidopsis* and OSR was observed. The NAC family member *JUB1*, which was identified as a longevity-promoting factor in *Arabidopsis* (Wu et al., 2012), was downregulated in N<sub>L</sub> leaves #4. Surprisingly, it was found to be induced in *Arabidopsis* during leaf senescence (Breeze et al., 2011). Noteworthy is also the downregulation of the OSR homolog of *WHY1* in N<sub>L</sub> leaves #4 (Figure 6), because this gene is involved in maintaining chloroplast stability. The *Arabidopsis* *WHY1* gene, which is one of only three Whirly family genes in this plant, is required for chloroplast genome stability (Marechal et al., 2009), and the

barley *WHIRLY1* ortholog is involved in premature senescence induction under photooxidative stress (Kucharewicz et al., 2017).

## Chloroplast Decomposition and Protein Degradation Pathway Activation in Senescing Leaves

The critical role of autophagy for the disassembly of chloroplasts, mitochondria and other cellular structures in the course of senescence has been extensively demonstrated in *Arabidopsis* (reviewed by Michaeli et al., 2016; Have et al., 2017). During developmental and starvation-induced senescence, entire chloroplasts can be degraded by autophagy (Minamikawa et al., 2001; Wada et al., 2009). Other than in *A. thaliana*, where 9 of 15 upregulated autophagy genes were activated in leaves that were not even fully expanded yet and showed no signs of senescence (Breeze et al., 2011), in OSR we do not observe activation of autophagy genes in N<sub>O</sub> leaves #4 or N<sub>L</sub> leaves #8, while senescence was initiated in these leaves. However, in the more advanced senescence stages in N<sub>L</sub> #4 and N<sub>O</sub> #8 leaves, more OSR autophagy genes appear to be upregulated than in *Arabidopsis* (Breeze et al., 2011), although we did not consider all statistically significantly regulated genes but only those exhibiting  $\geq 3$ -fold transcriptional changes. A possible explanation could be that autophagy is a generic, auto-cleaning process required to remove obsolete cell components and maintain cellular integrity. It is thus constitutively active at a low level which might be sufficient during early senescence. Only when senescence progresses it may become necessary to boost the autophagy pathway.

The more pronounced transcriptional activation of autophagy genes in N-deficient leaves #4 compared to N<sub>O</sub> leaves #8 could indicate that the cell death program, the last phase of senescence, has started in the lower canopy leaves, whereas the young upper canopy leaves #8 have to stay alive to serve as source leaves for nutrient remobilization toward the developing pods. This course of events is known from winter OSR genotypes (reviewed in Avice and Etienne, 2014). Consistent with this hypothesis is the much higher number of regulated protease genes in N<sub>O</sub> leaves #8 which may be involved in protein turnover and N remobilization, but not in executing cell death.

Differences between OSR and *Arabidopsis* are also apparent in the regulation of senescence-associated peptidase genes, which play a crucial role in providing nitrogen transport molecules like amino acids for developing sink organs (reviewed by Masclaux-Daubresse et al., 2010). Similar to the group of autophagy genes, in OSR more peptidase genes are differentially regulated than in *Arabidopsis* (Breeze et al., 2011), and a larger fraction of these genes is downregulated. These differences might indicate a partly different orchestration of the senescence course in OSR, which may reflect the more complex architecture and morphological development of OSR plants compared to *A. thaliana*. Recently, by protease activity profiling Poret et al. (2016) identified in senescing *B. napus* leaves after 23 days of N-starvation an activity increase relative to plants grown with ample N-supply of 17 serine- and cysteine-proteases with homology to 10 *Arabidopsis* proteases including *SAG12*, *AALP*, and *AARE*. In our study, both *AALP* and *AARE* are downregulated only in N-deficient



leaves #4 (Figure 7B). However, transcription data do not always reflect protein level or activity data, as has also been reported for metabolic flux data (Schwender et al., 2014), and especially proteases are frequently regulated at the post-transcriptional level.

## CONCLUSION

We found evidence that the sequence of senescence initiation and progression and also the effects of N-limitation are similar in the spring OSR cultivar 'Mozart' and in winter OSR cultivars. Like in winter OSR, long-term, mild N deficiency leads in spring OSR to premature shutdown of PSGs and senescence in lower canopy source leaves, whereas in upper canopy sink leaves senescence progression is delayed. The onset of senescence is accompanied by a massive reprogramming of the transcriptome. The affected regulatory and metabolic pathways are overall similar to those in *Arabidopsis*, but we identified transcription regulator and protein degradation genes that are specifically regulated in N-depleted lower canopy leaves or in upper leaves under ample N supply, and genes that are senescence-associated expressed in oilseed rape, but not in *Arabidopsis*. In future studies it will be interesting to address the question whether these genes fulfill specific tasks in N-remobilization during N deficiency-induced leaf senescence and if their regulation affects the nitrogen use efficiency of oilseed rape.

## REFERENCES

- Arvidsson, S., Kwasniewski, M., Riano-Pachon, D. M., and Mueller-Roeber, B. (2008). QuantPrime—a flexible tool for reliable high-throughput primer design for quantitative PCR. *BMC Bioinformatics* 9:465. doi: 10.1186/1471-2105-9-465
- Avice, J. C., and Etienne, P. (2014). Leaf senescence and nitrogen remobilization efficiency in oilseed rape (*Brassica napus* L.). *J. Exp. Bot.* 65, 3813–3824. doi: 10.1093/jxb/eru177
- Balazadeh, S., Riano-Pachon, D. M., and Mueller-Roeber, B. (2008). Transcription factors regulating leaf senescence in *Arabidopsis thaliana*. *Plant Biol. (Stuttg)* 10(Suppl. 1), 63–75. doi: 10.1111/j.1438-8677.2008.00088.x
- Balazadeh, S., Schildhauer, J., Araujo, W. L., Munne-Bosch, S., Fernie, A. R., Proost, S., et al. (2014). Reversal of senescence by N resupply to N-starved *Arabidopsis thaliana*: transcriptomic and metabolomic consequences. *J. Exp. Bot.* 65, 3975–3992. doi: 10.1093/jxb/eru119
- Balazadeh, S., Siddiqui, H., Allu, A. D., Matallana-Ramirez, L. P., Caldana, C., Mehrnia, M., et al. (2010). A gene regulatory network controlled by the NAC transcription factor ANAC092/AtNAC2/ORE1 during salt-promoted senescence. *Plant J.* 62, 250–264. doi: 10.1111/j.1365-313X.2010.04151.x
- Besseau, S., Li, J., and Palva, E. T. (2012). WRKY54 and WRKY70 co-operate as negative regulators of leaf senescence in *Arabidopsis thaliana*. *J. Exp. Bot.* 63, 2667–2679. doi: 10.1093/jxb/err450
- Bieker, S., and Zentgraf, U. (2013). "Plant senescence and nitrogen mobilization and signaling," in *Senescence and Senescence-Related Disorders*, eds Z. Wang and I. Hiroyuki (Rijeka: InTech), 53–83. doi: 10.5772/54392
- Breeze, E., Harrison, E., McHattie, S., Hughes, L., Hickman, R., Hill, C., et al. (2011). High-resolution temporal profiling of transcripts during *Arabidopsis* leaf senescence reveals a distinct chronology of processes and regulation. *Plant Cell* 23, 873–894. doi: 10.1105/tpc.111.083345
- Buchanan-Wollaston, V., Page, T., Harrison, E., Breeze, E., Lim, P. O., Nam, H. G., et al. (2005). Comparative transcriptome analysis reveals significant differences in gene expression and signalling pathways between developmental

## AUTHOR CONTRIBUTIONS

VS-R: acquisition, analysis, and interpretation of data; writing the manuscript. JF: acquisition of data and design of the work. AF: design of the work. RK: conception and design of the work; acquisition, analysis, and interpretation of data; writing the manuscript.

## FUNDING

This research was supported by the Deutsche Forschungsgemeinschaft (Forschergruppe FOR 948 grant no. KU715/10–2 to RK).

## ACKNOWLEDGMENTS

The authors thank Samuel Arvidsson for help with QuantPrime, Stefan Bieker, and Ulrike Zentgraf for help with plant harvesting, and Christine Rausch for instructing VS-R during the early phase of the project.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00048/full#supplementary-material>

- and dark/starvation-induced senescence in *Arabidopsis*. *Plant J.* 42, 567–585. doi: 10.1111/j.1365-313X.2005.02399.x
- Bussell, J. D., Keech, O., Fenske, R., and Smith, S. M. (2013). Requirement for the plastidial oxidative pentose phosphate pathway for nitrate assimilation in *Arabidopsis*. *Plant J.* 75, 578–591. doi: 10.1111/tjp.12222
- Chen, X., Truksa, M., Shah, S., and Weselake, R. J. (2010). A survey of quantitative real-time polymerase chain reaction internal reference genes for expression studies in *Brassica napus*. *Anal. Biochem.* 405, 138–140. doi: 10.1016/j.ab.2010.05.032
- Cheung, F., Trick, M., Drou, N., Lim, Y. P., Park, J. Y., Kwon, S. J., et al. (2009). Comparative analysis between homoeologous genome segments of *Brassica napus* and its progenitor species reveals extensive sequence-level divergence. *Plant Cell* 21, 1912–1928. doi: 10.1105/tpc.108.060376
- Desclos, M., Dubousslet, L., Etienne, P., Le Caherec, F., Satoh, H., Bonnefoy, J., et al. (2008). A proteomic profiling approach to reveal a novel role of *Brassica napus* drought 22 kD/water-soluble chlorophyll-binding protein in young leaves during nitrogen remobilization induced by stressful conditions. *Plant Physiol.* 147, 1830–1844. doi: 10.1104/pp.108.116905
- Desclos, M., Etienne, P., Coquet, L., Jouenne, T., Bonnefoy, J., Segura, R., et al. (2009). A combined N-15 tracing/proteomics study in *Brassica napus* reveals the chronology of proteomics events associated with N remobilisation during leaf senescence induced by nitrate limitation or starvation. *Proteomics* 9, 3580–3608. doi: 10.1002/pmic.200800984
- Drechsler, N., Zheng, Y., Bohner, A., Nobmann, B., von Wirén, N., Kunze, R., et al. (2015). Nitrate-dependent control of shoot K homeostasis by the nitrate transporter1/peptide transporter family member NPF7.3/NRT1.5 and the Stelar K<sup>+</sup> Outward Rectifier SKOR in *Arabidopsis*. *Plant Physiol.* 169, 2832–2847. doi: 10.1104/pp.15.01152
- Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* 7:191. doi: 10.1186/1471-2105-7-191
- Etienne, P., Desclos, M., Le Goua, L., Gombert, J., Bonnefoy, J., Maurel, K., et al. (2007). N-protein mobilisation associated with the leaf senescence process in

- oilseed rape is concomitant with the disappearance of trypsin inhibitor activity. *Funct. Plant Biol.* 34, 895–906. doi: 10.1071/Fp07088
- Franzaring, J., Gensheimer, G., Weller, S., Schmid, I., and Fangmeier, A. (2012). Allocation and remobilisation of nitrogen in spring oilseed rape (*Brassica napus* L. cv. Mozart) as affected by N supply and elevated CO<sub>2</sub>. *Environ. Exp. Bot.* 83, 12–22. doi: 10.1016/j.envexpbot.2012.03.015
- Franzaring, J., Weller, S., Schmid, I., and Fangmeier, A. (2011). Growth, senescence and water use efficiency of spring oilseed rape (*Brassica napus* L. cv. Mozart) grown in a factorial combination of nitrogen supply and elevated CO<sub>2</sub>. *Environ. Exp. Bot.* 72, 284–296. doi: 10.1016/j.envexpbot.2011.04.003
- Gironde, A., Poret, M., Etienne, P., Trouverie, J., Bouchereau, A., Le Caherec, F., et al. (2015). A profiling approach of the natural variability of foliar N remobilization at the rosette stage gives clues to understand the limiting processes involved in the low N use efficiency of winter oilseed rape. *J. Exp. Bot.* 66, 2461–2473. doi: 10.1093/jxb/erv031
- Gombert, J., Etienne, P., Ourry, A., and Le Dily, F. (2006). The expression patterns of SAG12/Cab genes reveal the spatial and temporal progression of leaf senescence in *Brassica napus* L. with sensitivity to the environment. *J. Exp. Bot.* 57, 1949–1956. doi: 10.1093/jxb/erj142
- Gombert, J., Le Dily, F., Lothier, J., Etienne, P., Rossato, L., Allirand, J. M., et al. (2010). Effect of nitrogen fertilization on nitrogen dynamics in oilseed rape using N-15-labeling field experiment. *J. Plant Nutr. Soil Sci.* 173, 875–884. doi: 10.1002/jpln.200800270
- Gregersen, P. L., Culetic, A., Boschian, L., and Krupinska, K. (2013). Plant senescence and crop productivity. *Plant Mol. Biol.* 82, 603–622. doi: 10.1007/s11103-013-0013-8
- Gregory, F. G. (1937). Mineral nutrition of plants. *Annu. Rev. Biochem.* 6, 557–578. doi: 10.1146/annurev.bi.06.070137.003013
- Guo, Y., and Gan, S. (2006). AtNAP, a NAC family transcription factor, has an important role in leaf senescence. *Plant J.* 46, 601–612. doi: 10.1111/j.1365-3113X.2006.02723.x
- Hanson, A. D., and Roje, S. (2001). One-carbon metabolism in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 52, 119–137. doi: 10.1146/annurev.arplant.52.1.119
- Have, M., Marmagne, A., Chardon, F., and Masclaux-Daubresse, C. (2017). Nitrogen remobilization during leaf senescence: lessons from Arabidopsis to crops. *J. Exp. Bot.* 68, 2513–2529. doi: 10.1093/jxb/erw365
- Hickman, R., Hill, C., Penfold, C. A., Breeze, E., Bowden, L., Moore, J. D., et al. (2013). A local regulatory network around three NAC transcription factors in stress responses and senescence in Arabidopsis leaves. *Plant J.* 75, 26–39. doi: 10.1111/tpj.12194
- Hinderhofer, K., and Zentgraf, U. (2001). Identification of a transcription factor specifically expressed at the onset of leaf senescence. *Planta* 213, 469–473. doi: 10.1007/s004250000512
- Hocking, P. J., Randall, P. J., and DeMarco, D. (1997). The response of dryland canola to nitrogen fertilizer: partitioning and mobilization of dry matter and nitrogen, and nitrogen effects on yield components. *Field Crops Res.* 54, 201–220. doi: 10.1016/S0378-4290(97)00049-X
- Homma, K., Suzuki, K., and Sugawara, H. (2011). The autophagy database: an all-inclusive information resource on autophagy that provides nourishment for research. *Nucleic Acids Res.* 39, D986–D990. doi: 10.1093/nar/gkq995
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Ishida, H., Izumi, M., Wada, S., and Makino, A. (2014). Roles of autophagy in chloroplast recycling. *Biochim. Biophys. Acta* 1837, 512–521. doi: 10.1016/j.bbabi.2013.11.009
- Izumi, M., Ishida, H., Nakamura, S., and Hidema, J. (2017). Entire photodamaged chloroplasts are transported to the central vacuole by autophagy. *Plant Cell* 29, 377–394. doi: 10.1105/tpc.16.00637
- Jensen, L. S., Christensen, L., Mueller, T., and Nielsen, N. E. (1997). Turnover of residual N-15-labelled fertilizer N in soil following harvest of oilseed rape (*Brassica napus* L.). *Plant Soil* 190, 193–202. doi: 10.1023/A:1004253611044
- Koeslin-Findeklee, F., Becker, M. A., van der Graaff, E., Roitsch, T., and Horst, W. J. (2015a). Differences between winter oilseed rape (*Brassica napus* L.) cultivars in nitrogen starvation-induced leaf senescence are governed by leaf-inherent rather than root-derived signals. *J. Exp. Bot.* 66, 3669–3681. doi: 10.1093/jxb/erv170
- Koeslin-Findeklee, F., Rizi, V. S., Becker, M. A., Parra-Londono, S., Arif, M., Balazadeh, S., et al. (2015b). Transcriptomic analysis of nitrogen starvation- and cultivar-specific leaf senescence in winter oilseed rape (*Brassica napus* L.). *Plant Sci.* 233, 174–185. doi: 10.1016/j.plantsci.2014.11.018
- Kruger, N. J., and von Schaewen, A. (2003). The oxidative pentose phosphate pathway: structure and organisation. *Curr. Opin. Plant Biol.* 6, 236–246. doi: 10.1016/S1369-5266(03)00039-6
- Kucharewicz, W., Distelfeld, A., Bilger, W., Muller, M., Munne-Bosch, S., Hensel, G., et al. (2017). Acceleration of leaf senescence is slowed down in transgenic barley plants deficient in the DNA/RNA-binding protein WHIRLY1. *J. Exp. Bot.* 68, 983–996. doi: 10.1093/jxb/erw501
- Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090
- Malagoli, P., Laine, P., Rossato, L., and Ourry, A. (2005a). Dynamics of nitrogen uptake and mobilization in field-grown winter oilseed rape (*Brassica napus*) from stem extension to harvest: I. Global N flows between vegetative and reproductive tissues in relation to leaf fall and their residual N. *Ann. Bot.* 95, 853–861. doi: 10.1093/aob/mci091
- Malagoli, P., Laine, P., Rossato, L., and Ourry, A. (2005b). Dynamics of nitrogen uptake and mobilization in field-grown winter oilseed rape (*Brassica napus*) from stem extension to harvest. II. An 15N-labelling-based simulation model of N partitioning between vegetative and reproductive tissues. *Ann. Bot.* 95, 1187–1198. doi: 10.1093/aob/mci131
- Marechal, A., Parent, J. S., Veronneau-Lafortune, F., Joyeux, A., Lang, B. F., and Brisson, N. (2009). Whirly proteins maintain plastid genome stability in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14693–14698. doi: 10.1073/pnas.0901710106
- Masclaux-Daubresse, C., Daniel-Vedele, F., Dechorgnat, J., Chardon, F., Gaufichon, L., and Suzuki, A. (2010). Nitrogen uptake, assimilation and remobilization in plants: challenges for sustainable and productive agriculture. *Ann. Bot.* 105, 1141–1157. doi: 10.1093/aob/mcq028
- Masclaux-Daubresse, C., Purdy, S., Lemaitre, T., Pourtau, N., Taconnat, L., Renou, J. P., et al. (2007). Genetic variation suggests interaction between cold acclimation and metabolic regulation of leaf senescence. *Plant Physiol.* 143, 434–446. doi: 10.1104/pp.106.091355
- Mei, H. S., and Thimann, K. V. (1984). The relation between nitrogen deficiency and leaf senescence. *Physiol. Plant.* 62, 157–161. doi: 10.1111/j.1399-3054.1984.tb00364.x
- Meier, U. (ed.) (2001). *BBCH Monograph - Growth Stages of Mono- and Dicotyledonous Plants*. Bonn: Federal Biological Research Centre for Agriculture and Forestry.
- Miao, Y., Laun, T., Zimmermann, P., and Zentgraf, U. (2004). Targets of the WRKY53 transcription factor and its role during leaf senescence in Arabidopsis. *Plant Mol. Biol.* 55, 853–867. doi: 10.1007/s11103-004-2142-6
- Michaeli, S., Galili, G., Genschik, P., Fernie, A. R., and Avin-Wittenberg, T. (2016). Autophagy in plants—what's new on the menu? *Trends Plant Sci.* 21, 134–144. doi: 10.1016/j.tplants.2015.10.008
- Michaeli, S., Honig, A., Levanony, H., Peled-Zehavi, H., and Galili, G. (2014). Arabidopsis ATG8-INTERACTING PROTEIN1 is involved in autophagy-dependent vesicular trafficking of plastid proteins to the vacuole. *Plant Cell* 26, 4084–4101. doi: 10.1105/tpc.114.129999
- Minamikawa, T., Toyooka, K., Okamoto, T., Hara-Nishimura, I., and Nishimura, M. (2001). Degradation of ribulose-bisphosphate carboxylase by vacuolar enzymes of senescing French bean leaves: immunocytochemical and ultrastructural observations. *Protoplasma* 218, 144–153. doi: 10.1007/BF01306604
- Minina, E. A., Bozhkov, P. V., and Hofius, D. (2014). Autophagy as initiator or executioner of cell death. *Trends Plant Sci.* 19, 692–697. doi: 10.1016/j.tplants.2014.07.007
- Noquet, C., Avice, J. C., Rossato, L., Beauclair, P., Henry, M. P., and Ourry, A. (2004). Effects of altered source-sink relationships on N allocation and vegetative storage protein accumulation in *Brassica napus* L. *Plant Sci.* 166, 1007–1018. doi: 10.1016/j.plantsci.2003.12.014

- Oda-Yamamizo, C., Mitsuda, N., Sakamoto, S., Ogawa, D., Ohme-Takagi, M., and Ohmiya, A. (2016). The NAC transcription factor ANAC046 is a positive regulator of chlorophyll degradation and senescence in Arabidopsis leaves. *Sci. Rep.* 6:23609. doi: 10.1038/srep23609
- Peng, M., Bi, Y. M., Zhu, T., and Rothstein, S. J. (2007). Genome-wide analysis of Arabidopsis responsive transcriptome to nitrogen limitation and its regulation by the ubiquitin ligase gene NLA. *Plant Mol. Biol.* 65, 775–797. doi: 10.1007/s11103-007-9241-0
- Poret, M., Chandrasekar, B., van der Hoorn, R. A., and Avice, J. C. (2016). Characterization of senescence-associated protease activities involved in the efficient protein remobilization during leaf senescence of winter oilseed rape. *Plant Sci.* 246, 139–153. doi: 10.1016/j.plantsci.2016.02.011
- Provart, N., and Zhu, T. (2003). A browser-based functional classification superviewer for Arabidopsis genomics. *Curr. Comput. Mol. Biol.* 2003, 271–272.
- Provart, N. J., Gil, P., Chen, W., Han, B., Chang, H. S., Wang, X., et al. (2003). Gene expression phenotypes of Arabidopsis associated with sensitivity to low temperatures. *Plant Physiol.* 132, 893–906. doi: 10.1104/pp.103.02.1261
- Rana, D., van den Boogaart, T., O'Neill, C. M., Hynes, L., Bent, E., Macpherson, L., et al. (2004). Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives. *Plant J.* 40, 725–733. doi: 10.1111/j.1365-3113X.2004.02244.x
- Rathke, G. W., Behrens, T., and Diepenbrock, W. (2006). Integrated nitrogen management strategies to improve seed yield, oil content and nitrogen efficiency of winter oilseed rape (*Brassica napus* L.): a review. *Agric. Ecosyst. Environ.* 117, 80–108. doi: 10.1016/j.agee.2006.04.006
- Rawlings, N. D., Barrett, A. J., and Finn, R. (2016). Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 44, D343–D350. doi: 10.1093/nar/gkv1118
- Roberts, I. N., Caputo, C., Criado, M. V., and Funk, C. (2012). Senescence-associated proteases in plants. *Physiol. Plant.* 145, 130–139. doi: 10.1111/j.1399-3054.2012.01574.x
- Rose, T. L., Bonneau, L., Der, C., Marty-Mazars, D., and Marty, F. (2006). Starvation-induced expression of autophagy-related genes in Arabidopsis. *Biol. Cell* 98, 53–67. doi: 10.1042/BC20040516
- Rossato, L., Laine, P., and Ourry, A. (2001). Nitrogen storage and remobilization in *Brassica napus* L. during the growth cycle: nitrogen fluxes within the plant and changes in soluble protein patterns. *J. Exp. Bot.* 52, 1655–1663. doi: 10.1093/jxb/52.361.1655
- Saeed, A. I., Bhagabati, N. K., Braisted, J. C., Liang, W., Sharov, V., Howe, E. A., et al. (2006). TM4 microarray software suite. *Methods Enzymol.* 411, 134–193. doi: 10.1016/S0076-6879(06)11009-5
- Scheible, W. R., Morcuende, R., Czechowski, T., Fritz, C., Osuna, D., Palacios-Rojas, N., et al. (2004). Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen. *Plant Physiol.* 136, 2483–2499. doi: 10.1104/pp.104.047019
- Schjoerring, J. K., Bock, J. G. H., Gammelvind, L., Jensen, C. R., and Mogensen, V. O. (1995). Nitrogen incorporation and remobilization in different shoot components of field-grown winter oilseed rape (*Brassica napus* L.) as affected by rate of nitrogen application and irrigation. *Plant Soil* 177, 255–264. doi: 10.1007/BF00010132
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.b2008.73
- Schulte auf'm Erley, G., Wijaya, K.-A., Ulas, A., Becker, H., Wiesler, F., and Horst, W. J. (2007). Leaf senescence and N uptake parameters as selection traits for nitrogen efficiency of oilseed rape cultivars. *Physiol. Plant.* 130, 519–531. doi: 10.1111/j.1399-3054.2007.00921.x
- Schwender, J., König, C., Klapperstuck, M., Heinzl, N., Munz, E., Hebbelmann, I., et al. (2014). Transcript abundance on its own cannot be used to infer fluxes in central metabolism. *Front. Plant. Sci.* 5:668. doi: 10.3389/fpls.2014.00668
- Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3, 1–25. doi: 10.2202/1544-6115.1027
- Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., et al. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 37, 914–939. doi: 10.1111/j.1365-3113X.2004.02016.x
- Thomas, H., and Ougham, H. (2014). The stay-green trait. *J. Exp. Bot.* 65, 3889–3900. doi: 10.1093/jxb/eru037
- Tilsner, J., Kassner, N., Struck, C., and Lohaus, G. (2005). Amino acid contents and transport in oilseed rape (*Brassica napus* L.) under different nitrogen conditions. *Planta* 221, 328–338. doi: 10.1007/s00425-004-1446-8
- Trick, M., Cheung, F., Drou, N., Fraser, F., Lobenhofer, E. K., Hurban, P., et al. (2009). A newly-developed community microarray resource for transcriptome profiling in Brassica species enables the confirmation of Brassica-specific expressed sequences. *BMC Plant Biol.* 9:50. doi: 10.1186/1471-2229-9-50
- Ülker, B., Shahid Mukhtar, M., and Somssich, I. E. (2007). The WRKY70 transcription factor of Arabidopsis influences both the plant senescence and defense signaling pathways. *Planta* 226, 125–137. doi: 10.1007/s00425-006-0474-y
- van der Graaff, E., Schwacke, R., Schneider, A., Desimone, M., Flügge, U. I., and Kunze, R. (2006). Transcription analysis of Arabidopsis membrane transporters and hormone pathways during developmental and induced leaf senescence. *Plant Physiol.* 141, 776–792. doi: 10.1104/pp.106.079293
- Wada, S., Ishida, H., Izumi, M., Yoshimoto, K., Ohsumi, Y., Mae, T., et al. (2009). Autophagy plays a role in chloroplast degradation during senescence in individually darkened leaves. *Plant Physiol.* 149, 885–893. doi: 10.1104/pp.108.130013
- Wang, R. C., Okamoto, M., Xing, X. J., and Crawford, N. M. (2003). Microarray analysis of the nitrate response in Arabidopsis roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. *Plant Physiol.* 132, 556–567. doi: 10.1104/pp.103.021253
- Wang, Y. H., Garvin, D. F., and Kochian, L. V. (2001). Nitrate-induced genes in tomato roots. Array analysis reveals novel genes that may play a role in nitrogen nutrition. *Plant Physiol.* 127, 345–359. doi: 10.1104/pp.127.1.345
- Ward, E. R., Uknes, S. J., Williams, S. C., Dincher, S. S., Wiederhold, D. L., Alexander, D. C., et al. (1991). Coordinate gene activity in response to agents that induce systemic acquired resistance. *Plant Cell* 3, 1085–1094. doi: 10.1105/tpc.3.10.1085
- Weltmeier, F., Maser, A., Menze, A., Hennig, S., Schad, M., Breuer, F., et al. (2011). Transcript profiles in sugar beet genotypes uncover timing and strength of defense reactions to *Cercospora beticola* infection. *Mol. Plant Microbe Interact.* 24, 758–772. doi: 10.1094/mpmi-08-10-0189
- Woo, H. R., Kim, H. J., Nam, H. G., and Lim, P. O. (2013). Plant leaf senescence and death - regulation by multiple layers of control and implications for aging in general. *J. Cell Sci.* 126(Pt 21), 4823–4833. doi: 10.1242/jcs.109116
- Wu, A., Allu, A. D., Garapati, P., Siddiqui, H., Dortay, H., Zanoor, M. I., et al. (2012). JUNGBRUNNEN1, a reactive oxygen species-responsive NAC transcription factor, regulates longevity in Arabidopsis. *Plant Cell* 24, 482–506. doi: 10.1105/tpc.111.090894
- Xu, G., Fan, X., and Miller, A. J. (2012). Plant nitrogen assimilation and use efficiency. *Annu. Rev. Plant Biol.* 63, 153–182. doi: 10.1146/annurev-arplant-042811-105532
- Yilmaz, A., Mejia-Guerra, M. K., Kurz, K., Liang, X., Welch, L., and Grotewold, E. (2011). AGRIS: the Arabidopsis gene regulatory information server, an update. *Nucleic Acids Res.* 39, D1118–D1122. doi: 10.1093/nar/gkq1120
- Yoshimoto, K., Hanaoka, H., Sato, S., Kato, T., Tabata, S., Noda, T., et al. (2004). Processing of ATG8s, ubiquitin-like proteins, and their deconjugation by ATG4s are essential for plant autophagy. *Plant Cell* 16, 2967–2983. doi: 10.1105/tpc.104.025395
- Zentgraf, U., Laun, T., and Miao, Y. (2010). The complex regulation of WRKY53 during leaf senescence of *Arabidopsis thaliana*. *Eur. J. Cell Biol.* 89, 133–137. doi: 10.1016/j.ejcb.2009.10.014

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Safavi-Rizi, Franzaring, Fangmeier and Kunze. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# *Metarhizium brunneum* (Ascomycota; Hypocreales) Treatments Targeting Olive Fly in the Soil for Sustainable Crop Production

Meelad Yousef<sup>1</sup>, Carmen Alba-Ramírez<sup>1</sup>, Inmaculada Garrido Jurado<sup>1</sup>, Jordi Mateu<sup>2</sup>,  
Silvia Raya Díaz<sup>1</sup>, Pablo Valverde-García<sup>1</sup> and Enrique Quesada-Moraga<sup>1\*</sup>

<sup>1</sup> Department of Agricultural and Forestry Sciences, ETSIAM, University of Cordoba, Cordoba, Spain, <sup>2</sup> Department of Agriculture, Livestock and Fisheries, Government of Catalonia, Catalonia, Spain

## OPEN ACCESS

### Edited by:

Leire Molinero-Ruiz,  
Instituto de Agricultura Sostenible  
(CSIC), Spain

### Reviewed by:

Fernando E. Vega,  
Agricultural Research Service (USDA),  
United States  
Paula Baptista,  
Polytechnic Institute of Bragança,  
Portugal

### \*Correspondence:

Enrique Quesada-Moraga  
equesada@uco.es

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 29 November 2017

**Accepted:** 01 January 2018

**Published:** 23 January 2018

### Citation:

Yousef M, Alba-Ramírez C,  
Garrido Jurado I, Mateu J, Raya  
Díaz S, Valverde-García P and  
Quesada-Moraga E (2018)  
*Metarhizium brunneum* (Ascomycota;  
Hypocreales) Treatments Targeting  
Olive Fly in the Soil for Sustainable  
Crop Production. *Front. Plant Sci.* 9:1.  
doi: 10.3389/fpls.2018.00001

Soil treatments with *Metarhizium brunneum* EAMa 01/58-Su strain conducted in both Northern and Southern Spain reduced the olive fly (*Bactrocera oleae*) population density emerging from the soil during spring up to 70% in treated plots compared with controls. A model to determine the influence of rainfall on the conidial wash into different soil types was developed, with most of the conidia retained at the first 5 cm, regardless of soil type, with relative percentages of conidia recovered ranging between 56 and 95%. Furthermore, the possible effect of UV-B exposure time on the pathogenicity of this strain against *B. oleae* adults coming from surviving preimaginals and carrying conidia from the soil at adult emergence was also evaluated. The UV-B irradiance has no significant effect on *M. brunneum* EAMa 01/58-Su pathogenicity with *B. oleae* adult mortalities of 93, 90, 79, and 77% after 0, 2, 4, and 6 of UV-B irradiance exposure, respectively. In a next step for the use of these *M. brunneum* EAMa 01/58-Su soil treatments within a *B. oleae* IPM strategy, its possible effect of on the *B. oleae* cosmopolitan parasitoid *Psytalia concolor*, its compatibility with the herbicide oxyfluorfen 24% commonly used in olive orchards and the possible presence of the fungus in the olive oil resulting from olives previously placed in contact with the fungus were investigated. Only the highest conidial concentration ( $1 \times 10^8$  conidia ml<sup>-1</sup>) caused significant *P. concolor* adult mortality (22%) with ensuing mycosis in 13% of the cadavers. There were no fungal propagules in olive oil samples resulting from olives previously contaminated by EAMa 01/58-Su conidia. Finally, the strain was demonstrated to be compatible with herbicide since the soil application of the fungus reduced the *B. oleae* population density up to 50% even when it was mixed with the herbicide in the same tank. The fungal inoculum reached basal levels 4 months after treatments ( $1.6 \times 10^3$  conidia g soil<sup>-1</sup>). These results reveal both the efficacy and environmental and food safety of this *B. oleae* control method, protecting olive groves and improving olive oil quality without negative effects on the natural enemy *P. concolor*.

**Keywords:** olive oil production, soil treatment, entomopathogenic fungi, microbial control, *Psytalia concolor*, *Bactrocera oleae*



## INTRODUCTION

There is a need to develop effective, economically viable, and environmentally friendly methods for pest control (Nicolopoulou-Stamati et al., 2016), which has become even more critical for those insect pests that have developed insecticide-resistance such as the olive fruit fly *Bactrocera oleae* Rossi (Diptera: Tephritidae) (Kakani et al., 2010; Hsu et al., 2015). This monophagous and multivoltine species is the most destructive to the olive crop worldwide (Daane and Johnson, 2010), not only reducing crop production, but even more important, olive oil quality (Mraïcha et al., 2010; Medjkouh et al., 2016; Caleca et al., 2017). The importance of this insect pest has been aggravated by irrational repeated aerial spray applications of chemical insecticides targeting *B. oleae* adults for more than 60 years (Haniotakis, 2005). Even if most *B. oleae* control efforts have targeted the adult stage, from mid-autumn onwards, larvae of the last generation of *B. oleae* fall to the ground to pupate ~3 cm below the soil surface beneath the tree canopy, which offers a great opportunity for an effective control of *B. oleae* (Dimou et al., 2003; Ekesi et al., 2007). Faced with this scenario, microbial control of soil dwelling stages of insect pests is among the most promising alternatives to synthetic chemical pesticides (Eilenberg and Hokkanen, 2006). In particular, entomopathogenic fungi have gained importance within the entomopathogenic microorganisms mainly due to their unique contact mode of action through the cuticle (Quesada-Moraga and Santiago-Álvarez, 2008). Besides entomopathogenic fungi are naturally distributed in a wide range of habitats and the soil is considered their natural reservoir (Quesada-Moraga et al., 2007; Pell et al., 2010; Garrido-Jurado et al., 2015); therefore, soil application of entomopathogenic fungi to target soil dwelling stages of insect pests could be a powerful and sustainable pest management strategy (Rogge et al., 2017). Yousef et al. (2017) have demonstrated the efficacy of soil treatments under olive tree canopy using the *M. brunneum* EAMa 01/58-Su strain (hereafter referred to as *M. brunneum*) for *B. oleae* control, as well as the compatibility of *M. brunneum* with commercial herbicides under laboratory conditions (Yousef et al., 2015), while Garrido-Jurado et al. (2011a) have demonstrated the lack of negative direct or indirect impact of such treatments on the olive crop soil-dwelling non-target arthropod population.

The use of entomopathogenic fungi has successfully reduced by 50–70% the adult *B. oleae* spring population, showing to be an effective, economically viable, and environmentally friendly method for *B. oleae* management. However, several aspects of the method related to efficacy, food safety and environmental sustainability still need to be addressed. In this study food safety, quality, and sustainability are targeted through (1) the compatibility of *M. brunneum* strain with the herbicide Oxiflourfen 24% EC that is commonly used in olive orchards in Andalusia (Spain) (2) the efficacy of this strain in different climatic conditions (North and South of Spain); (3) the effect of soil type and rainfall on the movement of conidia into the soil; (4) the effect of the exposure time to UV-B radiation on virulence of this strain against *B. oleae* adults emerging from surviving preimaginals and carrying conidia

from the soil at adult emergence; (5) the effect of the fungus on the cosmopolitan parasitoid *Psytalia concolor* (Szépligeti) (Hymenoptera: Braconidae); and (6) the presence of the fungus in the olive oil.

## MATERIALS AND METHODS

### Fungal Strain, Cultivation, and Inoculum Production

*Metarhizium brunneum* was obtained from the culture collection at the Agricultural and Forestry Sciences and Resources (AFSR) Department of the University of Cordoba, Spain. This strain was originally isolated from soil in a wheat plantation at Hinojosa del Duque, Cordoba, Spain. The strain was deposited in the Spanish collection of culture types (CECT) with accession number CECT 20764. The cultivation and inoculum production for the laboratory and field experiments were done as described by Yousef et al. (2017).

### Insects Used in the Laboratory Bioassays

*Bactrocera oleae* adults used in the laboratory bioassays were obtained from naturally infested fruit collected from September to December in the Cordoba area. The infested olives were maintained as described by Yousef et al. (2013) to obtain the adults. *P. concolor* were originally obtained from a population at the Technical University of Madrid (UPM, Madrid, Spain), and then a stock colony was maintained at the Department of Agricultural and Forestry Sciences of the University of Cordoba in a rearing chamber set at  $25 \pm 2^\circ\text{C}$ , 50–60% RH, and 16:8 h (L:D) following the standard procedure developed by Jacas and Viñuela (1994).

### Fungus-Herbicide Compatibility under Field Conditions

The field experiment was conducted in a commercial olive orchard to evaluate the compatibility of *M. brunneum* with the herbicide Oxiflourfen 24% EC. This experiment was performed in Castro del Río, Cordoba, Spain ( $37^\circ 41' 9.5''\text{N}$ ,  $4^\circ 29' 38.7''\text{W}$ ; 227 masl). The experimental site was divided into six 1-ha square sub-fields ( $\approx 98$  olive trees each; olive variety: Picual). Two of these sub-fields were the *M. brunneum*-treated plots, two were the fungi and oxyflourfen-treated plots and the others were the control plots. Soil application of the fungus was performed once in autumn (October–November) to target prepupating third-instar olive fruit fly larvae that exit from the fruits to the ground to pupate beneath the tree and spend the winter in the pupal stage (Santiago-Álvarez and Quesada-Moraga, 2007). The soil beneath each tree canopy in the olive orchards was sprayed with 1 l of *M. brunneum* suspension [which contained 1 g of conidia or  $1 \times 10^9$  conidia and the herbicide at recommended field concentration ( $21 \text{ l ha}^{-1}$ )].

To evaluate the compatibility of *M. brunneum* with the herbicide Oxiflourfen 24% EC, two types of monitoring were performed after the simultaneous treatment. The first one consisted in monitoring the fungal strain persistence in the soil from both fungus-treated plots and fungus and herbicide treated one. Before the treatment, six completely randomized

soil samples were collected using a soil corer (5 cm diam) to a depth of 15 cm to determine the natural presence of indigenous entomopathogenic fungi in the soil according to Goettel and Inglis (1996). After treatment, soil samples from fungus-treated and fungus and oxyflourfen treated plots were collected monthly beneath the canopy following the same procedure as mentioned above for 6 months. To assess the conidial density in each sample, the number of colony-forming units (CFU) per gram of dry soil was determined using Sabouraud chloramphenicol agar medium in petri dishes (Goettel and Inglis, 1996). Rainfall data was obtained from the climatological stations operated by Junta de Andalucía (Red de Alerta e Información Fitosanitaria-RAIF).

In the second monitoring, the adult population dynamics was compared in both treated and control plots using a combination of pheromone yellow and McPhail traps. A total of eight traps (five yellow traps and three McPhail traps) in each plot (treated and control) were randomly distributed and inspected weekly to count the number of *B. oleae* adults. McPhail traps were baited with diammonium phosphate 4%. To improve the count accuracy, each plot (both treated and control) was surrounded with 120 Olike traps (one per tree) (Caballero, 2002; Altolaguirre-Obrero et al., 2003) to reduce as much as possible the entry of adults from other farms. These traps were baited with protein-based attractant, which is the most effective for the *B. oleae* (Ruiz-Torres, 2010).

## Fungal Effectiveness in Northern Spain

The second experiment assessed the effectiveness of *M. brunneum* for *B. oleae* control under Northern Spain climatic conditions. The experiment was performed in Rodonyà located in Northern Spain (Tarragona, Spain) (41°16'38.438"N, 1°24'7.297"W; 312 masl). The experimental site was divided into four 0.5-ha square sub-fields (ca. 40 olive trees each; olive variety: Menya). Two of these sub-fields were the fungi-treated plots, and the others were the control plots. Soil application of the fungus was performed twice, once in autumn as described above and in late spring (Jun–Jul) to target the emerging adults. In both sites, the integrated pest management was applied as authorized by the Spanish Ministry of Agriculture, Food, and Environment (MAGRAMA, 2014). All treatments were performed following the same protocol described by Yousef et al. (2017).

The *B. oleae* adult monitoring described above was used to evaluate the effectiveness of the *M. brunneum* strain in soil applications in Northern Spain.

## Effect of Soil Type and Rain Volume on the Movement of Conidia in Soil

Four soils and one sandy substrate were used. The physicochemical properties of these soils are shown in Table 1. Further details of the analytical procedures used in this study have been previously described by Cañasveras et al. (2010). For the substrate preparation, quartz sand of aeolian origin was sieved to 0.2–0.5 mm, washed with a large volume of tap water enriched with Na<sub>2</sub>CO<sub>3</sub> (pH 9.5) to disperse clay, washed with deionized water to remove salts, and finally dried in an oven at 40°C. Previously, the field capacity of each soil was estimated by saturation with water. For this purpose, five glass

columns similar to a funnel (Figure 1) (140 mm height and 40 mm diameter) were packed with 100 g of soil, placing a piece of cotton in the bottom to keep the substrate in place. Deionized water was added slowly and at short intervals, ensuring that the water did not reach the lower edge of the funnel. The column was capped with Parafilm® and small holes were made with a needle. At 48 h, a portion of soil was taken from the central part of the funnel, discarding the top. In a weighing bottle of known weight (P1), the portion of soil collected was weighed (P2), placed in an oven at 105°C for 48 h, allowed to cool and reweighed (P3). Field capacity was determined using the following equation:

$$\text{Field capacity (\%)} = \frac{P2 - P3}{P3 - P1} \times 100$$

To determine the influence of the volume of effluent added on the adsorption of conidia, an assembly was performed as shown in Figure 1. The five glass columns were packed with 100 g of sterile soil (autoclaved for 60 min twice at an interval of 24 h) representing model combinations of texture (sandy or clay) and pH (acid or alkaline), a piece of cotton was placed in the bottom to keep the substrate in place. Then, 5 ml of a 10<sup>8</sup> conidia/ml suspension was added to the soil surface. The columns were then run with the test volume (i.e., 60, 85, 100, 140, 250, 300, or 400 ml) of 0.002 M CaCl<sub>2</sub>, added dropwise with a crystal burette. At the bottom, an Erlenmeyer flask was placed to collect percolated effluent and the collected volume was recorded.

To determine the number of CFU, each column was divided into three equal sections or depths (A, B, and C; ca. 5 cm each one). 1 g of soil from each section was dissolved in 10 ml of deionized water with Tween 80 (0.1% v/v) and shaken at 12 rpm in a Model 3000445 Orbit rotary stirrer (J.P. Selecta, Barcelona, Spain) for 90 min. Aliquots of 0.1 ml were then spread with a Drigalsky loop on Sabouraud dextrose agar supplemented with 0.5 g l<sup>-1</sup> of chloramphenicol (SDAC, Biolife, Italy). Aliquots of 0.1 ml of the collected effluent (E) were also spread on the SDAC medium. In some cases, dilutions were necessary before spreading. The plates were cultured at 25°C for 3–4 days and CFU were counted. Fungal growth was visually identified: *M. brunneum* colonies exhibited circular growth, were largely white and contained varying shades of green in the central mycelia (Humber, 1997).

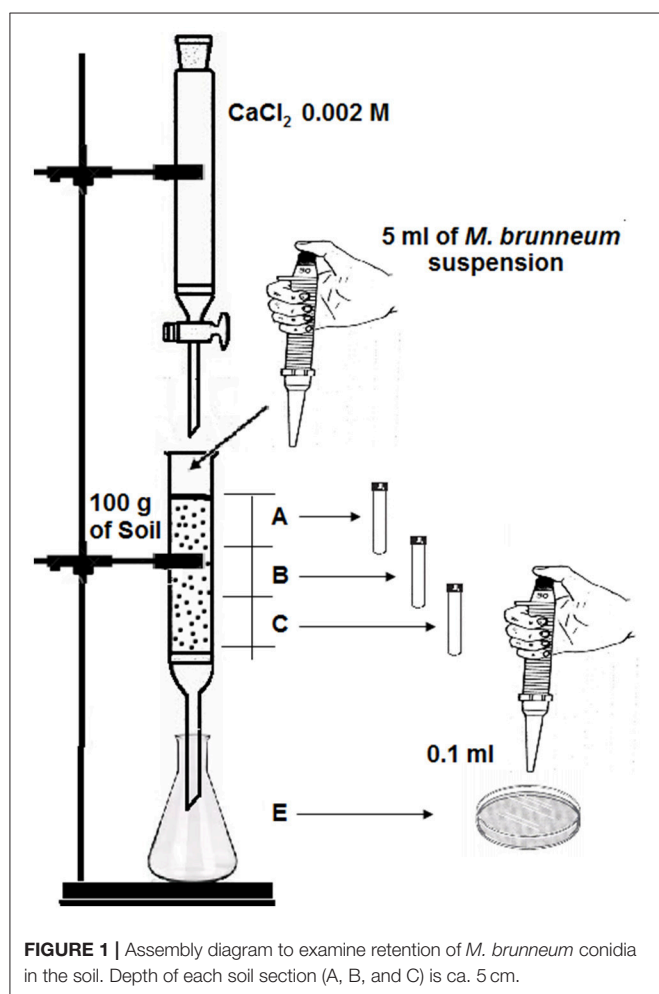
## Effect UV-B Radiation Exposure Time on the Pathogenicity of *M. brunneum* against *B. oleae* Adults

The irradiation experiment was conducted in a temperature-controlled chamber (Fitoclima S600PL, ARALAB, Portugal) at a constant temperature of 25 ± 1°C. To filter out the radiation below 290 nm, the irradiated samples (UV-B treatments) were protected with a 0.13 mm-thick cellulose-diacetate film (Clarifoil, Texas, USA), which allowed passage of most of the UV-B and UV-A radiation but prevented the UV-C radiation (<280 nm). In parallel a “no UV-B treatment” was performed, where cages were wrapped in aluminum foil to prevent UV-B exposure during the irradiation.

**TABLE 1** | Geographical location and physicochemical properties of the soil samples used in this work.

Soil name	Geographical location		Soil order	Soil factors										
	Locality	Province		Sand	Silt	Clay	Textural class	OM	CaCO <sub>3</sub>	pH	EC 1:5	CEC	Fe <sub>d</sub>	DC
				(g/kg)	(g/kg)	(g/kg)		(g/kg)	(g/kg)		(μS/cm)	(Cmolc/kg)	(g/kg)	(g/kg)
AG35	Pozoblanco	Córdoba	Alfisol	860	90	50	Sandy	14	2	6.0	17	7	3	23
AG51	La Luisiana	Sevilla	Alfisol	400	120	480	Clay	6	0	6.3	402	8	25	153
AG53	SantaCruz	Córdoba	Vertisol	159	250	592	Clay	5	215	8.5	105	13	5	187
INM9	Fuente de Piedra	Málaga	Alfisol	545	255	200	Sandy clay loam	24	674	8.4	330	9	4	124

OM, organic matter; EC, electrical conductivity; CEC, cation exchange capacity; Fe<sub>d</sub>, dithionite extractable iron (Mehra and Jackson, 1958); DC, dispersible clay.

**FIGURE 1** | Assembly diagram to examine retention of *M. brunneum* conidia in the soil. Depth of each soil section (A, B, and C) is ca. 5 cm.

A *M. brunneum* conidial suspension ( $10^9$  conidia  $\text{ml}^{-1}$ ) was prepared. Then, newly emerged (24 h) *B. oleae* adults were cold anesthetized and treated with 1 ml of conidial suspension by using a Potter Spray Tower (Burkard Rickmansworth Co, Rickmansworth, UK). There were two controls, the first one was treated with the same volume of a sterile 0.1% Tween 80 aqueous solution and then irradiated, and the other was treated with the fungus and not irradiated. Following treatment, the insects were placed in the above-mentioned methacrylate cages. Then, the

UV-B treatment (inoculated with the fungi and covered with cellulose-diacetate film) was irradiated at  $1200 \text{ mW m}^{-2}$  for 2, 4, or 6 h. The no UV-B treatment (inoculated with the fungi and covered with aluminum foil) was irradiated for 6 h. Adult diet and water were provided *ad libitum*. Three replicates of 10 insects each were used for the UV-B and no UV-B treatments. Mortality was monitored for 10 days. Dead flies were removed daily and processed to assess mycosis.

### Virulence Assay of *M. brunneum* against *P. concolor* Adults

Virulence of *M. brunneum* against *P. concolor* adults was evaluated under laboratory conditions. Four *M. brunneum* concentrations were prepared in a sterile aqueous solution that contained 0.1% Tween 80 [ $10^5$ ,  $10^6$ ,  $10^7$ , and  $10^8$  conidia  $\text{ml}^{-1}$ ], with the control consisting of no conidia. Newly emerged *P. concolor* adults that had previously been cold-anesthetized were treated with a Potter Spray Tower. Half a milliliter of conidial suspension was used for each replicate, with three replicates per treatment (10 adults each). After the treatment, adults were placed in methacrylate cages ( $80 \times 80 \times 60 \text{ mm}$ , Resopal, Madrid, SP) with lids containing a 20 mm diam circular hole covered with a net cloth. Adult (treated and control) parasitoids were fed on honey. Mortality data were registered for 10 days. Dead flies were removed daily and the development of mycosis on the cadavers was determined.

### Evaluation of the Presence of Fungal Structures in Olive Oil

To assess the possible presence of *M. brunneum* conidia in olive oil after the treatment, two laboratory experiments were performed. In the first one, the olives were directly exposed to the fungus, creating a “worst case” scenario. For that, three olive samples (1 kg each) were placed on trays ( $690 \text{ cm}^2$ ) and sprayed with  $5.75 \text{ ml}$  of  $10^9$  conidia  $\text{ml}^{-1}$  with an Aerograph 27085 (piston compressor of  $231 \text{ min}^{-1}$ , 15–50 PSI, nozzle diameter of 0.3 mm, China). After treatment, olives remained in the trays for 48 h, and then olives from different trays were mixed and processed in three different groups before milling: (1) olives from the first group were rinsed in water for 1 min; (2) olives from the second group were surface-sterilized with 1% sodium hypochlorite; and (3) olives from the third group remained unwashed. Finally, olives from each group were processed into oil



following the extraction process described by Vossen (2007) with some modifications. Olives were crushed for 2 min at 37°C using a Thermomix® TM31 (Vorwerk, Wuppertal, D). After grinding, the paste was warmed in the same Thermomix for 30 min at 27.5°C. Then, the oil was extracted through combination of pressing and centrifuge rotating at ~10000 rpm.

In the second experiment, the olives were indirectly exposed to the fungus, creating a “real case” scenario for *M. brunneum* soil application. The above-mentioned trays were prepared with 500 g of soil that covered the entire base. The soil used in this study was collected from a farm in Córdoba and was characterized as sandy loam (78.0% sand, 17.0% silt, 5.0% clay, and 0.2% organic matter) with a pH of 8.3. The soil was sieved (2-mm mesh) and stored in a dry place at ~25°C. The soil was then sterilized at 121°C for 20 min and dried in an oven at 105°C for 24 h. Then, soils were sprayed with 5.75 ml of the fungal suspension ( $10^9$  conidia  $\text{ml}^{-1}$ ) using the above-described aerograph. After treatment, soil samples were collected from the bioassay trays. The conidial density after treatment was assessed using the CFU per gram of dry soil. Then, olives were dropped onto soil (treated and control). Olive samples were taken from the trays every 4 days for 16 days. Three replicates (trays) per sampling date were used for treatment and control. Finally, olive oil of each sampling date was extracted following the above-described protocol.

For both experiments, the possible presence of the fungus was evaluated according to the CFU method: 100  $\mu\text{l}$  of each oil sample was spread onto medium and incubated at 25°C for 15 days.

## Statistical Analysis

The area under the *B. oleae* flight curves (AUBFC) in treated and control plots was calculated by trapezoidal integration method of SAS (Campbell and Madden, 1990). Then, the values of AUBFC were  $\log_{10}$  transformed and subjected to factorial analysis of variance using Statistix 9.0 (Analytical Software 2008). The same program was used to analyze mortality data. The values of average survival times obtained by the Kaplan-Meier method and compared using the log-rank test were calculated with SPSS 15.0 Software for Windows. Replicates in time, for all experiments, were analyzed as series of experiments with the model,  $y = \text{treatment} + \text{experiment} + \text{treatment} \times \text{experiment}$  (Littell et al., 2006). Since the effect of *experiment* and interaction *treatment*  $\times$  *experiment* was not significant, replicates from both experiments were combined in a model with only *treatment* as a factor (one-way ANOVA). Mortality data was transformed,  $y = \arcsin\left(\sqrt{\frac{\% \text{Mortality}}{100}}\right)$ , to improve normality and homogeneity of variance, both requirements for linear model analysis. Means from different treatments were compared using Tukey's test ( $P < 0.05$ ). The effect of soil type and rain volume on relative percentage of *M. brunneum* recovered was evaluated with a generalized linear model for ordinal data (proportional odds model). This proportional odds model is the standard generalized linear model for ordinal regression (Stroup, 2012) and is appropriate for this experiment since we are measuring the response as ordinal data type, expressed as relative or cumulative percentage of conidia in each section of the soil

column. The dependent variable or response is the four possible classes or soil sections (A, B, C, and E). This model calculates the cumulative probability or proportion of conidia at each soil section or in the sections above. i.e.,  $P(\text{conidia} \leq A)$  is the probability or proportion of conidia in section A;  $P(\text{conidia} \leq B)$  is the probability or proportion of conidia in sections A and B;  $P(\text{conidia} \leq C)$  is the probability or proportion of conidia in sections A, B, and C; and  $P(\text{conidia} \leq E)$  is the probability or proportion of conidia in the effluent (E) or in any of the sections = 100%. The proportion of conidia in one specific section can be calculated then as the difference between contiguous cumulative probabilities e.g., probability or proportion of conidia in section B,  $P(\text{conidia} = B) = P(\text{conidia} \leq B) - P(\text{conidia} \leq A)$ . The equation of the model is:

$$\eta_{cijk} = \eta_c + \text{Soil type}_i + \text{Rain volume}_j + \text{Soil type} \\ \times \text{Rain volume}_{ij}, c(\text{class}) = A, B, C, E$$

Where: Rain volume is modeled as continuous factor or covariate; k sub index refers to replicate; observations follow a multinomial distribution; and cumulative logit link function.

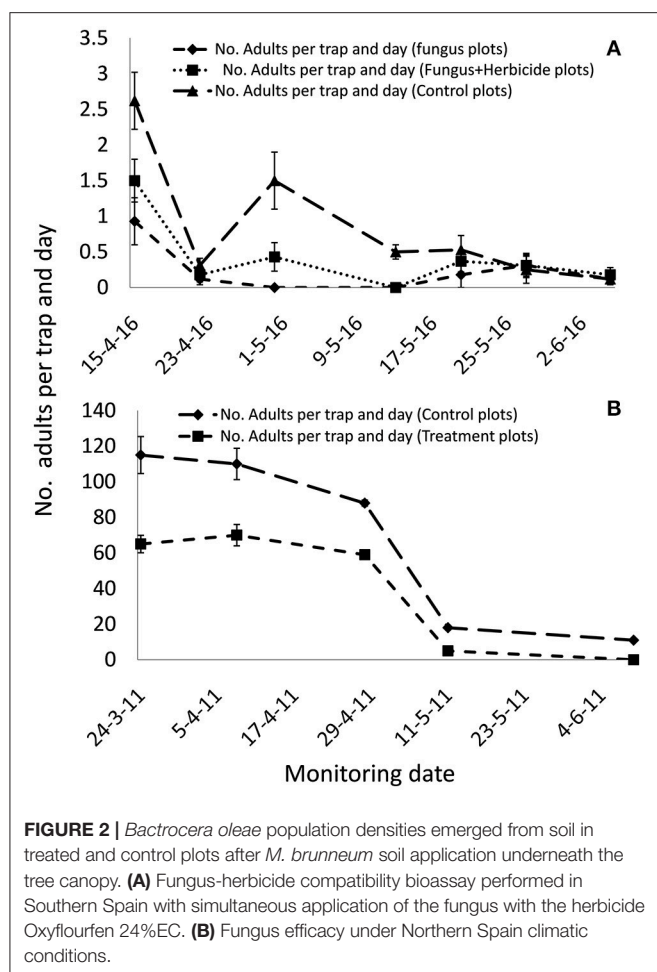
For the generalized linear model the estimation method was maximum likelihood with Laplace approximation. Model significance was evaluated with  $\chi^2$  test and the significance of the fixed effects was evaluated with F-approximate test ( $\alpha = 0.05$ ). Estimated cumulative probabilities for soil types were compared with odds ratio test. If the confidence interval for the ratio includes 1, the two soils are not significantly different (Stroup, 2012).

## RESULTS

### Fungus-Herbicide Compatibility and Fungal Effectiveness in Northern Spain

The soil applications of *M. brunneum* in both Northern and Southern Spain reduced the *B. oleae* population in treated plots compared with controls (Figure 2). The reduction in the *B. oleae* adult population emerging from the treated plots compared to the control plots was significant even if the fungus mixed with the herbicide ( $P < 0.05$ ). In the Southern Spain experiment, the maximum captures of adult flies was 0.9 and 1.5 flies per trap/day in the fungus and fungus and herbicide treated plots, respectively, compared with 2.6 flies per trap/day in the control plots. In Northern Spain, the maximum captures of adult flies were 65 flies per trap/day in the fungus treated plots compared with 115 flies per trap/day in the control plots. The *M. brunneum* soil density decreased after autumn treatments in relation with the monthly rainfall and date of fungus application (Figure 3). However, the fungus persisted in the soil in both fungus and fungus and herbicide treated plots. The fungal concentrations in the soil 4 months after treatment in both fungus and fungus and herbicide treated plots were  $1.2 \times 10^3$  and  $2.3 \times 10^3$  conidia per gram of soil, respectively (Figure 3).





**FIGURE 2 |** *Bactrocera oleae* population densities emerged from soil in treated and control plots after *M. brunneum* soil application underneath the tree canopy. (A) Fungus-herbicide compatibility bioassay performed in Southern Spain with simultaneous application of the fungus with the herbicide Oxyflourfen 24%EC. (B) Fungus efficacy under Northern Spain climatic conditions.

## Effect of Soil Type and Rain Volume on the Movement of Conidia in Soil

The analysis of the effect of soil type and rain volume on relative percentage of *M. brunneum* recovered showed a significant effect of soil type [ $F_{(4,130)} = 33.64$ ,  $P < 0.0001$ ], rain volume [ $F_{(1,130)} = 390.80$ ,  $P < 0.0001$ ], and interaction soil type x rain volume [ $F_{(4,130)} = 8.42$ ,  $P < 0.0001$ ]. The whole model was significant  $\chi^2_{9df} = 2347.08$ ,  $P < 0.0001$ .

The effect of rain volume on the relative percentage of conidia recovered for sections from A to E was inversely proportional for the series AG51, AG53, AG35, INM9, and FOCS.

The soils with more retention in the first section (A) were the clayey soils AG51 and AG53, with relative percentages of conidia recovered ranging between 83 and 97% of conidia, for the range of water volume evaluated. There were no significant differences between the profiles of retention for these two soil types (odds ratio test, 95% CI = 0.96–2.77) (Table 2 and Figure 4). The retention in the first section significantly decreased in the sandy soils AG35 and INM9 in comparison with the clayey soils AG51 and AG53. The relative percentage of conidia in section A ranged between 55 and 91%, with cumulative values for sections A and B ranging between 85 and 98%. There were no significant

differences between the profiles of AG35 and INM9 (odds ratio test, 95% CI = 0.77–1.44). The soil type with significant lower retention in the first sections was FOCS, with 23.4% of conidia in section C and effluent of 9.1% at the maximum rain volume of 400 ml.

## Effect UV-B Radiation Exposure Time on the Virulence of *M. brunneum* EAMa 01/58-Su Strain against *B. oleae* Adults

The virulence of *M. brunneum* was slightly higher on non-irradiated insects than UV-B treatment, with mortality ranging from 16.6 (non-treated and non-irradiated adults) to 92.9% (treated and non-irradiated adults), and fungal outgrowth from cadavers reaching 73.3% (Table 3). However, the average survival time (AST) of fungus-treated and irradiated adults were statistically equal to the AST of fungus-treated and non-irradiated adults, both, statically lower of AST of non-treated and irradiated adults or non-treated and non-irradiated adults (Table 3).

## Virulence Assay for *M. brunneum* against *Psytalia concolor* Adults

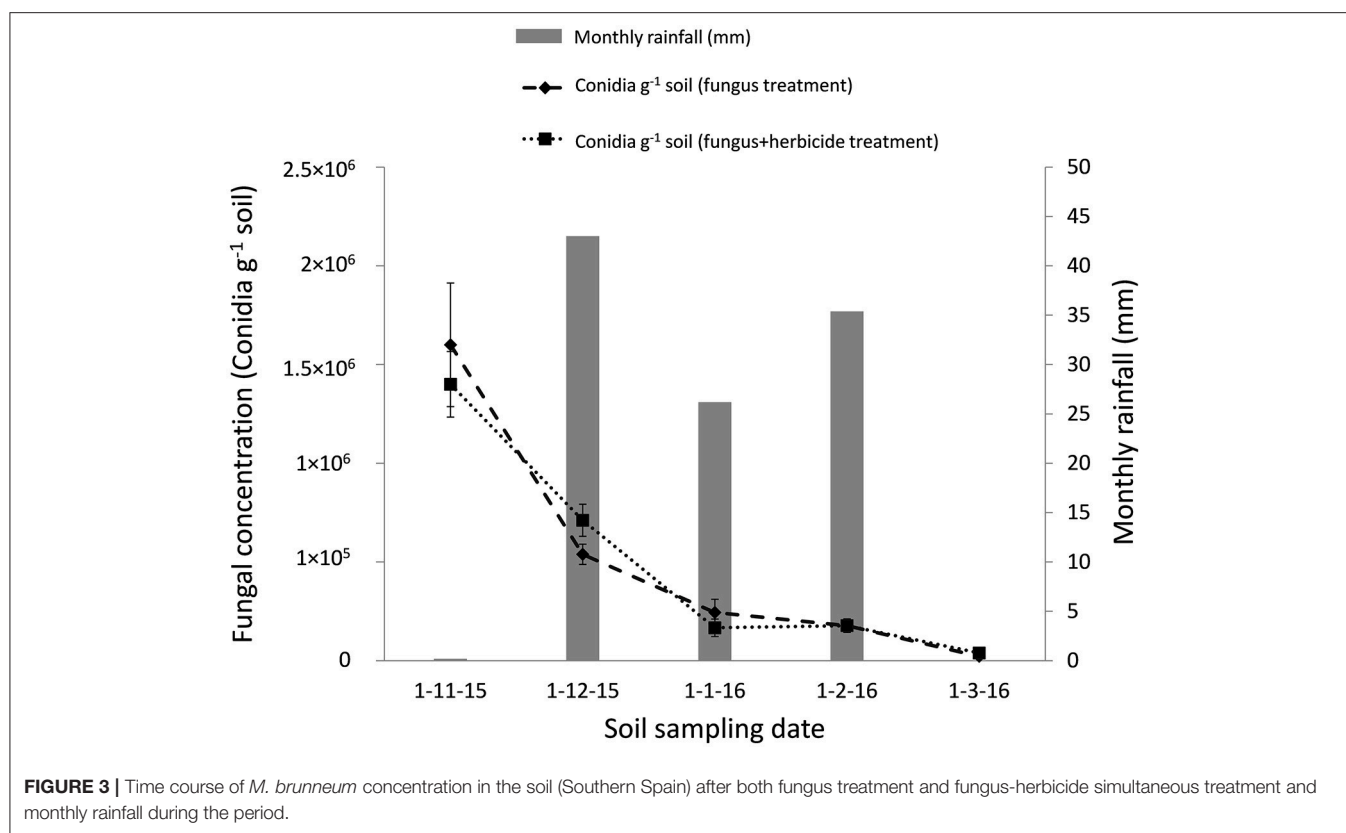
*P. concolor* adult mortality due to *M. brunneum* differed significantly at different conidial concentrations ( $P < 0.001$ ). However, only the highest concentration ( $10^8$  conidia  $\text{ml}^{-1}$ ) caused significant *P. concolor* adults mortality (21.6%) (Figure 5). *P. concolor* adult mortality caused by *M. brunneum* ranged between 1.6 and 6.6% at  $10^5$  to  $10^7$  conidia  $\text{ml}^{-1}$  compared with control mortality of 1.6%. Furthermore, the fungal treatment significantly influenced mycosis on the cadavers ( $P < 0.001$ ), with mycosis on 13.3, 1.6, 0, and 0% of the cadavers treated with  $10^8$ ,  $10^7$ ,  $10^6$ , and  $10^5$  conidia  $\text{ml}^{-1}$ , respectively. Only the highest concentration ( $10^8$  conidia  $\text{ml}^{-1}$ ) was statically different from the control (Figure 5).

## Evaluation of the Presence of Fungal Structures in the Olive Oil

No fungal propagules were detected in any of the olive oil samples obtained from both olive experiments, in which olives were directly or indirectly exposed to the fungus. There were no CFU in any of the Petri plates inoculated with the different olive oil samples.

## DISCUSSION

Throughout the last decades the development of biopesticides based on entomopathogenic fungi has become a very active line of work, since these products constitute an environmentally friendly alternative to synthetic/chemical pesticides (Sinha et al., 2016). Entomopathogenic fungi have been mainly investigated for aboveground pest control, whereas their potential for the control of soil-dwelling pests has been mostly ignored (Jackson et al., 2000). We had already developed a pioneer method based on soil application of *M. brunneum* for olive fruit fly control by targeting third-instar larvae in the soil during autumn and emerging adults from the soil during spring (Yousef et al., 2017).

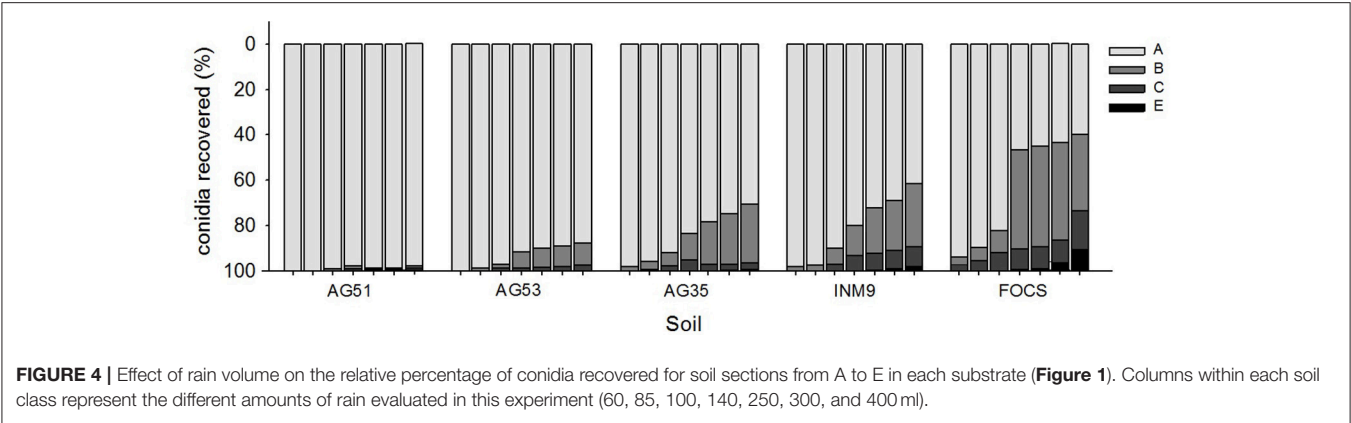


However, the results of the present work present additional evidence of the potential of this strategy to develop a safe and sustainable olive pest IPM.

*Metarhizium brunneum* mixed with the herbicide Oxyflourfen 24% EC in the atomizer tank and then applied beneath the olive trees canopy, reduced the *B. oleae* population that emerged during spring from the soil of treated plots compared to controls plots in the range of 43% (fungus+herbicide treatment) and 65% (fungus alone treatment). This *in vivo* compatibility of *M. brunneum* with the herbicide confirms our previous laboratory results (Yousef et al., 2015), and allows a fungus-herbicide simultaneous application reducing the application costs of both the fungus and herbicide. In addition, it is noteworthy that this reduction in *B. oleae* population density was a result of a single *M. brunneum* soil application in autumn. However, in our previous study, the soil application was performed twice, targeting preimagines in the soil in autumn and targeting adults at emergence from the soil in spring, providing between 50 and 70% of reduction in *B. oleae* population density (Yousef et al., 2017). The persistence of the fungus in the soil after treatment was not affected by the herbicide mixed with the fungus, with *M. brunneum* persisting in the soil, after both fungus treatment and fungus-herbicide simultaneous treatment during 4 months at the onset of the natural background concentration registered in bulk soil (Bruck, 2010; Scheepmaker and Butt, 2010). This is noteworthy because the decline of entomopathogenic fungi densities over time to

acceptable background level is requested by the EU regulations for registration purpose (Scheepmaker and Butt, 2010). This gradually decrease in time of the entomopathogenic fungi density in the soil may be influenced by edaphic, biotic, climatic, and cultural factors (Quesada-Moraga et al., 2007; Scheepmaker and Butt, 2010; Garrido-Jurado et al., 2011b). The effect of rainfall in the fate of fungal propagules in soil is the least understood, and has been shown to affect the foliar use of entomopathogenic fungi or the dispersion in soil, but in relation to crop residue (Bruck and Lewis, 2002; Jaronski, 2010). Many studies have suggested that rainfall is an important abiotic factor affecting conidia vertical mobility in the soil (Inglis et al., 2001; Garrido-Jurado et al., 2011b). However, this is first time that the fate of the fungal inoculum along the soil profile has been examined after the soil has been treated with entomopathogenic fungi. The results of the present work may serve as a model useful in estimating the fungal inoculum contained in each soil section according to soil type and amount of rain. More than 50% of the conidia was retained in the first 5 cm of the soil regardless of soil type and rain amount, which would guarantee the contact of the fungal propagules with *B. oleae* preimagines in the soil since around 80% of the third instar prepupating olive fruit fly larvae pupate ~3 cm below the soil surface (Dimou et al., 2003).

These results reveal both the high potential of *M. brunneum* and the key impact of the autumn soil application for the control of *B. oleae*. Our study validates the effectiveness use of



**FIGURE 4 |** Effect of rain volume on the relative percentage of conidia recovered for soil sections from A to E in each substrate (Figure 1). Columns within each soil class represent the different amounts of rain evaluated in this experiment (60, 85, 100, 140, 250, 300, and 400 ml).

**TABLE 2 |** Relative percentages (cumulative values) of conidia at different soil sections and soil types.

Soil type	Rain volume (ml)	Soil section A relative % (cumulative)			Soil section A + B relative % (cumulative)			Soil section A + B + C relative % (cumulative)			Comparison of soil types odds ratio test
		Estimate	Low 95%	Upp 95%	Estimate	Low 95%	Upp 95%	Estimate	Low 95%	Upp 95%	
AG51	140	97.2	96.4	97.8	99.3	99.2	99.5	99.8	99.8	99.8	a
AG51	400	95.3	93.4	96.7	98.9	98.4	99.2	99.7	99.6	99.8	
AG53	140	94.0	92.9	94.9	98.6	98.3	98.8	99.6	99.6	99.7	a
AG53	400	83.5	80.3	86.3	95.8	94.8	96.6	99.0	98.7	99.2	
AG35	140	88.2	86.8	89.5	97.1	96.7	97.5	99.3	99.2	99.4	b
AG35	400	66.1	62.4	69.7	89.8	88.1	91.3	97.5	96.9	97.9	
INM9	140	86.2	84.6	87.6	96.6	96.1	97.0	99.2	99.0	99.3	b
INM9	400	55.9	52.0	59.7	85.2	83.0	87.1	96.2	95.3	96.8	
FOCS	140	71.9	70.0	73.8	92.1	91.2	92.8	98.0	97.7	98.3	c
FOCS	400	31.3	28.3	34.5	67.4	64.2	70.5	90.1	88.3	91.6	

Values estimated with the generalized linear model for ordinal data for the exemplary cases of rain volume = 140 and 400 ml. Depth of each soil section from A to C is ca. 5 cm.

*M. brunneum*-soil applications as a pest management method in Northern Spain since both the climatic conditions and biological activity of *B. oleae* (up to 5 generations per year) are different to those in Southern Spain (Santiago-Álvarez and Quesada-Moraga, 2007). Surprisingly, similar results were obtained with a reduction percentage of almost 50% in the *B. oleae* population emerged from the soil of treated plots during spring compared with controls.

Our results also reveal that UV-B did not reduce the virulence of *M. brunneum* against *B. oleae* when adults were irradiated at 1200 mW m<sup>-2</sup> for 2, 4, and 6 h after being treated with the fungus. Even if the *B. oleae* adult mortality has been slightly reduced after 6 h exposure to UV-B, the AST of fungus-treated and irradiated adults were statically equal to the AST of fungus-treated and non-irradiated adults. However, the fungus applied to the soil during autumn targeting *B. oleae* larvae will be protected from environmental conditions, since the soil is the main natural reservoir for native entomopathogenic fungi and provides protection from extreme conditions (Ekesi et al., 2007; Zimmermann, 2007a,b; Garrido-Jurado et al., 2015). Nevertheless, it is interesting to know how UV-B may affect fungal virulence in *B. oleae* adults that emerge from the soil

during spring from surviving preimaginals. Generally, UV-B, can be particularly deleterious for entomopathogenic fungi due to its detrimental effects on various fungal biological processes (Inglis et al., 2001; Jaronski, 2010). Fernández-Bravo et al. (2017) have shown that UV-B radiation may potentially affect the virulence of *M. brunneum* against *Ceratitis capitata* adults, but the decrease of the conidial density due to UV-B exposure was not enough to avoid a viable number of conidia remaining that exceeded the threshold to cause disease.

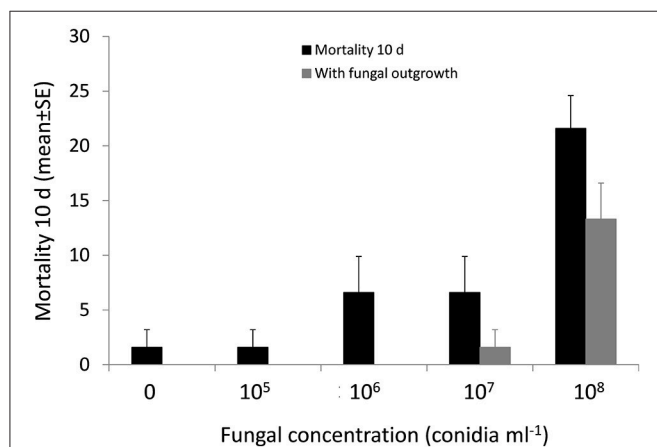
Once again, *M. brunneum* was shown to be safe for *P. concolor* in olive orchards. The highest *P. concolor* adult mortality caused by this strain was 21.6% at a concentration of 10<sup>8</sup> conidia ml<sup>-1</sup>. However, the same concentration (10<sup>8</sup> conidia ml<sup>-1</sup>) has caused more than 95% of *B. oleae* adult's mortality (Yousef et al., 2017). Nevertheless, at field conditions, the direct contact between *M. brunneum* applied to the soil and the parasitoid is not possible. Furthermore, studies of Ekesi et al. (2005) and Daane et al. (2015) have demonstrated that fungal application targeting tephritid fruit fly preimaginals in the soil can be compatible with the classic biological control in which *Psytalia* spp. larval parasitoids are field released. Our previous studies

**TABLE 3 |** Effect of the exposure time to UV-B ( $1200 \text{ mWm}^{-2}$ ) on the virulence of *M. brunneum* strain against adult *B. oleae*.

Treatment <sup>a</sup>	Kaplan-Meier survival analysis		Mortality % (mean $\pm$ SE)	
	AST (days) <sup>b</sup>	95% FL	Total	With fungal outgrowth
2 h	9.0 $\pm$ 0.2a	8.5–9.5	90.0 $\pm$ 5.7a	73.3 $\pm$ 12.0a
4 h	9.0 $\pm$ 0.2a	8.5–9.6	79.2 $\pm$ 5.8a	60.0 $\pm$ 10.0a
6 h	9.0 $\pm$ 0.3a	8.4–9.6	76.7 $\pm$ 7.3a	56.6 $\pm$ 12.0a
Control 1	10.6 $\pm$ 0.4b	9.8–11.4	20.0 $\pm$ 15.2b	0b
Control 2	8.7 $\pm$ 0.2a	8.1–9.2	92.9 $\pm$ 3.5a	63.3 $\pm$ 12.0a
Control 3	10.3 $\pm$ 0.3b	9.5–1.0	16.6 $\pm$ 6.6b	0b

<sup>a</sup>Control 1, Adult *B. oleae* were sprayed with  $1 \times 10^9$  conidia  $\text{ml}^{-1}$  and placed in methacrylate boxes, which were covered with cellulose-diacetate film (UV-B treatment) and aluminum foil for 6 h; Control 2, Treated and non-irradiated insects placed inside the chamber for 6 h; Control 3, non-treated and non-irradiated insects placed inside the chamber for 6 h.

<sup>b</sup>AST (mean  $\pm$  SE) limited to 10 days. Data in the same column followed by the same letter are not significantly different ( $\alpha = 0.05$ ) according to the log-rank test.

**FIGURE 5 |** Percentage mortality and fungal outgrowth (mean  $\pm$  SE) of newly emerged *P. concolor* adults after 10 days of exposure to suspensions of conidia at different concentrations.

have shown that soil applications of *M. brunneum* are safe to soil dwelling non-target arthropod communities, specially the formicid *Tapinoma nigerrimum* (Hymenoptera: Formicidae) used as a bioindicator in olive groves (Garrido-Jurado et al., 2011a).

## REFERENCES

- Altolaguirre-Obrero, M., López-Pérez, A., and Caballero-Jiménez, J. A. (2003). “Estrategia alternativa al control de mosca del olivo (*Bactrocera oleae* Gmelin) mediante “trampa OLIPE”. Ensayos en distintas zonas de la provincia de Córdoba,” in *Actas del XI Simposium Científico-Técnico Expoliva* (Jaén), 14–16.
- Bruck, D. J. (2010). Fungal entomopathogens in the rhizosphere. *Biocontrol* 55, 103–112. doi: 10.1007/s10526-009-9236-7

All olive oils obtained from olives previously exposed to *M. brunneum* were free of fungal propagules. The possibility of direct contact between the fungus and the olives is minimum, in spite of the direct contact experiment as a “worst case” scenario or the real conditions (soil application of the fungus). To date, this is the first study that addresses such possible presence of fungal inoculum in oils obtained from olives exposed to entomopathogenic fungi. These results demonstrate the food safety of this control method compared to the chemical insecticides used for more than 60 years for *B. oleae* control (Haniotakis, 2005), since pesticide residues have been detected in oils obtained from treated olives (Lentza-Rizos and Avramides, 1995). In addition, olive oil extraction at different exposure time to *M. brunneum* shows that there is no minimum waiting period between the application of the fungus and the harvest of the fruit.

This work highlights the adaptation of the olive fruit fly control method based on soil application of *M. brunneum* to the different climatic conditions and the possibility of a simultaneous soil application of the fungus with the herbicide beneath the tree canopy, which reduces application costs. This, together with the absence of negative effects on *P. concolor*, a *B. oleae* cosmopolitan parasitoid, demonstrates the environmental sustainability of this innovative method of control. In addition the use of entomopathogenic fungi leaves no residues in olive oil, in contrast to the use of chemical insecticides.

## AUTHOR CONTRIBUTIONS

MY, IG, and EQ-M designed the experiments. MY, CA-R, SR, and JM performed the experiments. MY, and PV-G analyzed the data. MY, IG, and EQ-M wrote the manuscript.

## ACKNOWLEDGMENTS

The authors thank the regional government of Andalusia for grant P11-AGR-7681, *Sustainable strategies for pest control based on the establishment of rhizosphere competent and endophyte entomopathogenic fungi*, and the Ministry of Economy and Competitiveness of the Spanish Government for grant, INNOLIVAR. We are also grateful to the Technical University of Madrid represented by Dra. Elisa Viñuela, Dra. Angeles Adan Del Rio, and Sr. Luis Quiros for kindly providing *P. concolor* pupae and the rearing standard procedure. I. Garrido-Jurado thanks the Ministry of Economy and Competitiveness of the Spanish Government for a Juan de la Cierva postdoctoral grant.

- Bruck, D. J., and Lewis, L. C. (2002). Rainfall and crop residue effects on soil dispersion and *Beauveria bassiana* spread to corn. *Appl. Soil Ecol.* 20, 183–190. doi: 10.1016/S0929-1393(02)00022-7
- Cañasveras, J. C., Barrón, V., del Campillo, M. C., Torrent, J., and Gómez, J. A. (2010). Estimation of aggregate stability indices in Mediterranean soils by diffuse reflectance spectroscopy. *Geoderma* 158, 78–84. doi: 10.1016/j.geoderma.2009.09.004
- Caballero, J. A. (2002). “Sistema de control de la mosca del olivo (*Bactrocera oleae*) en olivar ecológico. Experiencias en ‘Los Pedroches’,” in *Actas de la I Conferencia*



- Mundial de IFOAM Sobre Olivar Ecológico: Producciones y Culturas. Puente de Génave (Jaén), 421–424.
- Caleca, V., Antista, G., Campisi, G., Caruso, T., Verde, G. L., Maltese, M., et al. (2017). High quality extra virgin olive oil from olives attacked by the olive fruit fly, *Bactrocera Oleae* (Rossi) (Diptera Tephritidae): which is the tolerable limit? Data from experimental 'Nocellara Del Belice' and 'Cerasuola' olive groves in Sicily. *Chem. Eng. Trans.* 58, 451–456. doi: 10.3303/CET1758076
- Campbell, C. L., and Madden, L. V. (1990). *Introduction to Plant Disease Epidemiology*. New York, NY: John Wiley and Sons.
- Daane, K. M., and Johnson, M. W. (2010). Olive fruit fly: managing an ancient pest in modern times. *Annu. Rev. Entomol.* 55, 151–169. doi: 10.1146/annurev.ento.54.110807.090553
- Daane, K. M., Wang, X., Nieto, D. J., Pickett, C. H., Hoelmer, K. A., Blanchet, A., et al. (2015). Classic biological control of olive fruit fly in California, USA: release and recovery of introduce parasitoids. *Biocontrol* 60, 317–330. doi: 10.1007/s10526-015-9652-9
- Dimou, I., Koutsikopoulos, C., Economopoulos, A. P., and Lykakis, J. (2003). Depth of pupation of the wild olive fruit fly, *Bactocera* (Dacus) *oleae* (Gmel.) (Dipt., Tephritidae), as affected by soil abiotic factors. *J. Appl. Entomol.* 127, 12–17. doi: 10.1046/j.1439-0418.2003.00686.x
- Eilenberg, J., and Hokkanen, H. M. T. (2006). An ecological and societal approach to biological control. Dordrecht: Springer.
- Ekesi, S., Dimbi, S., and Maniania, N. K. (2007). "The role of entomopathogenic fungi in the integrated management of fruit flies (Diptera: Tephritidae) with emphasis on species occurring in Africa," in *Use of Entomopathogenic Fungi in Biological Pest Management*, eds S. Ekesi and N. K. Maniania (Trivandrum: Research SignPosts), 239–274.
- Ekesi, S., Maniania, N. K., Mohamed, S. A., and Lux, S. A. (2005). Effect of soil application of different formulations of *Metarhizium anisopliae* on African tephritid fruit flies and their associated endoparasitoids. *Biol. Control* 35, 83–91. doi: 10.1016/j.biocontrol.2005.06.010
- Fernández-Bravo, M., Flores-León, A., Calero-López, S., Gutiérrez-Sánchez, F., Valverde-García, P., and Quesada-Moraga, E. (2017). UV-B radiation-related effects on conidia inactivation and virulence against *Ceratitidis capitata* (Wiedemann) (Diptera: Tephritidae) of phylloplane and soil *Metarhizium* sp. strains. *J. Invertebr. Pathol.* 148, 142–151. doi: 10.1016/j.jip.2017.06.012
- Garrido-Jurado, I., Fernández-Bravo, M., Campos, C., and Quesada-Moraga, E. (2015). Diversity of entomopathogenic Hypocreales in soil and phylloplanes of five Mediterranean cropping systems. *J. Invertebr. Pathol.* 130, 97–106. doi: 10.1016/j.jip.2015.06.001
- Garrido-Jurado, I., Ruano, F., Campos, M., and Quesada-Moraga, E. (2011a). Effects of soil treatments with entomopathogenic fungi on soil dwelling non-target arthropods at a commercial olive orchard. *Biol. Control* 59, 239–244. doi: 10.1016/j.biocontrol.2011.07.001
- Garrido-Jurado, I., Torrent, J., Barrón, V., Corpas, A., and Quesada-Moraga, E. (2011b). Soil properties affect the availability, movement, and virulence of entomopathogenic fungi conidia against puparia of *Ceratitidis capitata* (Diptera: Tephritidae). *Biol. Control* 58, 277–285. doi: 10.1016/j.biocontrol.2011.05.017
- Goettel, M. S., and Inglis, G. D. (1996). "Fungi: hyphomycetes," in *Manual of Techniques in Insect Pathology*, ed L. Lacey (London: Academic Press), 213–249.
- Haniotakis, G. (2005). Olive pest control: present status and prospects. *IOBC/WPRS Bull.* 28, 1–9.
- Hsu, J.-C., Huang, L.-H., Feng, H. T., and Su, W.-Y. (2015). Do organophosphate based traps reduce control efficiency of resistant tephritid flies? *J. Pest Sci.* 88, 181–190. doi: 10.1007/s10340-014-0600-8
- Humber, R. (1997). "Identification of entomopathogenic fungi," in *Manual of Techniques in Insect Pathology*, ed L. Lacey (London: Academic Press), 153–185.
- Inglis, G. D., Goettel, M. S., Butt, T. M., and Strasser, H. (2001). "Use of hyphomycetous fungi for managing insect pest," in *Fungi as Biocontrol Agents. Progress, Problems and Potential*, eds T. M. Butt, C. Jackson, and N. Magan (Wallingford: CABI Publishing), 23–70.
- Jacas, J. A., and Viñuela, E. (1994). Analysis of a laboratory method to test the effects of pesticides on adult females of *Opius concolor* (Hym., Braconidae), a parasitoid of the olive fruit fly, *Bactrocera oleae* (Dip., Tephritidae). *Biocontrol Sci. Technol.* 4, 147–154. doi: 10.1080/09583159409355322
- Jackson, T. A., Alves, S. B., and Pereira, R. M. (2000). "Success in biological control of soil-dwelling insects by pathogens and nematodes," in *Biological Control: Measures of Success*, eds G. Gurr and S. Wratten (Dordrecht: Springer), 271–296.
- Jaronski, S. T. (2010). Ecological factors in the inundative use of fungal entomopathogens. *Biocontrol* 55, 159–185. doi: 10.1007/s10526-009-9248-3
- Kakani, E. G., Zygouridis, N. E., Tsoumani, K. T., Seraphides, N., Zalom, F. G., and Mathiopoulos, K. D. (2010). Spinosad resistance development in wild olive fruit fly *Bactrocera oleae* (Diptera: Tephritidae) populations in California. *Pest Manag. Sci.* 66, 447–453. doi: 10.1002/ps.1921
- Lentza-Rizos, C., and Avramides, E. J. (1995). Pesticide residues in olive oil. *Rev. Environ. Contam. Toxicol.* 141, 111–134. doi: 10.1007/978-1-4612-2530-0\_4
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006). *SAS for mixed models, 2nd edn*. Cary, NC: SAS Institute Inc.
- MAGRAMA. (2014). *Ministerio de Agricultura, Alimentación, y Medio Ambiente MAGRAMA*. Available online at: www.magrama.gob.es/es/
- Medjkouh, L., Tamendjari, A., Keciri, S., Santos, J., Nunes, M. A., and Oliveira, M. B. P. P. (2016). The effect of the olive fruit fly (*Bactrocera oleae*) on quality parameters, and antioxidant and antibacterial activities of olive oil. *Food Funct.* 7, 2780–2788. doi: 10.1039/C6FO00295A
- Mehra, O. P., and Jackson, M. L. (1958). "Iron oxide removal from soils and clays by a dithionite-citrate system buffered with sodium bicarbonate," in *Proceedings of the 7th National Conference on Clays and Clay Minerals* (Washington, DC), 317–327.
- Mraicha, F., Ksantini, M., Zouch, O., Avadi, M., Savadi, S., and Bouaziz, M. (2010). Effect of olive fruit fly infestation on the quality of the olive oil from *Chemlali* cultivar during ripening. *Food Chem. Toxicol.* 48, 3235–3241. doi: 10.1016/j.fct.2010.08.031
- Nicolopoulou-Stamati, P., Maipas, S., Kotampasi, C., Stamatis, P., and Hens, L. (2016). Chemical pesticides and human health: the urgent need for a new concept in agriculture. *Front. Public Health* 4:148. doi: 10.3389/fpubh.2016.00148
- Pell, J. K., Hannam, J. J., and Steinkraus, D. C. (2010). "Conservation biological control using fungal entomopathogens," in *The Ecology of Fungal Entomopathogens*, eds H. Roy, F. E. Vega, D. Chandler, M. Goettel, J. Pell, and E. Wajnberg (Heidelberg: Springer), 187–198.
- Quesada-Moraga, E., Navas-Cortés, J. A., Maranhão, E. A. A., Ortiz-Urquiza, A., and Santiago-Álvarez, C. (2007). Factors affecting the occurrence and distribution of entomopathogenic fungi in natural and cultivated soils. *Mycol. Res.* 111, 947–966. doi: 10.1016/j.mycres.2007.06.006
- Quesada-Moraga, E., and Santiago-Álvarez, C. (2008). "Hongos entomopatógenos," in *Control Biológico de Plagas Agrícolas*, eds J. A. Jacas and A. Urbaneja (San Jacinto, CA: VLC:Phytoma España), 98–120.
- Rogge, S. A., Mayerhofer, J., Enkerli, J., Bacher, S., and Grabenweger, G. (2017). Preventive application of an entomopathogenic fungus in cover crops for wireworm control. *Biocontrol* 5, 613–623. doi: 10.1007/s10526-017-9816-x
- Ruiz-Torres, M. (2010). Effectiveness of two attractants against the olive fruit fly (*Bactrocera oleae* Gmel.) and their impact on the arthropods community. *IOBC/WPRS Bull.* 59, 139–153.
- Santiago-Álvarez, C., and Quesada-Moraga, E. (2007). The olive fruit fly. *Oleae* 26, 60–61.
- Scheepmaker, J. W. A., and Butt, T. M. (2010). Natural and released inoculums levels of entomopathogenic fungal biocontrol agents in soil in relation to risk assessment and in accordance with EU regulations. *Biocontrol Sci. Technol.* 20, 503–552. doi: 10.1080/09583150903545035
- Sinha, K. K., Choudhary, A. K., and Kumari, P. (2016). "Entomopathogenic Fungi," in *Ecofriendly Pest Management for Food Security*, ed O. Omkar (London: Academic Press), 475–505.
- Stroup, W. W. (2012). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. Florida: CRC Press.
- Vossen, P. (2007). Olive oil: history, production, and characteristics of the world's classic oils. *Hortscience* 42, 1093–1100.
- Yousef, M., Garrido-Jurado, I., Ruiz-Torres, M., and Quesada-Moraga, E. (2017). Reduction of adult olive fruit fly populations by targeting preimaginals in the soil with the entomopathogenic fungus *Metarhizium brunneum*. *J. Pest. Sci.* 90, 345–354. doi: 10.1007/s10340-016-0779-y
- Yousef, M., Lozano-Tovar, M. D., Garrido-Jurado, I., and Quesada-Moraga, E. (2013). Biocontrol of *Bactrocera oleae* (Diptera: Tephritidae) with

- Metarhizium brunneum* and its extracts. *J. Econ. Entomol.* 106, 1118–1125. doi: 10.1603/EC12489
- Yousef, M., Quesada-Moraga, E., and Garrido-Jurado, I. (2015). Compatibility of herbicides used in olive orchards with a *Metarhizium brunneum* strain used for the control of preimaginal stages of tephritids in the soil. *J. Pest Sci.* 88, 605–612. doi: 10.1007/s10340-014-0632-0
- Zimmermann, G. (2007a). Review on safety of the entomopathogenic fungi *Beauveria bassiana* and *Beauveria brongniartii*. *Biocontrol Sci. Technol.* 17, 553–596. doi: 10.1080/09583150701309006
- Zimmermann, G. (2007b). Review on safety of the entomopathogenic fungus *Metarhizium anisopliae*. *Biocontrol Sci. Technol.* 17, 879–920. doi: 10.1080/09583150701593963

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Yousef, Alba-Ramírez, Garrido Jurado, Mateu, Raya Díaz, Valverde-García and Quesada-Moraga. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Sunflower Hybrid Breeding: From Markers to Genomic Selection

Aleksandra Dimitrijevic<sup>1</sup> and Renate Horn<sup>2\*</sup>

<sup>1</sup> Institute of Field and Vegetable Crops, Novi Sad, Serbia, <sup>2</sup> Institut für Biowissenschaften, Abteilung Pflanzengenetik, Universität Rostock, Rostock, Germany

## OPEN ACCESS

### Edited by:

Leire Molinero-Ruiz,  
Instituto de Agricultura Sostenible  
(CSIC), Spain

### Reviewed by:

Felicity Vear,  
INRA – Auvergne Rhône-Alpes  
Centre, France  
Lili Qi,  
Agricultural Research Service (USDA),  
United States  
Ruth Amelia Heinz,  
Instituto Nacional de Tecnología  
Agropecuaria (INTA), Argentina

### \*Correspondence:

Renate Horn  
renate.horn@uni-rostock.de

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 15 October 2017

**Accepted:** 20 December 2017

**Published:** 17 January 2018

### Citation:

Dimitrijevic A and Horn R (2018)  
Sunflower Hybrid Breeding: From  
Markers to Genomic Selection.  
Front. Plant Sci. 8:2238.  
doi: 10.3389/fpls.2017.02238

In sunflower, molecular markers for simple traits as, e.g., fertility restoration, high oleic acid content, herbicide tolerance or resistances to *Plasmopara halstedii*, *Puccinia helianthi*, or *Orobanche cumana* have been successfully used in marker-assisted breeding programs for years. However, agronomically important complex quantitative traits like yield, heterosis, drought tolerance, oil content or selection for disease resistance, e.g., against *Sclerotinia sclerotiorum* have been challenging and will require genome-wide approaches. Plant genetic resources for sunflower are being collected and conserved worldwide that represent valuable resources to study complex traits. Sunflower association panels provide the basis for genome-wide association studies, overcoming disadvantages of biparental populations. Advances in technologies and the availability of the sunflower genome sequence made novel approaches on the whole genome level possible. Genotype-by-sequencing, and whole genome sequencing based on next generation sequencing technologies facilitated the production of large amounts of SNP markers for high density maps as well as SNP arrays and allowed genome-wide association studies and genomic selection in sunflower. Genome wide or candidate gene based association studies have been performed for traits like branching, flowering time, resistance to *Sclerotinia* head and stalk rot. First steps in genomic selection with regard to hybrid performance and hybrid oil content have shown that genomic selection can successfully address complex quantitative traits in sunflower and will help to speed up sunflower breeding programs in the future. To make sunflower more competitive toward other oil crops higher levels of resistance against pathogens and better yield performance are required. In addition, optimizing plant architecture toward a more complex growth type for higher plant densities has the potential to considerably increase yields per hectare. Integrative approaches combining omic technologies (genomics, transcriptomics, proteomics, metabolomics and phenomics) using bioinformatic tools will facilitate the identification of target genes and markers for complex traits and will give a better insight into the mechanisms behind the traits.

**Keywords:** association panel, genome-wide association studies, genomic estimated breeding value, genomic selection, genome sequence, marker-assisted selection, sunflower, traits

## INTRODUCTION

Sunflower represents the second most important crop based on hybrid breeding, after maize (Seiler et al., 2017). It is mainly used for its seed oil, even though the seeds of confectionary sunflower also serve as snacks. With up to 12% of the global production of vegetable oils worldwide, sunflower takes position number four after palm oil, soybean and canola oil (Rauf et al., 2017). Apart from

its use for human nutrition, sunflower oil has a number of industrial applications as, e.g., basic component for polymer synthesis, biofuel, emulsifier or lubricants (Dimitrijevic et al., 2017).

Up until the beginning of the 1970s of the last century sunflower production was based on open-pollinated varieties (Vear, 2016). Events that led to changing sunflower production to hybrid breeding were the discoveries of the first cytoplasmic male sterility (CMS) source (Leclercq, 1969) and the identification of corresponding restorer genes (Kinman, 1970; Leclercq, 1971). Soon after, in 1972, the first sunflower commercial hybrid was available for production in United States (Putt, 1978). Exploitation of heterosis for hybrid development enabled farmers to obtain higher seed and oil yields, as well as increased uniformity (Bohra et al., 2016). The development of sunflower hybrids set up sunflower as a major viable crop worldwide and encouraged the founding of numerous public and private breeding centers (Skoric, 2012; Seiler et al., 2017). In recent years, public and private sector contributed to assemble huge plant genetic resources in sunflower, to identify markers for marker assisted selection (MAS) and to establish the use of new high-throughput technologies in sunflower. Today, the estimated value of global sunflower production reaches \$20 billion per year (FAO, 2016).

Basic directions in sunflower hybrid breeding include developing: (1) high seed and oil yield hybrids resistant to dominant diseases and tolerant to drought, (2) hybrids with changed oil properties, (3) confectionary hybrids, (4) herbicide resistant hybrids and (5) ornamental hybrids (Jocic et al., 2015). In addition, special markets have particular demands such as (1) achene and kernel properties as well as high protein content and lower oil content (lower than 40%) in confectionary sunflower production, (2) specific fatty acid and tocopherol composition in food and non-food industry or (3) plant height, ray and disk flower color, duration of flowering in ornamental sunflower hybrid breeding. The common needs for resistance against abiotic and biotic stress as well as the special needs of the various breeding purposes require the development of markers to facilitate the introduction of different traits.

Botanically, sunflower (*Helianthus annuus* L.) is a member of the Asteraceae family, one of the most diverse and largest families of flowering plants. Due to the economic importance of the cultivated sunflower and the ecophysiological variability within the genus *Helianthus*, sunflower became a model plant species for genome studies in the family (Bachlava et al., 2012). The sunflower genome with 3.6 Gb is quite large (Badouin et al., 2017), three times larger than the rapeseed genome (Chalhoub et al., 2014), or more than eight times larger than the one of rice (Arumuganathan and Earle, 1991). Due to its ability to grow in different agroecological conditions and its moderate drought tolerance, sunflower may become the oil crop of preference in the future, especially in the light of global environmental changes. Even though simulations showed an increase of sunflower yield for northern parts of Europe in view of predicted climate changes, negative effects on sunflower yield may occur in southern latitudes (Debaeke et al., 2017). Consequently, more attention should be paid to breeding for

better adaptation with regard to climate changes. These traits should include not only improvement in drought tolerance, but also introduction of pest resistance, salt tolerance and changes of plant architecture for better adaptation. Exploitation of available plant genetic resources in combination with the use of modern molecular tools for genome-wide association studies (GWAS) and application of genomic selection (GS) could lead to considerable improvements in sunflower. However, only in the recent years plant and genomic resources have become available in sunflower comparable to other crops (Figure 1). In this review we will talk about the long way that sunflower breeders and biotechnologists have to go and the future perspectives of using modern molecular tools in sunflower breeding.

## PLANT GENETIC RESOURCES IN SUNFLOWER

### Biparental and Wild Populations

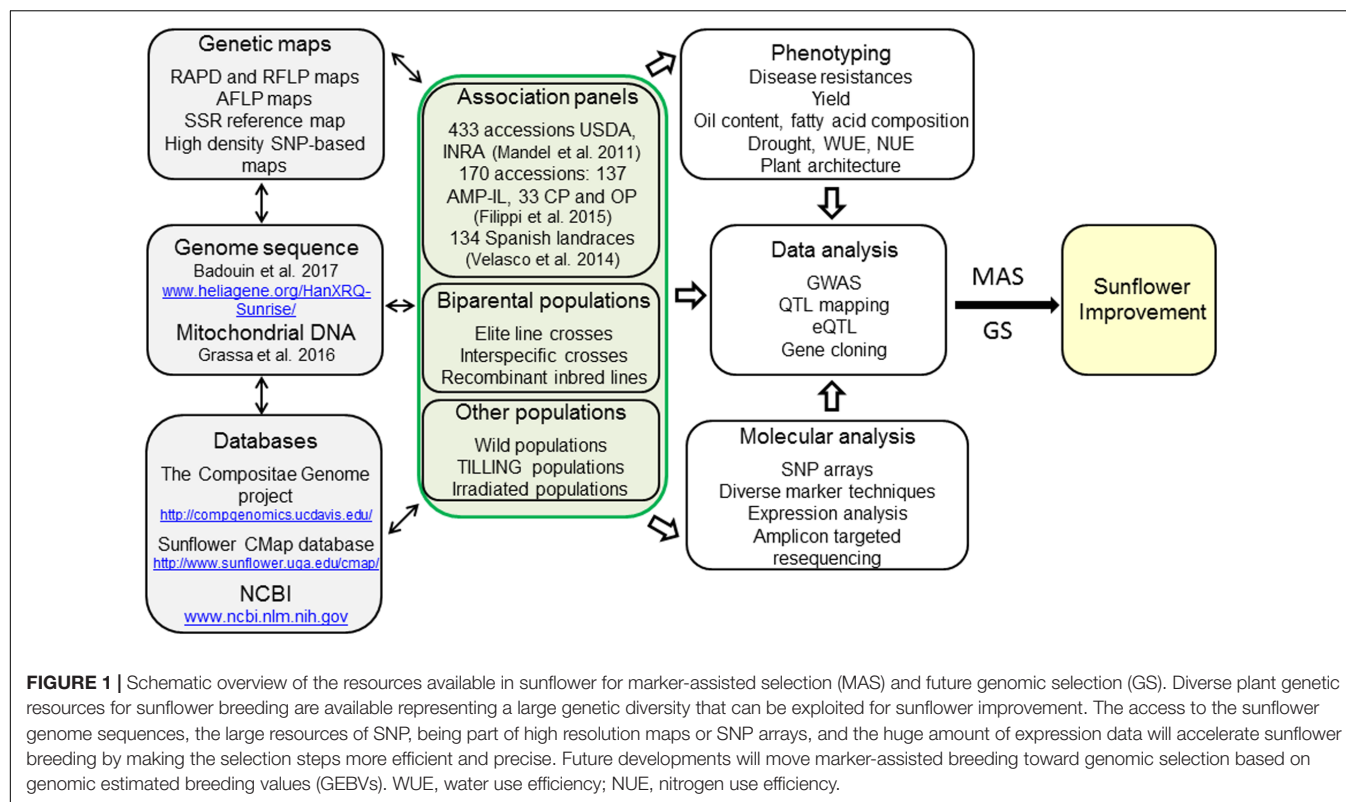
Biparental populations based on crosses between elite breeding, conventional, or introgressed lines (e.g., Berry et al., 1995; Horn et al., 2003; Vera-Ruiz et al., 2006; Kane et al., 2013; Livaja et al., 2016) as well as landraces and wild species (e.g., Quillet et al., 1995; Kim and Rieseberg, 1999; Brouillette et al., 2007; Ma et al., 2017) have been employed in sunflower for mapping of genes, marker detection, QTL analyses and gene cloning. In addition, recombinant inbred lines (RILs) have been developed that as immortals can be maintained forever by self-propagation (e.g., Berrios et al., 1999; Tang et al., 2002; Tang et al., 2006; Poormohammad Kiani et al., 2007a; Talukder et al., 2016). However, biparental populations have three major disadvantages: (1) these populations have to be individually established for each research project requiring time and resources, (2) only two alleles per locus can be evaluated and (3) due to missing recombination events populations show low resolutions in mapping (Bernardo, 2008; Patrick and Alfonso, 2013). The use of association panels overcomes these problems. To verify the usefulness of association panels the genetic diversity between wild populations and the genetic diversity fixed in association panels were compared (Mandel et al., 2011; Filippi et al., 2015). Even though alleles present only in the wild populations were detected, the majority of the alleles were present in the investigated association panels.

### Sunflower Collections

The largest sunflower collection is handled at the Institute of Field and Vegetable Crops, Novi Sad, Serbia consisting of over 7,000 sunflower inbred lines developed from different genetic sources and 21 perennial and 7 annual species (447 accessions in total)<sup>1</sup> (Atlagic and Terzic, 2014). The next largest collection of more than 5000 cultivated and wild *Helianthus* accessions is held at the USDA-ARS NPGS in Ames (Marek, 2016). About half of these, 2,519 accessions, represent the world's largest wild relatives sunflower collection, comprising 53 species – 39 perennial and 14 annual species (Seiler et al., 2017). Another large collection for sunflower (cultivated and wild) is maintained at the

<sup>1</sup><http://www.nsseme.com/about/inc/oilcrops/wild.php>





Vavilov Institute of Plant Industry, which consists of a total of 2,780 accessions from which 2,230 represent cultivated sunflower accessions and 550 wild sunflower accessions belonging to 24 species (19 perennials and 5 annuals) (Gavrilova et al., 2014). Some smaller numbers of 585 accessions of *H. annuus* are available via GRIN-CA, the Plant Gene Resources of Canada<sup>2</sup> and additional 613 sunflower accessions of diverse origin are distributed by the IPK Gatersleben<sup>3</sup>. These resources represent mostly uncharacterized plant material. In contrast to these, a well-defined collection of 400 open-pollinated varieties, landraces and breeding pools has been assembled by INRA to reflect the worldwide diversity present in sunflower (Mangin et al., 2017b). However, conservation of population diversity of sunflower populations represents a challenge in the maintenance process due to the self-incompatibility of wild sunflowers (Gandhi et al., 2005) and the possibility of genetic drifts occurring during the propagation of seed stocks (Mangin et al., 2017b). To study the preservation of the genetic variability, a set of 114 cultivated sunflower populations of the INRA collection were genotyped using a 384 Golden Gate SNP Assay. In conclusion, multiplication in isolation fields or use of cages is recommended to reduce loss of genetic variability in cultivated genetic resources.

These worldwide available collections of sunflower represent a valuable resource for the sunflower community. It could be of interest to include some additional accessions of these large collections to the existing association panels described below.

<sup>2</sup>[http://pgrc3.agr.gc.ca/order-ordre\\_e.html](http://pgrc3.agr.gc.ca/order-ordre_e.html)

<sup>3</sup>[https://gbis.ipk-gatersleben.de/GBIS\\_I/](https://gbis.ipk-gatersleben.de/GBIS_I/)

## Association Panels

Association panels have to be characterized by molecular markers like SSRs or SNPs to avoid false associations due to the population structure and family relationship. The review here focusses on association panels that are online available as the prior mentioned sunflower collections. To analyze the primary gene pool of sunflower an association panel consisting of 433 cultivated accessions from North America and Europe in addition to 24 wild sunflower populations distributed over the whole of United States were characterized by 34 selected EST-SSRs chosen on the presumptive neutrality toward domestication and breeding efforts (Chapman et al., 2008; Mandel et al., 2011). USDA cultivated accessions in this panel were assigned to the following categories: HA and RHA being either non-oil or oil, landrace, open-pollinated variety (OPV), non-oil introgressed, oil introgressed, other non-oil, and other oil. The INRA accessions could only be categorized into INRA-HA and INRA-RHA as the information on oil and non-oil was not always available. Analyses using the software STRUCTURE (Pritchard et al., 2000) and Principle Coordinates (PCO) analyses (Patterson et al., 2006) did not reveal deep genetic divisions within the germplasm (Mandel et al., 2011). The cultivated and the wild populations separated into two different groups and within the cultivated accessions the restorer-oil (RHA-oil) category stayed apart from the remaining gene pool (Mandel et al., 2011). This is not unexpected due to the hybrid history of sunflower in which the maintainer and restorer pools have been kept separate on purpose to maximize heterosis (Fick and Miller, 1997). A selection of 288 accessions still covers nearly 90% of the genetic diversity available in the

original larger panel. This association panel was named UGA-SAM1 and consists of 259 accessions, which are distributed by the Germplasm Resources Information Network (GRIN<sup>4</sup>) of the USDA National Plant Germplasm System (NPGS) and 29 accessions available through the French National Institute for Agricultural Research (INRA, France). This UGA-SAM1 population, which has been successfully employed in association studies (Mandel et al., 2013; Nambeesan et al., 2015), represents a very valuable tool for future association studies for the whole sunflower community. A minimal core set of 12 accessions (representing HA, RHA, oil and non-oil accessions as well as INRA material) capturing nearly 50% of the total allelic diversity might be ideal to build up a MAGIC (Multi-parent advanced generation intercross) population for sunflower. The MAGIC strategy is interesting for studies of multiple alleles in order to exploit higher recombination frequencies and better mapping resolution (Cavanagh et al., 2008). Development of MAGIC populations is in progress for numerous plant species (Bandillo et al., 2013) and would be interesting for sunflower as well.

Argentinean germplasm also represents a valuable genetic resource due to the long history of sunflower breeding in Argentina (de Bertero, 2003; Moreno et al., 2013). Filippi et al. (2015) characterized the population structure and calculated the genetic diversity of the association mapping population (AMP-IL), representing 137 INTA (National Institute of Agricultural Technology, Argentina) accessions and 33 accessions of open-pollinated and composite populations. The plant material is maintained by The Active Germplasm Bank of INTA Manfredi (AGB-IM). Using 42 SSR markers and/or SNP markers, detected by a 384 Illumina SNP-oligo pool array, estimated and observed heterozygosity as well as clustering using STRUCTURE and Discriminant Analysis of Principle Components (DAPC) were compared for the two marker types. As in other studies (Mandel et al., 2011, 2013) the population structure was dominated by the maintainer/restorer trait (Filippi et al., 2015).

A germplasm collection of 196 Spanish confectionary sunflower accessions is maintained at the Centre of Plant Genetic Resources of the National Institute for Agricultural and Food Research and Technology (CRF-INIA)<sup>5</sup>. A large genetic variation was revealed regarding hundred-seed weight, kernel percentage, seed oil content, fatty acid and tocopherol composition, phytosterols and other traits (Velasco et al., 2014; Pérez Vich et al., 2017).

In addition to well characterized association panels, considerable plant genetic resources are nowadays available in sunflower (cultivated as well as wild *H. annuus* accessions and accessions representing other species in the genus *Helianthus*).

## Mutagenized Populations

To increase the naturally available genetic variability sunflower has been mutagenized (Zambelli et al., 2015). Mutant populations have been successfully developed and used to screen for mutant phenotypes interesting for breeding purposes with regard to

flowering time, dwarf habitus, oil content, high oleic trait, herbicide resistance and branching (Soldatov, 1976; Gabard and Huby, 2001; Sala et al., 2008a; Cvejic et al., 2011b; Leon et al., 2013). Recently, a TILLING (Targeted Induced Local Lesion In Genomes) population for high throughput screening of EMS (ethyl methane sulfonate)-induced mutations in sunflower was established by Sabetta et al. (2011) and used for studies of genes involved in the fatty acid biosynthesis. Optimized mutagenesis using EMS was used to develop an additional sunflower TILLING platform (Kumar et al., 2013). Phenotypic characterization of 5,000 M2 lines was performed to estimate the mutation rates and to select interesting mutants. As seed oil biosynthesis is of major importance in sunflower, TILLING of *FatA* and *SAD* genes were investigated and revealed an overall mutation rate of one mutation every 480 kb (Kumar et al., 2013). Another possibility to develop mutant populations is to apply gamma irradiation or fast neutrons (Cvejic et al., 2011a). Optimal ranges for gamma irradiation and fast neutrons were explored in comparison to EMS concentrations.

Besides induced mutagenized populations, natural mutations that have occurred in wild sunflower populations have had significant impact on sunflower hybrid breeding, especially in the area of herbicide resistance. In the recent years intensive use of herbicides has led to the emergence of resistant wild sunflower populations. The first case was a population of common sunflower found in a soybean field in Rossville (KS, United States), in which imazethapyr that belongs the group of AHAS (acetohydroxy acid synthase) inhibitors was used over a time course of seven consecutive years for weed control. Thus, creating the first sunflower population, named ANN-PUR, resistant to one of the AHAS inhibitors (Al-Khatib et al., 1998). Resistance from this population was successfully introduced into commercial sunflower hybrids (Miller and Al-Khatib, 2000; Jocić et al., 2004). Sunflower production based on the use of this imidazolinone (IMI) resistance, which provides an efficient and easy control of post-emergence broadleaf weeds in Europe, is called Clearfield® technology. In addition to the discovery of IMI resistant sunflowers, another population of wild sunflowers (ANN-KAN), tolerant to another AHAS herbicide group called sulfonyleurea, was discovered in Kansas (United States) (Al-Khatib et al., 1999). The same tolerance was also obtained by EMS mutagenesis (Gabard and Huby, 2001). Later, more populations of wild sunflowers resistant to AHAS herbicides were found (e.g., White et al., 2002, 2003; Jacob et al., 2017). In addition, a new tolerance for imidazolinone called Clearfield Plus® was selected from an M2 population of 600,000 plants treated with EMS (Sala et al., 2008a).

Natural genetic diversity and naturally occurring or chemically/gamma-ray induced genetic variability represent a prerequisite for selection in breeding. The wide range of accessions maintained and made available by the germplasm banks for the research community is an extremely valuable starting point for successful breeding programs in sunflower allowing association studies and introduction of new traits into existing commercial breeding material. However, mutagenesis can create additional new genetic variability in traits where the natural variability is not sufficient.

<sup>4</sup><https://www.ars-grin.gov/>

<sup>5</sup><http://wwww.inia.es/coleccionescrf/PasaporteCRF.asp>

## GENETIC MAPS AND SUNFLOWER GENOME SEQUENCE

Different molecular markers, which have been applied in mapping genes and development of sunflower linkage maps (Figure 1), set the basis for the assessment of the genetic diversity present in the genus *Helianthus* as well as in cultivated and wild sunflower accessions. Positioning of desirable genes allowed the identification and development of more specific molecular markers. At the Sunflower CMap database<sup>6</sup> genetic maps available for sunflower have been listed and can be compared with each other by using the program CMAP (Kane et al., 2013).

The first map was developed on wild sunflower using RAPD markers (Rieseberg et al., 1993). A couple of years later maps were generated and published by using non-PCR based RFLP markers in different crosses of cultivated sunflower (Berry et al., 1995; Gentzbittel et al., 1995; Jan et al., 1998). These maps were published several years later than RFLP maps, e.g., in wheat, maize, barley, rice, and oilseed rape due to companies being involved in construction of the sunflower map (Hu, 2010). Later on, AFLP markers were added to the maps (Peerbolte and Peleman, 1996; Gedil et al., 2001). Most sunflower linkage maps contained 17 linkage groups (LG), representing the number of haploid chromosomes in sunflower. These maps were followed by genetic maps based on SSR markers (Tang et al., 2003b; Yu et al., 2003). The first composite genetic SSR map consisted of 278 single-locus SSR markers as well as additional 379 markers (public and proprietary), covering 1423 cM. This map that nowadays serves as reference genetic map for sunflower (Tang et al., 2003b) was then further saturated with additional SSR markers exploring three new mapping populations (Yu et al., 2003). In between more than 2,000 SSR have been derived from genomic sequences (gSSR) and EST (EST-SSR) and are now available for mapping and genotyping (Brunel, 1994; Dehmer and Friedt, 1998; Paniago et al., 2002; Tang et al., 2003b; Yu et al., 2003; Poormohammad Kiani et al., 2007b; Chapman et al., 2008; Heesacker et al., 2008). Existing sunflower maps were further enriched by these gSSRs, EST-SSRs, INDELs, TRAPs markers (Hu et al., 2007; Heesacker et al., 2008). These SSR markers (sequences and primers available through NCBI) represent a very valuable tool as they allow the localization of genes on individual linkage groups (Tang et al., 2003b) as well as on the recently published sunflower genome sequence of HanXRQ<sup>7</sup> (Badouin et al., 2017). About 3 gigabases (Gb) representing 80% of the whole genome size were assembled and represent an extremely useful tool for all different research programs that aim at the improvement of sunflower hybrids.

Finally, the step toward high-density maps was made possible by using SNP-based markers, starting with Lai et al. (2005) who derived SNPs from an EST database (as part of the Compositae Genome Project) and used them for mapping. An Infinium Beadchip including 9,480 SNPs based on transcriptome data was developed by Bachlava et al. (2012) and employed by Bowers et al.

(2012) to obtain four high-density genetic maps. Each of these maps contained 3,500–5,500 loci. Even though the maps were highly colinear, gaps in individual maps were observed. To solve this issue a consensus map of 10,080 loci was constructed from these data (Bowers et al., 2012). Talukder et al. (2014a) developed a high density map of 5,019 SNP markers obtained via RAD-sequencing. The rust resistance gene *R*<sub>12</sub> was fine-mapped using this SNP-based map. In addition, 118 SSR markers were included in the SNP map to address and orientate the linkage groups according to the sunflower reference genetic map. Celik et al. (2016) pioneered the use of genotyping-by-sequencing for large scale SNP detection in sunflower and developed a SNP-based linkage map of 817 SNP-markers covering all 17 LG by analyzing an F<sub>2</sub> obtained from the cross RHA 436 × H08 M1. Using the newly developed 25 K SNP array in sunflower Livaja et al. (2016) were able to construct a linkage map based on 6,355 SNP markers for the RIL population NDBLOSSel × CM625. The connection between genetic linkage maps and the sunflower karyotype was finally made by developing a molecular cytogenetic map for *H. annuus* (Feng et al., 2013). BAC and BIBAC clones with known genetic locations were used in fluorescence *in situ* hybridization (FISH) experiments to address the individual chromosomes.

The high resolution of the recently developed high-density maps in sunflower facilitates to narrow down the regions of interest, which should allow identification and cloning of genes for various relevant traits in the near future. In addition, SNP-based maps deliver markers closely linked to, e.g., resistance genes that can be applied in large scale marker-assisted breeding programs or can be integrated in SNP arrays.

## MARKER DEVELOPMENT BY LINKAGE MAPPING

### Resistance to Downy Mildew

Developed linkage maps set a good basis for localization and mapping of simply inherited traits. Most of the downy mildew resistance genes, conferring resistance to the oomycete *Plasmopara halstedii*, have been found to be dominantly inherited and consequently, relatively easy to map by using molecular markers. Identification of closely linked markers also represents a good basis for map-based cloning of the genes.

Great efforts were put into examination of downy mildew resistance genes (*R* genes), designated as *Pl* genes, that are distributed throughout sunflower genome: *Pl*<sub>13</sub>, *Pl*<sub>14</sub>, *Pl*<sub>16</sub>, and *Pl*<sub>Arg</sub> on LG1; *Pl*<sub>1</sub>, *Pl*<sub>2</sub>, *Pl*<sub>6</sub>, *Pl*<sub>7</sub>, *Pl*<sub>15</sub>, and *Pl*<sub>20</sub> on LG8; *Pl*<sub>5</sub>, *Pl*<sub>8</sub>, and *Pl*<sub>21</sub> on LG13; *Pl*<sub>17</sub> and *Pl*<sub>19</sub> on LG4; while *Pl*<sub>18</sub> is localized on LG2 of the sunflower SSR reference map (Mouzeyar et al., 1995; Roeckel-Drevet et al., 1996; Vear et al., 1997; Bert et al., 2001; Slabaugh et al., 2003; Yu et al., 2003; Mulpuri et al., 2009; Wieckhorst et al., 2010; Bachlava et al., 2011; Vincourt et al., 2012; Qi et al., 2015a; Qi L.L. et al., 2016; Zhang et al., 2016; Ma et al., 2017). Most of *Pl* genes are clustered, except for *Pl*<sub>Arg</sub> and *Pl*<sub>18</sub>.

The *Pl* cluster on LG8 was the first to be detected by molecular markers. Mouzeyar et al. (1995) used RAPD and RFLP markers for mapping the first downy mildew resistance gene, *Pl*<sub>1</sub>, which

<sup>6</sup><http://www.sunflower.uga.edu/cmap/>

<sup>7</sup><https://www.heliagene.org/HanXRQ-SUNRISE/>



is a part of a large *Pl* cluster (*Pl*<sub>1</sub>, *Pl*<sub>2</sub>, *Pl*<sub>6</sub>, *Pl*<sub>7</sub>). Of all genes in the cluster, *Pl*<sub>6</sub> gene was the most intensively examined since it conferred resistance to all the races present for a long time, except race 304. Different marker types were used for the introgression of *Pl*<sub>6</sub> into susceptible sunflower material including STS (Sequence-Tagged Sites) markers belonging to the TIR-NBS-LRR class of RGA (Resistance-Gene Analog) (Bouzidi et al., 2002) and co-dominant CAPS (Cleaved Amplified Polymorphic Sequence) markers (Table 1). The developed markers have been successfully used to introduce *Pl*<sub>6</sub> by MAS or to track the introduction of *Pl*<sub>6</sub> in backcrosses during conversion of downy mildew susceptible lines into resistant ones (Dimitrijevic et al., 2010; Jovic et al., 2010).

Two *Pl* genes originating from *H. argophyllus*, *Pl*<sub>8</sub> and *Pl*<sub>Arg</sub>, were also subject of numerous studies. *Pl*<sub>Arg</sub> confers resistance to all present downy mildew races and *Pl*<sub>8</sub> to 96% of all isolates collected in the north-central region of United States (Gilley et al., 2016). Radwan et al. (2004) developed STS markers for detection of the *Pl*<sub>5</sub>/*Pl*<sub>8</sub> locus that were later on also explored in other sunflower genotypes (Dimitrijevic et al., 2011). Bachlava et al. (2011) developed two resistance gene candidate (RGC) markers, RGC251 and RGC15/16, closely linked to *Pl*<sub>8</sub> that belong to the group of SSCP (Single-Strand Conformational Polymorphism) markers. However, SSCPs are labor-intensive and time-consuming in MAS. A comprehensive study of *Pl*<sub>8</sub> by Qi et al. (2017) explored previously published SNP markers as well as two SSR markers (Bowers et al., 2012; Talukder et al., 2014a) to genotype a F<sub>2</sub> population derived from the cross HA434 × RHA340. The three closest SNP markers, NSA\_000423, NSA\_002220, and NSA\_002251, were then investigated to check the specificity of the identified markers, concluding that NSA\_000423 and NSA\_002220 could serve as diagnostic markers in 87% of the tested sunflower lines when RHA 340 is used as donor for the *Pl*<sub>8</sub> gene. Validation of these markers across 548 sunflower lines proved their usefulness for MAS. However, a larger panel of sunflower lines should be tested.

Unlike, the *Pl*<sub>8</sub> gene, *Pl*<sub>Arg</sub> is not clustered. Several authors identified and developed different types of markers (SSRs, SNPs, RGCs) for MAS (Dußle et al., 2004; Wieckhorst et al., 2010; Imerovski et al., 2014b), some of these were also validated across a panel of sunflower lines. ORS716 was identified as the most useful marker in MAS (Table 1). Recently, Qi et al. (2017) combined available genomic data for the population obtained from the cross HA 89 × RHA 464 by use of SNP markers (Pegadaraju et al., 2013; Talukder et al., 2014a) with the phenotypic evaluation for resistance. The two nearest SNP markers (NSA\_007595 and NSA\_001835) narrowed the *Pl*<sub>Arg</sub> locus down to an area of 2.83 Mb. The nine identified SNP markers represent valuable diagnostic tools for introgression of *Pl*<sub>Arg</sub> into most genetic backgrounds in sunflower.

Other markers for use in MAS for downy mildew resistance include the identification of the tightly linked SSR marker ORS1008 to *Pl*<sub>13</sub> gene (Mulpuri et al., 2009), development of RGC markers tightly linked to *Pl*<sub>14</sub> gene (Bachlava et al., 2011) and the identification of one dominant co-segregating SSR marker (ORS1008) and one co-dominant tightly linked (EST)-SSR (HT636) to *Pl*<sub>16</sub>. Interestingly, HT636 and ORS1008 were

reported to be linked to both, *Pl*<sub>13</sub> and *Pl*<sub>16</sub>, indicating that these genes are in close vicinity to each other (Liu et al., 2012a). Qi et al. (2015a) used SSRs to place *Pl*<sub>17</sub> onto LG4 and then used SNPs identified by the National Sunflower SNP Consortium (Talukder et al., 2014a) and by Bowers et al. (2012) to saturate the region surrounding the *Pl*<sub>17</sub> gene. The authors identified SNP SFW04052 and ORS963 as the closest flanking markers linked to *Pl*<sub>17</sub>. A year later, Qi L.L. et al. (2016) used the same methodology to map *Pl*<sub>18</sub> to LG2 and found two SSRs and 10 SNPs flanking the *Pl*<sub>18</sub> gene. *Pl*<sub>18</sub> represents the first gene mapped to LG2. In 2017, two new *Pl* genes, *Pl*<sub>19</sub> and *Pl*<sub>20</sub>, were reported and mapped to LG4 and LG8, respectively (Ma et al., 2017; Zhang et al., 2017). Two SSRs and two SNPs were mapped in close vicinity to *Pl*<sub>19</sub>, while four SNP markers (SFW02745, SFW09076, S8\_11272025, and S8\_11272046) co-segregated with *Pl*<sub>20</sub>. All markers can be used in MAS and most importantly in pyramiding *Pl* genes in order to achieve long lasting resistance toward downy mildew. The development of SNP markers is of special interest because of the large number of markers generated that increase the likelihood to have markers available for any cross combination.

## Resistance to Sunflower Rust

Infections of sunflower plants with *Puccinia helianthi* Schwein lead to the rust disease. This fungus, which is mostly spread in North America, Argentina, South Africa, and Australia, can cause significant damage and yield reduction in infected fields. Genetic control of the disease can be effective; however, due to fast emergence of new races either by sexual or asexual reproduction, resistance achieved is short-termed. Consequently, a significant effort has been made into discovering rust resistance genes and the introduction into commercial lines and hybrids with a final goal of pyramiding several resistance genes in order to achieve long-term resistance. Most of the rust resistance genes (*R* genes), described so far, are monogenic dominant. *R* genes are located on different LGs of the sunflower genome with the majority being located on the LG13 [*R*<sub>4</sub>, *R*<sub>u6</sub>, *R*<sub>11</sub>, *R*<sub>adv</sub>, *R*<sub>13a</sub> (*R*<sub>HAR6</sub>), and *R*<sub>13b</sub>] (Bachlava et al., 2011; Qi et al., 2011b, 2012b; Gong et al., 2013b; Bulos et al., 2014).

First molecular studies were conducted on discovering markers for *R*<sub>1</sub> and *R*<sub>adv</sub> genes by use of RAPD and SCAR markers (Lawson et al., 1996, 1998). While *R*<sub>1</sub> gene was the first rust resistance gene present in a large number of sunflower lines, *R*<sub>adv</sub> is present in the line P2 owned by Pioneer Hi-Bred Australia (Lawson et al., 1998; Qi et al., 2011a). *R*<sub>adv</sub> is also present in the USDA line RHA 340, which Bachlava et al. (2011) used for mapping of the gene. Lawson et al. (1998) developed the SCAR marker SCT06950 linked to *R*<sub>1</sub> gene, which proved to be useful for detection of *R*<sub>1</sub> in different genetic backgrounds, except for the sunflower line MC29, which carries the *R*<sub>2</sub> and *R*<sub>10</sub> genes. For mapping of *R*<sub>2</sub>, Qi et al. (2015c) used a different MC29 line, called MC29 (USDA) as it was cultivated in the USDA-ARS Sunflower Research Unit, Fargo, North Dakota, which differs in term of resistance to NA race 6 in comparison to the MC29 line used by Lawson et al. (1998). Qi et al. (2015c) reported two SNP markers, NSA\_002316 and SFW01272, flanking the *R*<sub>2</sub> gene on LG14. Since, the closest marker, SFW01272, can only to a certain extent be used to detect the *R*<sub>2</sub> gene across different genetic



**TABLE 1** | Overview of resistance sources, locations, and markers of disease resistance genes in sunflower.

LG	Resistance gene	Source	Markers linked to the resistance gene	Distance (cM)	Reference
LG1	<i>Pl<sub>Arg</sub></i>	Arg1575-2 <sup>(1,2)</sup> RHA 419 <sup>(3)</sup> RHA 464 <sup>(4)</sup>	NSA_007595 <sup>(4)</sup>	0.01	DuBle et al., 2004 <sup>(1)</sup>
			NSA_001835 <sup>(4)</sup>	0.01	Wieckhorst et al., 2010 <sup>(2)</sup>
			ORS 716 <sup>(2,3)</sup>	0.3 <sup>(2)</sup> ; 0.0 <sup>(3)</sup>	
			ORS 662 <sup>(1,2,3)</sup>	1.9 <sup>(1)</sup> ; 0.3 <sup>(2)</sup> ; 0.0 <sup>(3)</sup>	Imerovski et al., 2014b <sup>(3)</sup>
			ORS 675 <sup>(3)</sup>	0.0	Qi et al., 2017 <sup>(4)</sup>
			RGC52a, RGC52b, RGC151 <sup>(2)</sup>	0.3	
	<i>Pl<sub>13</sub></i>	HA-R5	ORS1008 <sup>(1)</sup>	0.9 <sup>(1)</sup> ; 1.8 <sup>(2)</sup>	Mulpuri et al., 2009 <sup>(1)</sup>
	<i>Pl<sub>14</sub></i>	29004	HT636 <sup>(2)</sup>	2.2	Liu et al., 2012a <sup>(2)</sup>
			RGC203	1.6	Bachlava et al., 2011
	<i>Pl<sub>16</sub></i>	HA-R4	RGC188	2.6	
LG2	<i>R<sub>5</sub></i>	HA-R2	ORS1008	0.0	Liu et al., 2012a
			HT636	0.3	
			SFW03654 <sup>(2)</sup>	0.6	Qi et al., 2012a <sup>(1)</sup>
			ORS653a <sup>(1)</sup>	1.8	Qi et al., 2015b <sup>(2)</sup>
			NSA_000267 <sup>(2)</sup>	2.2	
	<i>Pl<sub>18</sub></i>	PI 494573	ORS1197-2 <sup>(1)</sup>	3.3	
			ORS203	0.4	Qi L.L. et al., 2016
			SFW03060	0.9	
			SFW03883	0.9	
			CRT214	1.1	
LG3	<i>Or<sub>5</sub></i>	RPG01 <sup>(1)</sup> PHD <sup>(2)</sup>	RTS05 <sup>(1)</sup>	5.6	Lu et al., 2000 <sup>(1)</sup>
			CRT 392 <sup>(2)</sup>	6.2	Tang et al., 2003a <sup>(2)</sup>
			ORS1036 <sup>(2)</sup>	7.5	
LG4	<i>Pl<sub>17</sub></i>	HA 458	ORS963	0.8	Qi et al., 2015a
	<i>Pl<sub>19</sub></i>	PI 435414	SFW04052	2.1	
			NSA_003564	0.6	Zhang et al., 2017
LG8	<i>Pl<sub>1</sub></i>	RHA 266	NSA_006089	0.6	
			S017H3-3	5.6	Mouzeyar et al., 1995
			S124EI-2	7.1	
	<i>Pl<sub>6</sub></i>	HA 335 <sup>(1)</sup> HA 336 <sup>(2)</sup>	Hap2R	0.0	Pankovic et al., 2007
			Hap3	0.0	
			CAPS HhaI	0.0	
	<i>Pl<sub>20</sub></i>	PI 494578	CAPS RsaI	0.0	
			SFW02745/SFW09076/S8_11272025/S8_11272046	0.0	Ma et al., 2017
			SCT06950	0.0	Lawson et al., 1998
LG11	<i>R<sub>1</sub></i>	RHA 279	NSA_003426/NSA_004155/NSA_000064 <sup>(2)</sup>	0.83 <sup>(2)</sup>	Gong et al., 2013a <sup>(1)</sup>
	<i>R<sub>12</sub></i>	RHA 464	CRT275 <sup>(2)</sup>	1.0	Talukder et al., 2014b <sup>(2)</sup>
			ORS1227 <sup>(1)</sup>	3.3	
			ZVG53 <sup>(1)</sup>	9.6	
	<i>R<sub>14</sub></i>	PH3	NSA_000064	0.7	Zhang et al., 2016
			ORS1227	1.6	
			ORS542	3.5	
			ZVG53	6.9	
LG13	<i>Pl<sub>5</sub></i>	XRQ	Ha-NT5R1,	4.8	Radwan et al., 2004
			Ha-NT5S3	6.4	
			Ha-NT5S1/Ha-NT5S2/ Ha-NT5R2	13.6	
	<i>Pl<sub>8</sub></i>	QIR8 ( <i>Pl<sub>8</sub></i> derived from RHA 340) <sup>(1)</sup>	RGC251 <sup>(2)</sup>	0.3	Radwan et al., 2004 <sup>(1)</sup>
			RGC15 <sup>(2)</sup>	0.4	Bachlava et al., 2011 <sup>(2)</sup>

(Continued)

TABLE 1 | Continued

LG	Resistance gene	Source	Markers linked to the resistance gene	Distance (cM)	Reference
		RHA 340 <sup>(2,3)</sup>	NSA_000423/SFW01497/ SFW08875 <sup>(3)</sup>	0.4	Qi et al., 2017 <sup>(3)</sup>
			Ha-NT8R3/Ha-NT8R4 <sup>(1)</sup>	3.5	
			Ha-NT8R5/Ha-NT8R6 <sup>(1)</sup>	4.8	
			Ha-NTIR11gEI <sup>(1)</sup>	6.3	
	<i>R<sub>4</sub></i>	HA-R3 <sup>(1,2)</sup>	SFW05240/SFW05630/SFW06095/ SFW08283 <sup>(2)</sup>	0.6	Qi et al., 2011b <sup>(1)</sup>
			SFW01497/SFW05630/SFW08875 <sup>(2)</sup>	0.7	Qi et al., 2015b <sup>(2)</sup>
			ORS581 <sup>(1)</sup>	0.8	
			ZVG61 <sup>(1)</sup>	2.1	
	<i>R<sub>HAR6</sub></i>	HA-R6	ZVG61	0.7	Bulos et al., 2013a
			ORS581	1.5	
	<i>R<sub>13a</sub></i>	HA-R6	SFW05743 <sup>(2)</sup>	0.2	Gong et al., 2013b <sup>(1)</sup>
			RGC15/16 <sup>(1)</sup>	0.3	Qi et al., 2015b <sup>(2)</sup>
			ORS316/ZVG61 <sup>(1)</sup>	0.4	
			ZVG61 <sup>(2)</sup>	0.4	
			SUN14 <sup>(1)</sup>	0.8	
	<i>R<sub>13b</sub></i>	RHA 397	SUN14 <sup>(1)</sup>	0.0	Gong et al., 2013b <sup>(1)</sup>
			SFW00757 <sup>(2)</sup>	0.0	Qi et al., 2015b <sup>(2)</sup>
			RGC15/16 <sup>(1)</sup>	0.1	
			ZVG61, ORS316 <sup>(1)</sup>	0.3	
			SFW04275/SFW04317/SFW05743 <sup>(2)</sup>	2.4	
			ZVG61, ORS316 <sup>(2)</sup>	5.9	
	<i>R<sub>adv</sub></i>	'Advance' <sup>(1)</sup>	SCX20 <sub>600</sub> <sup>(1)</sup>	0.0	Lawson et al., 1998 <sup>(1)</sup>
		RHA 340 <sup>(2)</sup>	RGC260 <sup>(2)</sup>	0.2	Bachlava et al., 2011 <sup>(2)</sup>
			ORS316 <sup>(2)</sup>	3.0	
	<i>P<sub>u6</sub></i>	P386	ORS316	2.5	Bulos et al., 2014
			ORS224	4.8	
	<i>R<sub>11</sub></i>	Rf ANN-1742	ORS728	0.3	Qi et al., 2012b
			ORS45	1.0	
	<i>Or<sub>ab-vl-8</sub></i>	AB-VL-8	ORS683	1.5	Imerovski et al., 2016
			ORS657	4.7	
LG14	<i>R<sub>2</sub></i>	MC 29 (USDA)	SFW01272	1.8	Qi et al., 2015c
			NSA_002316	2.8	
			SFW00211	2.9	

Numbers are brackets in the superscript in columns: "Source" and "Markers linked to the resistance gene" refer to the author citation list superscript numbers given in the "Reference" column concerning a specific resistance gene.

backgrounds; the authors recommend the use of two flanking SNP markers in order to minimize selection of false positives in MAS.

Further molecular studies of *R* genes include identification of molecular markers closely linked to *R<sub>4</sub>*, *R<sub>adv</sub>*, *P<sub>u6</sub>*, *R<sub>11</sub>*, *R<sub>13a</sub>* (*R<sub>HAR6</sub>*), and *R<sub>13b</sub>* genes that are located on LG13. Qi et al. (2011b) identified two markers flanking *R<sub>4</sub>* gene (ORS581 and ZVG61) in the cross HA 89 × HA-R3, which were later also reported to be linked to rust resistance genes *R<sub>13a</sub>* (*R<sub>HAR6</sub>*) and *R<sub>13b</sub>* located on the lower end of the LG13 (Bulos et al., 2013a; Gong et al., 2013b; Qi et al., 2015b) (Table 1). Further on, Gong et al. (2013b) saturated the region flanking the genes by analysis of RGC markers that were present in vicinity of downy mildew resistance gene *Pl<sub>8</sub>*, which was also mapped in the lower end of LG13. Another *R* gene that mapped in vicinity of *Pl<sub>8</sub>* and

fertility restorer gene *Rf<sub>1</sub>* was *R<sub>adv</sub>*. A completely co-segregating SCAR marker (Lawson et al., 1998) as well as RGC and SSR markers tightly linked to *R<sub>adv</sub>* were identified (Bachlava et al., 2011) (Table 1). Recently, Bulos et al. (2014) mapped *P<sub>u6</sub>* gene and identified closely linked SSRs to this gene in the sunflower line P386 on lower end of LG13. However, these markers are too far away to be useful in MAS (Table 1). *P<sub>u6</sub>* and *R<sub>4</sub>* map 6.25 cM apart from each other. Qi et al. (2012b) examined the *R<sub>11</sub>* gene and mapped it 1.6 cM from fertility restoration gene *Rf<sub>5</sub>* also on the lower end of LG13, hypothesizing the presence of a great rust R-gene cluster of *R<sub>adv</sub>*/*R<sub>11</sub>*/*R<sub>4</sub>*. SSR marker ORS45 was the closest to *R<sub>11</sub>* gene and was mapped 1 cM proximal to the gene, while ORS728 was shown to be a common marker for *R<sub>11</sub>* and *Rf<sub>5</sub>* genes. The results allow the conclusion that the lower end of LG13 harbors the second largest cluster of NBS-LRR encoding

genes: rust resistance and downy mildew resistance genes. Based on SSR and RGC markers used in this area, Gong et al. (2013b) proposed that this big cluster could be sub-divided into two clusters.  $R_{adv}$  and  $R_{11}$  form sub-cluster I, while  $R_4$ ,  $R_{13a/b}$ ,  $Pl_5$ ,  $Pl_8$  form subcluster II.  $Pl_{21}$  that was also positioned on LG13 mapped 8 cM proximal to  $Pl_5/Pl_8$  (Radwan et al., 2004; Vincourt et al., 2012).

Other rust resistance genes investigated by use of molecular markers include analysis of  $R_5$ . This is to date the only  $R$  gene discovered on LG2. Qi et al. (2012a, 2015b) identified two SSR and two SNP markers flanking the gene, with the closest being 0.6 cM away (Table 1). On LG11 two rust resistance genes have been mapped so far:  $R_{12}$  and  $R_{14}$ . Both genes were positioned in the middle of LG11, and were discovered in wild sunflower accessions, however, they have different origin:  $R_{12}$  from PI413047 and  $R_{14}$  from PI413038 (Gong et al., 2013a; Zhang et al., 2016). Both genes were mapped between the markers ORS1227 and ZVG53 (ORS1227 with 3.3 and 1.6 cM and ZVG53 with 9.6 and 6.9 cM from  $R_{12}$  and  $R_{14}$ , respectively). Talukder et al. (2014b) performed fine mapping of the  $R_{12}$  gene region by using SNP markers. Five SNP markers (NSA\_000064, NSA\_008884, NSA\_004155, NSA\_003320, and NSA\_003426) were linked with 0.83 cM to the gene, but only two markers (NSA\_003426 and NSA\_004155) proved to have diagnostic quality for  $R_{12}$  (Table 1). The nearest SNP marker to  $R_{14}$  was NSA\_000064, which was mapped with 0.7 cM from the gene in the  $F_2$  mapping population obtained from the cross HA 343  $\times$  PH3 (Zhang et al., 2016). However, this marker amplified the same banding pattern in RHA 464 ( $R_{12}$ ) and PH3 ( $R_{14}$ ). Zhang et al. (2016) identified thirteen SSR/InDel and two SNP markers that amplified different profiles between the two donors of  $R_{12}$  and  $R_{14}$  indicating polymorphisms between these regions.

One of the latest efforts in saturation mapping of  $R$  genes was published by Qi et al. (2015b) who used previously developed SFW and NSA SNP markers in order to saturate the regions surrounding  $R_4$ ,  $R_5$ ,  $R_{13a}$ , and  $R_{13b}$  genes and succeeded in identifying markers that are under 1 cM distant from all analyzed genes thus raising the efficiency of introduction of rust resistance in to susceptible material (Table 1). The authors used previously developed SSR markers and newly developed SNP markers for identification of homozygous “double-resistant”  $F_2$  individuals in a population obtained from a cross combination between a BC<sub>3</sub> $F_2$  plant harboring  $R_5$  and HA-R6 bearing  $R_{13a}$ . The  $F_4$  progeny obtained from chosen plants showed improved resistance toward races 336 and 777 in comparison to lines that possess only one resistance gene. Qi et al. (2015c) also performed marker-assisted pyramiding of  $R_2$  and  $R_{13a}$  in confectionary sunflower by use of SSR and SNP markers. Further pyramiding of  $R$  genes could lead to long-term improvements in sunflower rust resistance. The process of converting susceptible into resistant forms can be greatly facilitated and accelerated by use of the reported molecular markers.

## Resistance to Broomrape

Another constraint in sunflower production is broomrape (*Orobanche cumana*), a parasitic flowering plant, that can cause significant yield loss of up to 100%. Most of the genes that

confer resistance to broomrape were found to be monogenic dominant for broomrape races A to E and G (Vranceanu et al., 1980; Velasco et al., 2012), while resistance to race F was either inherited by a monogenic dominant gene (Pacureanu-Joita et al., 1998; Pérez-Vich et al., 2004) or by two recessive genes (Rodríguez-Ojeda et al., 2001) depending on the genetic background. Broomrape resistance genes are denoted as  $Or$  genes. Imerovski et al. (2014a) reported a single recessive resistance gene in the sunflower line HA-267 that carried a resistance gene higher than  $Or_6$ . The majority of molecular analyses were conducted in investigating and creating different types of molecular markers for detection of  $Or_5$  that conveys resistance to broomrape race E or lower (Lu et al., 2000; Tang et al., 2003a) (Table 1). The efficiency of RAPD and SSR primers in MAS for  $Or_5$  were tested by Iuoras et al. (2004), however, none of the primers proved to be efficient or accurate enough. Imerovski et al. (2013) identified SSR markers associated with  $Or_2$ ,  $Or_4$ , and  $Or_6$  genes that could be used in converting broomrape susceptible sunflower genotypes into resistant ones. However, *O. cumana* populations belonging to race F have shown different aggressiveness (Molinero-Ruiz et al., 2009). Imerovski et al. (2016) mapped newly identified broomrape resistant gene conferring resistance to broomrape races overcoming race F from sunflower inbred line AB-VL-8 on LG3. The authors named the gene  $Or_{ab-vl-8}$ , which was shown to be recessive and ORS683 mapped 1.5 cM from the gene. Further molecular analysis are needed in order to develop co-segregating markers for some of the  $Or$  genes. In addition, finding novel resistance sources is essential since broomrape races are emerging at a high speed. Recent work of Louarn et al. (2016) involved using 586, 985 SNPs from SUNRISE project<sup>8</sup> on GeneTitan<sup>®</sup> (Affymetrix) for identification of QTL for resistance to broomrape races F and G. The authors identified 17 QTL spread throughout 9 LGs. Among them was a stable QTL on LG13 that controlled the number of broomrape emergence that explained 15–30% of the phenotypic variability. This QTL was marked as the one that could be the most rapidly used. A molecular characterization of *O. cumana* populations in Europe using RAPD-PCR identified four groups (Molinero-Ruiz et al., 2014). These markers might be useful as molecular tools to detect first broomrape appearances in fields that had been free of virulent races (Molinero-Ruiz et al., 2014).

## Herbicide Tolerance

Different tolerances against herbicides inhibiting the large, catalytic subunit of the acetohydroxyacid synthase (AHASL) have become a very necessary tool in sunflower hybrid production and cultivation as these facilitate the application of either imidazolinones (IMIs) or sulfonylureas (SUs) against broadleaf weeds (Sala et al., 2012). It also allows a race independent control of broomrape (Skoric and Pacureanu, 2010). Three AHASL genes were isolated from sunflower: AHASL1 located on LG9, AHASL2 on LG6 and AHASL3 on LG2 (Kolkman et al., 2004). Only mutations in AHASL1 seem to be involved

<sup>8</sup><https://cnrgv.toulouse.inra.fr/fr/Projets/Analyse-de-genome-pour-l-amelioration-des-plantes/SUNRISE-SUNflower-Resources-to-Improve-yield-Stability-in-a-changing-Environment>

in the herbicide tolerance in sunflower. Four different mutated alleles have been explored for commercial use in sunflower hybrid breeding: Imisun/Clearfield®, Clearfield Plus®, Sures and ExpressSun® (Sala et al., 2012). Point mutations from C-T in codon 205 (*Ahasl1-1*) and in codon 197 (*Ahasl1-2*) (adopting the *Arabidopsis* nomenclature) confer moderate tolerance to IMIs and high tolerance to SUs, respectively. The allele *Ahasl1-3* is characterized by a G-A mutation in codon 122 and results in high levels of IMI tolerance (Sala et al., 2008b). The broadest range of herbicide tolerance is shown by allele *Ahasl1-4*, which has a G-T mutation in codon 574 (Sala and Bulos, 2012). A first SNP marker based on the C-T change in codon 205 proved to be very useful as it cosegregated with partially dominant herbicide tolerance for the Imisun/Clearfield® system (Kolkman et al., 2004), even though an additional non-target gene is required for the tolerance (Bruniard and Miller, 2001; Miller and Al-Khatib, 2002). One SSR marker exploiting the differences in the (ACC) repeats present in the *AHASL* gene allows the differentiation between the wild type *Ahasl1* allele and alleles *Ahasl1-1* and *Ahasl1-2* (Sures and ExpressSun®) (Kolkman et al., 2004; Bulos et al., 2013b). A CAPS marker developed by Bulos et al. (2013b) uses the A-T exchange to detect the *Ahasl1-3* allele (Clearfield Plus®) by digesting the PCR product with the restriction enzyme *BmgBI*. The markers can now help to select for herbicide tolerance. Nevertheless, the development of efficient screening tests for herbicide tolerance is crucial (e.g., Breccia et al., 2011; Vega et al., 2012).

## Seed Oil Quality

Several oil properties have been characterized as quantitative traits, however, some traits such as oleic acid content (OAC) could, to a certain extent, be considered a semi-qualitative trait since OAC is dependent not only on the environment, but also on the genetic background of the receiver line (Ferfuia et al., 2015; Regitano Neto et al., 2016). A partial duplication of the *FAD2-1* allele caused by chemical mutation leads to an increase in OAC by silencing the *FAD2-1* gene encoding *FAD2* (oleoyl-phosphatidylcholine desaturase) (Lacombe et al., 2002; Schuppert et al., 2006). This enzyme catalyzes the synthesis of linoleic acid from oleic acid and by silencing its activity oleic acid is accumulated. Soldatov (1976) created the Pervenets cultivar with elevated OAC, which has become the main source of elevated OAC in sunflower breeding programs worldwide due to the beneficial properties of high oleic sunflower oil (Allman-Farinelli et al., 2005; Vannozzi, 2006). Inheritance of the OAC trait has been a subject of numerous studies and different results were reported from a single dominant gene to several genes influencing OAC (Urie, 1984; Lacombe et al., 2004; Joksimovic et al., 2006; Bervillé, 2010; Ferfuia and Vannozzi, 2015; Premnath et al., 2016; Dimitrijevic et al., 2017). Gene/genes involved in inheritance of OAC have been denoted as *Ol* genes. Different markers were employed in mapping and detecting the mutation (*Ol* mutation) in sunflower. The two RAPD markers, F15-690 and AC10-765, were linked with 7.0 and 7.2 cM to *Ol<sub>1</sub>* gene, respectively (Dehmer and Friedt, 1998). Later on, the *Ol<sub>1</sub>-FAD2-1* locus was placed onto LG14 (Pérez-Vich et al., 2002; Schuppert et al., 2006). One major QTL identified by Pérez-Vich et al. (2002) explained 84.5% of the variation in the OAC. Schuppert et al. (2006)

provided dominant INDEL markers for tracking the *Ol* mutation in addition to identifying 49 SNPs and five INDELs in the 3'-region of *FAD2-1*. Three years later, a co-dominant SSR marker tightly linked to the *Ol* mutation and dominant markers specific for the mutation were published (Lacombe et al., 2009). Recently, Premnath et al. (2016) identified in addition to the QTL on LG14, two additional QTL for OAC on LG8 and LG9. The two markers HO\_Fsp\_b for the QTL on LG14 (Schuppert et al., 2006) and ORS762 for the QTL on LG8 explained about 60% of the phenotypic variation in OAC. Several of the markers have been used for validation across numerous sunflower lines (Nagarathna et al., 2011; Singchai et al., 2013; Bilgen, 2016; Dimitrijevic et al., 2016). Dimitrijevic et al. (2017) reported marker F4-R1 created by Schuppert et al. (2006) as the most efficient in MAS for OAC.

## Fertility Restoration

Development of reliable tools for detection of cytoplasmic male sterility (*cms*) and restorer of fertility (*Rf*) genes would significantly improve and accelerate the process of developing sunflower hybrids. In sunflower, CMS PET1 originating from an interspecific hybridization of *H. petiolaris* with *H. annuus* (Leclercq, 1969) is the only CMS cytoplasm worldwide used for hybrid breeding. Male sterility is caused by the co-transcription of the *atpA* gene with the new CMS-specific *orfH522* leading to the expression of a 16-kDa-protein (Horn et al., 1991; Köhler et al., 1991). Fertility restoration suppresses the co-transcription anther-specific (Monéger et al., 1994). In sunflower, the restorer genes for the PET1 cytoplasm represent the best characterized due to the commercial use of this cytoplasm in sunflower hybrid breeding. The restorer gene *Rf<sub>1</sub>*, which was originally discovered by Kinman (1970) in the line T66006-2-1-B, has since then been integrated into a number of USDA/ARS RHA lines like RHA 271, RHA 272, RHA 273, and others (Korell et al., 1992; Serieys, 2005). A second major dominant restorer gene *Rf<sub>2</sub>* was discovered in a test cross between T66006-2-1-B and MZ01398. However, this *Rf<sub>2</sub>* gene seems to be ubiquitously present in almost all cultivated sunflower lines, along with maintainer lines of CMS PET1 (Serieys, 2005). Only *Rf<sub>1</sub>* is responsible for restoring male fertility in sunflower hybrids (Leclercq, 1984). RAPD markers in combination with AFLP markers were very useful for mapping of the restorer gene *Rf<sub>1</sub>* (Horn et al., 2003), which was first positioned on LG6 of the RFLP sunflower map (Gentzbittel et al., 1995). Two RAPD markers OPK13\_454 and OPY10\_740, which mapped 0.8 and 2.0 cM from *Rf<sub>1</sub>*, respectively, were converted into more reliable, easier to handle SCAR markers HRG01 and HRG02 (Horn et al., 2003). A recent study of these SCAR markers for breeding practice proved that HRG01 is more efficient for *Rf<sub>1</sub>* detection in perennial species, whereas HRG02 gave better results for annual species (Markin et al., 2017). In addition a multiplex TaqMan assay was established that allowed the detection of HRG01 and *orfH522* at the same time (Markin et al., 2017). Using the SSR markers ORS1030, *Rf<sub>1</sub>* had been mapped to LG13 (Kusterer et al., 2005) of the sunflower reference map (Tang et al., 2003b). In addition, a CAPS marker H13, which mapped 7.7 cM from *Rf<sub>1</sub>* gene, was developed from the RAPD marker OPH13\_337 by digesting the PCR product with *HinfI* (Kusterer et al., 2005). The tight linkage between CAPS H13 and *Rf<sub>1</sub>* was



confirmed in *Xenia* hybrid combination (Port et al., 2013). An additional SSR marker ORS511 and a TRAP marker K11F05Sa12-160 were mapped to the *Rf*<sub>1</sub> gene with distances of 3.7 and 0.4 cM, respectively (Yue et al., 2010). A fertility restorer gene *Rf*<sub>3</sub>, which could be shown to be different from *Rf*<sub>1</sub> and *Rf*<sub>2</sub>, was identified in the confectionery restorer line RHA 280 (Jan and Vick, 2007). *Rf*<sub>3</sub> could be linked with eight markers to LG7, including five known SSR markers (ORS328, ORS331, ORS928, ORS966, and ORS1092) and three new SSR markers HT-619-1, HT619-2, and HT1013 derived from expressed sequence tags (Liu et al., 2012b). SSR ORS328, which mapped 0.7 cM distant from *Rf*<sub>3</sub>, represents so far the closest co-dominant marker to the gene (Liu et al., 2012b). Another restorer gene *Rf*<sub>3</sub> in RHA 340 has also been mapped to LG7 (Abratti et al., 2008). *Rf* ANN-1742, a restorer line derived from wild *H. annuus* showed resistance to rust (Qi et al., 2012b). The new rust resistance gene *R*<sub>11</sub> mapped with 1.6 cM closely to a restorer gene on the lower end of LG13. The SSR marker ORS728 was mapped 1.3 cM proximal from this restorer gene and 0.3 cM distal to *R*<sub>11</sub>. Marker analyses using HRG01, HRG02, STS115, and ORS728 indicated that this restorer gene, now called *Rf*<sub>5</sub>, might not be allelic to *Rf*<sub>1</sub> (Qi et al., 2012b).

So far, 72 new CMS sources have been described for sunflower (Serieys, 2005). However, only for very few of these CMS sources markers have been detected linked to the corresponding restorer genes (Horn et al., 2016). Feng and Jan (2008) tagged an additional restorer gene *Rf*<sub>4</sub> with molecular markers and assigned it to LG3 of the sunflower general reference map (Tang et al., 2003b). *Rf*<sub>4</sub> is restoring male fertility to a newly identified CMS cytoplasm GIG2. Schnabel et al. (2008) identified AFLP markers that mapped in close vicinity of the restorer gene *Rf*-*PEF1*, which represent a major restorer gene for the *PEF1* CMS cytoplasm, another potentially interesting CMS source for commercial sunflower hybrid breeding. In addition, markers were developed that allowed the distinction between the *PET1* cytoplasm and the *PEF1* cytoplasm. For CMS 514A, a *H. tuberosus* based male sterile cytoplasm, the restorer gene *Rf*<sub>6</sub> was located on LG3 with eight markers (Liu et al., 2013). Two SSR markers, ORS13 and ORS1114, mapped as close as 1.6 cM to *Rf*<sub>6</sub>. GISH showed *Rf*<sub>6</sub> to be present on a small translocation introgressed from *H. angustifolius*.

Further analyses are needed in order to develop more tightly linked molecular markers to *Rf* genes to locate them on the genetic map and to get an insight on the fertility restoration mechanisms in sunflower. In other species, most of the so far cloned restorer of fertility genes belong to the pentatricopeptide repeat gene family (PPR), however, also other types of restorer genes have been identified (Horn et al., 2014).

## ASSOCIATION MAPPING

For association mapping two approaches have been explored: (1) genome-wide association studies (GWAS) and (2) candidate gene approaches. For most plant species, the last strategy was predominantly applied because whole genome sequences have only recently become available (Fusari et al., 2008). However, high-throughput marker systems nowadays give full genome

coverage, which makes approaches as genome-wide association mapping, QTLSeq mapping and genomic selection possible (Mammadov et al., 2012). As the linkage disequilibrium (LD) in sunflower rapidly decays (Liu and Burke, 2006; Kolkman et al., 2007; Fusari et al., 2008) studies based on associations could result in resolution levels detecting genes underlying quantitative trait loci. However, it is important to analyze the population structure of the association mapping population to avoid false associations.

In sunflower, only one of the association mapping studies so far was performed genome-wide (Mandel et al., 2013), all others were candidate gene based (Fusari et al., 2012; Cadic et al., 2013; Talukder et al., 2014b; Nambeesan et al., 2015; McAssey et al., 2016). Genome-wide association mapping was performed in an association population of 271 lines (Mandel et al., 2011), using 5,359 SNP marker from the Illumina Infinium Beadchip (Mandel et al., 2013). Associations were studied regarding flowering time, branching and heterotic groups. LD showed considerable variability across the genome, but significant marker-trait associations were detected. Selection for disease resistance as well as initial domestication might be responsible for the genome-wide differences in the LD profile (Mandel et al., 2013). This first screen was followed by a more detailed, refined association mapping approach based on candidate genes for branching (Nambeesan et al., 2015). Shoot branching was differentiated in no branching, apical, mid-apical, mid, mid-basal, basal branching as well as whole plant branching or other phenotype. A total of 48 candidate genes described to be involved in branching in other plant species were used to detect homologs to 39 genes in sunflower. Up to eight of the highest BLAST hit for each gene were included in the analyses due to the recent triplication of the sunflower genome (Badouin et al., 2017). For 13 candidate genes for branching co-localization of SNPs associated with branching was observed (Nambeesan et al., 2015). Most of these were found on LG10, where previous QTL mapping had detected the B-Locus for recessive branching (Tang et al., 2006; Bachlava et al., 2009). With regard to flowering time, a SNP in *HaFT2* was identified that co-localized with a flowering time QTL (McAssey et al., 2016).

Association mapping and linkage mapping were combined with QTL detection to identify mutations responsible for changes in flowering time (Cadic et al., 2013). Associations with flowering time could be demonstrated for 11 regions distributed over 10 LGs. In addition, QTL for flowering time were detected on 11 LGs in a RIL population by linkage mapping. This large number of QTL is consistent with the polygenic pattern of inheritance of flowering time reported before (Leon et al., 2000). SNPs detected by association mapping were then investigated with regard to positional overlaps with QTL identified in the RIL population. The remaining eight regions contained five candidate genes potentially associated with flowering time in other species that showed SNPs in sunflower, one of the genes was the gibberellin receptor *GID1B* (Cadic et al., 2013). Thirty genes, including this gene had before been investigated as candidate genes for flowering time with regard to domestication and improvement in sunflower (Blackman et al., 2011). One major QTL, which was detected on LG14 by linkage mapping (Poormohammad Kiani et al., 2009), was not detected by the association study (Cadic et al., 2013). This can happen if alleles are present in

a low frequency in an association panel as one disadvantage of association mapping is that rare alleles are difficult to be associated with traits.

*Sclerotinia sclerotiorum*, a necrotrophic, fungal pathogen, is one of the most devastating diseases in sunflower. The fungus can cause three different types of diseases depending on which part of the plants gets infected and whether the infection occurs via ascospores or mycelia (Gulya et al., 1997). These are stalk rot, mid-stalk rot and head rot. In a panel consisting of 94 sunflower lines 16 candidate genes were screened for associations to *Sclerotinia* head rot using a Mixed Linear Model (MLM) that also considers family relationship as well as population structure (Fusari et al., 2012). These candidate genes had been derived from previous transcript profiling in sunflower (Peluffo, 2010) and *Brassica* (Zhao et al., 2007) after infecting the plants with *S. sclerotiorum*. Significant association of the haplotype 3 of the gene *HaRIC\_B*, representing a truncated gene, was detected and accounted for 20% reduction in *Sclerotinia* head rot. Candidate gene association mapping for *Sclerotinia* stalk rot was also performed in another association panel of 260 cultivated sunflower lines (Talukder et al., 2014b). Eight genes, which had been identified in defense response against *S. sclerotiorum* in *Arabidopsis* (Guimaraes and Stotz, 2004; Guo and Stotz, 2007), served as basis to identify the orthologous genes in sunflower. The panel was divided in two groups representing either the best resistance response or the most susceptible lines. Association studies found strong association of *HaCOII-1* and *HaCOII-2* with resistance against *Sclerotinia* stalk rot, explaining 7.4% of the observed phenotypic variation (Talukder et al., 2014b).

Association mapping studies in the recent years have shown that this approach represents an interesting alternative to linkage mapping especially regarding quantitative inherited traits.

## TOWARD GENOMIC SELECTION

Genomic selection (GS) is so far mostly used in animals, e.g., dairy cattle (Van Raden et al., 2009). However, application of genomic selection got started as well in plant breeding, e.g., in maize (Massman et al., 2013; Bandeira e Sousa et al., 2017; Cantelmo et al., 2017; Lyra et al., 2017), potato (Habyarimana et al., 2017), soybean (de Azevedo Peixoto et al., 2017), sugar beet (Würschum et al., 2013), and wheat (Bassi et al., 2016). Genomic selection was regarded as promising in hybrid breeding of self-pollinating crops as wheat (Longin and Reif, 2014; Zhao et al., 2015), especially if little is known about the heterotic pools. To implement GS into sunflower breeding programs some general aspects of genomic selection need to be emphasized.

Genomic selection selects the individuals based on genomic breeding values (GEBVs) (Meuwissen et al., 2001). The idea of GS is to use genome-wide molecular data to effectively select for quantitative trait loci (Bernardo, 2008; Massman et al., 2013; Würschum et al., 2013). More than 10,000 QTL have been detected by traditional mapping approaches considering 12 major crop species, but only very few have been successfully applied in marker-assisted breeding programs (Bernardo, 2008). Genomic selection is a concept that becomes more attractive

as high-throughput genotyping becomes feasible due to recent advances in genotyping platforms and to considerable price reductions in the last few years. As first step in GS, a training population has to be established that is genotyped and phenotyped. This training population is needed to adjust the statistical models, which are then applied to predict breeding and genotypic values of individuals that have not been phenotyped (Bassi et al., 2016). The breeding population consists of these not phenotyped individuals that are only genotyped. Selection is performed in the breeding population. Finally, a validation population serves to estimate the accuracy of the GS models (Bassi et al., 2016). Comparing traditional MAS and GS, three major differences are obvious: (1) within the training phase markers linked with a gene of interest and quantitative traits are identified in MAS, whereas in GS models are developed to predict GEBVs, (2) in the breeding phase only few markers are used in traditional MAS for genotyping, whereas in the GS genome-wide genome data are collected and (3) regarding the selection in the breeding phase traditional MAS uses only the identified markers to select the individuals by genotype, whereas selection in GS is performed based on the GEBV (Nakaya and Isobe, 2012). For the success of GS, the accuracy of the prediction of GEBV is the most important factor. The accuracy of prediction relies on the characteristics of the training population as size, marker density, trait heritability and kinship between training and breeding population as well as the ratio of training population : breeding population (Nakaya and Isobe, 2012; Bassi et al., 2016). In traditional MAS, markers tightly linked to a QTL could be applied in most other breeding population, so that the relationship between the mapping and the breeding population had not to be considered by the breeder. However, in GS the interrelationship between training and breeding population is crucial for the predictive power (Nakaya and Isobe, 2012).

In sunflower, prediction of hybrid performance was based on fingerprinting data in form of 572 AFLP markers (Reif et al., 2013). Intragroup (133) and intergroup hybrids and the parental lines were evaluated at two locations in 2 years for grain yield, oil content and oil yield. If no information on the General Combining Ability (GCA) of the parental lines was accessible, prediction of hybrid performance using genomic selection methods was accurate if the parents were closely related, but with genetically distant lines prediction proved challenging (Reif et al., 2013). However, prediction based on GCA could not be improved by genomic selection. In the recent years, large sets of markers were generated in sunflower by genotyping-by-sequencing (Baute et al., 2016; Celik et al., 2016; Talukder et al., 2016; Ma et al., 2017; Qi et al., 2017), application of the new 25 K SNP genotyping array (Livaja et al., 2016) and sequencing of parental lines (Mangin et al., 2017a). However, so far only SNP array data were used for genomic prediction of *Sclerotinia* resistance (Livaja et al., 2016) and sequencing data for the genomic prediction of sunflower hybrid oil content (Bonnafeous et al., 2016; Mangin et al., 2017a). In the latter case, an incomplete factorial design consisting of 36 CMS lines and 36 restorer lines was used to compare

prediction accuracy of GS and classical GCA modeling in sunflower. Multi-environmental field trials were performed to characterize 452 sunflower hybrids of the panel with regard to hybrid performance in oil content, which represents a primarily additive trait with high heritability. In addition, all 72 parental lines were sequenced to obtain genome-wide SNP markers (Mangin et al., 2017a). Genomic predictions were then made for missing hybrids and hybrid combinations lacking information about at least one parental line. In conclusion, GS led to considerable improvement in breeding efficiency compared to the conventional GCA modeling if little is known about one or both parental lines (Mangin et al., 2017a). For *Sclerotinia* midstalk rot, the prediction ability of a genome-based best linear unbiased prediction (GBLUP) model was evaluated in a biparental population genotyped with the 25 K SNP array (Livaja et al., 2016). High predictive abilities were obtained for “stem lesion length” and lower predictive abilities for “leaf lesion length” and “speed of fungal growth,” which represent traits with lower heritabilities. These first experimental trials for genomic predictions, using and comparing the results of different models, have shown the potential and the limitations for genomic selection in sunflower.

## FUTURE PERSPECTIVES

In this review the emphasis was given to plant genetic resources and molecular tools used to detect and exploit genetic diversity and to facilitate sunflower hybrid breeding. Traditional MAS has been successfully used to introduce monogenic traits into the breeding material, especially disease resistance as well as herbicide tolerance. Validation of identified molecular markers across different genotypes has also shown the limitation in markers to be used in different genetic backgrounds. However, sunflower researchers have put a lot of effort in the identification of markers linked to specific traits without gaining insight into the function of the involved genes, even though this would allow a better understanding of the metabolism and mechanisms behind traits. Breeding for complex polygenic traits is still challenging. With this regard, it is necessary to stress the importance of precise phenotypic evaluation, on which molecular biologists rely to correctly interpret the molecular and phenotypic data. High-throughput phenotyping as applied and tested in other crops would be also interesting for sunflower (Sankaran et al., 2015). There has been a first report on testing remote sensing on sunflower and maize in China with regard to future applications (Yu and Shang, 2017). In recent years, high throughput genotyping platforms, e.g., SNP arrays, GBS and whole genome sequencing have been established and successfully used in sunflower (Livaja et al., 2016; Qi L. et al., 2016; Talukder et al., 2016). GWAS (genome wide association study) and GS (genomic selection) using large amounts of markers across a wide range of genotypes provided by these techniques open up new possibilities to address complex traits in sunflower. However, GWAS is still expensive and unavailable for many researchers and breeders. Some initial steps have been made in order to create the

most appropriate models for prediction of hybrid performance based on GWAS and GS data (Bonnafous et al., 2016; Mangin et al., 2017a), yet there is still a need for further improvement of prediction models, which mostly take additive effects into account, whereas for heterosis also dominance and epistasis play an important role. As in conventional breeding, species-specific strategies will have to be developed for GS taking into account reproduction system, generation time, genome structure, harvested organs and breeding purposes (Nakaya and Isobe, 2012). However, first empirical GS studies in plants showed the potential for GS also in plant breeding. It could be demonstrated that the correct choice of population allows successful performance of GS even with lower numbers of markers and reasonable sizes of populations (Nakaya and Isobe, 2012).

Access to the recently published sunflower genome sequence (Badouin et al., 2017) should allow researchers and breeders to make sunflower breeding more efficient in the coming years. However, exploring the sunflower genome on its own is not enough. Extensive transcriptomics, proteomics and metabolomics data are required as only the combination of all Omics data will enable us to get to the bottom of some important physiological and molecular mechanisms unique to sunflower. This is especially important for quantitative traits such as drought tolerance or biotic stress resistance (e.g., against *Sclerotinia*, *Phoma*, *Phomopsis*). First results in this direction have been published. Transcriptional profiling has been done with regard to disease reactions of resistant and sensitive genotypes to pathogens as *S. sclerotiorum* (Muellenborn et al., 2011), *Plasmopara halstedii* (Livaja et al., 2013) and *Verticillium dahliae* (Guo et al., 2017). Identification of the differentially expressed genes now allows a better understanding of the mechanisms behind pathogen attacks and plant reactions. This knowledge will be helpful with regard to developing resistant cultivars. Earlier metabolome data of head rot between genotypes with different reactions to *S. sclerotiorum* also gave an indication to 63 metabolites involved in the attack of the pathogen (Peluffo et al., 2010). To analyze the response of sunflower to drought transcriptome analyses of sunflower genotypes under water-limited conditions in comparison to well-water plants have been performed by RNASeq or microarray analyses (Liang et al., 2017; Moschen et al., 2017; Sarazin et al., 2017). Combination of the transcriptomic and metabolic data made the identification of drought relevant hubs for transcription possible (Moschen et al., 2017). Leaf senescence is a naturally occurring process, but the onset and progress of senescence plays a major role for yield. Integration of transcriptomic and metabolomics data identified metabolites and transcription factors as applicable biomarkers (Moschen et al., 2016a,b). To explore the potential of other species in the genus *Helianthus* for sunflower breeding, transcriptomics have also been performed to address populations of, e.g., perennial sunflowers as *H. maximiliani* (Kawakami et al., 2014) and *H. tuberosus* (Jung et al., 2014) as well as interspecific hybrids of annuals in the first generation (Rowe and Rieseberg, 2013). Proteomic analyses in sunflower have been performed with regard to drought stress (Castillejo et al., 2008; Fulda et al., 2011; Ghaffari et al., 2013, 2017), cold acclimation



(Balbuena et al., 2011), response to metal-ion contamination (Garcia et al., 2006; Printz et al., 2013; Lopes Júnior et al., 2015), seed protein composition (De Sousa Barbosa et al., 2013), heterosis performance (Mohayjei et al., 2014), and resistance to *O. cumana* (Yang et al., 2017). In addition, the sunflower genome database represents a very valuable tool, which allows access to a wide range of transcriptome data, which have already been successfully used to address flowering time and oil metabolism (Badouin et al., 2017). However, further studies in sunflower are still needed in order to analyze in detail responses to different abiotic and biotic stress conditions and to prepare sunflower for future climatic challenges. Combining Omics data will allow system biology approaches to improve sunflower hybrids. Another aspect is the optimization of plant architecture to a more compact form, which would have an influence on photosynthesis, lodging, climatic adaptation and possible plant densities. This could also improve sunflower hybrid performance and increase yields per hectare by use of higher plant densities (Hall et al., 2010). Picheny et al. (2017) used the crop model SUNFLO to design sunflower ideotypes with optimized morphological and physiological traits for certain environments.

However, only a combined effort of the sunflower research community can make sunflower more competitive to other oil

crops. The new high-throughput technologies combined with new genomic-based breeding strategies give us the opportunity, as never before, to understand and mine genetic variation and to use it for improvement of sunflower hybrids.

## AUTHOR CONTRIBUTIONS

Both authors AD and RH have made an equal substantial, direct and intellectual contribution to writing the review and its revision, and approved it for publication. The table was prepared by AD, the figure by RH. The authors complied to the ethical standards.

## ACKNOWLEDGMENTS

The authors would like to thank the University of Rostock and the DFG for funding part of the research included (HO 1593/5-1, HO 1593/5-2, and HO 1593/6-1). They are also grateful for the University of Rostock and the DFG for funding the open access publication. Another part of the research presented was supported by project TR31025 (Ministry of Education, Science and Technological Development, R. Serbia).

## REFERENCES

- Abratti, G., Bazzalo, M. E., and León, A. (2008). "Mapping a novel fertility restoration gene in sunflower," in *Proceedings of the 17th International Sunflower Conference*, Córdoba, 617–621.
- Al-Khatib, K., Baumgartner, J. R., and Currie, R. S. (1999). "Survey of common sunflower (*Helianthus annuus*) resistance to ALS inhibiting herbicides in northeast Kansas," in *Proceedings of the 21st Sunflower Research Workshop*, Bismarck, ND, 210–215.
- Al-Khatib, K., Baumgartner, J. R., Peterson, D. E., and Currie, R. S. (1998). Imazethapyr resistance in common sunflower - *Helianthus annuus* L. *Weed Sci.* 46, 403–407.
- Allman-Farinelli, M. A., Gomes, K., Favaloro, E. J., and Petocz, P. (2005). A diet rich in high-oleic-acid sunflower oil favorably alters low-density lipoprotein cholesterol, triglycerides, and factor VII coagulant activity. *J. Am. Diet. Assoc.* 105, 1071–1079. doi: 10.1016/j.jada.2005.04.008
- Arumuganathan, K., and Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9, 208–219. doi: 10.1007/BF02672016
- Atlagic, J., and Terzic, S. (2014). "Sunflower genetic resources – interspecific hybridization and cytogenetics in prebreeding," in *Sunflowers: Growth and Development, Environmental Influences and Pests/Diseases*, ed. J. I. Arribas (New York, NY: Nova Science Publishers), 95–130.
- Bachlava, E., Radwan, O. E., Abratti, G., Tang, S., Gao, W., Heesacker, A. F., et al. (2011). Downy mildew (*Pl8* and *Pl14*) and rust (*RAdv*) resistance genes reside in close proximity to tandemly duplicated clusters of non-TIR-like NBS-LRR-encoding genes on sunflower chromosomes 1 and 13. *Theor. Appl. Genet.* 122, 1211–1221. doi: 10.1007/s00122-010-1525-0
- Bachlava, E., Tang, S., Pizarro, G., and Knapp, S. J. (2009). Pleiotropy of the branching locus (B) masks and unlinked quantitative trait loci affecting seed traits in sunflower. *Theor. Appl. Genet.* 120, 829–842. doi: 10.1007/s00122-009-1212-1
- Bachlava, E., Taylor, C. A., Tang, S., Bowers, J. E., Mandel, J. R., Burke, J. M., et al. (2012). SNP discovery and development of a high-density genotyping array for sunflower. *PLOS ONE* 7:e29814. doi: 10.1371/journal.pone.0029814
- Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., et al. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546, 148–152. doi: 10.1038/nature22380
- Balbuena, T. S., Salsa, J. J., Martinez-Force, E., Garces, R., and Thelen, J. J. (2011). Proteome analysis of cold acclimation in sunflower. *J. Proteome Res.* 10, 2330–2346. doi: 10.1021/pr101137q
- Bandeira e Sousa, M., Cuevas, J., de Oliveira Couto, E. G., Pérez-Rodríguez, P., Jarquin, D., Fritsche-Neto, R., et al. (2017). Genomic-enabled prediction in maize using kernel models with genotype x environment interaction. *G3* 7, 1995–2014. doi: 10.1534/g3.117.042341
- Bandillo, N., Raghvan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C., et al. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* 6:11. doi: 10.1186/1939-8433-6-11
- Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., and Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* sp.). *Plant Sci.* 242, 23–36. doi: 10.1016/j.plantsci.2015.08.021
- Baute, G. J., Owens, G. L., Bock, D. G., and Rieseberg, L. H. (2016). Genome-wide genotyping-by-sequencing data provide a high-resolution view of wild *Helianthus* diversity, genetic structure, and interspecies gene flow. *Am. J. Bot.* 103, 2170–2177. doi: 10.3732/ajb.1600295
- Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci.* 48, 1649–1664. doi: 10.2135/cropsci2008.03.0131
- Berrios, E. F., Gentzbittel, L., Alibert, G., Griveau, Y., Bervillé, A., and Sarrafi, A. (1999). Genetic control of in vitro-organogenesis in recombinant inbred lines of sunflower (*Helianthus annuus* L.). *Plant Breed.* 118, 359–361. doi: 10.1046/j.1439-0523.1999.00370.x
- Berry, S. T., León, A. J., Hanfrey, C. C., Challis, P., Burkholz, A., Barnes, S. R., et al. (1995). Molecular marker analysis of *Helianthus annuus* L. 2. Construction of an RFLP linkage map for cultivated sunflower. *Theor. Appl. Genet.* 91, 195–199. doi: 10.1007/BF00220877
- Bert, P. F., Tourvieille de Labrouhe, D., Philippon, J., Mouzeyar, S., Jouan, I., Nicolas, P., et al. (2001). Identification of a second linkage group carrying genes controlling resistance to downy mildew (*Plasmopara halstedii*) in sunflower



- (*Helianthus annuus* L.). *Theor. Appl. Genet.* 103, 992–997. doi: 10.1007/s001220100660
- Bervillé, A. (2010). “Oil composition variations,” in *Genetics, Genomics and Breeding of Sunflower*, eds J. Hu, G. Seiler, and C. Kole (Boca Raton, FL: Science Publisher), 253–277.
- Bilgen, B. (2016). “Characterization of sunflower inbred lines with high oleic acid content by DNA markers,” in *Proceedings of the 19th International Sunflower Conference*, Edirne, 662–668.
- Blackman, B. K., Rasmussen, D. A., Strasburg, J. L., Raduski, A. R., Burke, J. M., Knapp, S. J., et al. (2011). Contribution of flowering time genes to sunflower domestication and improvement. *Genetics* 187, 271–287. doi: 10.1534/genetics.110.121327
- Bohra, A., Jha, U. C., Adhimoolam, P., Bisht, D., and Singh, N. P. (2016). Cytoplasmic male sterility (CMS) in hybrid breeding in field crops. *Plant Cell Rep.* 35, 967–993. doi: 10.1007/s00299-016-1949-3
- Bonnafeous, F., Langdale, N., Sunrise Consortium and Mangin, B. (2016). “Inclusion of dominance effect in genomic selection model to improve predictive ability for sunflower hybrid performance,” in *Proceedings of the 19th International Sunflower Conference*, Edirne, 285.
- Bouzidi, M. F., Badaoui, S., Cambon, F., Vear, F., De Labrouhe, D., Nicolas, P., et al. (2002). Molecular analysis of a major locus for resistance to downy mildew in sunflower with specific PCR-based markers. *Theor. Appl. Genet.* 104, 592–600. doi: 10.1007/s00122-001-0790-3
- Bowers, J. E., Bachlava, E., Brunick, R. L., Knapp, S. J., and Burke, J. M. (2012). Development of a 10,000 locus genetic map of the sunflower genome based on multiple crosses. *G3* 2, 721–729. doi: 10.1534/g3.112.002659
- Breccia, G., Gil, M., Vega, T., Zorzoli, R., Picardi, L., and Nestares, G. (2011). Rapid test for detection of imidazolinone resistance in sunflower (*Helianthus annuus* L.). *Plant Breed.* 130, 109–113. doi: 10.1111/j.1439-0523.2009.01756.x
- Brouillette, L. C., Rosenthal, D. M., Rieseberg, L. H., Lexer, C., Malmberg, R. L., and Donovan, L. A. (2007). Genetic architecture of leaf ecophysiological traits in *Helianthus*. *J. Hered.* 98, 141–146. doi: 10.1093/jhered/esl063
- Brunel, D. (1994). A microsatellite marker in *Helianthus annuus* L. *Plant Mol. Biol.* 24, 397–400. doi: 10.1007/BF00020177
- Bruniard, J. M., and Miller, J. F. (2001). Inheritance of imidazolinone herbicide resistance in sunflower. *Helia* 24, 11–16.
- Bulos, M., Ramos, M. L., Altieri, E., and Sala, C. A. (2013a). Molecular mapping of a sunflower rust resistance gene from HAR6. *Breed. Sci.* 63, 141–146. doi: 10.1270/jsbbs.63.141
- Bulos, M., Sala, C. A., Altieri, E., and Ramos, M. L. (2013b). Marker assisted selection for herbicide resistance in sunflower. *Helia* 59, 1–16. doi: 10.2298/HEL1359001B
- Bulos, M., Vergani, P. N., and Altieri, E. (2014). Genetic mapping, marker assisted selection and allelic relationships for the *P<sub>u6</sub>* gene conferring rust resistance in sunflower. *Breed. Sci.* 64, 206–212. doi: 10.1270/jsbbs.64.206
- Cadic, E., Coque, M., Vear, F., Grezes-Besset, B., Puaquet, J., Piquemal, J., et al. (2013). Combined linkage and association mapping of flowering time in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 126, 1337–1356. doi: 10.1007/s00122-013-2056-2
- Cantelmo, F. N., Von Pinho, R. G., and Balestre, M. (2017). Genome-wide prediction for maize single-cross hybrids using the GBLUP model and validation in different crop seasons. *Mol. Breed.* 37:51. doi: 10.1007/s11032-017-0651-7
- Castillejo, M. A., Maldonado, A. M., Ogueta, S., and Jorin, J. V. (2008). Proteomic analysis of responses to drought stress in sunflower (*Helianthus annuus*) leaves by 2DE gel electrophoresis and mass spectrometry. *Open Proteomics J.* 1, 59–71. doi: 10.2174/1875039700801010059
- Cavanagh, C., Morell, M., Mackay, I., and Powell, W. (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol.* 11, 215–221. doi: 10.1016/j.pbi.2008.01.002
- Celik, I., Bodur, S., Frary, A., and Doganlar, S. (2016). Genome-wide SNP discovery and genetic linkage map construction in sunflower (*Helianthus annuus* L.) using a genotyping by sequencing (GBS) approach. *Mol. Breed.* 36:133. doi: 10.1007/s11032-016-0558-8
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.125343
- Chapman, M. A., Pashley, C. H., Wenzler, J., Hvala, J., Tang, S., Knapp, S. J., et al. (2008). A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus* L.). *Plant Cell* 20, 2931–2945. doi: 10.1105/tpc.108.059808
- Cvejić, S., Afza, R., Jocić, S., Prodanović, S., Miklić, V., Skoric, D., et al. (2011a). Radiosensitivity of sunflower inbred lines to mutagenesis. *Helia* 34, 99–106. doi: 10.2298/HEL1154099C
- Cvejić, S., Jocić, S., Prodanović, S., Terzić, S., Miladinović, D., and Balalić, I. (2011b). Creating new genetic variability in sunflower using induced mutations. *Helia* 34, 47–55. doi: 10.2298/HEL1155047C
- de Azevedo Peixoto, L., Moellers, T. C., Zhang, J., Lorenz, A. J., Bhering, L. L., Beavis, W. D., et al. (2017). Leveraging genomic prediction to scan germplasm collection crop improvement. *PLOS ONE* 12:e0179191. doi: 10.1371/journal.pone.0179191
- de Bertero, R. A. (2003). Origin of the Argentine sunflower. *Helia* 26, 127–136. doi: 10.2298/HEL0338127d
- De Sousa Barbosa, H., Quirino de Souza, D. L., Koolen, H. H. F., Gozzo, F. C., and Arruda, M. A. Z. (2013). Sample preparation focusing on plant proteomics: extraction, evaluation and identification of proteins from sunflower seeds. *Anal. Methods* 5, 116–123. doi: 10.1039/C2AY25503K
- Debaeke, P., Casadebaig, P., Flenet, F., and Langlade, N. (2017). Sunflower crop and climate change: vulnerability, adaptation, and mitigation potential from case-studies in Europe. *OCL* 24:D102. doi: 10.1051/ocl/2016052
- Dehmer, K. J., and Friedt, W. (1998). Development of molecular markers for high oleic acid content in sunflower (*Helianthus annuus* L.). *Ind. Crops Prod.* 7, 311–315. doi: 10.1016/S0926-6690(97)00063-0
- Dimitrijevic, A., Imerovski, I., Miladinovic, D., Cvejić, S., Jocić, S., Zeremski, T., et al. (2017). Oleic acid variation and marker-assisted detection of Pervenets mutation in high- and low-oleic sunflower cross. *Crop Breed. Appl. Biotechnol.* 17, 229–235. doi: 10.1590/1984-70332017v17n3a36
- Dimitrijevic, A., Imerovski, I., Miladinovic, D., Jocić, S., and Miklić, V. (2011). “Use of molecular markers in identification of non-TIR-NBS-LRR RGA linked to downy mildew resistance locus *P<sub>15</sub>/P<sub>18</sub>* in sunflower (*Helianthus annuus* L.),” in *Proceedings of the 19th Symposium of the Serbian Plant Physiology Society*, Banja Vrujci, 80.
- Dimitrijevic, A., Imerovski, I., Miladinovic, D., Jockovic, M., Cvejić, S., Jocić, S., et al. (2016). “Screening of the presence of *ol* gene in NS sunflower collection,” in *Proceedings of the 19th International Sunflower Conference*, Edirne, 661–667.
- Dimitrijevic, A., Imerovski, I., Miladinovic, D., Tancic, S., Dušanic, N., Jocić, S., et al. (2010). Use of SSR markers in identification of sunflower isogenic lines in late generations of backcrossing. *Helia* 33, 191–198. doi: 10.2298/hel1053191d
- Duflle, C. M., Hahn, V., Knapp, S. J., and Bauer, E. (2004). *Plarg* from *Helianthus argophyllus* is unlinked to other known downy mildew resistance genes in sunflower. *Theor. Appl. Genet.* 109, 1083–1086. doi: 10.1007/s00122-004-1722-9
- FAO (2016). FAOSTAT. Rome: FAO. Available at: <http://www.fao.org/faostat/en/#data/QC> (accessed October 5, 2016).
- Feng, J., and Jan, C.-C. (2008). Introgression and molecular tagging of *Rf<sub>4</sub>*, a new male fertility restoration gene from wild sunflower *Helianthus maximiliani* L. *Theor. Appl. Genet.* 117, 241–249. doi: 10.1007/s00122-008-0769-4
- Feng, J., Liu, Z., Cai, X., and Jan, C.-C. (2013). Towards a molecular cytogenetic map for cultivated sunflower (*Helianthus annuus* L.) by landed BAC/BIBAC clones. *G3* 3, 31–40. doi: 10.1534/g3.112.004846
- Ferfuaia, C., Turi, M., and Vannozzi, G. P. (2015). Variability of seed fatty acid composition to growing degree-days in high oleic acid sunflower genotypes. *Helia* 38, 61–78. doi: 10.1515/helia-2014-0022
- Ferfuaia, C., and Vannozzi, G. P. (2015). Maternal effect on seed fatty acid composition in a reciprocal cross of high oleic sunflower (*Helianthus annuus* L.). *Euphytica* 205, 325–336. doi: 10.1007/s10681-015-1378-3
- Fick, G. N., and Miller, J. F. (1997). “Sunflower breeding,” in *Sunflower Production and Technology*, ed. A. A. Scheiter (Madison, WI: ASA), 395–440.
- Filippi, C. V., Aguirre, N., Rivas, J. G., Zubrzycki, J., Puebla, A., Cordes, D., et al. (2015). Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. *BMC Plant Biol.* 15:52. doi: 10.1186/s12870-014-0360-x

- Fulda, S., Mikkat, S., Stegmann, H., and Horn, R. (2011). Physiology and proteomics of drought stress acclimation in sunflower (*Helianthus annuus* L.). *Plant Biol.* 13, 632–642. doi: 10.1111/j.1438-8677.2010.00426
- Fusari, C. M., Di Rienzo, J. A., Troglia, C., Nishinakamasu, V., Moreno, M. V., Maringolo, C., et al. (2012). Association mapping in sunflower for *Sclerotinia* head rot resistance. *BMC Plant Biol.* 12:93. doi: 10.1186/1471-2229-12-93
- Fusari, C. M., Lia, V. V., Hopp, H. E., Heinz, R. A., and Paniego, N. B. (2008). Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in sunflower elite inbred lines using the candidate gene approach. *BMC Plant Biol.* 8:7. doi: 10.1186/1471-2229-8-7
- Gabard, J. M., and Huby, J. P. (2001). Sulfonylurea-tolerant sunflower plants. U.S. Patent No. 20050044587.
- Gandhi, S. D., Heesacker, A. F., Freeman, C. A., Argyris, J., Bradford, K., and Knapp, S. J. (2005). The self-incompatibility locus (S) and quantitative trait loci for self-pollination and seed dormancy in sunflower. *Theor. Appl. Genet.* 111, 619–629. doi: 10.1007/s00122-005-1934-7
- Garcia, J. S., Gratao, P. L., Azevedo, R. A., and Arruda, M. A. Z. (2006). Metal contamination effects on sunflower (*Helianthus annuus* L.) growth and protein expression in leaves during development. *J. Agric. Food Chem.* 54, 8623–8630. doi: 10.1012/jf061593
- Gavrilo, V. A., Rozhkova, V. T., and Anisimova, I. N. (2014). Sunflower genetic collection at the Vavilov institute of plant industry. *Helia* 37, 1–16. doi: 10.1515/helia-2014-0001
- Gedil, M. A., Wye, C., Berry, S. T., Seger, B., Peleman, J., Jones, R., et al. (2001). An integrated RFLP-AFLP linkage map for cultivated sunflower. *Genome* 44, 213–221. doi: 10.1139/g00-111
- Gentzittel, L., Vear, F., Zhang, Y. X., and Berville, A. (1995). Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 90, 1079–1086. doi: 10.1007/BF00222925
- Ghaffari, M., Toorchi, M., Valizadeh, M., and Komatsu, S. (2013). Differential response of root proteome to drought stress in drought sensitive and tolerant sunflower inbred lines. *Funct. Plant Biol.* 40, 609–617. doi: 10.1071/FP12251
- Ghaffari, M., Toorchi, M., Valizadeh, M., and Shakiba, M. (2017). Proteomic prospects for tolerance of sunflower (*Helianthus annuus*) to drought stress during the flowering stage. *Crop Pasture Sci.* 68, 457–465. doi: 10.1071/CP17105
- Gilley, M. A., Misar, C. G., Gulya, T. J., and Markell, S. G. (2016). “Prevalence and virulence of *Plasmopara halstedii* (downy mildew) in sunflowers,” in *Proceedings of the 38th Sunflower Research Forum*, Fargo, ND.
- Gong, L., Guyla, T. J., Markell, S. G., Hulke, B. S., and Qi, L. L. (2013a). Genetic mapping of rust resistance genes in confection sunflower line HAR6 and oilseed line RHA 397. *Theor. Appl. Genet.* 126, 2039–2049. doi: 10.1007/s00122-013-2116-7
- Gong, L., Hulke, B. S., Gulya, T. J., Markell, S. G., and Qi, L. L. (2013b). Molecular tagging of a novel rust resistance gene  $R_{12}$  in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 126, 93–99. doi: 10.1007/s00122-012-1962-z
- Grassa, C. J., Ebert, D. P., Kane, N. C., and Rieseberg, L. H. (2016). Complete mitochondrial genome sequence of sunflower (*Helianthus annuus* L.). *Genome Announc.* 4:e00981-16. doi: 10.1128/genomeA.00981-16
- Guimaraes, R. L., and Stotz, H. U. (2004). Oxalate production by *Sclerotinia sclerotiorum* deregulates guard cells during infection. *Plant Physiol.* 136, 3703–3711. doi: 10.1104/pp.104.049650
- Gulya, T., Rashid, K. Y., and Masirevic, S. N. (1997). “Sunflower diseases,” in *Sunflower Technology and Production*, ed. A. A. Schneiter (Madison, WI: ASA), 263–379.
- Guo, S., Zuo, Y., Zhang, Y., Wu, C., Su, C., Jin, W., et al. (2017). Large-scale transcriptome comparison of sunflower genes responsive to *Verticillium dahliae*. *BMS Genomics* 18:42. doi: 10.1186/s12864-016-3386-7
- Guo, X., and Stotz, H. U. (2007). Defense against *Sclerotinia sclerotiorum* in *Arabidopsis* is dependent on jasmonic acid (JA), salicylic acid (SA), and ethylene (ET) signaling. *Mol. Plant Microbe Interact.* 20, 1384–1395. doi: 10.1094/MPMI-20-11-1384
- Habyarimana, E., Parisi, B., and Mandolino, G. (2017). Genomic prediction for yields, processing and nutritional quality traits in cultivated potato (*Solanum tuberosum* L.). *Plant Breed.* 136, 245–252. doi: 10.1111/pbr.12461
- Hall, A. J., Sporsaro, M. M., and Chimenti, C. A. (2010). Stem lodging in sunflower: variations in stem failure moment of force and structure across crop population densities and post-anthesis developmental stages in two genotypes of contrasting susceptibility to lodging. *Field Crops Res.* 116, 46–51. doi: 10.1016/j.fcr.2009.11.008
- Heesacker, A., Kishore, V. K., Gao, W., Tang, S., Kolkman, J. M., Gingle, A., et al. (2008). SSRs and INDELs mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. *Theor. Appl. Genet.* 117, 1021–1029. doi: 10.1007/s00122-008-0841-0
- Horn, R., Gupta, J. K., and Colombo, N. (2014). Mitochondrion role in molecular basis of cytoplasmic male sterility. *Mitochondrion* 19, 198–205. doi: 10.1016/j.mito.2014.04.004
- Horn, R., Köhler, R., and Zetsche, K. (1991). A mitochondrial 16 kDa protein is associated with cytoplasmic male sterility in sunflower. *Plant Mol. Biol.* 17, 29–36. doi: 10.1007/BF00036803
- Horn, R., Kusterer, B., Lazarescu, E., Prüfe, M., and Friedt, W. (2003). Molecular mapping of the  $Rf_1$  gene restoring pollen fertility in PET1-based F1-hybrids in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 106, 599–606. doi: 10.1007/s00122-002-1078-y
- Horn, R., Reddemann, A., and Drumeva, M. (2016). “Comparison of cytoplasmic male sterility based on PET1 and PET2 cytoplasm in sunflower (*Helianthus annuus* L.),” in *Proceedings of the 19th International Sunflower Conference*, Edirne, 620–629.
- Hu, J. (2010). “Genetic linkage maps: strategies, resources and achievements genetics,” in *Genomics and Breeding of Sunflower*, eds J. Hu, G. Seiler, and C. Kole (Enfield, NH: Science Publishers), 79–109. doi: 10.1201/b10192-4
- Hu, J., Yue, B., and Vick, B. A. (2007). Integration of trap markers onto a sunflower SSR marker linkage map constructed from 92 recombinant inbred lines. *Helia* 30, 25–36. doi: 10.2298/HEL0746025H
- Imerovski, I., Dimitrijevic, A., Miladinovic, D., Dedic, B., Jovic, S., and Cvejic, S. (2014a). “Preliminary SSR analysis of a novel broomrape resistance source,” in *Proceedings of the Third International Symposium on Broomrape in Sunflower*, Cordoba, 217–221.
- Imerovski, I., Dimitrijevic, A., Miladinovic, D., Dedic, B., Jovic, S., Kovacevic, B., et al. (2013). Identification of PCR markers linked to different *Or* genes in sunflower. *Plant Breed.* 132, 115–120. doi: 10.1111/pbr.12022
- Imerovski, I., Dimitrijevic, A., Miladinovic, D., Dedic, B., Jovic, S., Tubic, N. K., et al. (2016). Mapping of a new gene for resistance to broomrape races higher than F. *Euphytica* 209, 281–289. doi: 10.1007/s10681-015-1597-7
- Imerovski, I., Dimitrijevic, A., Miladinovic, D., Jovic, S., Dedic, B., Cvejic, S., et al. (2014b). Identification and validation of breeder-friendly DNA markers for *Plarg* gene in sunflower. *Mol. Breed.* 34, 779–788. doi: 10.1007/s11032-014-0074-7
- Iuoras, M., Stanciu, D., Ciuca, M., Nastase, D., and Geronzi, F. (2004). Preliminary studies related to the use of marker assisted selection for resistance of *Orobanche cumana* Wallr. in sunflower. *Rom. Agric. Res.* 21, 33–37.
- Jacob, J., Mulpuri, S., and Sivannarayana Kodeboyina, V. (2017). Screening of cultivated and wild *Helianthus* species reveals herbicide tolerance in wild sunflower and allelic variation at Ahas11 (acetohydroxyacid synthase 1 large subunit) locus. *Plant Genet. Res. C* 15, 421–429. doi: 10.1017/S1479262211600095
- Jan, C. C., and Vick, B. A. (2007). Inheritance and allelic relationship of fertility restoration genes for seven new sources of male-sterile cytoplasm in sunflower. *Plant Breed.* 126, 213–217. doi: 10.1111/j.1439-0523.2007.01350.x
- Jan, C. C., Vick, B. A., Miller, J. K., Kahler, A. L., and Butler, E. T. I. (1998). Construction of a RFLP linkage map for cultivated sunflower. *Theor. Appl. Genet.* 96, 15–22. doi: 10.1007/s001220050703
- Jovic, S., Cvejic, S., Hladni, N., Miladinovic, D., and Miklic, V. (2010). Development of sunflower genotypes resistant to downy mildew. *Helia* 33, 173–180. doi: 10.2298/hel1053173j
- Jocić, S., Malidža, G., and Škorić, D. (2004). Sunflower tolerant to imidazoline herbicides. *J. Sci. Agric. Res.* 229, 81–89.
- Jocic, S., Miladinovic, D., and Kaya, Y. (2015). “Breeding and genetics of sunflower,” in *Sunflower: Chemistry, Production, Processing, and Utilization*, eds E. Martinez-Force, N. T. Dunford, and J. J. Salas (Urbana, IL: AOCS), 1–26. doi: 10.2298/HEL0644033J
- Joksimovic, J., Atlagic, J., Marinkovic, R., and Jovanovic, D. (2006). Genetic control of oleic and linoleic acid contents in sunflower. *Helia* 29, 33–40. doi: 10.2298/hel0644033j

- Jung, W. Y., Lee, S. S., Kim, C. W., Kim, H. S., Min, S. R., Moon, J. S., et al. (2014). RNA-seq analysis and *de novo* transcriptome assembly of Jerusalem artichoke (*Helianthus tuberosus* Linne). *PLOS ONE* 9:e111982. doi: 10.1371/journal.pone.0111982
- Kane, N. C., Burke, J. M., Marek, L., Seiler, G., Vear, F., Baute, G., et al. (2013). Sunflower genetic, genomic and ecological resources. *Mol. Ecol. Resour.* 13, 10–20. doi: 10.1111/1755-0998.12023
- Kawakami, T., Darby, B. J., and Ungerer, M. C. (2014). Transcriptome resources for the perennial sunflower *Helianthus maximiliani* obtained from ecologically divergent populations. *Mol. Ecol. Resour.* 14, 812–819. doi: 10.1111/1755-0998.12227
- Kim, S.-C., and Rieseberg, L. H. (1999). Genetic architecture of species differences in annual sunflowers: implication for adaptive trait introgression. *Genetics* 153, 965–977.
- Kinman, M. L. (1970). “New developments in the USDA and state experiment station sunflower breeding programs,” in *Proceedings of the 4th International Sunflower Conference*, Memphis, TN, 181–183.
- Köhler, R. H., Horn, R., Lössl, A., and Zetsche, K. (1991). Cytoplasmic male sterility in sunflower is correlated with the co-transcription of a new open reading frame with the *atpA* gene. *Mol. Gen. Genet.* 227, 369–376. doi: 10.1534/genetics.107.074054
- Kolkman, J. M., Slabaugh, M. B., Bruniard, J. M., Berry, S., Bushman, B. S., Olungu, C., et al. (2004). Acetohydroxyacid synthase mutations conferring resistance to imidazolinone or sulfonylurea herbicides in sunflower. *Theor. Appl. Genet.* 109, 1147–1159. doi: 10.1007/s00122-004-1716-7
- Kolkman, J. M., Berry, S. T., Leon, A. J., Slabaugh, M. B., Tang, S., Gao, W., et al. (2007). Single nucleotide polymorphisms and linkage disequilibrium in sunflower. *Genetics* 177, 457–468. doi: 10.1534/genetics.107.074054
- Korell, M., Mösges, G., and Friedt, W. (1992). Construction of a sunflower pedigree map. *Helia* 15, 7–16.
- Kumar, A. P. K., Boualem, A., Bhattacharya, A., Parikh, S., Desai, N., Zambelli, A., et al. (2013). SMART – Sunflower mutant population and reverse genetic tools for crop improvement. *BMC Plant Biol.* 13:38. doi: 10.1186/1471-2229-13-38
- Kusterer, B., Horn, R., and Friedt, W. (2005). Molecular mapping of the fertility restoration locus *Rf<sub>1</sub>* in sunflower and development of diagnostic markers for the restorer gene. *Euphytica* 143, 35–43. doi: 10.1007/s10681-005-1795-9
- Lacombe, S., Kaan, F., Griveau, Y., and Bervillé, A. (2004). The Pervenets high oleic mutation: methodological studies. *Helia* 27, 41–54. doi: 10.2298/hel0440041l
- Lacombe, S., Leger, S., Kaan, F., Bervillé, A., and Sas, M. (2002). Genetic, molecular and expression features of the Pervenets mutant leading to high oleic acid content of seed oil in sunflower. *OCL* 9, 17–23. doi: 10.1051/ocl.2002.0017
- Lacombe, S., Souyris, I., and Bervillé, A. J. (2009). An insertion of oleate desaturase homologous sequence silences via siRNA the functional gene leading to high oleic acid content in sunflower seed oil. *Mol. Genet. Genomics* 281, 43–54. doi: 10.1007/s00438-008-0391-9
- Lai, Z., Livingstone, K., Zou, Y., Church, S. A., Knapp, S. J., Andrews, J., et al. (2005). Identification and mapping of SNPs from ESTs in sunflower. *Theor. Appl. Genet.* 111, 1532–1544. doi: 10.1007/s00122-005-0082-4
- Lawson, W. R., Goulter, K. C., Henry, R. J., Kong, G. A., and Kochman, J. K. (1996). RAPD markers for a sunflower rust resistance gene. *Aust. J. Agric. Res.* 47, 395–401. doi: 10.1071/AR9960395
- Lawson, W. R., Goulter, K. C., Henry, R. J., Kong, G. A., and Kochman, J. K. (1998). Marker-assisted selection for two rust resistance genes in sunflower. *Mol. Breed.* 4, 227–234. doi: 10.1023/A:1009667112088
- Leclercq, P. (1969). Une stérilité mâle cytoplasmique chez le tournesol. *Ann. Amélior. Plant.* 19, 99–106.
- Leclercq, P. (1971). La stérilité mâle cytoplasmique du tournesol. 1. Premières études sur la restauration de la fertilité. *Ann. Amélior. Plant.* 21, 45–54.
- Leclercq, P. (1984). Identification de gènes de restauration de fertilité sur cytoplasmic stérilisants chez le tournesol. *Agronomie* 4, 573–576. doi: 10.1051/agro:19840611
- Leon, A. J., Andrade, F. H., and Lee, M. (2000). Genetic mapping of factors affecting quantitative variation for flowering in sunflower. *Crop Sci.* 40, 404–407. doi: 10.2135/cropsci2000.402404x
- Leon, A. J., Zambelli, A. D., Reid, R. J., Morat, M. M., and Kaspar, M. (2013). Nucleotide sequences mutated by insertion that encode a truncated oleate desaturase protein, proteins, methods and uses. Patent No. WO 2013/004281 A1.
- Liang, C., Wang, W., Wang, J., Ma, J., Li, C., Zhou, F., et al. (2017). Identification of differentially expressed genes in sunflower (*Helianthus annuus*) leaves and roots under drought stress by RNA sequencing. *Bot. Stud.* 58:42. doi: 10.1186/s40529-017-0197-3
- Liu, A., and Burke, J. M. (2006). Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* 173, 321–330. doi: 10.1534/genetics.105.051110
- Liu, Z., Gulya, T. J., Seiler, G. J., Vick, B. A., and Jan, C. C. (2012a). Molecular mapping of the Pl(16) downy mildew resistance gene from HA-R4 to facilitate marker-assisted election in sunflower. *Theor. Appl. Genet.* 125, 121–131. doi: 10.1007/s00122-012-1820-z
- Liu, Z., Mulpuri, S., Feng, J., Vick, B. A., and Jan, C. C. (2012b). Molecular mapping of the *Rf<sub>3</sub>* fertility restoration gene to facilitate its utilization in breeding confection sunflower. *Mol. Breed.* 29, 275–284. doi: 10.1007/s11032-011-9563-0
- Liu, Z., Wang, D. M., Feng, J., Seiler, G. J., Cai, X., and Jan, C. C. (2013). Diversifying sunflower germplasm by integration and mapping of a novel male fertility restoration gene. *Genetics* 193, 727–737. doi: 10.1534/genetics.112.146092/-/DC1
- Livaja, M., Unterseer, S., Erath, W., Lehermeier, C., Wieseke, R., Plieske, J., et al. (2016). Diversity analysis and genomic prediction of *Sclerotinia* resistance in sunflower using a new 25 K SNP genotyping array. *Theor. Appl. Genet.* 129, 317–329. doi: 10.1007/s00122-015-2629-3
- Livaja, M., Wang, Y., Wieckhorst, S., Haseneyer, G., Seidel, M., Hahn, V., et al. (2013). BSTA: a targeted approach combines bulked segregant analysis with next-generation sequencing and de novo transcriptome assembly for SNP discovery in sunflower. *BMC Genomics* 14:628. doi: 10.1186/1471-2164-14-628
- Longin, C. F. H., and Reif, J. C. (2014). Redesigning the exploitation of wheat genetic resources. *Trends Plant Sci.* 19, 631–636. doi: 10.1016/j.tplants.2014.06.012
- Lopes Júnior, C. A., Barbosa Hde, S., Moretto Galazzi, R., Ferreira Koolen, H. H., Gozzo, F. C., and Arruda, M. A. (2015). Evaluation of proteome alterations induced by cadmium stress in sunflower (*Helianthus annuus* L.) cultures. *Ecotoxicol. Environ. Saf.* 119, 170–177. doi: 10.1016/j.ecoenv.2015.05.016
- Louarn, J., Boniface, M.-C., Pouilly, N., Velasco, L., Pérez-Vich, B., Vincourt, P., et al. (2016). Sunflower resistance to broomrape (*Orobanche cumana*) is controlled by specific QTLs for different parasitism stages. *Front. Plant Sci.* 7:590. doi: 10.3389/fpls.2016.00590
- Lu, Y. H., Melero-Vara, J. M., Garcia-Tejada, J. A., and Blanchard, P. (2000). Development of SCAR markers linked to the gene *Or5* conferring resistance to broomrape (*Orobanche cumana* Wallr.) in sunflower. *Theor. Appl. Genet.* 100, 625–632. doi: 10.1007/s001220050083
- Lyra, D. H., Mendonça, L. F., Galli, G., Alves, F. C., Granato, I. S. C., and Fritsche-Neto, R. (2017). Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. *Mol. Breed.* 37:80. doi: 10.1007/s11032-017-0681-1
- Ma, G. J., Markell, S. G., Song, Q. J., and Qi, L. L. (2017). Genotyping-by-sequencing targeting of a novel downy mildew resistance gene *Pl<sub>20</sub>* from wild *Helianthus argophyllus* for sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 130, 1519–1529. doi: 10.1007/s00122-017-2906-4
- Mammadov, J., Aggarwal, R., Buyyarapu, R., and Kumpatla, S. (2012). SNP markers and their impact on plant breeding. *Int. J. Plant Genomics* 2012:728398. doi: 10.1155/2012/728398
- Mandel, J. R., Dechaine, J. M., Marek, L. F., and Burke, J. M. (2011). Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. *Theor. Appl. Genet.* 123, 693–704. doi: 10.1007/s00122-011-1619-3
- Mandel, J. R., Nambeesan, S., Bowers, J. E., Marek, L., Ebert, D., Rieseberg, L. H., et al. (2013). Association mapping and the genomic consequences of selection in sunflower. *PLOS Genet.* 9:e1003378. doi: 10.1371/journal.pgen.1003378
- Mangin, B., Bonnafous, F., Blanchet, N., Boniface, M.-C., Bret-Mestries, E., Carrère, S., et al. (2017a). Genomic prediction of sunflower hybrids oil content. *Front. Plant Sci.* 8:1633. doi: 10.3389/fpls.2017.01633
- Mangin, B., Pouilly, N., Boniface, M.-C., Langdale, N. B., Vincourt, P., Vear, F., et al. (2017b). Molecular diversity of sunflower populations maintained as genetic resources is affected by multiplication processes and breeding for major traits. *Theor. Appl. Genet.* 130, 1099–1112. doi: 10.1007/s00122-017-2872-x



- Marek, L. F. (2016). "Sunflower genetic resources," in *Proceedings of the 19th International Sunflower Conference*, Edirne, 31–44.
- Markin, N., Usatov, A., Makarenko, M., Azarin, K., Gorbachenko, O., Kolokolova, N., et al. (2017). Study of informative DNA markers of the *Rf1* gene in sunflower for breeding practice. *Czech J. Plant Breed.* 53, 69–75. doi: 10.17221/108/2016-CJGPB
- Massman, J. M., Jung, H.-J. G., and Bernardo, R. (2013). Genome wide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. *Crop Sci.* 53, 58–66. doi: 10.2135/cropsci2012.02.0112
- McAssey, E., Corbi, J., and Burke, J. M. (2016). Range-wide phenotypic and genetic differentiation in wild sunflower. *BMC Plant Biol.* 16:249. doi: 10.1186/s12870-016-0937-7
- Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic values using genome-wide dense marker maps. *Genetics* 1157, 1819–1829.
- Miller, F. J., and Al-Khatib, K. (2000). "Development of herbicide resistant germplasm in sunflower," in *Proceedings of the 15th International Sunflower Conference*, Vol. 2, Toulouse, 12–15.
- Miller, J. F., and Al-Khatib, K. (2002). Registration of imidazolinone herbicide-resistant sunflower maintainer (HA 425) and fertility restorer (RHA 426 and RHA 427) germplasms. *Crop Sci.* 42, 988–989. doi: 10.2135/cropsci2002.988a
- Mohayejji, M., Capriotti, A. L., Cavaliere, C., Piovesana, S., Samperi, R., Stampachiachiere, S., et al. (2014). Heterosis profile of sunflower leaves: a label free proteomics approach. *J. Proteomics* 99, 101–110. doi: 10.1016/j.jpro.2014.01.028
- Molinero-Ruiz, L., García-Carneros, A. B., Collado-Romero, M., Raranciuc, S., Domínguez, J., and Melero-Vara, J. M. (2014). Pathogenic and molecular diversity in highly virulent populations of the parasitic weed *Orobanche cumana* (sunflower broomrape) from Europe. *Weed Res.* 54, 87–96. doi: 10.1111/wre.12056
- Molinero-Ruiz, M. L., García-Ruiz, R., Melero-Vara, J. M., and Domínguez, J. (2009). *Orobanche cumana* race F: performance of resistant sunflower hybrids and aggressiveness of populations of the parasitic weed. *Weed Res.* 49, 469–478. doi: 10.1111/j.1365-3180.2009.00708.x
- Monéger, F., Smart, C. J., and Leaver, C. J. (1994). Nuclear restoration of cytoplasmic male sterility in sunflower is associated with the tissue-specific regulation of a novel mitochondrial gene. *EMBO J.* 13, 8–17. doi: 10.1102/j.1460-2075.1994.tb06230x
- Moreno, M. V., Nishinakamasu, V., Loray, M. A., Alvarez, D., Gieco, J., Vicario, A., et al. (2013). Genetic characterization of sunflower breeding resources from Argentina: assessing diversity in key open-pollinated and composite populations. *Plant Genet. Resour.* 11, 238–249. doi: 10.1017/S1479262113000075
- Moschen, S., Bengoa Luoni, S., Di Rienzo, J. A., Caro Mdel, P., Watanabe, M., Hollmann, J., et al. (2016a). Integrating transcriptomic and metabolomic analysis to understand natural leaf senescence in sunflower. *Plant Biotechnol. J.* 14, 719–734. doi: 10.1111/pbi.12422
- Moschen, S., Di Rienzo, J. A., Higgins, J., Tohge, T., Watanabe, M., Gonzalez, S., et al. (2017). Integration of transcriptomic and metabolic data reveals hub transcription factors involved in drought stress response in sunflower (*Helianthus annuus* L.). *Plant Mol. Biol.* 94, 549–564. doi: 10.1007/s11103-017-0625-5
- Moschen, S., Higgins, J., Di Rienzo, J. A., Heinz, R. A., Paniego, N., and Fernandez, P. (2016b). Integration of sunflower transcriptomic and metabolic data by network and biosignature analysis. *BMC Bioinformatics* 17 (Suppl. 5):174. doi: 10.1186/s12859-016-1045-2
- Mouzeyar, S., Roeckel-Drevet, P., Gentzbittel, L., Philippon, J., De Labrouhe, D. T., Vear, F., et al. (1995). RFLP and RAPD mapping of the sunflower *Pl1* locus for resistance to *Plasmopara halstedii* race 1. *Theor. Appl. Genet.* 91, 733–737. doi: 10.1007/BF00220951
- Muellerborn, C., Krause, J.-H., and Cerboncini, C. (2011). Analysis of differential transcript expression reveals time-dependent leaf responses to *Sclerotinia sclerotiorum* in wild and cultivated sunflower. *Plant Mol. Biol. Rep.* 29, 597–608. doi: 10.1007/s11105-010-0265-2
- Mulpuri, S., Liu, Z., Feng, J., Gulya, T. J., and Jan, C. C. (2009). Inheritance and molecular mapping of a downy mildew resistance gene, *P13* in cultivated sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 119, 795–803. doi: 10.1007/s00122-009-1089-z
- Nagarathna, T. K., Shadakshari, Y. G., and Ramanappa, T. M. (2011). Molecular analysis of sunflower (*Helianthus annuus* L.) genotypes for high oleic acid using microsatellite markers. *Helia* 34, 63–68. doi: 10.2298/hel1155063n
- Nakaya, A., and Isobe, S. N. (2012). Will genomic selection be a practical method for plant breeding? *Ann. Bot.* 110, 1303–1316. doi: 10.1093/aob/mcs109
- Nambeesan, S. U., Mandel, J. R., Bowers, J. E., Marek, L. F., Ebert, D., Corbi, J., et al. (2015). Association mapping in sunflower (*Helianthus annuus* L.) reveals independent control of apical vs. basal branching. *BMC Plant Biol.* 15:84. doi: 10.1186/s12870-015-0458-9
- Pacureanu-Joita, M., Vranceanu, A. V., Soare, G., Marinescu, A., and Sandu, I. (1998). "The evaluation of the parasite-host interaction in the system *Helianthus annuus* L. - *Orobanche cumana* Wallr. in Romania," in *Proceedings of the 2nd Balkan Symposium in Field Crops*, Novi Sad, 153–158.
- Paniego, N., Echaide, M., Munoz, M., Fernandez, L., Torales, S., Faccio, P., et al. (2002). Microsatellite isolation and characterization in sunflower (*Helianthus annuus* L.). *Genome* 45, 34–43. doi: 10.1139/g01-120
- Pankovic, D., Radovanovic, N., Jovic, S., Satovic, Z., and Skoric, D. (2007). Development of co-dominant amplified polymorphic sequence markers for resistance of sunflower to downy mildew race 730. *Plant Breed.* 126, 440–444. doi: 10.1111/j.1439-0523.2007.01376.x
- Patrick, H., and Alfonso, C.-M. (2013). "New and renewed breeding methodology," in *Advance in Barley Science*, eds G. Zhang, C. Li, and X. Liu (Dordrecht: Springer), 349–357. doi: 10.1007/978-94-007-4682-4\_29
- Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and Eigen analysis. *PLOS Genet.* 2:e190. doi: 10.1371/journal.pgen.0020190
- Peerbolte, R. P., and Peleman, J. (1996). "The CARTISOL sunflower RFLP map (146 loci) extended with 291 AFLP markers," in *Proceedings of the 18th Sunflower Research Forum*, Jan 1996, Fargo, ND, USA (Bismarck, ND: National Sunflower Association), 174–178.
- Pegadaraju, V., Nipper, R., Hulke, B. S., Qi, L. L., and Schultz, Q. (2013). De novo sequencing of the sunflower genome for SNP discovery using the RAD (Restriction site Associated DNA) approach. *BMC Genomics* 14:556. doi: 10.1186/1471-2164-14-556
- Peluffo, L. (2010). *Characterization of the Defense Mechanisms against Sclerotinia sclerotiorum, Causal Agent of Sunflower Head Rot, through Metabolic and Transcriptional Profiles Analyses*. Ph.D. thesis, University of Buenos Aires, Buenos Aires.
- Peluffo, L., Lia, V., Troglia, C., Maringolo, C., Norma, P., Escande, A., et al. (2010). Metabolic profiles of sunflower genotypes with contrasting response to *Sclerotinia sclerotiorum* infection. *Phytochemistry* 71, 70–80. doi: 10.1016/j.phytochem.2009.09.018
- Pérez Vich, B., Aguirre, M. R., Guta, B., and Fernandez-Martinez, J. M. (2017). Genetic diversity of a germplasm collection of confectionery sunflower landraces from Spain. *Front. Plant Sci.*
- Pérez-Vich, B., Fernández-Martínez, J. M., Grondona, M., Knapp, S. J., and Berry, S. T. (2002). Stearoyl-ACP and oleoyl-PC desaturase genes cosegregate with quantitative trait loci underlying high stearic and high oleic acid mutant phenotypes in sunflower. *Theor. Appl. Genet.* 104, 338–349. doi: 10.1007/s001220100712
- Pérez-Vich, B., Velasco, L., and Fernandez-Martinez, J. M. (2004). "QTL mapping of resistance to races E and F of broomrape (*Orobanche cumana* Wallr.) in sunflower," in *Proceedings of the Parasitic Plant Management in Sustainable Agriculture Meeting on Breeding for Orobanche Resistance in Sunflower*, Bucharest, 6.
- Picheny, V., Trépos, R., and Casadebaig, P. (2017). Optimization of black-box models with uncertain climatic inputs – Application to sunflower ideotype design. *PLOS ONE* 12:e0176815. doi: 10.1371/journal.pone.0176815
- Poormohammad Kiani, S., Grieu, P., Maury, P., Hewezi, T., Gentzbittel, L., and Sarrafi, A. (2007a). Genetic variability for physiological traits under drought conditions and differential expression of water stress-associated genes in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 114, 193–207. doi: 10.1007/s00122-006-0419-7
- Poormohammad Kiani, S., Maury, P., Nouri, L., Ykhef, N., Grieu, P., and Sarrafi, A. (2009). QTL analysis of yield-related traits in sunflower under different



- water treatments. *Plant Breed.* 128, 363–373. doi: 10.1111/j.1439-0523.2009.01628.x
- Poormohammad Kiani, S., Talia, P., Maury, P., Grieu, R., Heinz, R., Perrault, A., et al. (2007b). Genetic analysis of plant water status and osmotic adjustment in recombinant inbred lines of sunflower under two water treatments. *Plant Sci.* 172, 773–787. doi: 10.1016/j.plantsci.2006.12.007
- Port, A., Nechifor, V., Midoni, A., and Duca, M. (2013). Usefulness of the diagnostic markers for the restorer gene *Rf1* in inheritance studies at sunflower. *Anal. Sci.* 14, 11–17. doi: 10.1007/s11032-016-0527-2
- Premnath, A., Narayana, M., Ramakrishnan, C., Kuppasamy, S., and Chockalingam, V. (2016). Mapping quantitative trait loci controlling oil content, oleic acid and linoleic acid content in sunflower (*Helianthus annuus* L.). *Mol. Breed.* 36:106. doi: 10.1007/s11032-016-0527-2
- Printz, B., Sergeant, K., Guignard, C., and Renaut, J. (2013). Physiological proteome study of sunflowers exposed to a polymetallic constraint. *Proteomics* 13, 1993–2015. doi: 10.1002/pmic.201200400
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Putt, E. D. (1978). “History and present world status,” in *Sunflower Science and Technology*, ed. J. F. Carter (Madison, WI: ASA), 1–25.
- Qi, L., Long, Y., Talukder, Z. I., Seiler, G., Block, C. C., and Gulya, T. J. (2016). Genotyping-by-Sequencing uncovers the introgression alien segments associated with *Sclerotinia* basal stalk rot resistance from species – I. *Helianthus argophyllus* and *H. petiolaris*. *Front. Genet.* 7:219. doi: 10.3389/fgene.2016.00219
- Qi, L. L., Foley, M. E., Cai, X. W., and Gulya, T. J. (2016). Genetics and mapping of a novel downy mildew resistance gene, *Pl18*, introgressed from wild *Helianthus argophyllus* into cultivated sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 129, 741–752. doi: 10.1007/s00122-015-2662-2
- Qi, L. L., Gulya, T. J., Hulke, B. S., and Vick, B. A. (2012a). Chromosome location, DNA markers and rust resistance of the sunflower gene *R5*. *Mol. Breed.* 30, 745–756. doi: 10.1007/s11032-011-9659-6
- Qi, L. L., Gulya, T. J., Seiler, G. J., Hulke, B. S., and Vick, B. A. (2011a). Identification of resistance to new virulent races of rust in sunflowers and validation of DNA markers in the gene pool. *Phytopathology* 101, 241–249. doi: 10.1094/PHYTO-06-10-0162
- Qi, L. L., Hulke, B. S., Vick, B. A., and Gulya, T. J. (2011b). Molecular mapping of the rust resistance gene *R4* to a large NBS-LRR cluster on linkage group 13 of sunflower. *Theor. Appl. Genet.* 123, 351–358. doi: 10.1007/s00122-011-1588-6
- Qi, L. L., Long, Y. M., Jan, C. C., Ma, G. J., and Gulya, T. J. (2015a). *Pl17* is a novel gene independent of known downy mildew resistance genes in the cultivated sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 128, 757–767. doi: 10.1007/s00122-015-2470-8
- Qi, L. L., Long, Y. M., Ma, G. J., and Markell, S. G. (2015b). Map saturation and SNP marker development for the rust resistance genes (*R4*, *R5*, *R13a*, and *R13b*) in sunflower (*Helianthus annuus* L.). *Mol. Breed.* 35:196. doi: 10.1007/s11032-015-0380-8
- Qi, L. L., Ma, G. J., Long, Y. M., Hulke, B. S., Gong, L., and Markell, S. G. (2015c). Relocation of a rust resistance gene *R2* and its marker-assisted gene pyramiding in confection sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 128, 477–488. doi: 10.1007/s00122-014-2446-0
- Qi, L. L., Seiler, G. J., Vick, B. A., and Gulya, T. J. (2012b). Genetics and mapping of the *R11* gene conferring resistance to recently emerged rust races, tightly linked to male fertility restoration, in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 125, 921–932. doi: 10.1007/s00122-012-1883-x
- Qi, L. L., Talukder, Z. I., Hulke, B. S., and Foley, M. E. (2017). Development and dissection of diagnostic SNP markers for the downy mildew resistance genes *PlArg* and *Pl8* and marker-assisted gene pyramiding in sunflower (*Helianthus annuus* L.). *Mol. Genet. Genomics* 292, 1–13. doi: 10.1007/s00438-017-1290-8
- Quillet, M. C., Madjidian, N., Griveau, Y., Seriesys, H., Tersac, M., Lorieux, M., et al. (1995). Mapping genetic factors controlling pollen viability in an interspecific cross in *Helianthus* section *Helianthus*. *Theor. Appl. Genet.* 91, 1195–1202. doi: 10.1007/BF00220929
- Radwan, O., Bouzidi, M. F., Nicolas, P., and Mouzeyar, S. (2004). Development of PCR markers for the *Pl5/Pl8* locus for resistance to *Plasmopara halstedii* in sunflower, *Helianthus annuus* L. from complete CC-NBS-LRR sequences. *Theor. Appl. Genet.* 109, 176–185. doi: 10.1007/s00122-004-1613-0
- Rauf, S., Jamil, N., Ali Tariq, S., Khan, M., and Kausar, M. (2017). Progress in modification of sunflower oil to expand its industrial value. *J. Sci. Food Agric.* 97, 1997–2006. doi: 10.1002/jsfa.8214
- Regitano Neto, A., de Oliveira Miguel, A. M. R., Mourad, A. L., Henriques, E. A., and Alves, R. M. V. (2016). Environmental effect on sunflower oil quality. *Crop Breed. Appl. Biotechnol.* 16, 197–204. doi: 10.1590/1984-7032016v16n3a30
- Reif, J. C., Zhao, Y., Würschum, T., Gowda, M., and Hahn, V. (2013). Genomic prediction of sunflower hybrid performance. *Plant Breed.* 132, 107–114. doi: 10.1111/pbr.12007
- Rieseberg, L. H., Choi, H., Chan, R., and Spore, C. (1993). Genomic map of a diploid hybrid species. *Heredity* 70, 285–285. doi: 10.1038/hdy.1993.41
- Rodríguez-Ojeda, M. I., Fernández-Escobar, J., and Alonso, L. C. (2001). “Sunflower inbred line (KI-374) carrying two recessive genes for resistance against a highly virulent Spanish population of *Orobancha cernua* Loefl/O. *cumana* Wallr. race F” in *Proceedings of the 7th International Parasitic Weed Symposium*, eds L. J. Muselman, C. Parker, and J. A. C. Verkleij (Nantes: Université de Nantes), 208–211.
- Roedel-Drevet, P., Gagne, G., Mouzeyar, S., Gentzbittel, L., Philippon, J., Nicolas, P., et al. (1996). Colocation of downy mildew (*Plasmopara halstedii*) resistance genes in sunflower (*Helianthus annuus* L.). *Euphytica* 91, 225–228. doi: 10.1007/BF00021074
- Rowe, H. C., and Rieseberg, L. H. (2013). Genome-scale transcriptional analyses of first-generation interspecific sunflower hybrids reveals broad regulatory compatibility. *BMC Genomics* 14:342. doi: 10.1186/1471-2164-14-342
- Sabetta, W., Alba, V., Blanco, A., and Montemurro, C. (2011). sunTILL: a TILLING resource for gene function analysis in sunflower. *Plant Methods* 7:20. doi: 10.1186/1746-4811-7-20
- Sala, C. A., and Bulos, M. (2012). Inheritance and molecular characterization of broad range tolerance to herbicides targeting acetohydroxyacid synthase in sunflower. *Theor. Appl. Genet.* 124, 355–364. doi: 10.1007/s00122-011-1710-9
- Sala, C. A., Bulos, M., Alteri, E., and Ramos, M. L. (2012). Genetics and breeding of herbicide tolerance in sunflower. *Helia* 35, 57–70. doi: 10.2298/HEL1257057S
- Sala, C. A., Bulos, M., and Echarte, A. M. (2008a). Genetic analysis of an induced mutation conferring imidazolinone resistance in sunflower. *Crop Sci.* 48, 1817–1822. doi: 10.2135/cropsci2007.11.0625
- Sala, C. A., Bulos, M., Echarte, A. M., Whitt, S. R., and Ascenzi, R. (2008b). Molecular and biochemical characterization of an induced mutation conferring imidazolinone resistance in sunflower. *Theor. Appl. Genet.* 108, 105–112. doi: 10.1007/s00122-008-0880-6
- Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., et al. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: a review. *Eur. J. Agron.* 70, 112–123. doi: 10.1016/j.eja.2015.07.004
- Sarazin, V., Duclercq, J., Guillot, X., Sangwan, B., and Sangwan, R. S. (2017). Water-stressed sunflower transcriptome analysis revealed important molecular markers involved in drought stress response and tolerance. *Environ. Exp. Bot.* 142, 45–63. doi: 10.1016/j.jenvexpbot.2017.08.005
- Schnabel, U., Engelmann, U., and Horn, R. (2008). Development of markers for the use of the PEF1-cytoplasm in sunflower hybrid breeding. *Plant Breed.* 127, 587–591. doi: 10.1111/j.1439-0532.2008.01516
- Schuppert, G. F., Tang, S., Slabaugh, M. B., and Knapp, S. J. (2006). The sunflower high-oleic mutant Ol carries variable tandem repeats of *FAD2-1*, a seed-specific oleoyl-phosphatidyl choline desaturase. *Mol. Breed.* 17, 241–256. doi: 10.1007/s11032-005-5680-y
- Seiler, G. J., Qi, L. L., and Marek, L. F. (2017). Utilization of sunflower crop wild relatives for cultivated sunflower improvement. *Crop Sci.* 57, 1083–1101. doi: 10.2135/cropsci2016.10.0856
- Seriesys, H. (2005). “Identification, study, utilization in breeding programs of new CMS sources in the FAO Subnetwork,” in *Proceedings of the Sunflower Subnetwork Progress Report*, Rome, 47–53.
- Singchai, A., Muangsang, N., and Machikowa, T. (2013). Evaluation of SSR markers associated with high oleic acid in sunflower. *Int. J. Biol. Food Vet. Agric. Eng.* 7, 631–634.
- Skoric, D. (2012). “Sunflower breeding,” in *Sunflower Genetics and Breeding*, ed. Z. Kovacevic (Novi Sad: Serbian Academy of Sciences), 165–354.

- Skoric, D., and Pacureanu, M. (2010). "Sunflower breeding for resistance to broomrape (*Orobanche cumana* Wallr.)," in *Proceedings of the International Symposium "Breeding of Sunflower on Resistance to Diseases"*, Krasnodar, 19–29.
- Slabaugh, M. B., Yu, J.-K., Tang, S., Heesacker, A., Hu, X., Lu, G., et al. (2003). Haplotyping and mapping a large cluster of downy mildew resistance gene candidates in sunflower using multilocus intron fragment length polymorphisms. *Plant Biotechnol. J.* 1, 167–185. doi: 10.1046/j.1467-7652.2003.00016.x
- Soldatov, K. I. (1976). "Chemical mutagenesis in sunflower breeding," in *Proceedings of the 7th International Sunflower Conference*, Krasnodar, 352–357.
- Talukder, Z. I., Gong, L., Hulke, B. S., Pegadaraju, V., Song, Q., Schultz, Q., et al. (2014a). A high-density SNP map of sunflower derived from RAD-sequencing facilitating fine-mapping of the rust resistance gene R12. *PLOS ONE* 9:e98628. doi: 10.1371/journal.pone.0098628
- Talukder, Z. I., Hulke, B. S., Qi, L., Scheffler, B. E., Pegadaraju, V., McPhee, K., et al. (2014b). Candidate gene association of *Sclerotinia* stalk rot resistance in sunflower (*Helianthus annuus* L.) uncovers the importance of COII homologs. *Theor. Appl. Genet.* 127, 193–209. doi: 10.1007/s00122-013-2210
- Talukder, Z. I., Seiler, G. J., Song, Q., Ma, G., and Qi, L. (2016). SNP discovery and QTL mapping of *Sclerotinia* basal stalk rot resistance in sunflower using genotyping-by-sequencing. *Plant Genome* 9, 1–16. doi: 10.3835/plantgenome2016.03.0035
- Tang, S., Heesacker, A., Kishore, V. K., Fernandez, A., Sadik, E. S., Cole, G., et al. (2003a). Genetic mapping of the Or<sub>5</sub> gene for resistance to race E in sunflower. *Crop Sci.* 43, 1021–1028. doi: 10.2135/cropsci2003.1021
- Tang, S., Kishore, V. K., and Knapp, S. J. (2003b). PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. *Theor. Appl. Genet.* 107, 6–19. doi: 10.1007/s00122-003-1233-0
- Tang, S., Leon, A., Bridges, W. C., and Knapp, S. J. (2006). Quantitative trait loci for genetically correlated seed traits are tightly linked to branching and pericarp pigment loci in sunflower. *Crop Sci.* 46, 721–734. doi: 10.2135/cropsci2005.0006-7
- Tang, S., Yu, J. K., Slabaugh, M. B., Shintani, D. K., and Knapp, S. J. (2002). Simple sequence repeat map of the sunflower genome. *Theor. Appl. Genet.* 105, 1124–1136. doi: 10.1007/s00122-002-0989-y
- Urie, A. L. (1984). "Inheritance of very high oleic acid content in sunflower," in *Proceedings of the 6th Sunflower Research Workshop Sunflower Association*, Bismarck, ND, 8–9.
- Van Raden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstenaar, T. S., Schnabel, R. D., Taylor, J. F., et al. (2009). Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92, 16–24. doi: 10.3168/jds.2008-1514
- Vannozzi, G. P. (2006). The perspectives of use of high oleic sunflower for oleochemistry and energy raws. *Helia* 29, 1–24. doi: 10.2298/hel0644001v
- Vear, F. (2016). Changes in sunflower breeding over the last fifty years. *OCL* 23:D202.
- Vear, F., Gentzbittel, L., Philippon, J., Mouzeyar, S., Mestries, E., Roedel-Drevet, P., et al. (1997). The genetics of resistance to five races of downy mildew (*Plasmopara halstedii*) in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 95, 584–589. doi: 10.1007/s001220050599
- Vega, T., Breccia, G., Gil, M., and Nestares, G. (2012). Acetohydroxyacid synthase (AHAS) in vivo assay for screening imidazolinone-resistance in sunflower (*Helianthus annuus* L.). *Plant Physiol. Biochem.* 61, 103–107. doi: 10.1016/j.plaphy.2012.09.013
- Velasco, L., Fernandez-Cuesta, A., and Fernandez-Martinez, J. M. (2014). Variability of seed quality traits in a collection of Spanish landraces of confectionary sunflower. *Crop Pasture Sci.* 65, 242–249. doi: 10.1071/CP13390
- Velasco, L., Pérez-Vich, B., Yassin, A. A., Jan, C. C., and Fernandez-Martínez, J. M. (2012). Inheritance of resistance to sunflower broomrape (*Orobanche cumana* Wallr.) in an interspecific cross between *Helianthus annuus* and *Helianthus debilis* subsp. *tardiflorus*. *Plant Breed.* 131, 220–221. doi: 10.1111/j.1439-0523.2011.01915.x
- Vera-Ruiz, E. M., Velasco, L., Leon, A. J., Fernández-Martínez, J. M., and Pérez-Vich, B. (2006). Genetic mapping of the *Tph1* gene controlling beta-tocopherol accumulation in sunflower seeds. *Mol. Breed.* 17, 291–296. doi: 10.1007/s11032-005-5678-5
- Vincourt, P., As-sadi, F., Bordat, A., Langlade, N. B., Gouzy, J., Pouilly, N., et al. (2012). Consensus mapping of major resistance genes and independent QTL for quantitative resistance to sunflower downy mildew. *Theor. Appl. Genet.* 125, 909–920. doi: 10.1007/s00122-012-1882-y
- Vranceanu, A. V., Tudor, V. A., Stoenescu, F. M., and Pirvu, N. (1980). "Virulence groups of *Orobanche cumana* Wallr. different hosts and resistance sources and genes in sunflower," in *Proceedings of the 9th International Sunflower Conference*, Torremolinos, 74–82.
- White, A. D., Graham, M. A., and Owen, M. D. K. (2003). Isolation of acetolactate synthase homologs in common sunflower. *Weed Sci.* 51, 845–853. doi: 10.1614/P2002-136
- White, A. D., Owen, M. D., Hartzler, R. G., and Cardina, J. (2002). Common sunflower resistance to acetolactate-inhibiting herbicides. *Weed Sci.* 50, 432–437. doi: 10.1614/0043-1745(2002)050[0432:CSRTAS]2.0.CO;2
- Wieckhorst, S., Bachlava, E., Dušle, C. M., Tang, S., Gao, W., Saski, C., et al. (2010). Fine mapping of the sunflower resistance locus *Plarg* introduced from the wild species *Helianthus argophyllus*. *Theor. Appl. Genet.* 121, 1633–1644. doi: 10.1007/s00122-010-1416-4
- Würschum, T., Reif, J. C., Kraft, G., Janssen, G., and Zha, Y. (2013). Genomic selection in sugar beet breeding populations. *BMC Genet.* 14:85. doi: 10.1186/1471-2156-14-85
- Yang, C., Xu, L., Zhang, N., Islam, F., Song, W., Hu, L., et al. (2017). iTRAQ-based proteomics of sunflower cultivars differing in resistance to parasitic weed *Orobanche cumana*. *Proteomics* 17, 13–14. doi: 10.1002/pmic.201700009
- Yu, B., and Shang, S. (2017). Multi-year mapping of maize and sunflower in Hetao irrigation district of China with high spatial and temporal resolution vegetation index series. *Remote Sens.* 9:855. doi: 10.3390/rs9080855
- Yu, J. K., Tang, S., Slabaugh, M. B., Heesacker, A., Cole, G., Herring, M., et al. (2003). Towards a saturated molecular genetic linkage map for sunflower. *Crop Sci.* 43, 367–387. doi: 10.2135/cropsci2003.3670
- Yue, B., Vick, B. A., Cai, X., and Hu, J. (2010). Genetic mapping for the *Rf1* (fertility restoration) gene in sunflower (*Helianthus annuus* L.) by SSR and TRAP markers. *Plant Breed.* 129, 24–28. doi: 10.1111/j.1439-0523.2009.01661.x
- Zambelli, A., Leon, A., and Garcés, R. (2015). "Mutagenesis in sunflower," in *Sunflower Oilseed, Chemistry, Production, Processing and Utilization*, eds E. Martínez Force, N. T. Dunford, and J. J. Salsa (Urbana, IL: AOCS), 27–52. doi: 10.1016/B978-1-893997-94-3.50008-8
- Zhang, M., Liu, Z., and Jan, C. C. (2016). Molecular mapping of a rust resistance gene *R14* in cultivated sunflower line PH 3. *Mol. Breed.* 36:32. doi: 10.1007/s11032-016-0456-0
- Zhang, Z. W., Ma, G. J., Zhao, J., Markell, S. G., and Qi, L. L. (2017). Discovery and introgression of the wild sunflower-derived novel downy mildew resistance gene *Pl19* in confection sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 130, 29–39. doi: 10.1007/s00122-016-2786-z
- Zhao, J., Wang, J., An, L., Doerge, R. W., Chen, Z. J., Grau, C. R., et al. (2007). Analysis of gene expression profiles in response to *Sclerotinia sclerotiorum* in *Brassica napus*. *Planta* 227, 13–24. doi: 10.1007/s00425-007-0586-z
- Zhao, Y., Mette, M. F., and Reif, J. C. (2015). Genomic selection in hybrid breeding. *Plant Breed.* 134, 1–10. doi: 10.1111/pbr.12231

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Dimitrijevic and Horn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Genetic Architecture of Capitulate Glandular Trichome Density in Florets of Domesticated Sunflower (*Helianthus annuus* L.)

Qing-Ming Gao<sup>1</sup>, Nolan C. Kane<sup>2</sup>, Brent S. Hulke<sup>1\*</sup>, Stephan Reinert<sup>2</sup>, Cloe S. Pogoda<sup>2</sup>, Silas Tittes<sup>2</sup> and Jarrad R. Prasifka<sup>1</sup>

<sup>1</sup> USDA-ARS Red River Valley Agricultural Research Center, Fargo, ND, United States, <sup>2</sup> Ecology and Evolutionary Biology Department, University of Colorado, Boulder, CO, United States

## OPEN ACCESS

### Edited by:

Dragana Miladinović,  
Institute of Field and Vegetable Crops,  
Serbia

### Reviewed by:

Begoña Pérez Vich,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain  
Jadranka Z. Luković,  
Faculty of Science, University of Novi  
Sad, Serbia

### \*Correspondence:

Brent S. Hulke  
brent.hulke@ars.usda.gov

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 19 September 2017

**Accepted:** 18 December 2017

**Published:** 09 January 2018

### Citation:

Gao Q-M, Kane NC, Hulke BS, Reinert S, Pogoda CS, Tittes S and Prasifka JR (2018) Genetic Architecture of Capitulate Glandular Trichome Density in Florets of Domesticated Sunflower (*Helianthus annuus* L.). *Front. Plant Sci.* 8:2227. doi: 10.3389/fpls.2017.02227

Capitate glandular trichomes (CGT), one type of glandular trichomes, are most common in Asteraceae species. CGT can produce various secondary metabolites such as sesquiterpene lactones (STLs) and provide durable resistance to insect pests. In sunflower, CGT-based host resistance is effective to combat the specialist pest, sunflower moth. However, the genetic basis of CGT density is not well understood in sunflower. In this study, we identified two major QTL controlling CGT density in sunflower florets by using a F<sub>4</sub> mapping population derived from the cross HA 300 × RHA 464 with a genetic linkage map constructed from genotyping-by-sequencing data and composed of 2121 SNP markers. One major QTL is located on chromosome 5, which explained 11.61% of the observed phenotypic variation, and the second QTL is located on chromosome 6, which explained 14.06% of the observed phenotypic variation. The QTL effects and the association between CGT density and QTL support interval were confirmed in a validation population which included 39 sunflower inbred lines with diverse genetic backgrounds. We also identified two strong candidate genes in the QTL support intervals, and the functions of their orthologs in other plant species suggested their potential roles in regulating capitulate glandular trichome density in sunflower. Our results provide valuable information to sunflower breeding community for developing host resistance to sunflower insect pests.

**Keywords:** sesquiterpenes, capitulate glandular trichome, glandular trichome, sunflower, *Helianthus annuus* L., heat shock transcription factor, WRKY transcription factor

## INTRODUCTION

Plant trichomes, the hair-like structures on above-ground plant surfaces, are key features governing interaction with the environment, including biotic and abiotic factors. Plant trichomes vary greatly in morphology and function, with at least 300 known types of plant trichomes (Wagner, 1991; Spring, 2000; Werker, 2000). Based on their metabolic activity, plant trichomes are classified into two broad groups: glandular trichomes (GTs) and non-glandular trichomes (non-GTs), which co-exist on plant surfaces like leaves, flowers, stems, and bracts (Hare and Elle, 2002; Rautio et al., 2002). A well-studied example of non-GTs is from the model plant *Arabidopsis thaliana*, which possesses trichomes that are unicellular, unbranched or with two to five branches (Oppenheimer et al., 1991; Larkin, 1994; Szymanski et al., 2000). In contrast, GTs are usually multicellular and



consist of differentiated basal, stalk, and secretory cells. GTs have been found on ~30% of all vascular plants, particularly in dicots.

Numerous studies have shown plant trichomes are involved in water protection, absorbing UV radiation, and attracting pollinators by releasing specific chemicals (Skaltsa et al., 1994; Moraes et al., 1998; Paré and Tumlinson, 1999; Benz and Martin, 2006; Lusa et al., 2014). However, the most notable function of plant trichomes is their role in plant defense. Non-GTs serve as physical obstacles to create an unfavorable microenvironment for herbivores, parasites, and pathogens. Some non-GTs form spine- or thorn-like structures that are able to injure and trap insects and other pests during their searching and feeding activities (Pillemer and Tingey, 1976; Riddick and Wu, 2011). On the other hand, GTs are metabolic factories and mainly function in chemical defense in plants (Peiffer et al., 2009; Tian et al., 2012). The biochemical pathways and metabolic profiles in GTs have been studied in great detail in plants (Tissier, 2012). GTs can synthesize and release a wide range of chemicals, including flavonoids, terpenes, alkaloids, acyl-sugars, methyl-ketones and surface proteins (Glas et al., 2012). Recent molecular studies confirm, in different plant species, that trichome development is under the control of a conserved regulatory network, and have identified the function of several key genes in the network (Serna and Martin, 2006; Yang and Ye, 2012; Pattanaik et al., 2014). Compared with leaf tissues, more than 200 genes involved in secondary metabolism and defense-related genes are differentially expressed in trichomes in *Nicotiana tabacum*, and some of these were only expressed in trichomes (Cui et al., 2011). These secondary metabolites are either effective toxins against the pests or attractors to the natural enemies of the pests (Howe and Jander, 2008; War et al., 2012). In some cases, GTs actively release resinous or sticky chemicals on plant surfaces to limit insect movement and function as a physical defense (Simmons and Gurr, 2005; Glas et al., 2012). Studies on wild species of *Medicago* have shown that erect GTs are effective against stem-, leaf-, and fruit-eating pests (Kitch et al., 1985; Danielson et al., 1990), demonstrating their wide role in defense of many types of tissues.

In Asteraceae, GTs have been the subject of many studies for their ability to synthesize various secondary metabolites of ecological and medical importance (Wang et al., 2009; Chadwick et al., 2013; Lv et al., 2017). GTs on tarweed (*Madia elegans*) entrap insects by releasing sticky chemicals and provide carrion as food resource for predators. This indirect defense efficiently decreases the herbivore's activities and increases plant fitness under natural field conditions (Krimmel and Pearse, 2013). Among all the secondary metabolites produced by GTs, sesquiterpene lactones (STLs) are the most prevalent within Asteraceae species. Chemically, STLs derive from the isoprenoid biosynthetic pathway and contain a basic backbone of 15 carbon atoms (Zidorn, 2008). STLs show great structural variety in the arrangement of the basic skeleton and in the composition of the side chain. More than 7,000 distinct molecules of STL have been reported so far (Fraga, 2005). Besides of their ecological role in nature, STLs provide benefits to human health (Chadwick et al., 2013). Artemisinin, a unique STL molecule isolated from sweet wormwood (*Artemisia annua*), is the most effective treatment

for malaria (Covello et al., 2007). Other studies also showed STLs have anti-carcinogenic and anti-inflammatory effects (Li-Weber et al., 2002; Ghantous et al., 2010; Nasim et al., 2011). Consequently, there is great interest in understanding the anatomy, development, biochemistry and molecular biology of GTs in Asteraceae, and the biosynthetic pathways of artemisinin and other STLs have been well studied (Spring and Bienert, 1987; Covello et al., 2007; Zidorn, 2008; Wang et al., 2009; Duarte and Empinotti, 2012; Menin et al., 2012; Michalska et al., 2013; Eljounaidi et al., 2014; Aschenbrenner et al., 2016; Bombo et al., 2016; Lv et al., 2017).

In sunflower (*Helianthus annuus*), there are three main types of trichomes: non-GTs, linear glandular trichomes (LGTs), and capitate glandular trichomes (CGTs). As early as 60–70 h post-germination, these three types of trichomes form on the primordia of the first true leaves, and LGTs and CGTs begin active biosynthesis and secretion of secondary metabolites such as flavonoids, terpenoids, and sesquiterpenes at 72–96 h post-germination (Aschenbrenner et al., 2014). The LGTs are often found among non-GTs and are present on the most parts of sunflower plant, while CGTs only occur on the leaf surfaces and the anther appendages (distal ends) of sunflower florets (Aschenbrenner et al., 2013). In fact, CGTs are the only type of trichome found on anther appendages, where several important sunflower pests feed on pollen and florets, especially the most damaging insect pest of North American sunflower, the sunflower moth (*Homoeosoma electellum* Hulst.). As seen in other Asteraceae, sunflower CGTs synthesize STLs as the primary chemical defense component against herbivores (Göpfert et al., 2005; Rowe et al., 2012).

Since STL are phytotoxic, they are mainly synthesized in the stalk cells of CGTs and secreted and accumulated in the cuticle globes (Göpfert et al., 2005; Amrehn et al., 2013, 2015). Three sesquiterpene synthases (*HaGAS1*, *HaGAS2*, and *HaTPS2*) and one monooxygenase (*HaGAO*) involved in the STL biosynthesis pathway have been identified in sunflower plants (Göpfert et al., 2009; Amrehn et al., 2015). The expression pattern of these genes and the STL biosynthesis activity accompanied anther development, and STL levels in CGTs reached a peak before disk flower opening and remained more or less constant for several days (Göpfert et al., 2009; Amrehn et al., 2015). While feeding or visiting, the insects cause the rupture of the CGTs on anther florets and lead to direct contact with STL. *In vitro* studies have shown that both polar and non-polar extracts of CGTs caused high mortality and retarded growth for the larvae of sunflower moth, and the purified STL species also showed similar effects (Rossiter et al., 1986; Rogers et al., 1987; Prasifka, 2015; Prasifka et al., 2015). Interestingly, the extracts of CGTs are also effective against other floret-feeding insects like yellowstriped armyworm (*Spodoptera ornithogalli*) and migratory grasshopper (*Melanoplus sanguinipes*), indicating a broader spectrum of CGT-mediated defense (Mabry et al., 1977; Isman, 1985).

Three observations have lead us to perform this study: (1) the number of plant trichomes is heritable. In plant taxonomy, trichomes are used as a grouping character since it is a heritable trait across different plant species (Spring, 2000; Reis et al., 2002). Studies in other crop plants have also shown GT density



is a quantitatively-inherited trait (Kitch et al., 1985; Agren and Schemske, 1992; Maliepaard et al., 1995; Andrade et al., 2017). However, such information is not available in sunflower. (2) Studies in sunflower have suggested the possibility of developing inbreds or hybrids with high CGT density. Previously, Prasifka (2015) has shown that public sunflower inbred lines (both male and female) present a large range in mean CGT number (>50-fold differences), but wild sunflower and commercial hybrids showed much less variation in the mean of CGT number (~5-fold differences). Thus, breeding a sunflower line with high CGT density is plausible. (3) A positive correlation between CGT number and STL levels in sunflower florets has been documented (Prasifka et al., 2015). Increasing CGT density in sunflower florets could be a strategy for reducing the fitness of insect pests like sunflower moth, thereby reducing damage from insect pests. Therefore, the objective of this study is to identify the QTL responsible for CGT density of anther florets in sunflower. We believe this study will advance our understanding of CGT-mediated resistance to insect pests in sunflower and also benefit sunflower breeding programs by providing the first marker-assisted resource for insect resistance.

## MATERIALS AND METHODS

### Plant Materials

Plants of *H. annuus* genotype HA 300 (PI 552938) and RHA 464 (PI 655015; Hulke et al., 2010) were grown in the one-gallon plastic pots under greenhouse conditions with a photoperiod of 16-h light and 8-h dark. High CGT number per floret was observed in maintainer inbred HA300, and low CGT number per floret was observed in restorer inbred RHA 464 (Prasifka, 2015). To map the genetic factors which control CGT number per floret in sunflower, crosses were made between the parents HA 300 (female) and RHA 464 (male) in the 2013 winter greenhouse, and F<sub>1</sub> plants were grown in a field environment near Fargo, ND, in the summer of 2013. The 300 selfed F<sub>2</sub> plants and 280 F<sub>3</sub> plants were grown in the 2014 winter greenhouse and 2014 Fargo field, respectively. A total of 239 F<sub>4</sub> plants were grown in the 2015 winter greenhouse in the conditions described above and subjected to CGT counting and DNA sampling.

To validate the QTL mapping results of the F<sub>4</sub> population, a separate diversity panel of 39 inbred lines was selected, which provided a distinct genetic background based on breeding history and pedigree information. These lines were grown in 2016 under greenhouse conditions for CGT counting.

CGTs were sampled and counted in single environments because of the very high heritability of the trait and high cost of evaluation. Prasifka (2015) assessed CGT number from public inbred lines and sampled three replicate plants with three subsamples (florets) from each plant in field environments in 2012 and 2013, finding a very high correlation between the two years (environments;  $R^2 = 0.98$ ). Similarly, Spring and Bienert (1987) found that while compounds within glands on leaves were affected by lighting, the number of glands per unit area were not. These published data suggest this trait does not possess genotype-by-environment variation generally typical of quantitative traits.

### CGT Imaging and Counting

Unopened sunflower florets from 239 F<sub>4</sub> plants were removed with forceps from the outermost 1 cm of the capitulum 1 day before anthesis. After storage at  $-20^{\circ}\text{C}$ , one floret per plant was dissected by removing the corolla and making a latitudinal cut through the fused anther tube, after which the unfurled anther tube was attached to an aluminum mount with double-sided carbon adhesive tape (Ted Pella Inc., Redding CA, USA). Mounted groups of 15 florets were sputter-coated with a conductive layer of gold/palladium (Balzers SCD 030, Balzers Union Ltd., Liechtenstein). Scanning electron micrographs (SEM) were obtained using a JEOL JSM-6490LV scanning electron microscope (JEOL USA, Inc., Peabody MA, USA) operating at an accelerating voltage of 15 kV and a magnification of  $45\times$ . Since the floret area of HA 300 (mean = 14.5,  $SE = 0.75$ ,  $n = 25$ ) and RHA 464 (mean = 13.3,  $SE = 0.2$ ,  $n = 25$ ) are similar, micrographs subsequently were used to quantify the total number of CGT per floret. To estimate the counting error between florets on the same plant (natural variation or trichome loss due to handling), a second floret of each of 30 plants in the sampled F<sub>4</sub> population was analyzed by SEM and compared to the initial result.

### DNA Sampling and Sequencing

Leaf tissues were collected from each F<sub>4</sub> plant, and genomic DNA (gDNA) was extracted from lyophilized leaf material according to the Qiagen DNeasy 96 Plant Kit protocol. Subsequently, library prep was performed using a modified Genotyping by Sequencing (GBS) protocol (Meyer and Kircher, 2010). This protocol used restriction enzymes *MseI* and *EcoRI* to first digest and fragment the gDNA. The result was a pool of fragments with sticky-ends from restriction cut sites that provide the template for adaptor ligation. Illumina adaptors and barcodes were ligated to the digested fragments. A subset of fragments were then amplified using Illumina PCR primers. The samples were normalized and pooled following PCR, using SequelPrep normalization kit (Thermo Fisher Scientific, USA). Next, the pooled, normalized samples were run out on a 2.5% agarose gel. Using a 1,000 bp ladder, we selected 300–400 bp fragments by cutting out this segment from the gel. The gel segment was then purified using Qiagen gel purification kit according to accompanying protocol. Samples were sent to the University of Texas, Austin, for sequencing on an Illumina HiSeq 2500 sequencer. The results were multiplexed 150 bp single-end barcoded GBS reads.

For HA 300, RHA 464, and a portion of the validation panel, genomic DNA was extracted from lyophilized leaf, and whole genome sequencing was performed. Genomic libraries were prepared using Nextera<sup>®</sup> XT DNA library prep kits (Illumina<sup>®</sup>) according to the protocol. Each gDNA sample was diluted to the appropriate concentration using a Qubit 3.0 fluorometer (Thermo Fisher Scientific, USA). Each sample was barcoded by the unique dual index adapters Nextera<sup>®</sup> i5 and i7. Resulting libraries were cleaned using solid-phase reversible immobilization (SPRI) to remove fragment sizes less than 300 base pairs via an epMotion 5075TMX automated liquid handling system (Eppendorf North America). Sample quality control (QC) was conducted to ensure appropriate sample concentration and

fragment size using a Qubit 3.0 fluorometer and an Agilent 2100 Bioanalyzer prior to normalizing the loading concentration to 1.8–2.1 pM with 1% PhiX control v3 added (Illumina®). Samples that passed QC were processed for paired-end 150 base pair reads on the Illumina NextSeq® sequencer. Whole genome sequencing was conducted at the BioFrontiers Institute Next-Generation Sequencing Facility at University of Colorado. The remainder of the validation panel was sequenced as part of the sunflower SAM population by the Genome Quebec Innovation Centre at McGill University, Montreal, QC, Canada (Mandel et al., 2013; Burke, pers. comm.).

## SNP Marker Calling

All GBS reads were demultiplexed using `process_radtags` (v1.45) command from the Stacks software suite (Catchen et al., 2013). Default parameters were used with the exception of adding the “r” and “disable\_rad\_check” flags. After demultiplexing, each sample was checked for overrepresented sequences with FastQC (v0.11.5; Andrews, 2010). Trimmomatic (v0.35; Bolger et al., 2014) was used with default parameters to trim and clean the demultiplexed sequences. Reads with overrepresented sequences as identified by FastQC were also trimmed by appending them to the “TruSeq3-SE.fa” adapters file distributed with Trimmomatic. Reads were aligned to the sunflower reference genome (HA 412HO\_v1.1) using the BWA mem (v0.7.12-r1039; Li and Durbin, 2009) algorithm with default parameters and the duplicates in alignments were marked with Picard (v2.8.1; Broad Institute). A realignment and base call recalibration were performed with GATK (Genome Analysis Toolkit) and followed the GATK Best Practices. Finally, the alignment quality was checked with Qualimap (v2.2.1; Okonechnikov et al., 2015), which rendered the data ready for variant calling. The variant calling was performed with freebayes (v1.1.0-1-gf15e66e; Garrison and Marth, 2012), and WGS data from the parental lines (HA 300 and RHA 464) were jointly genotyped to improve the variant calling quality. The variant sites were filtered using vcftools (v0.1.13; Danecek et al., 2011). We kept only bi-allelic single nucleotide polymorphic (SNP) sites, and required a minimum site quality of 30 and a minor allele frequency greater than 0.05.

In addition to filtering sites with vcftools, we applied a protocol to filter sites that are likely to be misassembled multi-copy portions of the genome. Repetitive sites are susceptible to misassembly with respect to the reference sequence and may result in several reads erroneously mapping to the same locus. To reduce the number of these sites in our variant calls, we assessed sequencing depth of whole genome shotgun sequence reads from four modern sunflower cultivars and one landrace (Table S1) aligned to the reference genome. We used the samtools depth command to produce the number of mapped sequences per site for each individual. We summed the depths at each position, and calculated the frequency of summed depth values across samples. We then plotted the summed depth values against their frequencies, providing a visual means to choose a range of summed depth values that are most likely to correspond to well-assembled single copy sites of the genome. We then filtered our vcf table to only include the identified single-copy sites. Out of 3,200,466 diagnostic sites differing between the parental

genomes, we identified 1,138,669 single-copy diagnostic SNPs. After filtering the vcf table, missing genotypic information from the GBS data of the F<sub>4</sub> population was trio imputed using our own custom phaser2.pl and genotyper7.pl software (Kane, 2017), and leveraging the whole genome sequence of the parent lines HA 300 and RHA 464. These programs are designed to accurately impute for high-quality single-copy, diagnostic SNPs. Because of our extensive filtering, and high quality data, these assumptions were met.

## QTL Analysis

QTL analyses were carried out using the R/qtl package in R version 3.2.3 (Broman et al., 2003). The genetic map was constructed with the `est.map` function (assuming a genotyping error rate of 0.001). Single QTL analysis and LOD scores were calculated by a single QTL genome scan (`scanone` function) with standard interval mapping (0.1 cM steps, assuming a genotyping error rate of 0.001). Pairwise QTL interactions were calculated using a two-dimensional QTL scan (`scantwo` function) via standard interval mapping (0.1 cM steps, assuming a genotyping error rate of 0.001). LOD significance thresholds for type I error rates of  $\alpha < 0.05$  were determined by running 1,000 permutations on the single- and two-dimensional QTL scan. In addition, the `stepwiseqtl` function (`max.qtl = 4`, `additive.only = FALSE`) was used to perform forward/backward stepwise search, and find the QTL model which has the highest LOD score. The `fitqtl` function was used to create a QTL project to fit the phenotypic data with the selected model, and one QTL was omitted at a time to obtain an ANOVA table.

## Validation of the Mapped QTL

To validate the QTL results, the SNP markers were called from the validation population together with HA 300 and RHA 464, as described above. The SNP markers derived from the QTL support interval were extracted. Monomorphic SNPs between the two parent lines were removed, and the heterozygous SNPs and markers with missing data in one of the parent lines were also filtered out. The SNP markers were treated as fixed effects and all possible marker pairs were tested by fitting a regression function in R as  $y \sim \text{SNP}_m + \text{SNP}_n$ , where  $y$  is the CGT number,  $\text{SNP}_m$  is the effect of the  $m$ th marker from chromosome 5, and  $\text{SNP}_n$  is the effect of the  $n$ th marker from chromosome 6. The significance threshold for multiple comparisons was determined by the adjusted  $p$ -values with the false discovery rate at 0.05. (Benjamini and Yekutieli, 2001).

## Identification and Phylogenetic Analysis of Candidate Genes within the QTL Support Interval

To identify the candidate genes with the highest likelihood of influence on the phenotype, we first manually checked through the QTL support interval region in the sunflower reference genome (INRA Sunflower Bioinformatics Resources, 2014) and listed all the annotated genes (Table S5). We then filtered the list based on two criteria: (1) significant association between a co-located SNP and CGT number in the validation population; and (2) putative gene function. The sequences of selected candidate

genes were further characterized by two approaches, NCBI (National Center for Biotechnology Information) ORFfinder search and BLAST homology search. All ORFs within these sequences were identified using NCBI ORFfinder with the following parameters: a minimal ORF length of 150 nucleotides, the standard genetic code, ATG and alternative start codons, and ignore nested ORFs. BLASTp searches were carried out with the given ORFs to obtain the most similar protein sequences in Asteraceae, Solanaceae, Malvaceae and Brassicaceae. The sequences with an E-value less than  $10^{-5}$  were selected, combined, and searched for the presence of conserved protein domains using ScanProsite (Sigrist et al., 2012). Comparisons were made among HA 412HO (reference sequence), HA 300, and RHA 464 for genomic sequence differences, alternative splice sites, and changes in protein domains at the selected loci. For phylogenetic analysis, the conserved protein domain sequences with 10–15 additional amino acids on the N-terminal and C-terminal ends were used for alignment. The alignment was done with the L-INS-i strategy implemented in MAFFT (version 7.310; Katoh, 2002). The BLOSUM62 scoring matrix and a gap opening penalty of 1.5 were selected to assess the phylogenetic relationship among the protein sequences. After the alignment, a neighbor-joining tree was constructed with PhyML including the Smart Model Selection approach (version 3; Guindon et al.,

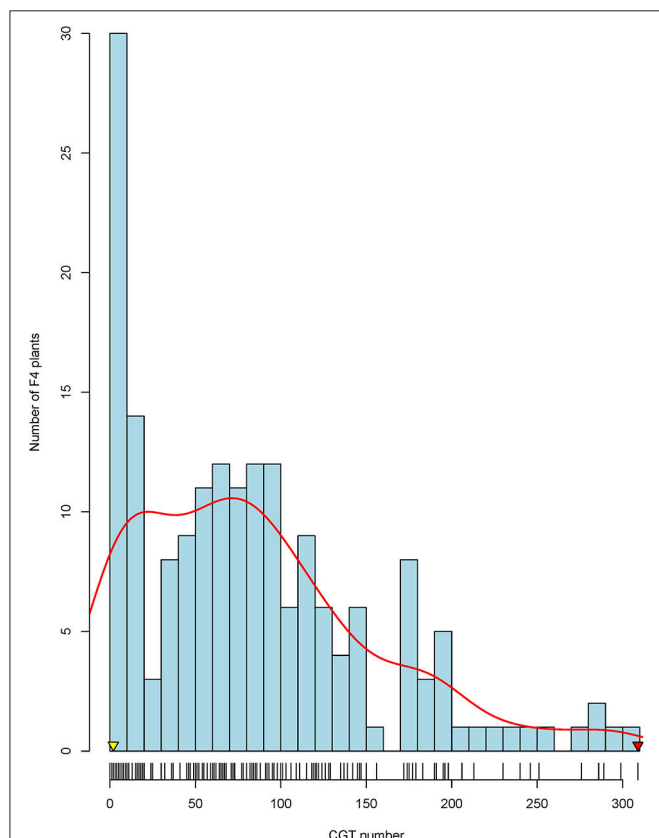
2010; Lefort et al., 2017), bootstrap reassembling with 1,000 bootstrap samples. The phylogenetic tree was colorized according to protein families in Archaeopteryx (version 0.9920 beta; Han and Zmasek, 2009).

## RESULTS

### Capitate Glandular Trichome Density in Mapping Population

The two parents presented a significant difference in CGT density, with the male parent RHA 464 having  $\sim 2$  CGT per floret and the female parent HA 300 having  $\sim 300$  CGT per floret (**Figure 1**). The CGT numbers per floret were counted for 239  $F_4$  plants, and 179 plants with good genotypic data quality were selected as the final mapping population. Based on frequency distribution, 179  $F_4$  plants were classified into three groups (**Figure 1**). Of these 179  $F_4$  plants, 25 plants (13.9%) had high CGT density which is more than 150 CGT per floret, 109 plants (60.9%) had medium CGT density which is 25–150 CGT per floret and 45 plants (25.2%) had low CGT density which is less than 25 CGT per floret. Shapiro-Wilk normality test (Shapiro and Wilk, 1965) indicated that the CGT number in the  $F_4$  population was not normally distributed ( $W = 0.915$ ,  $p < 1.1 \times 10^{-8}$ ), and the distribution of CGT number was moderately skewed toward the lower CGT number (Skewness = 0.98). The original mapping population with 239 individuals showed a similar distribution of CGT number (Skewness = 0.97;  $W = 0.916$ ,  $p < 1.1 \times 10^{-10}$ ).

To estimate the potential error in CGT counting (natural variation or trichome loss due to handling), 30 plants from the  $F_4$  population were randomly selected for recounting. A second



**FIGURE 1 |** Frequency distribution of capitate glandular trichomes (CGTs) in the  $F_4$  mapping population. The arrowheads indicate the CGT number of the parental lines, HA 300 (red arrow) and RHA 464 (yellow arrow).

**TABLE 1 |** Summary of SNP loci on the HA 300  $\times$  RHA 464 linkage map.

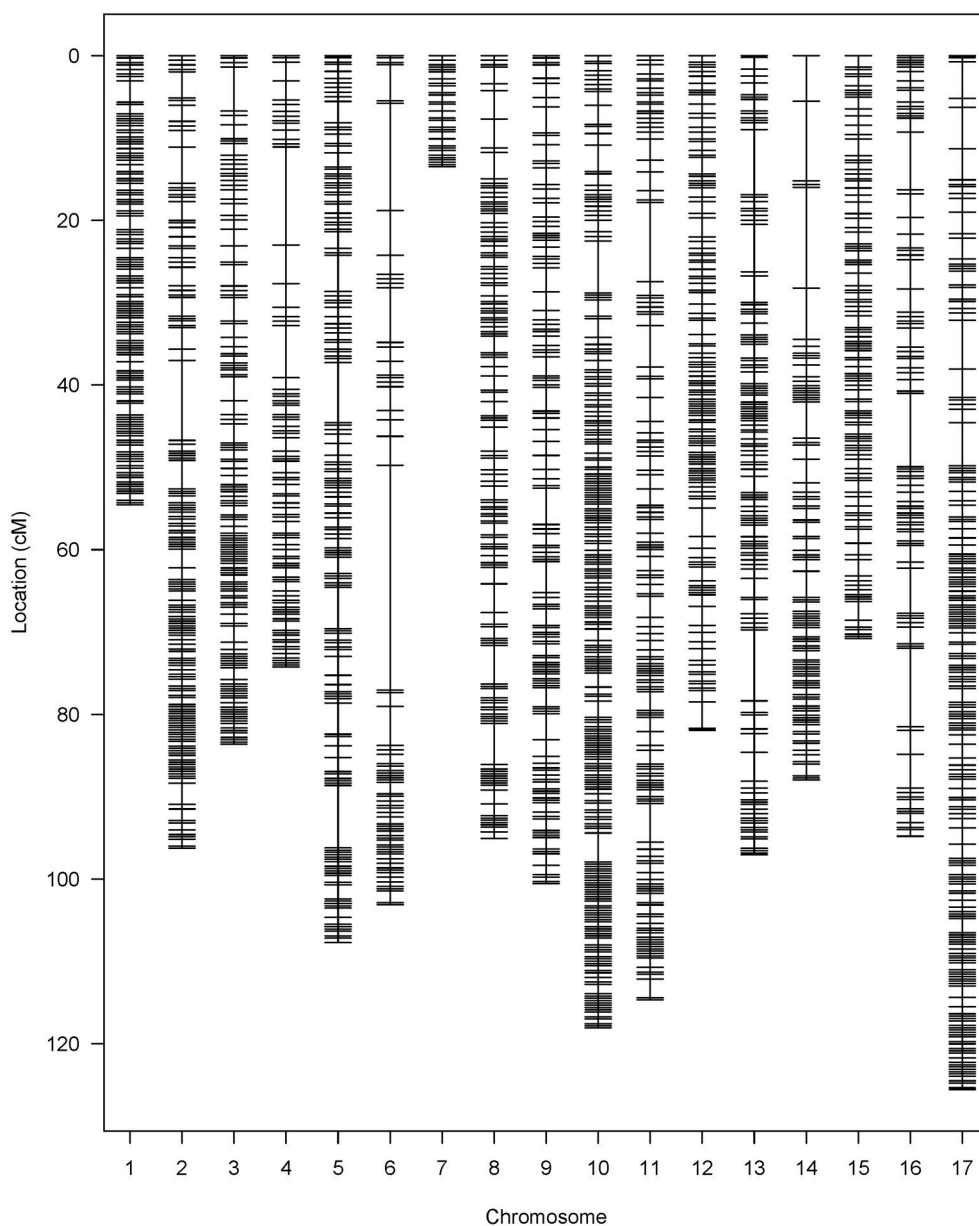
Linkage groups	Number of SNPs	Map length (cM)	Density (cM/marker)	Largest gap
1	129	54.56	0.43	2.58
2	149	96.23	0.65	9.71
3	134	83.62	0.63	5.36
4	86	74.24	0.87	11.91
5	147	107.71	0.74	7.58
6	73	103.31	1.43	27.32
7	32	13.48	0.43	0.85
8	130	95.07	0.74	5.01
9	129	100.54	0.79	4.43
10	238	118.06	0.50	6.30
11	141	114.65	0.82	9.63
12	129	81.91	0.64	3.49
13	123	97.05	0.80	8.63
14	95	87.95	0.94	12.24
15	112	70.77	0.64	2.29
16	86	94.79	1.12	9.51
17	188	125.58	0.67	5.94
Total	2121	1519.52	0.72	/

floret from each of these 30 plants was counted, and the CGT numbers from first floret and second floret were highly correlated ( $r = 0.98$ ,  $p < 2.2 \times 10^{-16}$ ) (Figure S1).

## Genetic Map Construction

A total of 443.2 M raw GBS reads were generated from 239  $F_4$  plants with the Illumina HiSeq 2500 sequencing system. After trimming the adapters and overrepresented sequences with Trimmomatic (v0.35; Bolger et al., 2014), the number of remaining reads was 210.1 M. From these 210.1 M reads, 97.9% (205.7 M) of the reads were successfully aligned to the sunflower reference genome HA 412HO\_v1.1., and the mean

length of reads was 74.3 bp. Before performing variant calling, five individuals were removed due to poor genotyping quality. A total of 16,028,511 SNPs was produced in the initial variant calling, and 1,138,669 single-copy diagnostic SNP were retained after applying several filters ( $\text{minQ} > 30$ ,  $\text{MAF} > 0.05$ , biallelic only). To further control genotypic data quality, we also checked each individual in the  $F_4$  population, and the individuals with low coverage on one or more chromosomes (less than 10 SNPs per chromosome) were dropped. A total of 179 individuals with acceptable sequence quality were kept to conduct imputation. A total of 1.09 million SNPs was generated from the imputation, and monomorphic markers and markers with LD above 0.9



**FIGURE 2 |** SNP distribution in the HA 300 × RHA 464 genetic linkage map.



were filtered out first, followed by removing SNPs with high missing rate and significant distortion from expected Mendelian ratio. Finally, 2,121 high quality SNPs were kept and used for constructing the genetic linkage map (**Figure 2**).

The 2121 SNP markers were distributed in all 17 chromosomes, and the number of markers per chromosome varies from 32 on chromosome 7 to 238 on chromosome 10 (**Table 1**). The total length of the map was 1,519.52 cM, and the size of chromosomes varies from 13.48 cM (chromosome 7) to 125.58 cM (chromosome 17). Chromosome 7 had the fewest SNP markers and shortest map distance. The average density per marker was 0.72 cM, and the largest gap between two adjacent loci was 27.32 cM on chromosome 6.

## QTL Analysis

After performing a single QTL scan on the data collected from the  $F_4$  population, two putative QTL were revealed. One was mapped to chromosome Ha5 at the 14.64 cM position, and the other one was mapped to chromosome Ha6 at the 60.72 cM position (**Figure 3**). Both QTL had LOD scores greater than the LOD threshold (4.70,  $\alpha = 0.05$ ), which was 95 percentile of the distribution of genome-wide max LOD obtained by 1,000 permutations (**Table S2**). Therefore, a two-QTL model scan was performed with the *scantwo* function, and the same pair of

QTL on Ha5 and Ha6 were identified (**Table S3**). In this two-QTL model, the additive effect of these two QTL was strongly supported by the data, but no epistasis interaction was detected. The two-QTL additive model (Ha5@14.6, Ha6@60.5) explained 24.58% of CGT number variation for the  $F_4$  mapping population, while the single QTL model explained 11.61 and 14.06% of the CGT number variation, respectively (**Table 2**). To confirm the QTL model, we also performed a forward/backward stepwise search, allowing four QTL at maximum, and found the two-QTL model had the best LOD score. The closest SNPs to each QTL were identified, and the phenotypes were plotted against genotypes at the putative QTL (**Figure S2**, **Table S4**). The 1.5—LOD support interval for SNP marker Ha5\_11356218 was 12.1–15.7 cM, which extended from 8.07 to 12.65 Mbp on the physical map, and the support interval of SNP marker Ha6\_8364901 was 41.3–76.3 cM, which covered from 7.42 to 8.58 Mbp on the physical map (**Table 2**, **Figure S3**).

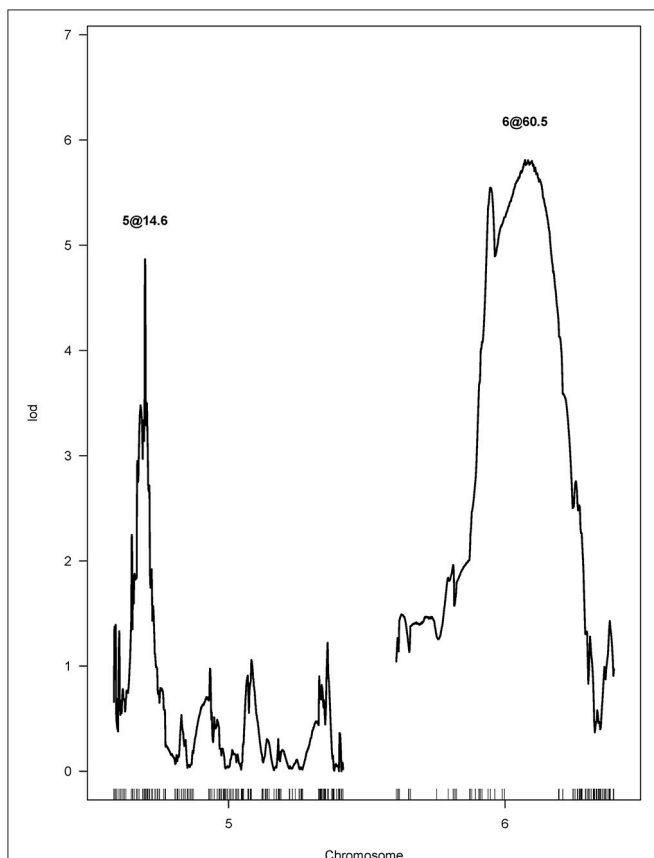
## Validation of the Mapped QTL

To validate the QTL analysis, we selected 39 inbred lines which were whole genome sequenced with  $10 \times$  coverage, and had phenotypic data from a previous study (Prasifka, 2015). With the WGS reads, 609914 SNPs were produced from variant calling and 485516 good quality SNPs were kept after applying the additional filters (minQ > 30, MAF > 0.05, biallelic only, max missing rate < 10%). A phylogenetic tree was constructed from these 485516 SNPs using the SNPhylo pipeline (Lee et al., 2014; **Figure S4**). As shown in **Figure S4**, the two parent lines, RHA 464 and HA 300, were clustered into two distinct subgroups, and the other 37 lines in the validation population presented additional genetic backgrounds.

To test the two-QTL model, which was suggested by the QTL mapping results, we retrieved all the SNP markers from the QTL support intervals. A total of 990 SNP markers were extracted from the QTL support interval on chromosome 5, with a marker density of 0.22 SNP per kb, and 363 SNP markers were extracted from the QTL support interval on chromosome 6, with a marker density of 0.31 SNP per kb. After filtering these SNP markers, as described above, we kept 385 SNP markers from the QTL support interval on chromosome 5 and 33 SNP markers from the QTL support interval on chromosome 6. Next, we tested pairs of SNP markers, one from each QTL, with a linear regression function in R [ $\text{lm}(y \sim \text{SNP}_m + \text{SNP}_n)$ ], where  $y$  is the CGT number,  $\text{SNP}_m$  is the effect of the  $m$ th marker on chromosome 5 and  $\text{SNP}_n$  is the effect of the  $n$ th marker on chromosome 6]. In total, 12705 two-QTL models were tested, and 414 of these models were detected as significant with the adjusted  $p < 0.05$  ( $p = 1.6 \times 10^{-4}$ ) (**Figures 4A,B**). The best two-QTL model explained 59.9% of total phenotypic variation.

## Identification of Putative Genes Controlling CGT Number

To identify putative genes which control CGT number in sunflower, we manually checked through the physical regions, which included the Ha5\_11356218 marker support interval and the Ha6\_8364901 marker support interval, with sunflower genome JBrowse (INRA Sunflower Bioinformatics Resources,



**FIGURE 3** | LOD curves for the capitule glandular trichome (CGT) phenotype in the  $F_4$  mapping population, calculated by two-QTL scan.

**TABLE 2 |** Two QTL for capitate glandular trichome (CGT) density identified in the HA 300 × RHA 464 mapping population.

QTL	Closest SNP	LOD	Explained variance (%)	p-value	1.5-LOD support interval
Ha5@14.6	Ha5_11356218	10.06	11.61	$1.85 \times 10^{-5***}$	12.1–15.7 cM
Ha6@60.5	Ha6_8364901	12.19	14.06	$2.18 \times 10^{-6***}$	41.3–76.3 cM
Ha5@14.6 + Ha6@60.5 <sup>a</sup>	/	10.97	24.58	$4.89 \times 10^{-10***}$	/
Ha5@14.6*Ha6@60.5 <sup>b</sup>	/	1.62	0.935	0.45	/

\*\*\*Significant at  $p < 0.001$  as determined by permutation testing.

<sup>a</sup>Additive effect of two QTL.

<sup>b</sup>Interaction effect of two QTL.

2014). A total of 105 genes were predicted within the Ha5\_11356218 marker support interval on chromosome 5, and 35 genes were predicted within the Ha6\_8364901 marker support interval on chromosome 6. A list of these genes with SNP marker data and expression profile in flower tissues is shown in Table S5.

Since some of the SNP markers in the QTL support intervals were within these genes, we found 16 genes on chromosome 5 and five genes on chromosome 6 with variation that had significant associations with CGT number (**Figures 4A,B**). Based on gene annotations that suggested plausible functional association, the genes co-localized with Ha5\_10149906 and Ha6\_7633946 were selected for further characterization. This pair of markers had strong association with CGT number (adjusted  $p < 0.05$ ,  $p = 3.6 \times 10^{-5}$ ) and the two-QTL model explained 47.2% of phenotypic variation in the validation data. The genotypic data at these loci together with phenotypic data are shown in **Table 3**. A single factor ANOVA was also used to test the association between each SNP marker and CGT number. The marker Ha5\_10149906 showed strong association with CGT number ( $p = 0.004$ ), explaining 20.3% of the phenotypic variation, and the single SNP marker Ha6\_7633946 also showed significant association with CGT number ( $p = 0.0004$ ), explaining 31.1% of the phenotypic variation (**Figures 4C,D**).

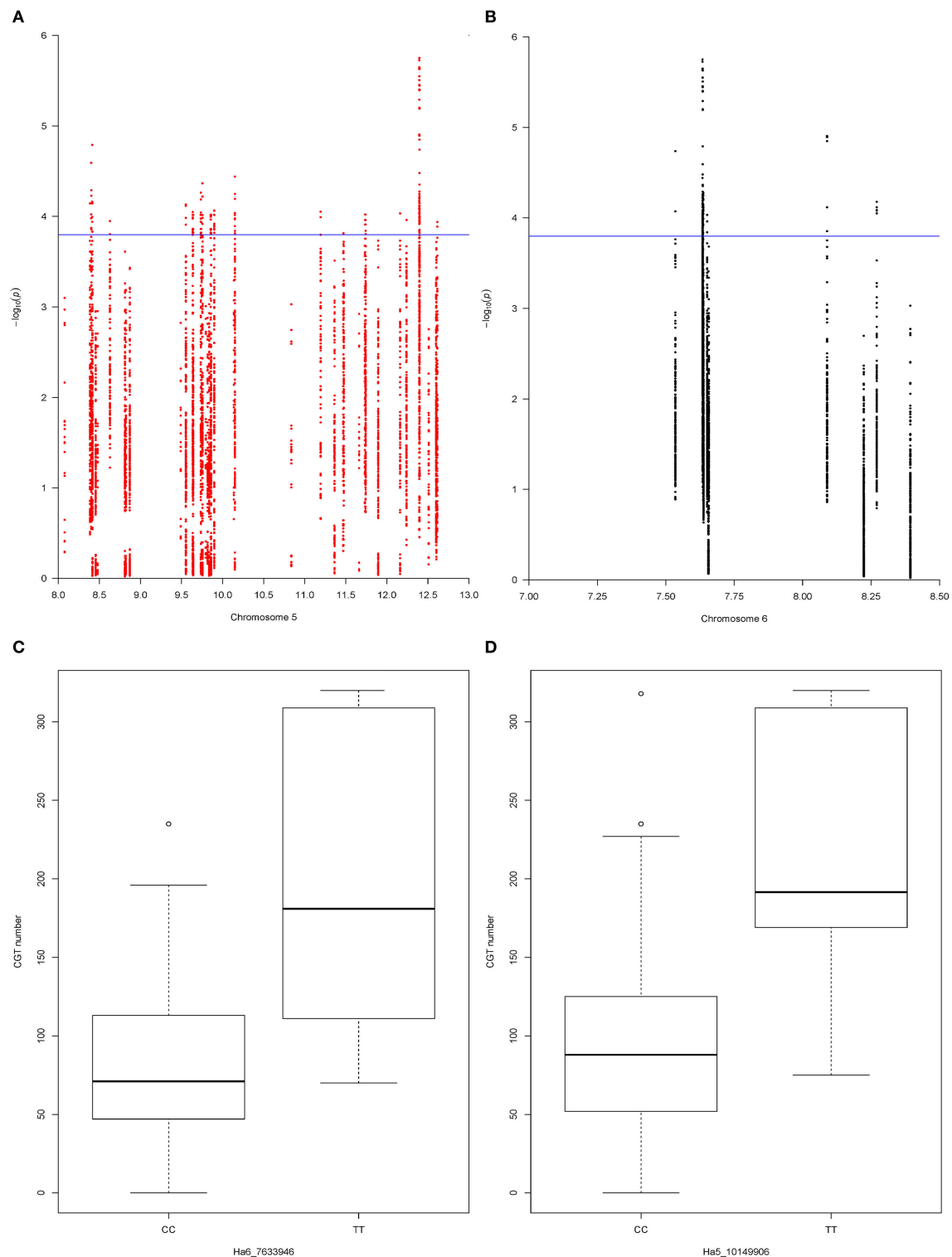
The SNP marker Ha5\_10149906 is located within the second exon of the gene Ha5g003120, which is annotated as a member of heat shock transcription factor (HSF) family (Table S5). The HSF proteins are not only involved in heat stress responses but also participate in many other developmental/physiological activities such as cell division and root growth (Westerheide et al., 2012). As shown in the phylogenetic tree (Figure S5), sunflower Ha5g003120 gene is orthologous to *Arabidopsis thaliana* gene *HSF2A*. Furthermore, we obtained the genomic sequence of gene Ha5g003120 from the two parents, HA 300 and RHA 464, and the reference genome HA 412HO, and performed alignments. Interestingly, the genomic sequence in RHA 464 has a 326 bp deletion in the promoter region of Ha5g003120, but this deletion is not present in HA 300 and HA 412HO (**Figure 5A**).

The SNP marker Ha6\_7633946 is located in the 3'-UTR (untranslated region) of the gene Ha6g003560, which is annotated as a member of the WRKY transcription factor family (Table S5). WRKY proteins are key regulators of many biotic, abiotic and physiological responses in plants (Phukan et al., 2016). Similarly, the ORFs within the gene sequence of Ha6g003560 were confirmed with NCBI ORFfinder. The

confirmed ORF sequence was used for BLASTp searches to obtain the most similar protein sequences in other phyla (Figure S6). As shown in the phylogenetic tree, Ha6g003560 closely resembles *Arabidopsis thaliana* gene *WRKY44/TTG2*. It has been shown that the *WRKY44/TTG2* gene regulates trichome development in *Arabidopsis* (Ishida et al., 2007). In addition, we also performed alignments with the genomic sequence of gene Ha6\_7633946 from HA 300, RHA 273 and reference genome XRQ. Since the sequence quality was poor in this region in RHA 464 we used the sequence from RHA 273 instead because these two inbred lines are phylogenetically close, they share the same phenotype for CGT, and they share the same haplotype of chromosome 6 (**Table 3**, Figure S4). As a result, we found a 51 bp deletion in the intron of gene Ha6\_7633946 in RHA 273, and this deletion results in an alternative splicing (**Figures 5B, 6**). These results make the putative HSF gene Ha5g003120 and WRKY gene Ha6g003560 good candidates as the regulators of CGT density in sunflower florets.

## DISCUSSION

Trichomes, especially glandular trichomes, play an important role in plant defense. CGTs are most common in Asteraceae species (Bombo et al., 2016). CGTs are well-known for their ability to produce various secondary metabolites such as STLs. In sunflower, however, the genetics of CGT density is not well understood. In this study, we reported for the first time that CGT density in sunflower florets is a quantitative trait, and two major QTL were identified in a biparental mapping population. The F<sub>4</sub> mapping population derived from high CGT density parent HA 300 and low CGT density parent RHA 464, showed segregation in CGT number per floret (**Figure 1**). Of 179 F<sub>4</sub> plants, only 25 plants (13.9%) showed high CGT density (greater than 150 CGT per floret) and 45 plants (25.2%) showed low CGT density (less than 50 CGT per floret). The variation of CGT density from CGT counting errors was marginal since the CGT numbers from first floret and second floret were highly correlated (Figure S1). As shown in **Figure 1**, the frequency of CGT density in the F<sub>4</sub> mapping population showed a moderately skewed distribution (Skewness = 0.98) toward the low CGT density parent. For this reason, we attempted non-parametric interval mapping, which is an alternative method for analyzing non-normal phenotypic data, but it exhibited less power to detect QTL in our population (Broman, 2003; Fernandes et al., 2007; data not shown). In



**FIGURE 4 | (A)** The  $-\log_{10} p$ -value for each SNP marker in the chromosome 5 QTL support interval. Values were plotted against their respective physical positions (Mbp). The blue line represents the significance threshold ( $p = 1.6 \times 10^{-4}$ ). **(B)** The  $-\log_{10} p$ -value for each SNP marker in the chromosome 6 QTL support interval. Values were plotted against their respective physical positions (Mbp). The blue line represents the significance threshold ( $p = 1.6 \times 10^{-4}$ ). **(C)** Effect of SNP marker Ha6\_7633946 on CGT number. The box plot of each allele at SNP marker Ha6\_7633946 shows the significant association with CGT number in the validation panel ( $p = 0.0004$ ). The horizontal bar (bold) indicates the median, discontinuous lines represent the upper and lower quartile, and the outlier data are labeled by a circle. **(D)** Effect of SNP marker Ha5\_10149906 on CGT number. The box plot of each allele at SNP marker Ha5\_10149906 shows the significant association with CGT number in the validation panel ( $p = 0.004$ ). The horizontal bar (bold) indicates the median, discontinuous lines represent the upper and lower quartile, and the outlier data are labeled by a circle.

**TABLE 3 |** Genotypes of two QTL and phenotypic data in the validation population (/ represents missing data).

Inbred lines	CGT number	Ha5_10149906	Ha6_7633946
RHA 464 (male parent)	2	CC	CC
HA 300 (female parent)	309	TT	TT
HA 124	28	CC	CC
HA 133	188	TT	TT
HA 207	195	TT	CC
HA 232	92	/	/
HA 248	334	/	/
HA 291	60	CC	CC
HA 292	108	CC	CC
HA 301	30	/	/
HA 305	235	CC	CC
HA 321	174	CC	TT
HA 349	133	CC	CC
HA 371	109	TT	CC
HA 383	75	TT	/
HA 404	318	CC	TT
HA 441	169	TT	CC
HA 445	136	/	/
HA 456	88	CC	CC
HA 467	140	/	/
HA 821	320	TT	TT
HA 851	117	CC	CC
HA 853	111	CC	TT
HA 89	71	CC	CC
ND-NONOIL M3	70	CC	TT
RHA 270	196	CC	CC
RHA 272	49	CC	CC
RHA 273	0	CC	CC
RHA 274	91	CC	CC
RHA 278	227	CC	TT
RHA 331	108	CC	CC
RHA 374	101	CC	TT
RHA 377	45	CC	CC
RHA 386	55	CC	CC
RHA 388	65	CC	CC
RHA 396	0	CC	CC
RHA 400	160	CC	TT
RHA 408	55	CC	CC
RHA 468	118	/	/
RHA 855	13	CC	CC

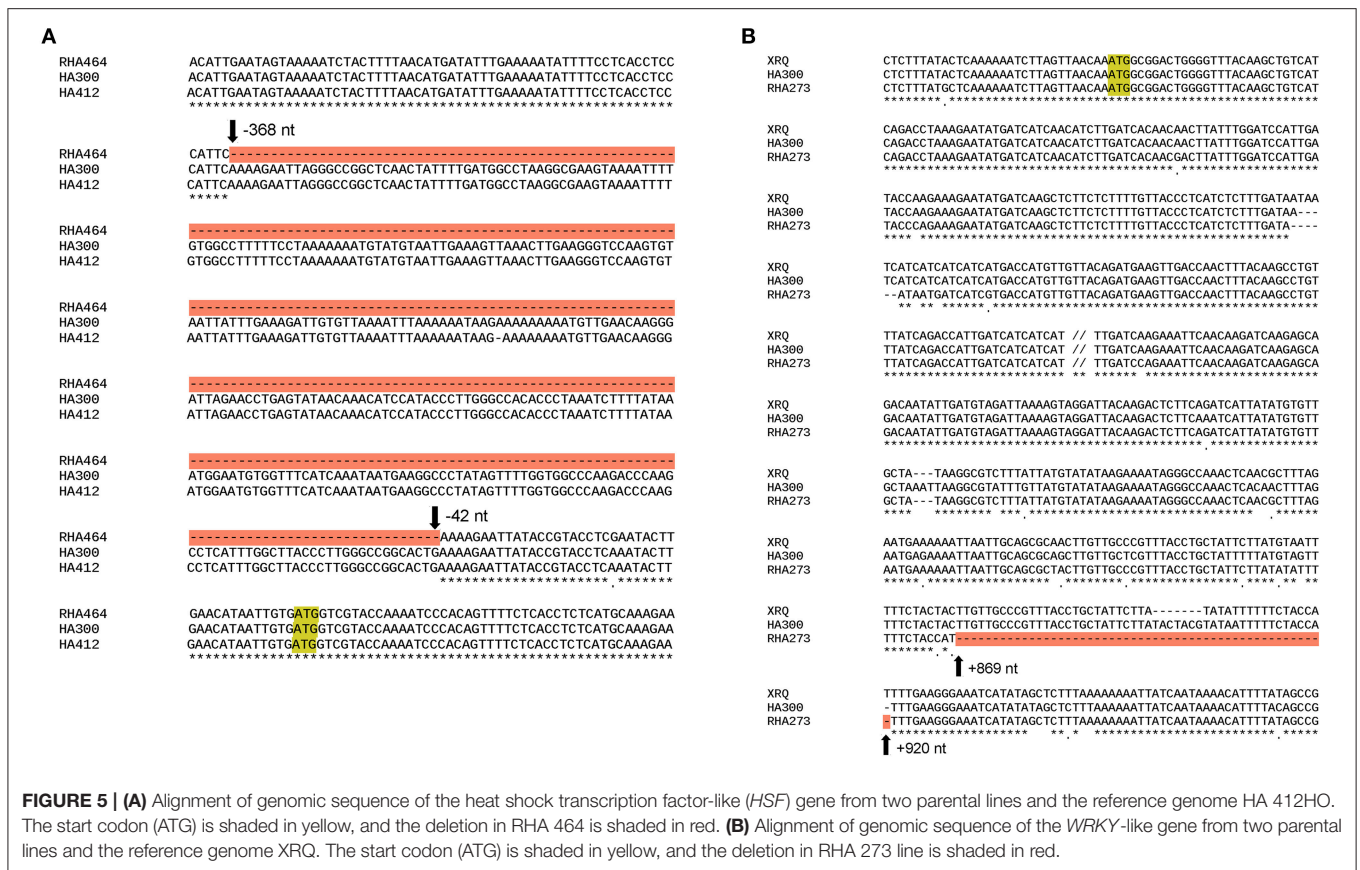
general, the simple interval mapping method can give reasonable results if the phenotypic data are not highly skewed (Broman and Sen, 2009). We also determined statistical significance on the basis of a genome-wide permutation test, which utilizes the same empirical distribution (Table S3). For these reasons, we used the original CGT numbers to perform QTL analysis without any data transformation.

The linkage map in the QTL analysis was constructed with SNP markers generated through GBS. GBS is a high-throughput and highly cost-effective genotyping and SNP

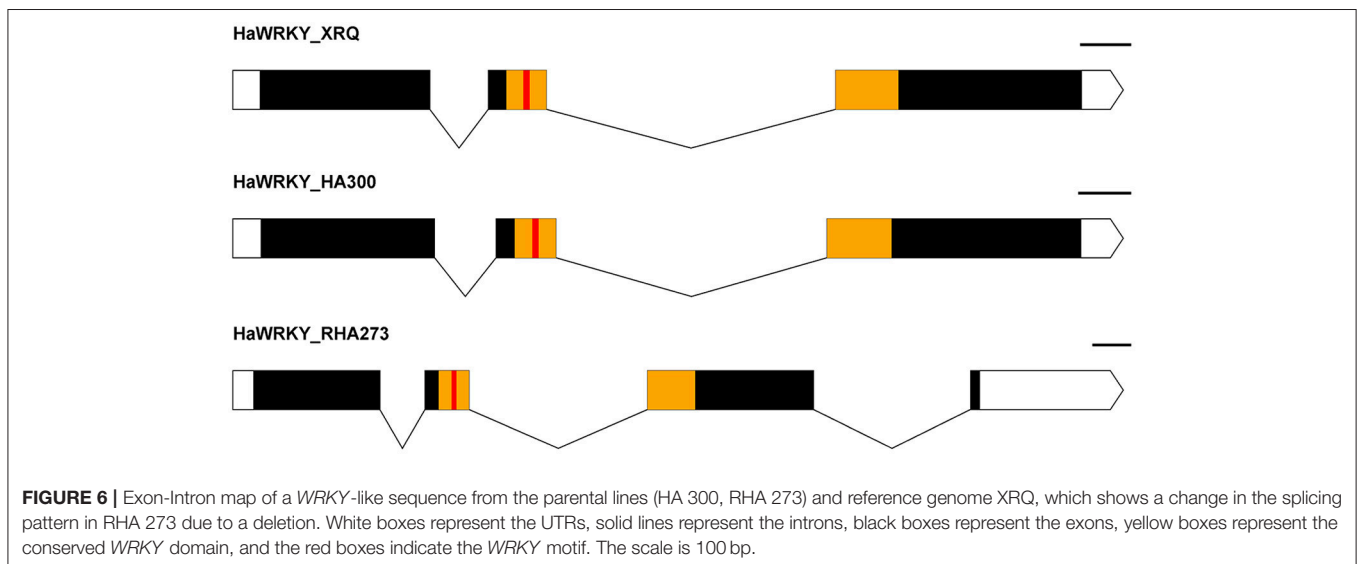
discovery approach, and has been applied to many plant species (Elshire et al., 2011; Kumar et al., 2012; Melo et al., 2016; Torkamaneh et al., 2016). To date, two GBS-SNP marker maps have been reported in sunflower (Celik et al., 2016; Talukder et al., 2016). In our study, we successfully mapped 2121 SNP markers, which is more than twice the number of unique SNP markers mapped in previous studies (Table 1). There are two possible reasons to explain why we were able to discover more unique SNP markers by GBS. First, we used a customized bioinformatic pipeline to process the GBS data. As described in our methods, we followed the GATK Best Practices to recalibrate the alignment, and used FreeBayes, a haplotype-based variant caller, to detect SNP markers (Garrison and Marth, 2012). The pipeline we used in our study could be valuable for discovering SNP markers, in addition to the TASSEL-GBS pipeline which was used in previous studies (Glaubitz et al., 2014; Celik et al., 2016; Talukder et al., 2016). Second, for our biparental mapping population, we carried out whole genome sequencing for the parent lines (HA 300 and RHA 464). These were included in the SNP marker calling and imputation steps together with their offspring, and this greatly helped to discover more reliable SNP markers. It is also a good practice to extract DNA from fresh young leaves or seedlings to obtain high quality DNA for GBS. With these modifications, we think GBS is still a good option for genotyping and SNP discovery in sunflower, especially considering the advantage of high-throughput and cost-effectiveness. The length of the genetic linkage map in this study was 1,519.52 cM, which is comparable with a sunflower consensus linkage map of 1,443.84 cM developed from three F2 mapping populations (Talukder et al., 2014), and also a recent SNP-based linkage map of 1,401.36 cM (Talukder et al., 2016). The marker density in this study (0.72 cM/SNP) is higher than the other two SNP marker maps (1.33 cM/SNP; Talukder et al., 2016; 3.03 cM/SNP; Celik et al., 2016). Despite good whole genome coverage, several regions with gaps greater than 10 cM were observed on chromosome 4, 6, and 14 (Table 1). Notably, the largest gap is 27.32 cM on chromosome 6, which is likely due to chromosome structure differences between the parents. Consequently, we were unable to detect SNP markers from these regions.

Two major QTL controlling CGT density in sunflower were identified in a biparental mapping population (Figure 3). One QTL was mapped to chromosome 5 (Ha5@14.6), which explained 11.61% of the CGT number variation. The second QTL was mapped to chromosome 6 (Ha6@60.5), which explained 14.06% of the CGT number variation (Table 2). No epistasis interaction was detected, and this suggests that these two QTL mediate CGT density in an additive manner. Similarly, two QTL associated with type VI glandular trichome density were identified in cultivated tomato, and seven QTL associated with trichome density were detected in soybean (Maliepaard et al., 1995; Du et al., 2009). These findings suggest that the trichome density in plants is regulated by multiple genes. A large portion of the CGT number variation is still unaccounted for in the current study, and this could be due to some error in genotyping, or a confluence of several genes of minor effect.





**FIGURE 5 | (A)** Alignment of genomic sequence of the heat shock transcription factor-like (*HSF*) gene from two parental lines and the reference genome HA 412HO. The start codon (ATG) is shaded in yellow, and the deletion in RHA 464 is shaded in red. **(B)** Alignment of genomic sequence of the *WRKY*-like gene from two parental lines and the reference genome XRQ. The start codon (ATG) is shaded in yellow, and the deletion in RHA 273 line is shaded in red.



**FIGURE 6 |** Exon-Intron map of a *WRKY*-like sequence from the parental lines (HA 300, RHA 273) and reference genome XRQ, which shows a change in the splicing pattern in RHA 273 due to a deletion. White boxes represent the UTRs, solid lines represent the introns, black boxes represent the exons, yellow boxes represent the conserved *WRKY* domain, and the red boxes indicate the *WRKY* motif. The scale is 100 bp.

To validate the QTL results, we selected 39 sunflower inbred lines with diverse genetic backgrounds as the validation population (Figure S4). We used multiple linear regression analysis to validate the QTL-trait association and also estimate the QTL effects. A total of 385 SNP markers were selected from the 1.5-LOD support interval on chromosome 5 and 33

SNP markers were selected from the QTL support interval on chromosome 6. Based on annotated gene functions, one SNP marker pair, Ha5\_10149906 and Ha6\_7633946, was identified as the best two-QTL model and corresponded to genes with a plausible effect on the phenotype. This two-QTL model explained 47.2% of phenotypic variation in the validation population.

Moreover, the significant associations between single SNP markers (Ha5\_10149906 and Ha6\_7633946) and CGT number were also supported by single factor ANOVA (**Figures 4C,D**). Thus, the QTL effects and the association between CGT density and QTL support intervals were confirmed in validation.

The SNP marker Ha5\_10149906 is located within the second exon of the gene Ha5g003120. A 326 bp deletion was observed in the promoter region of the gene Ha5g003120 in RHA 464, and this deletion might lead to substantial change in gene expression pattern (**Figure 5A**). The gene Ha5g003120 is annotated as a member of heat shock transcription factor (HSF) family. Although no experimental results have shown that HSF proteins are involved in trichome development, some studies indicate these proteins are required for cell division and root growth (Westerheide et al., 2012). The SNP marker Ha6\_7633946 is located in the 3'-UTR of the gene Ha6g003560 and is associated with an alternative splice site in RHA 273, due to a 51 bp deletion in the second intron. This alternative splicing alters the *WRKY* domain, which determines DNA-binding specificity (Llorca et al., 2014). The phylogenetic tree showed that Ha6g003560 is grouped together with *Arabidopsis thaliana* gene *WRKY44/TTG2*. *WRKY44/TTG2* is a key regulatory gene for trichome development, and a mutation in *WRKY44/TTG2* causes significantly reduced trichome number and unbranched trichomes in *Arabidopsis* (Johnson, 2002; Ishida et al., 2007; Pesch et al., 2014). Taken together, the *HSF* gene Ha5g003120 and *WRKY* gene Ha6g003560 are strong candidates for regulating CGT density in sunflower florets. It is possible, however, that other adjacent genes are contributing to the phenotype. Further studies are required to characterize gene functions in these regions.

In summary, we successfully detected SNP markers by GBS and constructed a genetic linkage map with these SNP markers. We also identified two major QTL controlling CGT density in sunflower florets by using the F<sub>4</sub> population derived from the cross HA 300 × RHA 464. In addition, we found two plausible candidate genes in the QTL support intervals. Future work will focus on optimizing STL chemical composition in CGT for enhancing host resistance to sunflower insect pests.

## REFERENCES

- Agren, J., and Schemske, D. (1992). Artificial selection on trichome number in *Brassica rapa*. *Theor. Appl. Genet.* 83, 673–678. doi: 10.1007/BF00226683
- Amrehn, E., Aschenbrenner, A. K., Heller, A., and Spring, O. (2015). Localization of sesquiterpene lactone biosynthesis in cells of capitate glandular trichomes of *Helianthus annuus* (Asteraceae). *Protoplasma* 253, 447–455. doi: 10.1007/s00709-015-0823-4
- Amrehn, E., Heller, A., and Spring, O. (2013). Capitate glandular trichomes of *Helianthus annuus* (Asteraceae): ultrastructure and cytological development. *Protoplasma* 251, 161–167. doi: 10.1007/s00709-013-0534-7
- Andrade, M. C., Silva, A. A. D., Neiva, I. P., Oliveira, I. R. C., Castro, E. M. D., Francis, D. M., et al. (2017). Inheritance of type IV glandular trichome density and its association with whitefly resistance from *Solanum galapagense* accession LA1401. *Euphytica* 213:52. doi: 10.1007/s10681-016-1792-1

## AUTHOR NOTE

Mention of trade names or commercial products in this report is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. The USDA is an equal opportunity provider and employer.

## AUTHOR CONTRIBUTIONS

Q-MG analyzed the data and wrote the paper. NK, BH, and JP designed and led the genomics analyses, population development and genetics analyses, and trait evaluations, respectively. SR conducted phylogenetic analyses of the candidate genes. CP and ST conducted GBS and whole genome shotgun sequencing and bioinformatic analyses. All authors have read and agree with the contents of the manuscript.

## FUNDING

National Sunflower Association, award # 14-E03.

## ACKNOWLEDGMENTS

The authors would like to acknowledge Jayma Moore and Scott Payne at NDSU Electron Microscopy Center for their help in obtaining SEM images of florets, and Jamie Miller-Dunbar for help in quantifying CGT. We sincerely thank Brady Koehler, Mike Grove and Brian Smart for collecting DNA samples, greenhouse operations and support in the laboratory. We thank Brian Smart and Dr. Gerald Seiler for critically reviewing the manuscript. We also thank Jamie Prior Kershner at the BioFrontiers NGS Core Facility for assistance with Illumina sequencing.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.02227/full#supplementary-material>

- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Aschenbrenner, A. K., Amrehn, E., Bechtel, L., and Spring, O. (2014). Trichome differentiation on leaf primordia of *Helianthus annuus* (Asteraceae): morphology, gene expression and metabolite profile. *Planta* 241, 837–846. doi: 10.1007/s00425-014-2223-y
- Aschenbrenner, A. K., Horakh, S., and Spring, O. (2013). Linear glandular trichomes of *Helianthus* (Asteraceae): morphology, localization, metabolite activity and occurrence. *AoB Plants* 5:plt028. doi: 10.1093/aobpla/plt028
- Aschenbrenner, A.-K., Kwon, M., Conrad, J., Ro, D.-K., and Spring, O. (2016). Identification and characterization of two bisabolene synthases from linear glandular trichomes of sunflower (*Helianthus annuus* L., Asteraceae). *Phytochemistry* 124, 29–37. doi: 10.1016/j.phytochem.2016.01.009
- Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188. doi: 10.1214/aos/1013699998

- Benz, B. W., and Martin, C. E. (2006). Foliar trichomes, boundary layers, and gas exchange in 12 species of epiphytic *Tillandsia* (Bromeliaceae). *J. Plant Physiol.* 163, 648–656. doi: 10.1016/j.jplph.2005.05.008
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bombo, A. B., Appezzato-Da-Glória, B., Aschenbrenner, A. K., and Spring, O. (2016). Capitate glandular trichomes in *Aldama discolor* (Heliantheae - Asteraceae): morphology, metabolite profile and sesquiterpene biosynthesis. *Plant Biol.* 18, 455–462. doi: 10.1111/plb.12423
- Broman, K. W. (2003). Mapping quantitative trait loci in the case of a spike in the phenotype distribution. *Genetics* 163, 1169–1175. Available online at: <http://www.genetics.org/content/163/3/1169>
- Broman, K. W., and Sen, S. (2009). *A Guide to QTL Mapping with R/qtl*. Dordrecht: Springer.
- Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19, 889–890. doi: 10.1093/bioinformatics/btg112
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354
- Celik, I., Bodur, S., Frary, A., and Doganlar, S. (2016). Genome-wide SNP discovery and genetic linkage map construction in sunflower (*Helianthus annuus* L.) using a genotyping by sequencing (GBS) approach. *Mol. Breed.* 36, 133–142. doi: 10.1007/s11032-016-0558-8
- Chadwick, M., Trewin, H., Gawthrop, F., and Wagstaff, C. (2013). Sesquiterpenoids lactones: Benefits to plants and people. *Int. J. Mol. Sci.* 14, 12780–12805. doi: 10.3390/ijms140612780
- Covello, P. S., Teoh, K. H., Polichuk, D. R., Reed, D. W., and Nowak, G. (2007). Functional genomics and the biosynthesis of artemisinin. *Phytochemistry* 68, 1864–1871. doi: 10.1016/j.phytochem.2007.02.016
- Cui, H., Zhang, S. T., Yang, H. J., Ji, H., and Wang, X. J. (2011). Gene expression profile analysis of tobacco leaf trichomes. *BMC Plant Biol.* 11:76. doi: 10.1186/1471-2229-11-76
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Danielson, S. D., Manglitz, G. R., and Sorensen, E. L. (1990). Field resistance of perennial glandular-haired *Medicago* strains and alfalfa cultivars to the alfalfa weevil. *Euphytica* 49, 177–182. doi: 10.1007/BF00027268
- Du, W. J., Yu, D. Y., and Fu, S. X. (2009). Analysis of QTLs for the trichome density on the upper and downer surface of leaf blade in soybean [*Glycine max* (L.) Merr.]. *Agric. Sci. China* 8, 529–537. doi: 10.1016/S1671-2927(08)60243-6
- Duarte, M. d. R., and Empinotti, C. B. (2012). Leaf and stem microscopic identification of *Tithonia diversifolia* (Hemsl.) A. Gray (Asteraceae). *Brazilian J. Pharm. Sci.* 48, 109–116. doi: 10.1590/S1984-82502012000100013
- Eljounaidi, K., Cankar, K., Comino, C., Moglia, A., Hehn, A., Bourgaud, F., et al. (2014). Cytochrome P450s from *Cynara cardunculus* L. CYP71AV9 and CYP71BL5, catalyze distinct hydroxylations in the sesquiterpene lactone biosynthetic pathway. *Plant Sci.* 223, 59–68. doi: 10.1016/j.plantsci.2014.03.007
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. doi: 10.1371/journal.pone.0019379
- Fernandes, E., Pacheco, A., and Penha-Gonçalves, C. (2007). Mapping of quantitative trait loci using the skew-normal distribution. *J. Zhejiang Univ. Sci. B* 8, 792–801. doi: 10.1631/jzus.2007.B0792
- Fraga, B. M. (2005). Natural sesquiterpenoids. *Nat. Prod. Rep.* 22, 465. doi: 10.1039/b501837b
- Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907* [q-bio.GN]
- Ghantous, A., Gali-Muhtasib, H., Vuorela, H., Saliba, N. A., and Darwiche, N. (2010). What made sesquiterpene lactones reach cancer clinical trials? *Drug Discov. Today* 15, 668–678. doi: 10.1016/j.drudis.2010.06.002
- Glas, J., Schimmel, B., Alba, J., Escobar-Bravo, R., Schuurink, R., and Kant, M. (2012). Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores. *Int. J. Mol. Sci.* 13, 17077–17103. doi: 10.3390/ijms131217077
- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., et al. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 9:e90346. doi: 10.1371/journal.pone.0090346
- Göpfert, J. C., Heil, N., Conrad, J., and Spring, O. (2005). Cytological development and sesquiterpene lactone secretion in capitate glandular trichomes of Sunflower. *Plant Biol.* 7, 148–155. doi: 10.1055/s-2005-837575
- Göpfert, J. C., Macnevin, G., Ro, D.-K., and Spring, O. (2009). Identification, functional characterization and developmental regulation of sesquiterpene synthases from sunflower capitate glandular trichomes. *BMC Plant Biol.* 9:86. doi: 10.1186/1471-2229-9-86
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Han, M. V., and Zmasek, C. M. (2009). phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* 10:356. doi: 10.1186/1471-2105-10-356
- Hare, J. D., and Elle, E. (2002). Variable impact of diverse insect herbivores on dimorphic *Datura wrightii*. *Ecology* 83, 2711. doi: 10.1890/0012-9658(2002)083[2711:VIODIH]2.0.CO;2
- Howe, G. A., and Jander, G. (2008). Plant immunity to insect herbivores. *Annu. Rev. Plant Biol.* 59, 41–66. doi: 10.1146/annurev.arplant.59.032607.092825
- Hulke, B. S., Miller, J. F., and Gulya, T. J. (2010). Registration of the restorer oilseed sunflower germplasm RHA 464 possessing genes for resistance to downy mildew and sunflower rust. *J. Plant Reg.* 4, 249–254. doi: 10.3198/jpr2010.02.0114crg
- INRA Sunflower Bioinformatics Resources (2014). *HA412-HO Genome Portal (Bronze version)*. Available online at: <https://www.heliagene.org/HA412.v1.1.bronze.20141015/>
- Ishida, T., Hattori, S., Sano, R., Inoue, K., Shirano, Y., Hayashi, H., et al. (2007). Arabidopsis TRANSPARENT TESTA GLABRA2 is directly regulated by R2R3 MYB transcription factors and is involved in regulation of GLABRA2 transcription in epidermal differentiation. *Plant Cell Online* 19, 2531–2543. doi: 10.1105/tpc.107.052274
- Isman, M. B. (1985). Toxicity and tolerance of sesquiterpene lactones in the migratory grasshopper, *Melanoplus sanguinipes* (Acrididae). *Pestic. Biochem. Physiol.* 24, 348–354. doi: 10.1016/0048-3575(85)90146-4
- Johnson, C. S. (2002). TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of arabidopsis, encodes a WRKY transcription factor. *Plant Cell Online* 14, 1359–1375. doi: 10.1105/tpc.001404
- Kane, N. (2017). *Tri Imputation Tools for GBS Data: Phaser2.pl and genotyper7.pl*. Available online at: [https://github.com/KaneLab/Ref\\_guided\\_genotyping](https://github.com/KaneLab/Ref_guided_genotyping)
- Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kitch, L. W., Shade, R. E., Nyquist, W. E., and Axtell, J. D. (1985). Inheritance of density of erect glandular trichomes in the Genus *Medicago*. *Crop Sci.* 25:607. doi: 10.2135/cropsci1985.0011183X002500040007x
- Krimmel, B. A., and Pearse, I. S. (2013). Sticky plant traps insects to enhance indirect defence. *Ecol. Lett.* 16, 219–224. doi: 10.1111/ele.12032
- Kumar, S., Banks, T. W., and Cloutier, S. (2012). SNP discovery through Next-Generation Sequencing and its applications. *Int. J. Plant Genomics* 2012, 1–15. doi: 10.1155/2012/831460
- Larkin, J. C. (1994). Roles of the GLABROUS1 and TRANSPARENT TESTA GLABRA genes in Arabidopsis trichome development. *Plant Cell* 6, 1065–1076. doi: 10.1105/tpc.6.8.1065
- Lee, T. H., Guo, H., Wang, X., Kim, C., and Paterson, A. H. (2014). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* 15:162. doi: 10.1186/1471-2164-15-162
- Lefort, V., Longueville, J. E., and Gascuel, O. (2017). SMS: smart model selection in PhyML. *Mol. Biol. Evol.* 34, 2422–2424. doi: 10.1093/molbev/msx149
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li-Weber, M., Gaiasi, M., Treiber, M. K., and Krammer, P. H. (2002). The anti-inflammatory sesquiterpene lactone parthenolide suppresses CD95 gene expression in peripheral blood T cells. *Cell Death Differ.* 9, 1256–1265. doi: 10.1038/sj.cdd.4401102



- Llorca, C. M., Potschin, M., and Zentgraf, U. (2014). bZIPs and WRKYs: two large transcription factor families executing two different functional strategies. *Front. Plant Sci.* 5:169. doi: 10.3389/fpls.2014.00169
- Lusa, M. G., Cardoso, E. C., Machado, S. R., and Appezzato-Da-Gloria, B. (2014). Trichomes related to an unusual method of water retention and protection of the stem apex in an arid zone perennial species. *AoB Plants* 7:plu088. doi: 10.1093/aobpla/plu088
- Lv, Z., Zhang, L., and Tang, K. (2017). New insights into artemisinin regulation. *Plant Signal. Behav.* 12:e1366398. doi: 10.1080/15592324.2017.1366398
- Mabry, T. J., Gill, J. E., Burnett, W. C., and Jones, S. B. (1977). "Antifeedant sesquiterpene lactones in the Compositae," in *ACS Symposium Series Host Plant Resistance to Pests* (Washington, DC: American Chemical Society), 179–184.
- Maliepaard, C., Bas, N., Heusden, S. V., Kos, J., Pet, G., Verkerk, R., et al. (1995). Mapping of QTLs for glandular trichome densities and *Trialeurodes vaporariorum* (greenhouse whitefly) resistance in an F2 from *Lycopersicon esculentum* × *Lycopersicon hirsutum* f. *glabratum*. *Heredity* 75, 425–433. doi: 10.1038/hdy.1995.155
- Mandel, J. R., Nambeesan, S., Bowers, J. E., Marek, L. F., Ebert, D., Rieseberg, L. H., et al. (2013). Association mapping and the genomic consequences of selection in sunflower. *PLoS Genet.* 9:e1003378. doi: 10.1371/journal.pgen.1003378
- Melo, A. T. O., Bartaula, R., and Hale, I. (2016). GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC Bioinformatics* 17:29. doi: 10.1186/s12859-016-0879-y
- Menin, B., Comino, C., Portis, E., Moglia, A., Cankar, K., Bouwmeester, H. J., et al. (2012). Genetic mapping and characterization of the globe artichoke (+)-germacrene A synthase gene, encoding the first dedicated enzyme for biosynthesis of the bitter sesquiterpene lactone cynaropicrin. *Plant Sci.* 190, 1–8. doi: 10.1016/j.plantsci.2012.03.006
- Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 2010:pdb.prot5448. doi: 10.1101/pdb.prot5448
- Michalska, K., Kisiel, W., and Zidorn, C. (2013). Sesquiterpene lactones from *Crepis aurea* (Asteraceae, Cichorieae). *Biochem. Syst. Ecol.* 46, 1–3. doi: 10.1016/j.bse.2012.08.005
- Moraes, C. M. D., Lewis, W. J., Paré, P. W., Alborn, H. T., and Tumlinson, J. H. (1998). Herbivore-infested plants selectively attract parasitoids. *Nature* 393, 570–573. doi: 10.1038/31219
- Nasim, S., Pei, S., Hagen, F. K., Jordan, C. T., and Crooks, P. A. (2011). Melampomagnolide B: a new antileukemic sesquiterpene. *Bioorganic Med. Chem.* 19, 1515–1519. doi: 10.1016/j.bmc.2010.12.045
- Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2015). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32, 292–294. doi: 10.1093/bioinformatics/btv566
- Oppenheimer, D. G., Herman, P. L., Sivakumaran, S., Esch, J., and Marks, M. (1991). A myb gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell* 67, 483–493. doi: 10.1016/0092-8674(91)90523-2
- Paré, P. W., and Tumlinson, J. H. (1999). Plant volatiles as a defense against insect herbivores. *Plant Physiol.* 121, 325–332. doi: 10.1104/pp.121.2.325
- Pattanaik, S., Patra, B., Singh, S. K., and Yuan, L. (2014). An overview of the gene regulatory network controlling trichome development in the model plant, *Arabidopsis*. *Front. Plant Sci.* 5:259. doi: 10.3389/fpls.2014.00259
- Peiffer, M., Tooker, J. F., Luthe, D. S., and Felton, G. W. (2009). Plants on early alert: glandular trichomes as sensors for insect herbivores. *New Phytologist* 184, 644–656. doi: 10.1111/j.1469-8137.2009.03002.x
- Pesch, M., Dartan, B., Birkenbihl, R., Somssich, I. E., and Hulskamp, M. (2014). *Arabidopsis* TTG2 regulates TRY expression through enhancement of activator complex-triggered activation. *Plant Cell* 26, 4067–4083. doi: 10.1105/tpc.114.129379
- Phukan, U. J., Jeena, G. S., and Shukla, R. K. (2016). WRKY transcription factors: molecular regulation and stress responses in plants. *Front. Plant Sci.* 7:760. doi: 10.3389/fpls.2016.00760
- Pillemer, E. A., and Tingey, W. M. (1976). Hooked trichomes: a physical plant barrier to a major agricultural pest. *Science* 193, 482–484. doi: 10.1126/science.193.4252.482
- Prasifka, J. R. (2015). Variation in the number of capitate glandular trichomes in wild and cultivated sunflower germplasm and its potential for use in host plant resistance. *Plant Genet. Resour.* 13, 68–74. doi: 10.1017/S1479262114000653
- Prasifka, J. R., Spring, O., Conrad, J., Cook, L. W., Palmquist, D. E., and Foley, M. E. (2015). Sesquiterpene lactone composition of wild and cultivated sunflowers and biological activity against an insect pest. *J. Agric. Food Chem.* 63, 4042–4049. doi: 10.1021/acs.jafc.5b00362
- Rautio, P., Markkola, A., Martel, J., Tuomi, J., Harma, E., Kuikka, K., et al. (2002). Developmental plasticity in birch leaves: defoliation causes a shift from glandular to nonglandular trichomes. *Oikos* 98, 437–446. doi: 10.1034/j.1600-0706.2002.980308.x
- Reis, C. D., Sajo, M. D. G., and Stehmann, J. R. (2002). Leaf structure and taxonomy of *Petunia* and *Calibrachoa* (Solanaceae). *Braz. Arch. Biol. Technol.* 45, 59–66. doi: 10.1590/S1516-89132002000100010
- Riddick, E. W., and Wu, Z. (2011). Lima bean–lady beetle interactions: hooked trichomes affect survival of *Stethorus punctillum* larvae. *BioControl* 56, 55–63. doi: 10.1007/s10526-010-9309-7
- Rogers, C. E., Gershenzon, J., Ohno, N., Mabry, T. J., Stipanovic, R. D., and Kreitner, G. L. (1987). Terpenes of wild sunflowers (*Helianthus*): an effective mechanism against seed predation by larvae of the sunflower moth, *Homoeosoma electellum* (Lepidoptera: Pyralidae). *Environ. Entomol.* 16, 586–592. doi: 10.1093/ee/16.3.586
- Rossiter, M., Gershenzon, J., and Mabry, T. J. (1986). Behavioral and growth responses of specialist herbivore, *Homoeosoma electellum*, to major terpenoid of its host, *Helianthus* SPP. *J. Chem. Ecol.* 12, 1505–1521. doi: 10.1007/BF01012368
- Rowe, H. C., Ro, D. K., and Rieseberg, L. H. (2012). Response of sunflower (*Helianthus annuus* L.) leaf surface defenses to exogenous methyl jasmonate. *PLoS ONE* 7:e37191. doi: 10.1371/journal.pone.0037191
- Serna, L., and Martin, C. (2006). Trichomes: different regulatory networks lead to convergent structures. *Trends Plant Sci.* 11, 274–280. doi: 10.1016/j.tplants.2006.04.008
- Shapiro, S. S., and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611. doi: 10.1093/biomet/52.3.4.591
- Sigrist, C. J. A., Castro, E. D., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., et al. (2012). New and continuing developments at PROSITE. *Nucleic Acids Res.* 41, D344–D347. doi: 10.1093/nar/gks1067
- Simmons, A. T., and Gurr, G. M. (2005). Trichomes of *lycopersicon* species and their hybrids: effects on pests and natural enemies. *Agric. For. Entomol.* 7, 265–276. doi: 10.1111/j.1461-9555.2005.00271.x
- Skaltsa, H., Vervikokidou, E., Harvala, C., Karabourniotis, G., and Manetasi, Y. (1994). UV-B protective potential and flavonoid content of leaf hairs of *Quercus ilex*. *Phytochemistry* 37, 987–990. doi: 10.1016/S0031-9422(00)89514-X
- Spring, O. (2000). Chemotaxonomy based on metabolites from glandular trichomes. *Adv. Bot. Res.* 31, 153–174. doi: 10.1016/S0065-2296(00)31009-6
- Spring, O., and Bienert, U. (1987). Capitate glandular hairs from sunflower leaves: development, distribution and sesquiterpene lactone content. *J. Plant Physiol.* 130, 441–448. doi: 10.1016/S0176-1617(87)80209-2
- Szymanski, D. B., Lloyd, A. M., and Marks, M. (2000). Progress in the molecular genetic analysis of trichome initiation and morphogenesis in *Arabidopsis*. *Trends Plant Sci.* 5, 214–219. doi: 10.1016/S1360-1385(00)01597-1
- Talukder, Z. I., Gong, L., Hulke, B. S., Pegadaraju, V., Song, Q., Schultz, Q., et al. (2014). A high-density SNP map of sunflower derived from RAD-sequencing facilitating fine-mapping of the rust resistance gene R12. *PLoS ONE* 9:e98628. doi: 10.1371/journal.pone.0098628
- Talukder, Z. I., Seiler, G. J., Song, Q., Ma, G., and Qi, L. (2016). SNP discovery and QTL mapping of *Sclerotinia* basal stalk rot resistance in Sunflower using Genotyping-by-Sequencing. *Plant Genome* 9, 1–16. doi: 10.3835/plantgenome2016.03.0035
- Tian, D., Tooker, J., Peiffer, M., Chung, S. H., and Felton, G. W. (2012). Role of trichomes in defense against herbivores: comparison of herbivore response to woolly and hairless trichome mutants in tomato (*Solanum lycopersicum*). *Planta* 236, 1053–1066. doi: 10.1007/s00425-012-1651-9
- Tissier, A. (2012). Glandular trichomes: what comes after expressed sequence tags? *Plant J.* 70, 51–68. doi: 10.1111/j.1365-313x.2012.04913.x
- Torkamaneh, D., Laroche, J., and Belzile, F. (2016). Genome-Wide SNP Calling from genotyping by sequencing (GBS) data: a comparison of



- seven pipelines and two sequencing technologies. *PLoS ONE* 11:e0161333. doi: 10.1371/journal.pone.0161333
- Wagner, G. J. (1991). Secreting glandular trichomes: more than just hairs. *Plant Physiol.* 96, 675–679. doi: 10.1104/pp.96.3.675
- Wang, W., Wang, Y., Zhang, Q., Qi, Y., and Guo, D. (2009). Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing. *BMC Genomics* 10:465. doi: 10.1186/1471-2164-10-465
- War, A. R., Paulraj, M. G., Ahmad, T., Buhroo, A. A., Hussain, B., Ignacimuthu, S., et al. (2012). Mechanisms of plant defense against insect herbivores. *Plant Signal. Behav.* 7, 1306–1320. doi: 10.4161/psb.21663
- Werker, E. (2000). Trichome diversity and development. *Adv. Bot. Res.* 31, 1–35. doi: 10.1016/S0065-2296(00)31005-9
- Westerheide, S. D., Raynes, R., Powell, C., Xue, B., and Uversky, V. N. (2012). HSF transcription factor family, heat shock response, and protein intrinsic disorder. *Curr. Protein Pept. Sci.* 13, 86–103. doi: 10.2174/138920312799277956
- Yang, C., and Ye, Z. (2012). Trichomes as models for studying plant cell differentiation. *Cell. Mol. Life Sci.* 70, 1937–1948. doi: 10.1007/s00018-012-1147-6
- Zidorn, C. (2008). Sesquiterpene lactones and their precursors as chemosystematic markers in the tribe Cichorieae of the Asteraceae. *Phytochemistry* 69, 2270–2296. doi: 10.1016/j.phytochem.2008.06.013

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Gao, Kane, Hulke, Reinert, Pogoda, Tittes and Prasifka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# An Oleuropein $\beta$ -Glucosidase from Olive Fruit Is Involved in Determining the Phenolic Composition of Virgin Olive Oil

David Velázquez-Palmero<sup>1,2</sup>, Carmen Romero-Segura<sup>1</sup>, Rosa García-Rodríguez<sup>1</sup>, María L. Hernández<sup>1</sup>, Fabián E. Vaistij<sup>2</sup>, Ian A. Graham<sup>2</sup>, Ana G. Pérez<sup>1</sup> and José M. Martínez-Rivas<sup>1\*</sup>

<sup>1</sup> Department of Biochemistry and Molecular Biology of Plant Products, Instituto de la Grasa (CSIC), Sevilla, Spain, <sup>2</sup> Centre for Novel Agricultural Products, Department of Biology, University of York, York, United Kingdom

## OPEN ACCESS

### Edited by:

Mariela Torres,  
Instituto Nacional de Tecnología  
Agropecuaria (INTA), Argentina

### Reviewed by:

Rosario Muleo,  
Università degli Studi della Tuscia, Italy  
Benedetto Ruperti,  
Università degli Studi di Padova, Italy

### \*Correspondence:

José M. Martínez-Rivas  
mrivas@cica.es

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

Received: 28 July 2017

Accepted: 20 October 2017

Published: 07 November 2017

### Citation:

Velázquez-Palmero D,  
Romero-Segura C,  
García-Rodríguez R, Hernández ML,  
Vaistij FE, Graham IA, Pérez AG and  
Martínez-Rivas JM (2017) An  
Oleuropein  $\beta$ -Glucosidase from Olive  
Fruit Is Involved in Determining  
the Phenolic Composition of Virgin  
Olive Oil. *Front. Plant Sci.* 8:1902.  
doi: 10.3389/fpls.2017.01902

Phenolic composition of virgin olive oil is determined by the enzymatic and/or chemical reactions that take place during olive fruit processing. Of these enzymes,  $\beta$ -glucosidase activity plays a relevant role in the transformation of the phenolic glycosides present in the olive fruit, generating different secoiridoid derivatives. The main goal of the present study was to characterize olive fruit  $\beta$ -glucosidase genes and enzymes responsible for the phenolic composition of virgin olive oil. To achieve that, we have isolated an olive  $\beta$ -glucosidase gene from cultivar Picual (*OepGLU*), expressed in *Nicotiana benthamiana* leaves and purified its corresponding recombinant enzyme. Western blot analysis showed that recombinant *OepGLU* protein is detected by an antibody raised against the purified native olive mesocarp  $\beta$ -glucosidase enzyme, and exhibits a deduced molecular mass of 65.0 kDa. The recombinant *OepGLU* enzyme showed activity on the major olive phenolic glycosides, with the highest levels with respect to oleuropein, followed by ligstroside and demethyloleuropein. In addition, expression analysis showed that olive *GLU* transcript level in olive fruit is spatially and temporally regulated in a cultivar-dependent manner. Furthermore, temperature, light and water regime regulate olive *GLU* gene expression in olive fruit mesocarp. All these data are consistent with the involvement of *OepGLU* enzyme in the formation of the major phenolic compounds present in virgin olive oil.

**Keywords:**  $\beta$ -Glucosidase, *Olea europaea*, oleuropein, olive fruit, phenolic compounds, virgin olive oil

## INTRODUCTION

Olive (*Olea europaea* L.) is one of the first plants grown as oil crop. Consequently, olive oil is one of the oldest known plant oils and it can be consumed as virgin olive oil (VOO). In the Mediterranean diet, this oil constitutes the main lipid source and it has been related with several beneficial nutritional properties which are mainly associated to its phenolic components (Konstantinidou et al., 2010; Visioli and Bernardini, 2011). However, phenolic compounds are relevant not only because their nutritional properties, but also due to their organoleptic characteristics. In fact, phenolic components are involved in the pungent and bitter sensory notes of VOO (Andrewes et al., 2003; Mateos et al., 2004). Phenolic compounds are being currently used as a trait in new

cross breeding programs (León et al., 2011), and also as VOO quality markers, because of their health promoting and organoleptic properties.

Oleuropein, demethyloleuropein and ligstroside, the most significant phenolic glycosides detected in the olive fruit, belong to the secoiridoids class, a group of monoterpenoids typical of the Oleaceae family with a cleaved methylcyclopentane skeleton (Obied et al., 2008). On the contrary, the main phenolic compounds detected in VOO are the secoiridoid derivatives, resulting from the enzymatic hydrolysis of these olive fruit glycosides. Specifically, the aldehydic forms of oleuropein and ligstroside aglycones (3,4-DHPEA-EA and *p*-HPEA-EA, respectively), and the dialdehydic forms of decarboxymethyloleuropein and ligstroside aglycones (3,4-DHPEA-EDA and *p*-HPEA-EDA, respectively) (Montedoro et al., 2002). Oleuropein derivatives exhibit the highest antioxidant activity (Ramos-Escudero et al., 2015), protein-denaturing/protein-cross-linking properties (Konno et al., 1999), cytotoxic effects (Bernini et al., 2011) and effectivity as chronic disease preventive agents (Pinto et al., 2011).

The phenolic profile of VOO is mainly derived from the amount of phenolic glycosides originally found in the tissues of olive fruit and the activity of oxidative and hydrolytic enzymes operating on these glycosides during VOO processing (García-Rodríguez et al., 2011; Romero-Segura et al., 2012). Although secoiridoids biosynthesis and degradation pathways are still not fully understood (Obied et al., 2008), hydrolysis by highly specific  $\beta$ -glucosidases seems to be critical for the diverse roles attributed to secoiridoid derivatives. In this sense, the wide array of physiological roles assigned to plants  $\beta$ -glucosidases ( $\beta$ -d-glucoside glucohydrolases, EC 3.2.1.21), such as functions in plant secondary metabolism, symbiosis, defense, signaling, and cell wall lignification and catabolism, are determined by their tissue and subcellular localization, and their substrate-specificities (Cairns and Esen, 2010).

The existence of various  $\beta$ -glucosidase isoforms in olive was first reported by Mazzuca et al. (2006), who described the localization of two isoforms of oleuropein-degradative  $\beta$ -glucosidases in the oil droplets and in the chloroplasts of mesocarp of green olive fruits. Transcriptomic (Alagna et al., 2012) and proteomic (Bianco et al., 2013) studies confirm that olive, similar to most plants, possesses several distinct  $\beta$ -glucosidases. Recently, the isolation and characterization of a defense-related  $\beta$ -glucosidase gene from olive (cv. Koroneiki) has been described (Koudounas et al., 2015). Nearly all these earlier reports on this enzyme have been centered in its physiological function as a defense mechanism which specifically generates oleuropein-derived compounds with established antimicrobial activities. In contrast, no similar studies have been carried out on the  $\beta$ -glucosidase genes/enzymes in relation to the VOO quality; despite that this knowledge may be very valuable to enhance marker assisted breeding programs to obtain new varieties with tailored oil quality characteristics.

We have previously isolated and purified to apparent homogeneity a protein with  $\beta$ -glucosidase activity from olive fruit mesocarp which exhibits high activity with the main phenolic glycoside in olive fruit, oleuropein, and gives rise to one of the

most important phenolic compounds in VOO (3,4-DHPEA-EA) as the main reaction product (Romero-Segura et al., 2009). Data on the  $\beta$ -glucosidase activity during ripening of olive fruit from cultivars Arbequina and Picual are in good agreement with the phenolic composition of the oils obtained from fruits with different degrees of maturity (Romero-Segura et al., 2012).

The objective of this study was to characterize olive fruit  $\beta$ -glucosidase genes and enzymes responsible for the phenolic composition of VOO. Thus, we have isolated an olive  $\beta$ -glucosidase gene from cultivar Picual, which codes for an enzyme that displays the highest activity toward oleuropein. The immunological and catalytic properties of this olive  $\beta$ -glucosidase enzyme, together with its expression data, are in agreement with its participation in the biosynthesis of the major phenolic compounds found in VOO.

## MATERIALS AND METHODS

### Plant Material

Olive (*Olea europaea* L. cv. Picual and Arbequina) trees were cultivated in the experimental orchard of Instituto de la Grasa, Sevilla (Spain), with drip irrigation and fertirrigation from the time of flowering to fruit ripening. In the case of non-irrigated treatment, the olive trees received only natural rainfall.

Young drupes, developing seeds, and mesocarp tissue were harvested at different weeks after flowering (WAF) corresponding to different developmental stages of the olive fruit: green (9, 12, 16, and 19 WAF); yellow-green (23 WAF); turning or veraison (28 and 31 WAF); and mature or fully ripe (35 WAF). Immediately after harvesting, olive tissues were frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$ .

Stress treatments were carried out according to Hernández et al. (2011). Olive branches with approximately 100 olive fruit at turning stage (28 WAF) were taken from olive trees and incubated in a growth chamber at  $25^{\circ}\text{C}$  with a 12 h light/12 h dark cycle to imitate physiological conditions of the tree. The light intensity was  $11.5 \mu\text{mol m}^{-2} \text{s}^{-1}$ . For stress experiments, standard conditions were modified according to the effect studied. For low and high temperature treatments, the branches with olive fruit were incubated at the standard light intensity, at 15 or  $35^{\circ}\text{C}$ , respectively. To evaluate the effect of the darkness, the standard temperature was maintained, and light was turned off. To study the effect of wounding, the whole surface of the olive fruit was mechanically damaged with pressure at zero time using forceps with serrated tips, affecting mesocarp tissue. To maintain the natural photoperiod day/night of the olive fruit, the zero time of each experiment was selected 2 h after the start of the light period. When indicated, olive mesocarp tissues were collected, frozen in liquid nitrogen, and kept at  $-80^{\circ}\text{C}$ .

### Isolation of a $\beta$ -Glucosidase Full-Length cDNA Clone

Candidate sequences for olive  $\beta$ -glucosidases were identified in the olive ESTs database (Muñoz-Mérida et al., 2013) by means of the tblastn algorithm together with amino acid sequences of known plant  $\beta$ -glucosidase proteins. One of

them, which showed high expression levels in mesocarp tissue according to *in silico* expression analysis, was selected for cloning. Based on this sequence, a specific pair of primers CR3 5'-AAGAGCACCAAAGTCTGCAATG-3' and CR4 5'-GGAGCCCAACTCCTTTATTGG-3' was designed. These primers, together with an aliquot of an olive Uni-ZAP XR cDNA library constructed with mRNA isolated from 13 WAF olive fruit of cultivar Picual (Haralampidis et al., 1998), were used for PCR amplification. The generated DNA fragment was subcloned into the vector pSpark® I (Canvax, Spain) and sequenced in both directions. DNA sequence determination and analysis was carried out as described in Hernández et al. (2016).

## Total RNA Extraction and cDNA Synthesis

1–2 g of frozen olive tissues harvested from at least three different olive trees, were used for total RNA isolation as described by Hernández et al. (2005). Verification of RNA quality, removal of genomic DNA and cDNA synthesis were performed according to Hernández et al. (2009).

## Quantitative Real-Time PCR (qRT-PCR)

Gene expression analysis was performed by qRT-PCR using an Mx3000P™ real-time PCR System and the “Brilliant® SYBR® Green Q-PCR Master Mix (Stratagene, La Jolla, CA, United States) as previously described (Hernández et al., 2009). Primer3 program<sup>1</sup> was used to design primers for gene-specific amplification (Supplementary Table S1). The housekeeping olive ubiquitin2 gene (*OeUBQ2*, AF429430) was used as an endogenous reference to normalize. The real-time PCR data were calibrated relative to the corresponding gene expression level in 12 WAF mesocarp tissue from Picual in the case of tissue and developmental expression studies, whereas for the stress studies the data were calibrated relative to the corresponding gene expression level at zero time for each treatment and cultivar. In both cases, the  $2^{-\Delta\Delta C_T}$  method for relative quantification was followed (Livak and Schmittgen, 2001). The data are presented as means  $\pm$  standard deviation (SD) of three different qRT-PCR reactions carried out in three different 96-well plates. Each reaction was performed in duplicate in each plate.

## Transient Expression of *OepGLU* Gene in *Nicotiana benthamiana*

For functional *Agrobacterium*-mediated CaMV35S-driven transient expression, the *OepGLU* coding sequence was PCR-amplified using the specific primers YDV1F (5'-CACC ATGGATATCCAAAGCAAC-3') and YDV1R+His (5'-CTAG TGATGGTGTGGTGTATGCCCGGTGCTGCCTCTAAGCCTT TTAC-3'), and subcloned into the GATEWAY®-compatible binary vector pH2GW7 (Karimi et al., 2002). The resulting purified pH2GW7-*OepGLU* construct was used to transform *Agrobacterium tumefaciens* strain GV3101 using freeze-thaw method described by Höfgen and Willmitzer (1988). *Nicotiana benthamiana* leaves were pressure infiltrated with *A. tumefaciens*

cultures (OD<sub>600</sub> approximately 1.0) as described by Popescu et al. (2007). Samples were collected 3 days after infiltration, frozen in liquid nitrogen and kept at  $-80^{\circ}\text{C}$ .

## Purification of *OepGLU* Recombinant Isoenzyme Expressed in *N. benthamiana* Leaves

To obtain the crude extract, 4 g of infiltrated leaf tissue were thawed and homogenized in 30 ml of 20 mM Na-phosphate buffer pH 7.4 containing 500 mM NaCl, 20 mM imidazole, 5% (w/v) polyvinyl polypyrrolidone, and 1 mM phenylmethanesulfonyl fluoride using an Ultraturrax homogenizer at  $4^{\circ}\text{C}$ . The resulting homogenate was centrifuged at 27000 g for 20 min at  $4^{\circ}\text{C}$ . The clear supernatant was filtered through three layers of Miracloth (Calbiochem, United States) and was used as the crude extract.

To purify the recombinant *OepGLU* protein containing the C-terminal 6xHis tag motif, 30 ml of crude extract was loaded onto a 1-ml His GraviTrap column (GE Healthcare, United Kingdom), and *OepGLU* protein was eluted with 3 ml of 20 mM Na-phosphate pH 7.4 containing 500 mM NaCl and 500 mM imidazole. Remaining NaCl and imidazole were removed by means of a PD-10 column (GE Healthcare, United Kingdom). The enzymatic solution was concentrated in 30 kDa microcentrifuge filters Vivaspin® (GE Healthcare, United Kingdom) at 2000 g and  $4^{\circ}\text{C}$  to a final volume of 250  $\mu\text{l}$ . This purified and concentrated preparation was used for *OepGLU* biochemical characterization.

## $\beta$ -Glucosidase Assay

Two methods for *in vitro* assaying  $\beta$ -glucosidase activity were used in this study (Romero-Segura et al., 2009). A spectrophotometric method, in which the  $\beta$ -glucosidase activity was determined by continuously monitoring the increase in absorbance at 405 nm related to the increasing amount of *p*-nitrophenol liberated from the synthetic glucoside pNPG, and a second method based on the direct determination of the hydrolyzed natural olive glucoside, oleuropein, by HPLC analysis.

## HPLC Analysis

Analytical HPLC of phenolic compounds was performed in a Beckman Coulter liquid chromatographic system equipped with a System Gold 168 detector, a solvent module 126 and a Mediterranean Sea 18 column (4.0 mm i.d. x 250 mm, particle size = 5  $\mu\text{m}$ ) (Teknokroma, Spain). Quantification and identification of phenolic compounds was performed following a previously described methodology (Luaces et al., 2007).

## Protein Determination and Electrophoresis

The protein concentration was estimated using the Bio-Rad (United States) Bradford protein reagent dye with BSA as standard. SDS-PAGE was performed as previously described (Romero-Segura et al., 2009).

<sup>1</sup><http://primer3.ut.ee/>



## Preparation of Anti- $\beta$ -Glucosidase Polyclonal Antibodies and Immunoblot Analysis

Polyclonal antibodies against the native olive  $\beta$ -glucosidase protein purified from olive mesocarp according to the method described by Romero-Segura et al. (2009) were prepared in rabbit by Production and Animal Experimentation General Service of the University of Seville.

For Western blotting, 5–6  $\mu$ g of protein samples were separated by SDS-PAGE as described above and electro-blotted onto nitrocellulose membrane using the Mini Trans-Blot® system (Bio-Rad). Bound anti- $\beta$ -glucosidase primary antibody was detected using an anti-rabbit alkaline phosphatase-conjugated secondary antibody (Sigma-Aldrich, United States). When a mouse monoclonal anti-6xHis antibody (GE Healthcare) was used as primary antibody, an anti-mouse alkaline phosphatase-conjugated antibody (Invitrogen) was employed as secondary antibody. To detect alkaline phosphatase activity after antibodies incubation, nitrocellulose membrane was submerged in a solution obtained by dissolving a SIGMAFAST™ BCIP®/NBT tablet (Sigma-Aldrich) in 10 ml of distilled water.

## RESULTS

A number of contigs with a high degree of similarity to plant  $\beta$ -glucosidases were selected from the olive EST database (Muñoz-Mérida et al., 2013). Among them, one which showed high expression levels in mesocarp tissue according to *in silico* expression analysis was chosen for cloning. Two specific primers were designed on the basis of this contig sequence, and used for PCR amplification together with an aliquot of an olive cDNA library of cultivar Picual. We obtained a full-length cDNA clone of 1848 bp, which was designated *OepGLU*, and contained an open reading frame encoding a predicted protein of 551 amino acids (Supplementary Figure S1), with a calculated molecular mass of 62.8 kDa and a *pI* of 6.6. The deduced *OepGLU* amino acid sequence from cultivar Picual displayed a 98% identity to an olive  $\beta$ -glucosidase cDNA clone (AY083162) from cultivar Koroneiki (Koudounas et al., 2015).

Alignment of the deduced amino acid sequence of *OepGLU* from cultivar Picual with other plant  $\beta$ -glucosidase protein sequences (Supplementary Figure S1) suggests that it codes for a  $\beta$ -glucosidase enzyme because it showed the conserved sequence motifs characteristic of the glycosyl hydrolases family 1 (GH1) T(F/L/M)NEP and Y(I/V)TENG, which include the two glutamic acid residues involved in the catalytic mechanism (Esen, 2003). In addition, the conserved amino acids Gln, His, Asn, Glu and two Trp which have been shown to be essential for the binding of the glucose (Cairns and Esen, 2010) were also found. A putative *N*-glycosylation site (N83) has also been detected in the *OepGLU* sequence, and the presence of a conserved GH1 family domain has been identified by NCBI Conserved Domain Search and the Pfam software. Analysis of *OepGLU* deduced protein sequence with target prediction software such as WolfPSORT and TargetP did not give rise to a clear subcellular localization. In fact, a

putative nuclear localization signal (RRKR) could be found at amino acids 543–546 and a 25 amino acid N-terminal signal peptide was also predicted (Supplementary Figure S1), generating after its proteolytic cleavage a mature protein of 526 amino acids, with a calculated molecular mass of 60.3 kDa and a *pI* of 7.0.

## Purification and Immunological Characterization of *OepGLU* Recombinant Enzyme

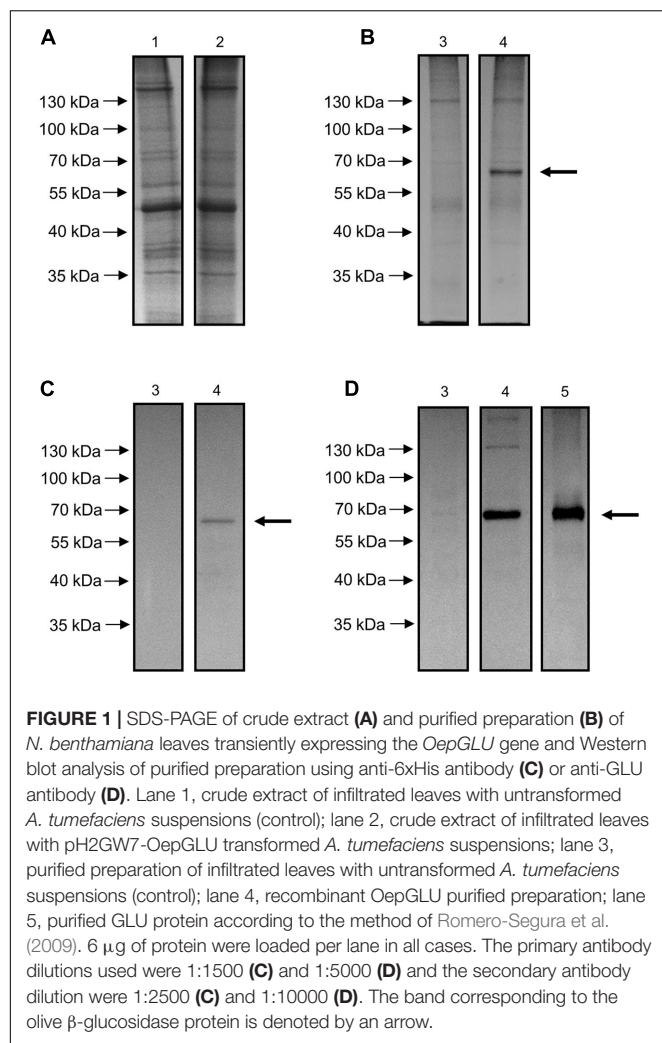
To verify the functional identity of the *OepGLU* gene, *Agrobacterium*-mediated transient expression in *N. benthamiana* leaves was carried out. To that end, the *OepGLU* coding region, including a C-terminal 6xHis tag, was subcloned into the vector pH2GW7 using the GATEWAY™ technology. The resultant plasmid designated pH2GW7-*OepGLU*+His was used to transform *A. tumefaciens* GV3101, and tobacco leaves were infiltrated with bacterial cells carrying this plasmid. Expression of the recombinant protein was optimal at 3 days after infiltration. SDS-PAGE analysis of the crude extract did not show the protein band with the expected molecular mass for the recombinant *OepGLU* (Figure 1A). However, when the enzyme preparation purified by affinity chromatography was used, an intense band with a molecular mass of 65.5 kDa was detected (Figure 1B, lane 4). In contrast, this protein band was not observed in purified preparations isolated from tobacco leaves infiltrated with untransformed *Agrobacterium* cells (Figure 1B, lane 3).

Purified preparations of recombinant *OepGLU* were also analyzed by western blot using the anti-6xHis antibody (Figure 1C) and the antibody raised against the native olive  $\beta$ -glucosidase (Figure 1D). In both cases, a protein band with the same molecular mass as deduced from the SDS-gel was observed.

## Kinetic Properties of Recombinant *OepGLU* in Comparison to the Native Enzyme

In the present study, purified preparations of recombinant *OepGLU*, but not crude extracts, were able to hydrolyze the artificial substrate *p*-nitrophenyl- $\beta$ -D-glucopyranoside (pNPG) and the major natural olive phenolic glycoside oleuropein in a time-dependent manner. The enzymatic hydrolysis of oleuropein by purified recombinant *OepGLU* was monitored for up to 60 min by HPLC analysis (Figure 2). More than 50% of the oleuropein initially present was hydrolyzed after 5 min, producing as the first reaction product a mixture of oleuropein aglycone isomers (OA-isomers). After 15 min, the broad peak of OA-isomers was reduced, whereas 3,4-DHPEA-EA and hydroxytyrosol began to accumulate. The purified recombinant olive  $\beta$ -glucosidase exhibited a specific activity of 67.6 U/mg using oleuropein as substrate.

Once it was established the capacity of the purified recombinant olive  $\beta$ -glucosidase to hydrolyze oleuropein, its activity was also measured using the synthetic glucoside pNPG as substrate, exhibiting a much lower specific activity (1.4 U/mg). Hence, the natural substrate oleuropein was used to perform the biochemical characterization of the recombinant olive  $\beta$ -glucosidase.



The purified recombinant olive  $\beta$ -glucosidase displayed an optimum pH of 5.5 with a fast decrease of activity above it (Figure 3A), and showed > 80% of its highest activity at 25–45°C, with an optimum temperature when assayed at 40°C and a strong decline over 45°C (Figure 3B). Thermal inactivation kinetics showed that recombinant *OepGLU* was active up to 40°C with a significant decline over this temperature (Figure 3C).

In addition,  $\beta$ -glucosidase activity was assayed with a mixture of oleuropein, demethyloleuropein and ligstroside, the three most important phenolic glycosides identified in olive fruit, in order to mimic what happens during the milling step of the industrial VOO extraction process, when the enzyme and possible substrates meet as olive fruit tissues are disrupted. Substrate selectivity experiments (Table 1) showed that the highest activity level was reached using oleuropein as substrate, followed by ligstroside (25.6%) and demethyloleuropein (15.6%). In particular, after a 5 min reaction time it could be observed in the corresponding chromatogram that most of the oleuropein was hydrolyzed followed by ligstroside, with the consequent appearance of OA-isomers and ligstroside aglycone isomers (LA-isomers), respectively (Supplementary Figure S2).

Verbascoside was used as negative control, since its chemical structure is significantly different from that of the other three phenolic glucosides and does not contain a non-reducing terminal  $\beta$ -D-glucosyl residue.

To obtain the kinetic parameters of recombinant olive  $\beta$ -glucosidase, enzyme activity was measured over a range of concentrations of oleuropein as substrate (Supplementary Figure S3). The calculated  $K_m$  for oleuropein was 26.8 mM and the  $V_{max}$  263.2 U/mg, with a catalytic efficiency ( $V_{max}/K_m$ ) of 9.8.

## Tissue Specificity and Developmental Expression of Olive *GLU* Gene Is Cultivar-Dependent

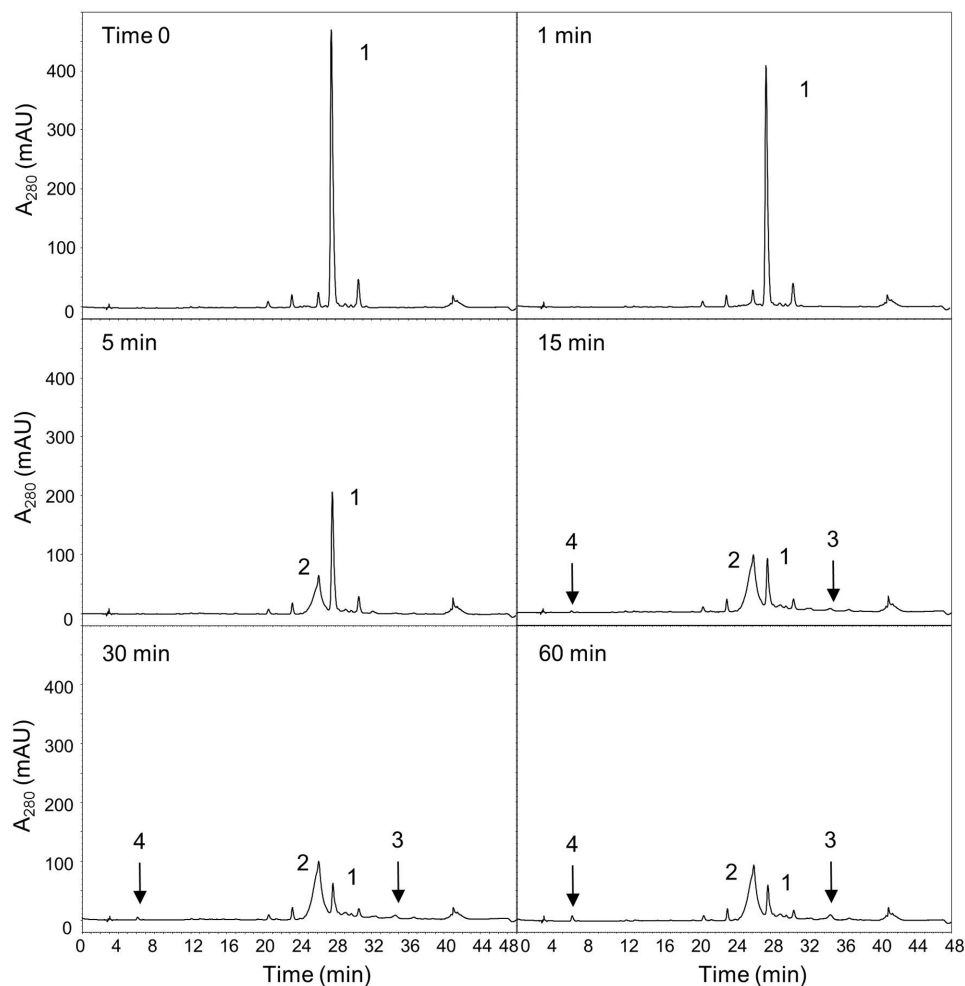
Olive *GLU* gene expression levels were analyzed in different Picual and Arbequina olive tissues by qRT-PCR using specific primers (Figure 4A), with the aim of investigating its physiological role and its possible contribution to the content of the different phenolic compounds present in the VOO. In both cultivars, higher expression levels were detected in young drupes and mesocarp compared to seeds, where transcript levels were negligible. Young drupes of 9 WAF from cultivar Arbequina showed the highest expression level. Interestingly, transcript levels in mesocarp at turning stage (28 WAF) were much higher in Picual than in Arbequina cultivar.

Besides, olive *GLU* transcript levels were analyzed at different times during olive fruit development and ripening in Picual and Arbequina mesocarp and seed tissues (Figure 4B). A maximum transcription level was observed in green mesocarp (16 WAF) from both cultivars. In the case of cultivar Arbequina, the high expression level detected at 16 WAF dramatically decreased after this maximum, reaching constant low levels during the rest of the olive fruit development and ripening periods. On the contrary, in the cultivar Picual, olive *GLU* expression level showed a second maximum, lower than the first one, once the olive fruit ripening period has started (28 WAF). Unlike mesocarp tissue, olive *GLU* gene exhibited almost undetectable transcript levels in seeds from Picual and Arbequina (Figure 4B), which is consistent with the very low enzyme activity levels observed in both cultivars (Romero-Segura et al., 2011).

The same study was also performed in Picudo, Hojiblanca and Manzanilla cultivars, using olive fruit mesocarp at the three main stages in which olive fruit are harvested for olive oil production: yellow-green (23 WAF), turning or veraison (31 WAF), and mature or fully ripe (35 WAF) (Supplementary Figure S4). The olive *GLU* gene expression levels in the cultivars Picudo and Hojiblanca remained low, showing no significant changes. On the contrary, the transcript levels in the cultivar Manzanilla showed an increase during fruit ripening, and then decreased at the end of the ripening period.

## Transcriptional Regulation in Olive Fruit Mesocarp of *GLU* Gene in Response to Abiotic Stresses

To examine the effect of different abiotic stresses on the expression level of the *GLU* gene in mesocarp tissue, olive tree



**FIGURE 2 |** Time-course of the oleuropein hydrolysis catalyzed by the purified recombinant OepGLU. Enzyme activity was assayed as indicated in Material and Methods. Identification of peaks: 1, oleuropein; 2, OA-isomers; 3, 3,4-DHPEA-EA; 4, hydroxytyrosol.

branches from Picual and Arbequina cultivars with olive fruit at turning stage (28 WAF) were incubated for 24 h modifying the standard conditions (25°C with a 12 h light / 12 h dark cycle) dependent on the effect to be tested. No changes in the olive *GLU* gene expression levels were observed in olive fruit mesocarp when standard conditions were used (**Figures 5, 6**).

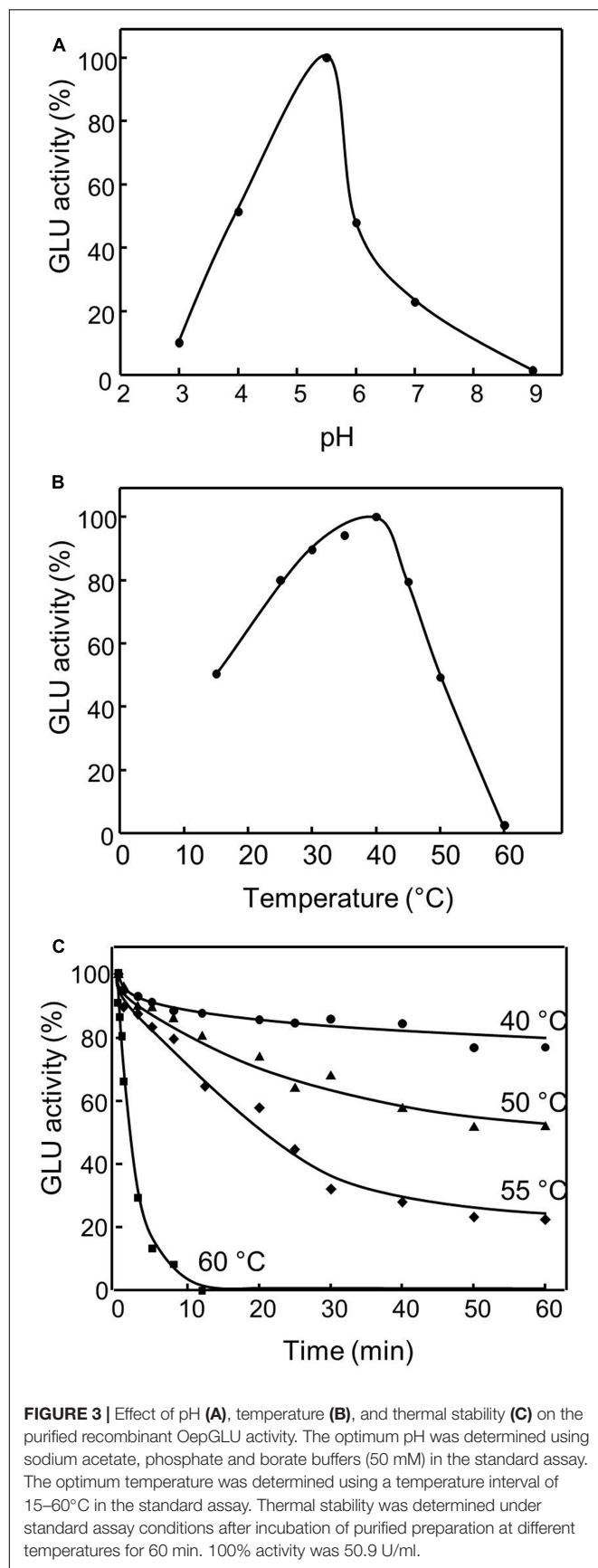
When low temperature (15°C) was used to incubate the olive fruit, a significant transient increase in the expression levels of olive *GLU* was observed in both cultivars, with a maximum after 3 or 6 h of treatment for Picual and Arbequina cultivars, respectively (**Figure 5A**). On the contrary, the incubation at high temperature (35°C) of olive fruit brought about a reduction in the olive *GLU* gene transcript levels in both cultivars especially after 1 h of treatment, reaching almost undetectable transcript levels after 24 h (**Figure 5B**).

To examine the effect of darkness on the transcript levels of the olive *GLU* gene in Picual and Arbequina mesocarp tissues, branches were incubated for 24 h at 25°C in the darkness. A decrease in the olive *GLU* gene expression levels was detected in

both cultivars, mainly in Picual during the first 3 h of incubation (**Figure 6A**).

In addition, the potential involvement of olive *GLU* in the transcriptional response to wounding was tested in olive fruit subjected to mechanical damage from olive branches incubated at standard conditions. In this case, olive *GLU* gene expression levels declined progressively in both cultivars (**Figure 6B**).

On the other hand, since a number of studies point out that different water regimes could affect the phenolics content of VOO, showing a negative correlation between the content of secoiridoid derivatives and the water amount used for olive growing (Gómez-Rico et al., 2006; Servili et al., 2007), the effect of water regime on the transcript levels of olive *GLU* was investigated in olive fruit mesocarp of Picual and Arbequina cultivars grown with natural rainfall or irrigation. A higher transcript level was detected for the olive *GLU* gene when Picual and Arbequina were cultivated with natural rainfall only (**Figure 7**), except for late ripening stages (35 WAF) where transcript levels were very low in both watering conditions.



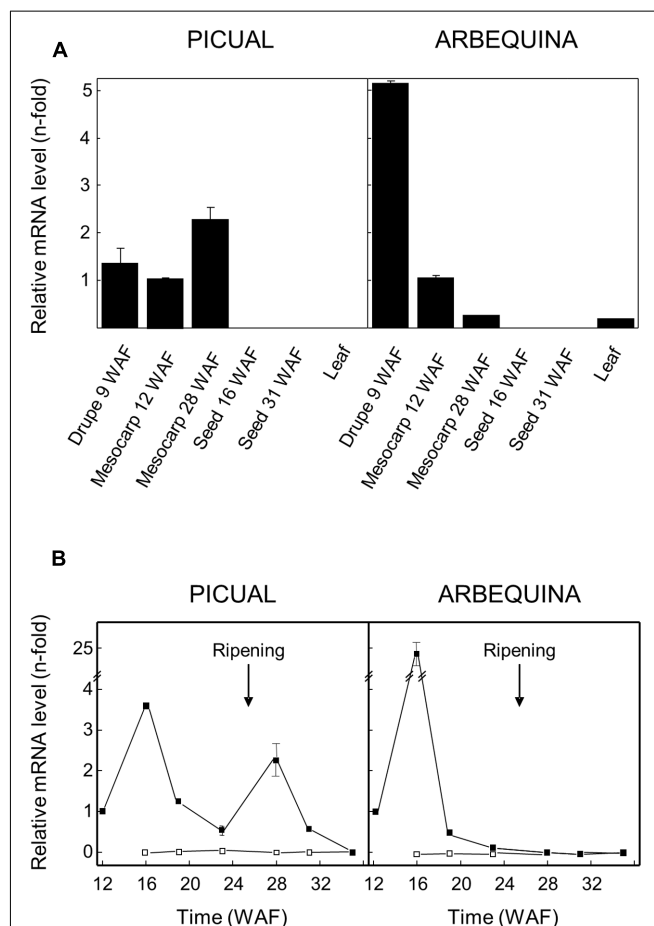
**TABLE 1 |** Substrate selectivity of purified recombinant OepGLU on a mixture of various natural olive glycosides.

Substrate	Relative activity (%)
Oleuropein	100.0
Ligstroside	25.6
Demethyloleuropein	15.6

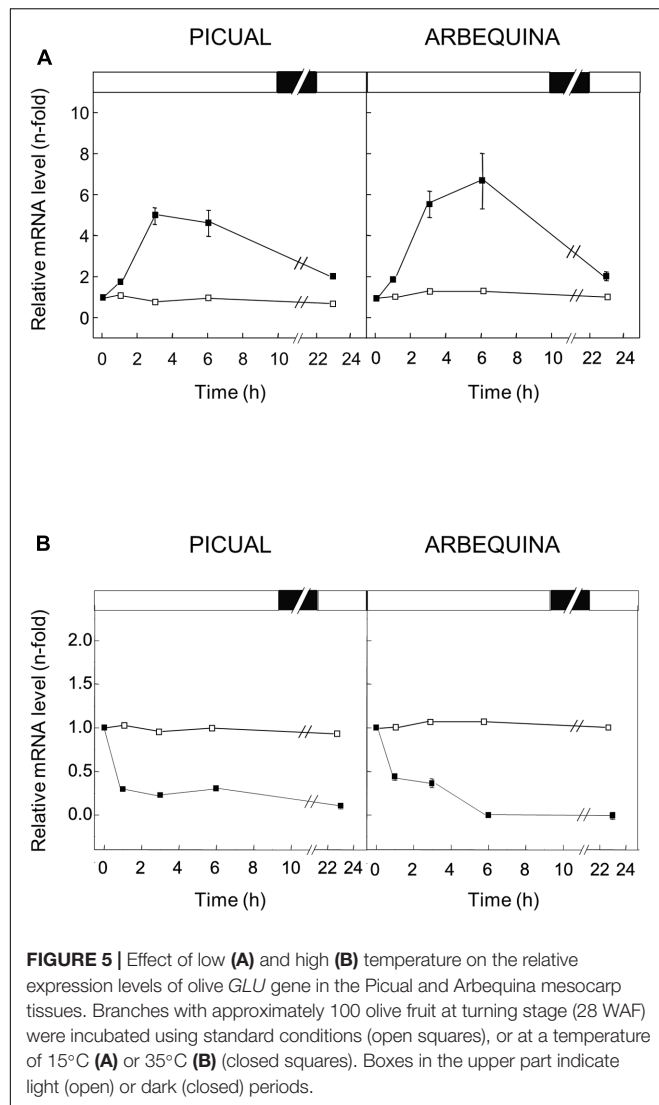
Activity was determined by measuring the corresponding hydrolyzed natural olive glycoside by HPLC. Initial concentration of the substrates in the assay mixture was 5 mM each.

## DISCUSSION

Several candidate olive  $\beta$ -glucosidase sequences were identified from an olive ESTs database (Muñoz-Mérida et al., 2013). This is consistent with the numerous  $\beta$ -glucosidase genes usually detected in the same plant, as reported in Arabidopsis (Xu et al., 2004). Since higher  $\beta$ -glucosidase activity levels have been observed in olive fruit mesocarp in comparison to seeds

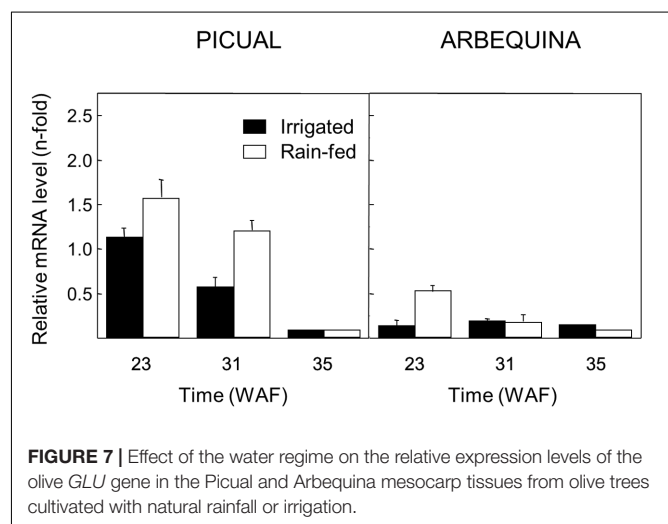
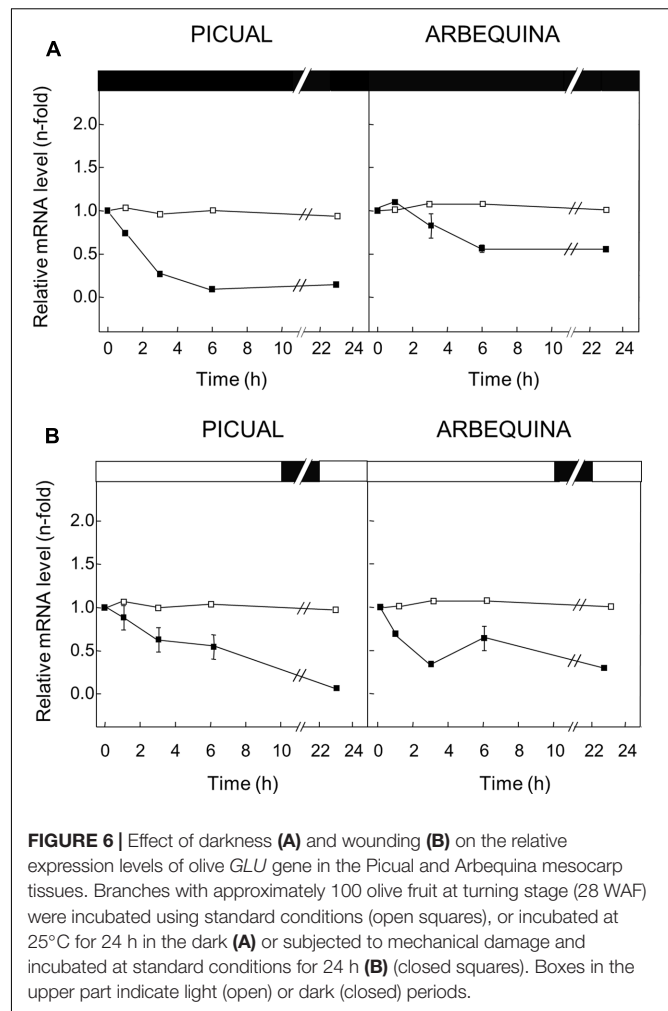






(Romero-Segura et al., 2011), one of the contigs which showed high expression levels in mesocarp tissue compared to seeds according to *in silico* expression analysis, was selected for cloning. Sequence analysis of the  $\beta$ -glucosidase gene isolated from olive (cv. Picual) showed that its deduced amino acid sequence contains the conserved sequence motifs and domains characteristic of the glycosyl hydrolases family 1 (GH1) (Esen, 2003; Cairns and Esen, 2010), and suggests that it codes for a  $\beta$ -glucosidase enzyme.

In order to characterize the immunological and kinetic properties of the olive  $\beta$ -glucosidase recombinant enzyme, transient expression in *N. benthamiana* leaves was performed. The band corresponding to the recombinant olive  $\beta$ -glucosidase was observed by SDS-PAGE only when the enzymatic preparation purified by affinity chromatography was applied, but not in the case of the purified preparation isolated from tobacco leaves infiltrated with untransformed *Agrobacterium* cells used as control, or when the crude extract was loaded onto the gel. In addition, western blot analysis of the purified preparations was



performed using two types of antibodies. In the first case, an anti 6xHis antibody detected a band with identical molecular mass (65.0 kDa) than that observed by SDS-PAGE, confirming that it corresponded to the recombinant enzyme. This molecular mass

of OepGLU is in the range of 55–65 kDa described for almost all plant  $\beta$ -glucosidase monomers (Esen, 2003), and is identical to that of 65.4 kDa reported for the  $\beta$ -glucosidase protein purified from olive fruit mesocarp (Romero-Segura et al., 2009). In the second case, an antibody raised against the native olive  $\beta$ -glucosidase protein purified from olive mesocarp (Romero-Segura et al., 2009) detected a band with identical molecular mass. Furthermore, when a preparation corresponding to native enzyme purified according to the method of Romero-Segura et al. (2009) was used as positive control, an intense band with a similar apparent size was detected. All these data strongly suggest that the recombinant OepGLU corresponds to the native  $\beta$ -glucosidase enzyme previously purified by our group from olive fruit mesocarp, which has been demonstrated to play a key role in shaping the VOO phenolic composition (Romero-Segura et al., 2012). Interestingly, although native and recombinant olive  $\beta$ -glucosidase enzymes exhibit similar molecular masses, the occurrence of post-translational modifications cannot be discarded. In fact, a unique *N*-glycosylation site was predicted in the OepGLU amino acid sequence.

Although the capacity of the recombinant olive  $\beta$ -glucosidase to hydrolyze oleuropein has been previously demonstrated using crude extracts from *N. benthamiana* leaves (Koudounas et al., 2015), a quantitative *in vitro* enzymatic assay using purified recombinant protein, to avoid interferences of metabolites and enzyme activities present in the tobacco leaves crude extract, has not been reported so far. Furthermore, a comprehensive characterization of the kinetic properties of the enzyme has not been carried out up to date. Purified preparations of recombinant OepGLU exhibited  $\beta$ -glucosidase activity with both, the artificial substrate pNPG and the major natural olive phenolic glucoside oleuropein. This result demonstrates that the *OepGLU* gene code for a  $\beta$ -glucosidase enzyme, confirming its functional identity. Interestingly, purified recombinant OepGLU showed very low activity levels when pNPG was used as substrate, as previously reported for the native enzyme from olive (Romero-Segura et al., 2009) and privet tree (Konno et al., 1999). It has been described for plant  $\beta$ -glucosidases that there is not a correspondence between the activity levels exhibited using non-physiological substrates such as pNPG, and those obtained using their natural substrates (Cairns et al., 2015). The comparison of the data obtained from the time-course of the oleuropein hydrolysis catalyzed by the purified recombinant OepGLU, with those previously reported using the purified native enzyme (Romero-Segura et al., 2012), shows that the native olive  $\beta$ -glucosidase hydrolyzes oleuropein more efficiently. These discrepancies on the relative activity of  $\beta$ -glucosidase proteins purified from the native plant and the corresponding recombinant proteins have been previously described (Himeno et al., 2013). OA-isomers are the first reaction products formed by the recombinant OepGLU after the hydrolysis of the glucoside. The elimination of the glucose molecule could destabilize the phenolic aglucone and during the reaction course the formed isomers tend to its stabilization yielding 3,4-DHPEA-EA, which simultaneously produce hydroxytyrosol by chemical hydrolytic reactions, since the recombinant  $\beta$ -glucosidase is the only enzyme present in the reaction mixture.

The purified recombinant OepGLU displays the highest activity at pH 5.5 and 40°C, similar values to those reported for the olive native  $\beta$ -glucosidase enzyme (Romero-Segura et al., 2009). Optimum pH in the range of 4.5–5.5 has been described for other  $\beta$ -glucosidases from plants such as rice (Akiyama et al., 1998) and *Citrus sinensis* (Cameron et al., 2001). With respect to optimum temperature, data in the interval of 40–50°C has been previously reported for the  $\beta$ -glucosidases from maize (Esen, 1992) and *Citrus sinensis* (Cameron et al., 2001). The recombinant olive  $\beta$ -glucosidase exhibits a high thermostability up to 40°C, as reported for the native enzyme from olive (Romero-Segura et al., 2009). Taking into account its thermal resistance profile, the OepGLU enzyme could act during the malaxation step of the industrial process to obtain VOO, where temperatures higher than 30°C are not unusual. However, it has been reported that after 15 min of malaxation at this temperature, no  $\beta$ -glucosidase activity could be detected in the paste, likely due to the presence of enzyme inhibitors (García-Rodríguez, 2014). In the same way, the thermal resistance of OepGLU could explain why the thermal treatment of olive fruit at temperatures of 56–68°C just before the milling step causes a high decrease in the content of the secoiridoid derivatives in the VOO (Yousfi et al., 2010), since at those temperatures the olive  $\beta$ -glucosidase enzyme should be inactivated, highly reducing the degree of hydrolysis of the phenolic glucosides. Substrate selectivity experiments showed that the recombinant olive  $\beta$ -glucosidase exhibits a higher preference for oleuropein as substrate, followed by ligstroside and demethyloleuropein. These results are in agreement to those reported for the native enzyme (Romero-Segura et al., 2009), and demonstrate the capacity of the recombinant OepGLU to hydrolyze the three main phenolic glucosides of olive fruit, which are the precursors of the main secoiridoid derivatives present in the VOO. The recombinant olive  $\beta$ -glucosidase showed kinetic parameters of  $K_m$  for oleuropein (26.8 mM) and  $V_{max}$  (263.2 U/mg), different of those described (3.8 mM and 2,500 U/mg, respectively) for the native enzyme (Romero-Segura et al., 2009), and indicate a lower catalytic efficiency of the recombinant enzyme as previously mentioned.

Plant *GLU* genes are developmentally regulated (Morant et al., 2008), and exhibit different spatial expression patterns depending on their physiological functions. In this sense, olive *GLU* gene showed different transcript levels in the studied tissues, showing its spatial regulation. Besides, changes in the *GLU* transcript level reveal that this gene is also temporally regulated, and moderately correlate with changes in the  $\beta$ -glucosidase activity levels in Picual and Arbequina previously reported (Romero-Segura et al., 2012). This minor discrepancy observed between transcript and activity levels could be explained by the occurrence of post-translational modifications of the olive  $\beta$ -glucosidase enzyme such as *N*-glycosylation, as previously mentioned. In fact, *N*-glycosylation of plant  $\beta$ -glucosidases has been widely described (Morant et al., 2008). Furthermore, the contribution of other olive  $\beta$ -glucosidases isoforms to the enzyme activity levels observed cannot be discarded. Interestingly, previous studies on these two cultivars have demonstrated significant differences not only in terms of  $\beta$ -glucosidase activity but also in their

phenolic profiles along fruit ripening, with Picual oils being described as a VOO with medium-high phenolic content at any ripening stage while Arbequina oils typically have medium-low concentration of phenolic compounds (Romero-Segura et al., 2012). Moreover, expression data from cultivars Picudo, Hojiblanca and Manzanilla confirm the cultivar-dependent transcriptional regulation of the olive *GLU* gene during olive fruit development and ripening. Hence, knowledge of the specific *GLU* expression profile for each olive cultivar is critical to determine the optimum harvesting time in order to obtain VOO with the highest phenolic content.

In our study, we have also found that the transcript level of the olive *GLU* gene in olive fruit mesocarp is transcriptionally regulated in response to different abiotic stresses. Low and high temperatures brought about the induction and repression of olive *GLU* gene, respectively. These data are in agreement with those described for the Arabidopsis  $\beta$ -glucosidase gene *AtBG1*, since its expression levels increase when leaves are subjected to cold stress (Lee et al., 2006). The decrease of the expression levels of olive *GLU* gene observed at 35°C is also consistent with the lower content of secoiridoid derivatives in oils extracted from olive fruit pre-treated for 24 h at 30–50°C (García et al., 2001). Darkness treatment of olive fruit from both cultivars produces a decrease in the expression levels of olive *GLU* gene, indicating that light may be implicated in the regulation of its transcription. On the other hand, although plant  $\beta$ -glucosidases have been involved in the defense against herbivores and pathogen attacks (Minic, 2008), this primary response acts at enzyme activity level, being regulated by compartmentalization, since enzyme and natural substrates are differentially located at subcellular level and only meet when cell integrity is disrupted (Morant et al., 2008). Therefore, it is not surprising that olive *GLU* gene is not induced after wounding, which indicates that the regulation at transcriptional level is not operating. In contrast to our data, the transcript increase of olive *GLU* gene in olive fruit mesocarp after olive fruit fly attack has been reported (Corrado et al., 2012), although only one olive fruit developmental stage was used in that study. Finally, higher expression levels of the olive *GLU* gene were detected under water deficit conditions in Picual and Arbequina cultivars, mainly at turning stage. Similar results have been recently reported for the *OeGLU12-like2* gene from cultivar Frantoio (Cirilli et al., 2017). The increase in the olive *GLU* transcript level detected under water deficit conditions, with the corresponding increase in the enzyme activity, could also significantly contribute to explain the higher content of secoiridoid derivatives reported in VOO obtained from olive fruit under water stress conditions (Artajo et al., 2006; Stefanoudaki et al., 2009).

## CONCLUSION

We have purified the olive recombinant  $\beta$ -glucosidase enzyme (OepGLU). Immunological detection, molecular mass determination and kinetic properties of the recombinant OepGLU strongly indicates that it corresponds to the native olive  $\beta$ -glucosidase enzyme previously purified from olive fruit

mesocarp (Romero-Segura et al., 2009), which has been shown as the main enzyme involved in the transformation during VOO processing, of oleuropein and other phenolic glycosides from olive fruit onto their corresponding secoiridoid derivatives present in VOO. However, the contribution of other olive fruit  $\beta$ -glucosidase isoenzymes to oleuropein hydrolysis cannot be discarded. Our results have also shown that olive *GLU* gene expression is not only spatially and temporally regulated in olive fruit, but also is cultivar-dependent and regulated by temperature, light and water regime. This study represents a significant step to elucidate the factors responsible for the phenolic content and profile of VOO. In addition, this information will help in the design of molecular markers for the marker-assisted selection of novel olive cultivars with improved phenolic content and composition in their oils.

## ACCESSION NUMBERS

The nucleotide sequence reported in this paper for *OepGLU* gene has been submitted to the GenBank/EMBL/DDBJ database with the accession number KX278417.

## AUTHOR CONTRIBUTIONS

AP and JM-R conceived and designed the study, DV-P and CR-S performed the cloning and qRT-PCR experiments, DV-P and RG-R carried out the immunological studies, DV-P performed the transient expression experiments and the analytical and biochemical studies, MH and JM-R supervised the cloning, qRT-PCR and immunological studies, FV and IG supervised the transient expression experiments, AP supervised the analytical and biochemical studies, JM-R wrote the manuscript. All authors discussed and commented the manuscript.

## FUNDING

This work was supported by the Spanish Ministry of Economy and Competitiveness [grants no. AGL2008-00258, AGL2011-24442]. Fellowship from the JAE-Predoctoral CSIC program to DV-P. Fellowships from the FPI predoctoral program to CR-S and RG-R. Contract from the JAE-Postdoctoral CSIC program to MH.

## ACKNOWLEDGMENT

We thank Rosario Sánchez (Instituto de la Grasa, CSIC) for technical assistance.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.01902/full#supplementary-material>

## REFERENCES

- Akiyama, T., Kaku, H., and Shibuya, N. (1998). A cell wall-bound  $\beta$ -glucosidase from germinated rice: purification and properties. *Phytochemistry* 48, 49–54. doi: 10.1016/S0031-9422(97)01099-6
- Alagna, F., Mariotti, R., Panara, F., Caporali, S., Urbani, S., Veneziani, G., et al. (2012). Olive phenolic compounds: metabolic and transcriptional profiling during fruit development. *BMC Plant Biol.* 12:162. doi: 10.1186/1471-2229-12-162
- Andrews, P., Busch, J. L. H. C., de Joode, T., Groenewegen, A., and Alexandre, H. (2003). Sensory properties of virgin olive oil polyphenols: identification of deacetoxy-ligstroside aglycon as a key contributor to pungency. *J. Agric. Food Chem.* 51, 1415–1420. doi: 10.1021/jf026042j
- Artajo, L. S., Romero, M. P., Tovar, M. J., and Motilva, M. J. (2006). Effect of irrigation applied to olive trees (*Olea europaea* L.) on phenolic compound transfer during olive oil extraction. *Eur. J. Lipid Sci. Technol.* 108, 19–27. doi: 10.1002/ejlt.200500227
- Bernini, R., Crisante, F., Merendino, N., Molinari, R., Soldatelli, M. C., and Velotti, F. (2011). Synthesis of a novel ester of hydroxytyrosol and  $\alpha$ -lipoic acid exhibiting an antiproliferative effect on human colon cancer HT-29 cells. *Eur. J. Med. Chem.* 46, 439–446. doi: 10.1016/j.ejmech.2010.10.028
- Bianco, L., Alagna, F., Baldoni, L., Finnie, C., Svensson, B., and Perrotta, G. (2013). Proteome regulation during *Olea europaea* fruit development. *PLOS ONE* 8:e53563. doi: 10.1371/journal.pone.0053563
- Cairns, J. R. K., and Esen, A. (2010).  $\beta$ -glucosidases. *Cell. Mol. Life Sci.* 67, 3389–3405. doi: 10.1007/s00018-010-0399-2
- Cairns, J. R. K., Mahong, B., Baiya, S., and Jeon, J. S. (2015).  $\beta$ -Glucosidases: multitasking, moonlighting or simply misunderstood? *Plant Sci.* 241, 246–259. doi: 10.1016/j.plantsci.2015.10.014
- Cameron, R. G., Manthey, J. A., Baker, R. A., and Grohmann, K. (2001). Purification and characterization of a beta-glucosidase from *Citrus sinensis* var. Valencia fruit tissue. *J. Agric. Food Chem.* 49, 4457–4462. doi: 10.1021/jf0101010z
- Cirilli, M., Caruso, G., Gennai, C., Urbani, S., Frioni, E., Ruzzi, M., et al. (2017). The role of polyphenoloxidase, peroxidase, and  $\beta$ -glucosidase in phenolics accumulation in *Olea europea* L. fruits under different water regimes. *Front. Plant Sci.* 8:717. doi: 10.3389/fpls.2017.00717
- Corrado, G., Alagna, F., Rocco, M., Renzone, G., Varrichio, P., Coppola, V., et al. (2012). Molecular interactions between the olive and the fruit fly *Bactrocera oleae*. *BMC Plant Biol.* 12:86. doi: 10.1186/1471-2229-12-86
- Esen, A. (1992). Purification and partial characterization of maize (*Zea mays* L.)  $\beta$ -glucosidase. *Plant Physiol.* 98, 174–182. doi: 10.1104/pp.98.1.174
- Esen, A. (2003). “ $\beta$ -Glucosidase,” in *Handbook of Food Enzymology*, eds J. Whitaker, A. Voragen, and D. Wong (New York City, NY: Marcel Dekker), 791–802.
- García, J. M., Yousfi, K., Mateos, R., Olmo, M., and Cert, A. (2001). Reduction of oil bitterness by heating of olive (*Olea europaea*) fruits. *J. Agric. Food Chem.* 49, 4231–4235. doi: 10.1021/jf001302n
- García-Rodríguez, R. (2014). *Factores Bioquímicos Implicados en la Degradación Oxidativa de Compuestos Fenólicos Durante el Proceso de Obtención del Aceite de Oliva Virgen*. Ph.D. thesis, Universidad de Sevilla, Seville.
- García-Rodríguez, R., Romero-Segura, C., Sanz, C., Sánchez-Ortiz, A., and Pérez, A. (2011). Role of polyphenol oxidase and peroxidase in shaping the phenolic profile of virgin olive oil. *Food Res. Int.* 44, 629–635. doi: 10.1016/j.foodres.2010.12.023
- Gómez-Rico, A., Salvador, M. D., La Greca, M., and Fregapane, G. (2006). Phenolic and volatile compounds of extra virgin olive oil (*Olea europaea* L. Cv. Cornicabra) with regard to fruit ripening and irrigation management. *J. Agric. Food Chem.* 54, 7130–7136. doi: 10.1021/jf060798r
- Haralampidis, K., Milioni, D., Sánchez, J., Baltrusch, M., Heinz, E., and Hatzopoulos, P. (1998). Temporal and transient expression of stearoyl-ACP carrier protein desaturase gene during olive fruit development. *J. Exp. Bot.* 49, 1661–1669. doi: 10.1093/jxbbot/49.327.1661
- Hernández, M. L., Mancha, M., and Martínez-Rivas, J. M. (2005). Molecular cloning and characterization of genes encoding two microsomal oleate desaturases (*FAD2*) from olive. *Phytochemistry* 66, 1417–1426. doi: 10.1016/j.phytochem.2005.04.004
- Hernández, M. L., Padilla, M. N., Mancha, M., and Martínez-Rivas, J. M. (2009). Expression analysis identifies *FAD2-2* as the olive oleate desaturase gene mainly responsible for the linoleic acid content in virgin olive oil. *J. Agric. Food Chem.* 57, 6199–6206. doi: 10.1021/jf900678z
- Hernández, M. L., Padilla, M. N., Sicardo, M. D., Mancha, M., and Martínez-Rivas, J. M. (2011). Effect of different environmental stresses on the expression of oleate desaturase genes and fatty acid composition in olive fruit. *Phytochemistry* 72, 178–187. doi: 10.1016/j.phytochem.2010.11.026
- Hernández, M. L., Sicardo, M. D., and Martínez-Rivas, J. M. (2016). Differential contribution of endoplasmic reticulum and chloroplast  $\omega$ -3 fatty acid desaturase genes to the linolenic acid content of olive (*Olea europaea*) fruit. *Plant Cell Physiol.* 57, 138–151. doi: 10.1093/pcp/pcv159
- Himeno, N., Saburi, W., Wakuta, S., Takeda, R., Matsuura, H., Nabeta, K., et al. (2013). Identification of rice  $\beta$ -glucosidase with high hydrolytic activity towards salicylic acid  $\beta$ -D-glucoside. *Biosci. Biotech. Biochem.* 77, 934–939. doi: 10.1271/bbb.120889
- Höfgen, R., and Willmitzer, L. (1988). Storage of competent cells for *Agrobacterium* transformation. *Nucleic Acids Res.* 16:9877. doi: 10.1093/nar/16.20.9877
- Karimi, M., Inze, D., and Depicker, A. (2002). GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7, 193–195. doi: 10.1016/S1360-1385(02)02251-3
- Konno, K., Hirayama, C., Yasui, H., and Nakamura, M. (1999). Enzymatic activation of oleuropein: a protein crosslinker used as chemical defense in the privet tree. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9159–9164. doi: 10.1073/pnas.96.16.9159
- Konstantinidou, V., Covas, M.-I., Muñoz-Aguayo, D., Khymenets, O., de la Torre, R., Saez, G., et al. (2010). In vivo nutrigenomic effects of virgin olive oil polyphenols within the frame of the Mediterranean diet: a randomized controlled trial. *FASEB J.* 24, 2546–2557. doi: 10.1096/fj.09-148452
- Koudounas, K., Banilas, G., Michaelidis, C., Demoliou, C., Rigas, S., and Hatzopoulos, P. (2015). A defence-related *Olea europaea*  $\beta$ -glucosidase hydrolyses and activates oleuropein into a potent protein cross-linking agent. *J. Exp. Bot.* 66, 2093–2106. doi: 10.1093/jxb/erv002
- Lee, K. H., Piao, H. L., Kim, H.-Y., Choi, S. M., Jiang, F., Hartung, W., et al. (2006). Activation of glucosidase via stress-induced polymerization rapidly increases active pools of abscisic acid. *Cell* 126, 1109–1120. doi: 10.1016/j.cell.2006.07.034
- León, L., Beltran, G., Aguilera, M. P., Rallo, L., and Barranco, D. (2011). Oil composition of advanced selections from an olive breeding program. *Eur. J. Lipid Sci. Technol.* 113, 870–875. doi: 10.1002/ejlt.201000535
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_T}$  method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Luaces, P., Romero, C., Gutiérrez, F., Sanz, C., and Pérez, A. G. (2007). Contribution of olive seed to the phenolic profile and related quality parameters of virgin olive oil. *J. Sci. Food Agric.* 87, 2721–2727. doi: 10.1002/jsfa.3049
- Mateos, R., Cert, A., Pérez-Camino, M. C., and García, J. M. (2004). Evaluation of virgin olive oil bitterness by quantification of secoiridoid derivatives. *J. Am. Oil Chem. Soc.* 81, 71–75. doi: 10.1007/s11746-004-0859-x
- Mazzuca, S., Spadafora, A., and Innocenti, A. M. (2006). Cell and tissue localization of  $\beta$ -glucosidase during the ripening of olive fruit (*Olea europaea*) by *in situ* activity assay. *Plant Sci.* 171, 726–733. doi: 10.1016/j.plantsci.2006.07.006
- Minic, Z. (2008). Physiological roles of plant glycoside hydrolases. *Planta* 227, 723–740. doi: 10.1007/s00425-007-0668-y
- Montedoro, G., Baldioli, M., Selvaggini, R., Begliomini, A. L., and Taticchi, A. (2002). Relationship between phenolic compounds of olive fruit and olive oil: importance of the endogenous enzymes. *Acta Hort.* 586, 551–556. doi: 10.17660/ActaHortic.2002.586.115
- Morant, A. V., Jørgensen, K., Jørgensen, C., Paquette, S. M., Sánchez-Pérez, R., Møller, B. L., et al. (2008).  $\beta$ -Glucosidases as detonators of plant chemical defense. *Phytochemistry* 69, 1795–1813. doi: 10.1016/j.phytochem.2008.03.006
- Muñoz-Mérida, A., González-Plaza, J. J., Cañada, A., Blanco, A. M., García-López, M. C., Rodríguez, J. M., et al. (2013). *De novo* assembly and functional annotation of the olive (*Olea europaea*) transcriptome. *DNA Res.* 20, 93–108. doi: 10.1093/dnares/dss036
- Obied, H. K., Prenzler, P. D., Ryan, D., Servili, M., Taticchi, A., Esposto, S., et al. (2008). Biosynthesis and biotransformations of phenol-conjugated oleosidic secoiridoids from *Olea europaea* L. *Nat. Prod. Rep.* 25, 1167–1179. doi: 10.1039/b719736e
- Pinto, J., Paiva-Martins, F., Corona, G., Debnam, E. S., Oruna-Concha, M. J., Vauzour, D., et al. (2011). Absorption and metabolism of olive oil



- secoiridoids in the small intestine. *Br. J. Nutr.* 105, 1607–1618. doi: 10.1017/S000711451000526X
- Popescu, S. C., Popescu, G. V., Bacham, S., Zhang, Z., Seay, M., Gerstein, M., et al. (2007). Differential binding of calmodulin-related proteins to their targets revealed through high-density *Arabidopsis* protein microarrays. *Proc. Natl. Acad. Sci. U.S.A.* 104, 4730–4735. doi: 10.1073/pnas.0611615104
- Ramos-Escudero, F., Morales, M. T., and Asuero, A. (2015). Characterization of bioactive compounds from monovarietal virgin olive oils: relationship between phenolic compounds-antioxidant capacities. *Int. J. Food Prop.* 18, 348–358. doi: 10.1080/10942912.2013.809542
- Romero-Segura, C., García-Rodríguez, R., Sánchez-Ortiz, A., Sanz, C., and Pérez, A. G. (2012). The role of olive  $\beta$ -glucosidase in shaping the phenolic profile of virgin olive oil. *Food Res. Int.* 45, 191–196. doi: 10.1016/j.foodchem.2014.11.033
- Romero-Segura, C., García-Rodríguez, R., Sanz, C., and Pérez, A. G. (2011). Virgin olive oil phenolic profile as a result of the anabolic and catabolic enzymes status in the olive fruit. *Acta Hort.* 924, 379–384. doi: 10.17660/ActaHortic.2011.924.47
- Romero-Segura, C., Sanz, C., and Pérez, A. G. (2009). Purification and characterization of an olive fruit  $\beta$ -glucosidase involved in the biosynthesis of virgin olive oil phenolics. *J. Agric. Food Chem.* 57, 7983–7988. doi: 10.1021/jf901293c
- Servili, M., Esposto, S., Lodolini, E., Selvaggini, R., Taticchi, A., Urbani, S., et al. (2007). Irrigation effects on quality, phenolic composition, and selected volatiles of virgin olive oils cv. Leccino. *J. Agric. Food Chem.* 55, 6609–6618. doi: 10.1021/jf070599n
- Stefanoudaki, E., Williams, M., Chartzoulakis, K., and Harwood, J. (2009). Effect of irrigation on quality attributes of olive oil. *J. Agric. Food Chem.* 57, 7048–7055. doi: 10.1021/jf900862w
- Visioli, F., and Bernardini, E. (2011). Extra virgin olive oil's polyphenols: biological activities. *Curr. Pharm. Des.* 17, 786–804. doi: 10.2174/138161211795428885
- Xu, Z., Escamilla-Trevino, L., Zeng, L., Lalgondar, M., Bevan, D., Winkel, B., et al. (2004). Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. *Plant Mol. Biol.* 55, 343–367. doi: 10.1007/s11103-004-0790-1
- Yousfi, K., Moyano, M. J., Martínez, F., Cayuela, J. A., and García, J. M. (2010). Postharvest heat treatment for olive oil debittering at the industrial scale. *J. Am. Oil Chem. Soc.* 87, 1053–1061. doi: 10.1007/s11746-010-1588-8

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Velázquez-Palmero, Romero-Segura, García-Rodríguez, Hernández, Vaistij, Graham, Pérez and Martínez-Rivas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Variability in Susceptibility to Anthracnose in the World Collection of Olive Cultivars of Cordoba (Spain)

Juan Moral<sup>1,2\*</sup>, Carlos J. Xavier<sup>3</sup>, José R. Viruega<sup>3</sup>, Luis F. Roca<sup>3</sup>, Juan Caballero<sup>4</sup> and Antonio Trapero<sup>3</sup>

<sup>1</sup> Departamento de Agronomía, Universidad de Córdoba, Córdoba, Spain, <sup>2</sup> Department of Plant Pathology, Kearney Agricultural Research and Extension Center, University of California, Davis, Davis, CA, United States, <sup>3</sup> Departamento de Agronomía, ETSIAM, Universidad de Córdoba, Córdoba, Spain, <sup>4</sup> Departamento de Olivicultura, IFAPA Centro Alameda del Obispo, Córdoba, Spain

## OPEN ACCESS

### Edited by:

Leire Molinero-Ruiz,  
Instituto de Agricultura Sostenible  
(CSIC), Spain

### Reviewed by:

Franco Nigro,  
Università degli studi di Bari Aldo  
Moro, Italy  
Natalia Peres,  
University of Florida, United States

### \*Correspondence:

Juan Moral  
juanmoralmoral@yahoo.es;  
jmoral@ucdavis.edu

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

Received: 31 July 2017

Accepted: 18 October 2017

Published: 06 November 2017

### Citation:

Moral J, Xavier CJ, Viruega JR,  
Roca LF, Caballero J and Trapero A  
(2017) Variability in Susceptibility to  
Anthracnose in the World Collection of  
Olive Cultivars of Cordoba (Spain).  
Front. Plant Sci. 8:1892.  
doi: 10.3389/fpls.2017.01892

Anthracnose of olive (*Olea europaea* ssp. *europaea* L.), caused by *Colletotrichum* species, is a serious disease causing fruit rot and branch dieback, whose epidemics are highly dependent on cultivar susceptibility and environmental conditions. Over a period of 10 years, there have been three severe epidemics in Andalusia (southern Spain) that allowed us to complete the assessment of the World Olive Germplasm Bank of Córdoba, one of the most important cultivar collections worldwide. A total of 308 cultivars from 21 countries were evaluated, mainly Spain (174 cvs.), Syria (29 cvs.), Italy (20 cvs.), Turkey (15 cvs.), and Greece (16 cvs.). Disease assessments were performed using a 0–10 rating scale, specifically developed to estimate the incidence of symptomatic fruit in the tree canopy. Also, the susceptibility of five reference cultivars was confirmed by artificial inoculation. Because of the direct relationship between the maturity of the fruit and their susceptibility to the pathogen, evaluations were performed at the end of fruit ripening, which forced coupling assessments according to the maturity state of the trees. By applying the cluster analysis to the 308 cultivars, these were classified as follows: 66 cvs. highly susceptible (21.4%), 83 cvs. susceptible (26.9%), 66 cvs. moderately susceptible (21.4%), 61 cvs. resistant (19.8%), and 32 cvs. highly resistant (10.4%). Representative cultivars of these five categories are “Ocal,” “Lechín de Sevilla,” “Arbequina,” “Picual,” and “Frantoio,” respectively. With some exceptions, such as cvs. Arbosana, Empeltre and Picual, most of the Spanish cultivars, such as “Arbequina,” “Cornicabra,” “Hojiblanca,” “Manzanilla de Sevilla,” “Morisca,” “Picudo,” “Farga,” and “Verdial de Huévar” are included in the categories of moderately susceptible, susceptible or highly susceptible. The phenotypic evaluation of anthracnose reaction is a limiting factor for the selection of olive cultivars by farmers, technicians, and breeders.

**Keywords:** olive, diseases, anthracnose, *Colletotrichum*, fruit rot

## INTRODUCTION

Olive (*Olea europaea* ssp. *europaea* L.) is the most extensively planted fruit crop in the world, covering more than 10.2 million hectares of land, mainly in the Mediterranean Basin. Almost 25% of the total olive trees are grown in Spain, where more than 45% of the world's olive oil is produced (FAO, 2015). Olive industry (oil and table) is a vital sector of Spanish agro-food system with a total

value of 3 billion euros, more than 385,000 agricultural holdings, and over half million farmers (INE, 2015)<sup>1</sup> In fact, the global price of olive oil yearly depends to a large extent to the Spanish production.

Olive oil has numerous beneficial properties, mainly associated with its high content of monounsaturated oleic acid (Espósito et al., 2004) albeit other minor components, such as phenolic compounds (i.e., hydroxytyrosol, oleocanthal, and squalene) have also shown substantial benefits for consumer health (Beauchamp et al., 2005). For this reason, olive oil consumption and, concomitantly the olive-growing area, have notably increased worldwide in recent years. Projects to expand the olive-growing surface affect new areas for this crop, such as different provinces of China, India, Saudi Arabia, or the States of Florida and Hawaii in the USA. In many of these plantation areas, the adaptation of the different olive cultivars and the pests and diseases of this crop are unknown. Fortunately, the technicians and farmers have a broad range of cultivars that can be selected according to their adaptation to the target agro-environment area.

Olive tree was probably domesticated over 6,000 years ago in the Middle East, from which it spread until covering the entire Mediterranean Basin. It is also accepted the existence of other diversification centers across Mediterranean Basin (Besnard et al., 2013; Díez et al., 2015). The first farmers selected the most outstanding individuals in each olive-growing area according to their adaptation to the soil and climate and their agronomic characteristics. These original cultivars were subsequently maintained by vegetative propagation and, in general, have remained confined to small areas (Rallo et al., 2005; Díez et al., 2015). The presence of homonymy (different cultivars with the same name in different zones), synonymy (a given cultivar with various names in the areas that it occupies), and wrong denominations is frequent in this crop (Ganino et al., 2006; Trujillo et al., 2013). Thereafter, the exact number of olive cultivars is unknown but it is likely that this number reaches around 2000 (Bartolini et al., 1998). For this reason, and due to the absence of systematic studies, there is a reduced classification of the olive cultivars according to their susceptibility to many diseases, particularly in the case of anthracnose (Moral and Trapero, 2009).

Olive anthracnose, caused by the fungal complex species *Colletotrichum acutatum sensu lato* (s. lat.), *C. boninense* s. lat., and *C. gloeosporioides* s. lat. is the most destructive disease of olive fruit and is widely distributed in many olive-growing regions of the world (Martín and García-Figueres, 1999; Talhinhos et al., 2005; Cacciola et al., 2012; Moral et al., 2014; Schena et al., 2014). About 13 *Colletotrichum* species, belonging to these three complex species have been described affecting this crop (Talhinhos et al., 2005; Schena et al., 2014; Chattaoui et al., 2016). In general, several *Colletotrichum* species coexist in each olive-growing region with one or two dominant species and several secondary (Faedda et al., 2011; Moral et al., 2014). For example, the species *C. godetiae* (syn. *C. clavatum*) and *C. acutatum sensu stricto* (s. str.) are dominant in olive orchards

of southern Italy (Schena et al., 2017); while *C. acutatum* s. str. and *C. gloeosporioides* s. s. are dominant and secondary species, respectively, in Tunisia (Chattaoui et al., 2016).

Olive anthracnose is also called soapy fruit due to its characteristic fruit-rot syndrome with a profuse production of spores in a gelatinous matrix under wet conditions (Moral et al., 2009). In addition to the direct losses due to the premature fall of affected fruit, phytotoxins produced by the pathogen in the rotten fruit cause a second syndrome, the dieback of shoots and branches (Ballio et al., 1969; Moral et al., 2009). Furthermore, even with low fruit-rot incidence (5%), the olive oil coming from affected orchards shows poor chemical and organoleptic characteristics that restrict or impede its commercialization as extra virgin olive oil (Moral et al., 2014).

Both cultivar susceptibility and weather conditions profoundly influence on the olive anthracnose severity (Moral and Trapero, 2012). For example, in the central provinces of Andalusia, where the susceptible cultivars Hojiblanca and Picudo grow, severe epidemics occur if the weather conditions are conducive during the autumn despite the farmers typically applying a 2–3 copper-based fungicide treatments during this period. In southern Portugal, where autumn-winter is more humid than in central Andalusia, severe epidemics occur in super-high-density (hedgerow system) olive orchards planted with the moderately susceptible cultivar Arbequina (Moral and Trapero, 2012; Moral et al., 2014).

Although satisfactory results controlling olive anthracnose can be obtained using inorganic and organic fungicides, field application is not always effective for different reasons, such as: (i) the number of registered fungicides in post-bloom is very small; (ii) *Colletotrichum* shows a high tolerance to copper, the basic ingredient of the main fungicides used; (iii) and the optimum period for fungicide application is relatively short (Roca et al., 2007; Cacciola et al., 2012; Moral et al., 2014). Since the unripe fruit are resistant to the pathogen, the use of early harvesting before the fruit reaches full ripening or the selection of late-maturing cultivars are efficient and environmentally friendly control measures (Moral et al., 2008). However, these practices have some agronomical inconveniences: (i) if the fruit is immature (not completely black), it usually shows less oil content than mature fruit; (ii) the immature fruit show a high fruit retention force being difficult its mechanical harvest; and (iii) when the fruit still immature (green) during the winter, it is highly sensitive to frost damage (Rallo et al., 2005). Therefore, the use of resistant olive cultivars to anthracnose is the most effective control method, which does not show the previous inconvenience, and can be combined with other measures, such as biological and chemical methods or cultural practices (Moral et al., 2008; Moral and Trapero, 2009; Preto et al., 2017).

The World Olive Germplasm Bank of Córdoba (WOGBC) from Andalusia region, southern Spain, is a magnificent setting to evaluate the susceptibility of the olive cultivars to aerial diseases, including anthracnose, for several reasons: (i) the WOGBC is currently one of the largest olive germplasm banks with more than 900 accessions and 411 cultivars from 24 countries (A. Belaj, unpublished data); (ii) it is located in an endemic anthracnose area (Moral et al., 2015); and (iii) the whole group

<sup>1</sup><http://www.ine.es..>

of olive trees of WOGBC has been identified using Simple Sequence Repeats (SSR) markers, resolving the discrepancies due to misidentification of the trees (Trujillo et al., 2013). During the last 10 years, we have developed and validated laboratory and field methods to evaluate the susceptibility of olive cultivars to anthracnose (Moral et al., 2008; Moral and Trapero, 2009). These methods have been used to assess the resistance of traditional olive cultivars and new ones, coming from the breeding program of the University of Córdoba (UCO) and the Andalusian Institute for Research and Formation in Agriculture and Fishery (IFAPA) (Moral and Trapero, 2009; Moral et al., 2015). While the disease reaction of some olive cultivars is well-known (Moral and Trapero, 2009; Talhinas et al., 2015), most of cultivars are still unclassified for their resistance to this pathogen (Moral et al., 2014).

Because the olive cultivars from WOGBC have not been systematically screened for resistance to *Colletotrichum*, the objectives of this study were the following: (i) to assess the susceptibility of cultivars in the WOGBC to anthracnose caused by *Colletotrichum* spp.; (ii) to identify cultivars representative of each of the susceptibility categories determined in this study; and (iii) to correlate the anthracnose susceptibility with other phenotypic characteristics of cultivars. The phenotypic evaluation of disease reaction is a limiting factor for the selection of olive cultivars by farmers, technicians, and breeders.

## MATERIALS AND METHOD

### Plant Material and Orchard

We have evaluated 308 accessions of cultivated olive trees from 22 countries of origin (Table 1). These accessions are conserved as a live collection in the WOGBC located in a 5.2-ha flat and uniform field (37.51°N, 4.18°W, altitude 113 m) at the IFAPA, Center “Alameda del Obispo.” The soil of the orchard was classified as a Typic Xerofluvent with a sandy-loam texture, and the climatic conditions were typical of the Mediterranean area (García-López et al., 2016). The experimental orchard is located ≈1 Km from the main river of Andalucía, Guadalquivir River, in a humid area where anthracnose is an endemic disease (Moral et al., 2015). Because these accessions are fully authenticated by Trujillo et al. (2013), we will refer to them as cultivars. The WOGBC currently contains more than 411 cultivars (Caballero et al., 2006; A. Belaj, unpublished data). Olive trees used were planted between 1982 and 1992, so they were at least 5-years-old trees when evaluated. Because the experimental plot has a completely randomized design, several anthracnose susceptible cultivars (Table 1) are randomly distributed in the WOGBC resulting in a homogeneous distribution of *Colletotrichum s. lat.* inoculum, as with other pathogens, such as *Venturia oleaginea* (López-Doncel, 2003; Moral, 2009). In this study, which covers the period 1997–2008, we present the result of the evaluation of 308 well-identified cultivars according to their reaction to the pathogen. The remaining cultivars were not evaluated for different reasons: tree loss by *Verticillium dahliae*, absence of yield, absence of replicated trees of the same cultivar, etcetera. Each of the 308 evaluated cultivars has from two to 22 replicated trees randomly distributed in the experimental plot.

The olive trees were planted in a 7 × 7 m square using one trunk per plant. The trees were initially pruned to select three or four main branches to form the canopy structure: a free open vase. After that, the olive trees were periodically pruned to renewal branches or to eliminate dead branches. The experimental orchard was irrigated during spring-summer applying over 2,000 m<sup>3</sup> of water per year using drip irrigation. Three Bordeaux mixture treatments (Caldo Bordelés Vallés, IQV, 2 kg active Cu++ per ha and treatment) were applied throughout the year at the end of winter (February–March), spring (April–May), and autumn (October) (Roca et al., 2007). According to olive pest, it should be noted that the population of the olive fruit fly (*Bactrocera oleae*) is well established in the WOGBC area albeit there are not specific studies about it (Caballero J., unpublished data).

In previous studies, we identified the population of the pathogen in the WOGBC as *C. acutatum* group A4 (Moral et al., 2015), which is the only group described in the central provinces of Andalusia (Moral et al., 2014). The molecular group A4 was reclassified as *C. clavatum species nova* (sp. nov.) by Faedda et al. (2011). However, Damm et al. (2012) showed then that *C. clavatum* is synonym of the previous species *C. godetiae*. Because the latter name is rarely used in non-etiology studies, we maintain the generic name *C. acutatum s. lat.* throughout the present article.

### Assessment of Disease Incidence in the Field

Fruit-rot incidence was assessed in each olive tree using a previously described and validated 0–10 rating scale (Moral and Trapero, 2009). The rating scale is the logistic transformation of the proportion of symptomatic fruit and it based on the sigmoidal equation:

$$Y = \frac{100}{1 + 3^{(7-X)}} \quad (1)$$

in which  $Y$  = percentage of affected fruit and  $X$  = scale value (0–10). The fruit-rot incidence data are normalized using this scale, and the scale values can be analyzed directly using parametric methods (Moral and Trapero, 2009; Table 2). When the olive trees showed a high percentage (>5%) of symptomatic fruit, the assessor directly quantified the percentage of symptomatic fruit on the tree canopy. The percentage of affected fruit was then transformed in scaling rating values (0–10) rewriting the above equation as:

$$X = 7 - \frac{\ln\left(\frac{100}{Y} - 1\right)}{\ln 3} = 7 - \log_3\left(\frac{100}{Y} - 1\right) \quad (2)$$

Disease incidence was assessed at different times when most the fruit showed a value 4 in the 0 to 4 rating scale (from green to black, respectively) for olive fruit ripening (Rallo et al., 2005). Disease incidence was assessed from mid-December to the end of January in 1997–1998, 2005–2006, and 2006–2007 when severe epidemics of olive anthracnose occurred in the experimental orchard (Moral and Trapero, 2009). Data from low-level epidemic years were not considered because these data only



**TABLE 1** | Reaction of olive cultivars to anthracnose, caused by *Colletotrichum acutatum*, in the World Olive Germplasm Bank of Córdoba (Spain).

Reaction	No	Cultivar <sup>a</sup>
Highly Susceptible (HS)	66	Abbadí, Acebuchera, Adkam, Agouromanakolia, Azapa, Bolvino, Borriolenca, Bouteillan, Cañivano Negro-55, Chalchali, Chorro de Castro del Río, Corbella, Cornicabra de Mérida, Dulzal de Carmona, Farga, Forastera de Tortosa, Fulla de Salze, Gatuno, Gerboui, Gordal Sevillana, Grosal de Cieza, Habichuelero de Baena, Hojiblanca, Imperial, Imperial de Jaén, Jabaluna, Jaropo, Limoncillo, Llorón de Atarfe, Loaime, Lucio, Lucques, Machorrón, Mahati-1010, Manzanilla de Almería, Manzanillera de Huércal, Manzanillo de Cabra, Merhavia, Meski, Morisca, Morisca de Mancor, Negrillo de Iznalloz, Negrillo de la Carlota, Negrillo Redondo, Nevadillo de Santisteban del Puerto, Nevado Azul, <b>Ocal</b> , Ocal-25, Ojo de Liebre, Palomar, Pavo, Picudo, Picudo Blanco de Estepa, Rachati, Rechino, Safrawi, Salgar Redondo, Salonenque, Sant Agostino, Sayfi, Sevillana de Abia, Sevillanca, Temprano, Uovo di Piccione, Uslu, and Verdial de Cádiz.
Susceptible (S)	83	Abbadí Abou Gabra-84, Abbadí Shalal, Abou Kanani, Adramitini, Alameño Blanco, Alameño de Cabra, Alameño de Montilla, Amargoso, Amygdalolia Nana, Argudell, Ascolana Tenera, Asnal, Ayvalik, Barri, Blanqueta, Carolea, Castellana, Chorreo de Montefrío, Cipressino, Cirujal, Cordobés de Arroyo de la luz, Cordovil de Castelo Branco, Cornezuelo de Jaén, Cornicabra, Doebli, Dulzal, Escarabajillo, Escarabajuelo de Atarfe, Escarabajuelo de Posadas, Figueretes, Galega Vulgar, Gemlik, Gordal de Granada, Gordal de Hellín, Gordal de Vélez Rubio, Grosal Vimbodí, Izmir Sofralik, Jlot-841, Kaesi, Kalamon, Kolybada, <b>Lechín de Sevilla</b> , Lentisca, Levantinka, Macho de Jaén, Mahati-846, Manzanilla de Piedra Buena, Manzanilla de Agua, Manzanilla de Sevilla, Manzanilla Prieta, Mastoidis, Mawi, Mollar Basto, Mollar de Cieza, Morona Negral de Sabiñan, Negrillo de Estepa, Negro del Carpio, Nevadillo Blanco de Jaén, Nevado Basto, Nevado Rizado, Olivo de Mancha Real, Pajarero, Perafort, Pico Limón de Grazaema, Picual de Almería, Polinizador, Racimal, Shami, Sorani, Tempranillo de Yeste, Torcio de Huelma, Toruno, Varudo, Vera, Verdal de Manresa, Verdala, Verdale, Verde Verdelho, Verdial de Badajoz, Verdial de Huévar, Verdiell, and Zarza.
Moderately Susceptible (MS)	66	Abou Satl Mohazam, <b>Arbequina</b> , Arroniz, Azulejo, Barnea, Beyaz Yaglik, Bodoquera, Caballo, Cañivano Negro, Carrasquenho de Elvas, Carrasqueño de Alcaudete, Carrasqueño de la Sierra, Carrasquillo, Cerezuela, Changlot Real, Chorro, Çobrancosa, Cordovil de Serpa, Cornicabra de Jerez, Cornicabra-1, Domat, Dwarf D, Erbek Yaglik, Escarabajuelo de Úbeda, Gaydoyrelia, Genovesa, Habichuelero de Grazaema, Itrana, Kelb et Ter, Khashabi, Kiraz, Klon-14-1081-1, Klon-14-1081-2, Konservolia, Lastovka, Lechín de Granada, Maarri, Majhol-1059, Manzanilla Cacereña, Manzanilla de Abia, Manzanilla del Piquito, Mari, Masabi, Mission, Moojeski, Morrut, Myrtolia, Negrillo de Arjona, Oblica, <i>Olea ferruginea</i> , Olivo de Maura, Picholine, Plementa Bjelica, Pulazeqin, Rapasayo, Redondilla de Logroño, Reixonenca, Royal de Calatayud, Royal de Cazorla, Royal de Sabiñan, Sabatera, Sandalio, Tanche, Toffahi, Valanolia, Villalonga, and Vinyols.
Resistant (R)	61	Abbadí Abou Gabra-10, Acebuche de Caravaca, Alfafara, Aloreña de Iznalloz, Arbosana, Athalassa, Beladi, Belluti, Bent al Kadi, Biancollilla, Bical, Buga, Buidiego, Canetera, Carrasqueño de Jumilla, Carrasqueño de Porcuna, Chalkidikis, Chemlal de Kabilie, Chemlali, Chetoui, Coratina, Corralones de Andújar, Curivell, Datilero, Dokkar, Dolce, Elmacik, Enagua de Arenas, Fishomi, Grit Eytini, Grossanne-67, Joanenca, Kalokerida, Kotruvsi, Leccino, Llumeta, Majhol-1013, Majhol-152, Manzanilla de Hellín, Manzanilla de Montefrío, Manzanilla de San Vicente, Maurino, Memecik, Mixani, Moraiolo, Patronet, Pecoso, Picholine Marocaine, Pico Limón, <b>Picual</b> , Real Sevillana, Redondilla de Grazaema, Sevillano de Jumilla, Shengue, Sinop, Sollana, Vallesa, Vaneta, Verdial de Vélez-Málaga, Zaity, and Zard.
Highly Resistant (HR)	32	Ayrouni, Azul, Bosana, Callosina, Caninese, Crnica, Dolce Agogia, Empeltre, <b>Frantoio</b> , Frantoio A. Corsini, Grappolo, Istarska Bjelica, Kan Çelebi, Kato Drys, Koroneiki, Mavreya, Megaritiki, Menya, Ouslati, Pendolino, Pequeña de Casas Ibañez, Perillo de Jaén, Piñonera-76, Racimal de Jaén, Rosciola, Royal de Calatayud-4, Selvatico, Toffahi-1000, Ulliri i Bardhe i Ti, Ulliri i Kuq, Warden, and Zalmati.

<sup>a</sup> Reference cultivars appear in bold.

divide the olive cultivars into two groups, the highly susceptible cultivars, which show a low-medium incidence of fruit rot, and the rest of cultivars showing no symptomatic fruit (Moral and Trapero, 2009; Xaviér, 2009).

Because at the beginning of the spring 2007 there was an important peak of the dieback syndrome (chlorosis and wilting of leaves and dieback of shoots and branches, i.e., green and lignified shoots, respectively) after two epidemic years, we assessed this second syndrome according to volume of tree canopy affected using the following rating scale: 0 < 10%, 1 = 10–24%, 2 = 25–49%, 3 = 50–74%, 4 = 75–90%, and 5 ≥ 90% (Moral and Trapero, 2009). In addition, we evaluated the presence of the pathogen on symptomatic tissues (leaves and dieback shoots) by culturing small pieces on acidified Potato Dextrose Agar plus 100 mg of copper sulfate (CuSO<sub>4</sub>·5H<sub>2</sub>O) per liter (Moral et al., 2009) from May to August during 2006 and 2007.

## Inoculation of Detached Fruit

Apparently asymptomatic yellowish-green—value of 2 on the olive ripening scale (Rallo et al., 2005)—fruit of five cultivars (“Arbequina,” “Frantoio,” “Lechín de Sevilla,” “Ocal,” and “Picual”) were collected at the onset of ripening from olive trees in the experimental orchard during the non-epidemic year 2012. These reference cultivars were selected by their well-known response to anthracnose in the field (Moral et al., 2014, 2015). The fruit were inoculated and incubated according to Moral et al. (2008). Briefly, fruit were washed, disinfested, and sprayed with a conidial suspension (10<sup>5</sup> conidia per ml or sterile water for the control) of isolate Col-104 of *C. acutatum* s. lat. Inoculated and control fruit were incubated in humid chambers (plastic containers) at 22–24°C under fluorescent lights (12 h alternating photoperiod, 40 μmol m<sup>-2</sup>s<sup>-1</sup>). Disease severity was periodically assessed for 80 days using a 0–5 rating scale where 0 = no

**TABLE 2 |** Rating scale values, average, and interval of percentage of olive fruit affected by anthracnose.

Scale value (0–10)	Description	Average (Interval [%] <sup>a</sup> )	Range <sup>b</sup>
0	Affected fruit not observed	<0.04	0.04
1	From one to three affected fruit per olive tree	0.14 (0.04–0.23)	0.19
2	From one to three affected fruit per each quadrant of the tree canopy	0.41 (0.24–0.70)	0.46
3	From four to nine affected fruit per each quadrant of the tree canopy	1.22 (0.71–2.09)	1.38
4	Affected fruit easily detected (from 20 to 36 affected fruit per each half of the tree canopy)	3.57 (2.10–6.02)	3.92
5	Direct quantification of affected fruit	10 (6.03–16.13)	10.1
6	Direct quantification of affected fruit	25 (16.14–36.60)	20.46
7	Direct quantification of affected fruit	50 (36.61–63.39)	26.78
8	Direct quantification of affected fruit	75 (63.40–83.86)	20.46
9	Direct quantification of affected fruit	90 (83.87–93.97)	10.1
10	<36 asymptomatic fruit per each half of the tree canopy	97 (93.98–100)	3.92

<sup>a</sup>Values taken from the logistic function  $Y = \frac{100}{1+3^{(7-X)}}$  in which  $Y$  = percentage of affected fruit and  $X$  = scale value.

<sup>b</sup>Range of affected fruit (%) for each scale value.

<sup>c</sup>Detection limit of visual assessments in the field (one affected fruit from 2,500 observed fruit per tree; Moral and Trapero, 2009).

visible symptoms, 1 = visible symptoms affecting <25% of the fruit surface, 2 = 25–49%, 3 = 50–74%, 4 = 75–100%, and 5 = soapy fruit (Moral et al., 2008). There were two replicates (moist chambers) per treatment and 25 fruit per replicate arranged in a completely randomized design. The experiment was repeated once and data analyses were performed on the pooled data from the two replicates.

## Statistical Analysis

To estimate the phenotypic stability of the olive cultivar on anthracnose reaction during the three studied seasons, we calculated the proportion of variability explained by the cultivar, season, and the interaction cultivar-season. For that, analysis of variance (ANOVA) was performed on the rating scale data of fruit-rot incidence of the cultivars, which were evaluated during the three epidemic seasons with a representative number of repetitions (at least two olive trees of the same cultivar). Subsequently, we calculated eta squared ( $\eta^2$ ) as the ratio the variability associated with an effect and the total variability of our analysis ( $\eta^2 = \frac{SS_{\text{effect}}}{SS_{\text{Total}}}$ ). Because  $\eta^2$  overestimates the effect size, we also calculated partial omega squared ( $\omega^2$ ) using the equation (Fritz et al., 2012):

$$\omega^2 = \frac{SS_{\text{effect}} - df_{\text{effect}} \times MS_{\text{error}}}{SS_{\text{total}} + MS_{\text{error}}} \quad (3)$$

in which  $SS$  = sum of squares,  $df$  = degrees of freedom, and  $MS$  = mean square. Likewise, ANOVA was performed on the rating

scale data of fruit-rot incidence and branch dieback severity of the five reference cultivars.

In previous work, we have classified 18 cultivars into three categories of susceptibility by comparing their reaction according to the disease response of moderately resistant cultivar Arbequina (Moral et al., 2015). In this study, due to the high number of studied cultivars and, for establishing groups more homogeneous according to their reaction to the pathogen, the cultivars are classified roughly into five categories. These five susceptibility categories were determined according to the reaction of the reference cvs. Arbequina, Frantoio, Lechín de Sevilla, Ocal, and Picual, which were selected by their well-known response to anthracnose under field conditions (Moral et al., 2005, 2014, 2015). Subsequently, a non-hierarchical K-Means cluster analysis with the initial cluster center method under the assumption of five groups ( $k = 5$ , one group for each of the reference cultivars) was applied to the classification of the whole of the cultivars using the average value of the epidemic seasons due to the fact that it is the best parameter to discriminate the susceptibility/resistance reaction of the cultivars (Moral and Trapero, 2009; Table 2). The relationship between fruit-rot incidence and branch-dieback severity were analyzed by the non-parametric Spearman rank correlation. A chi-squared test was used to determine whether the country of origin of the cultivars had any effect on the frequency of resistant or susceptible cultivars.

In the artificial inoculation test, disease severity values of inoculated fruit were used to calculate the McKinney's Index (MKI; McKinney, 1923), in which disease severity is expressed as a percentage of the maximum possible level according to the following formula:

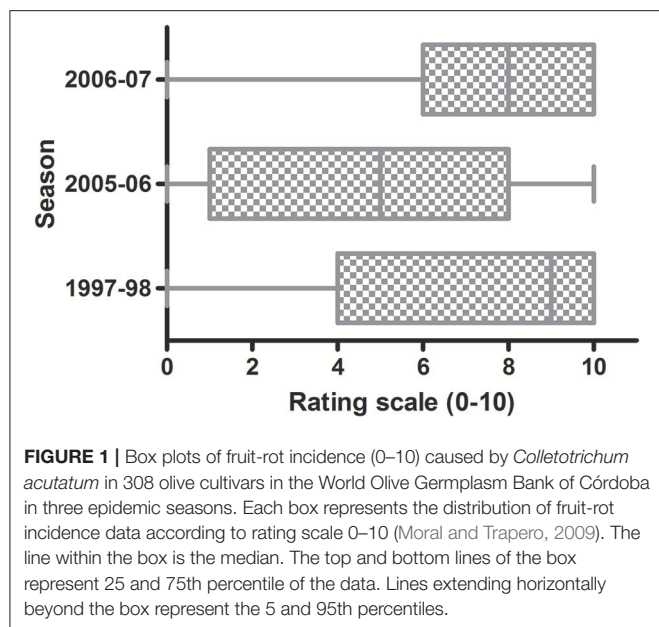
$$\text{McKinney's Index} = \frac{\sum (n_i \times i)}{5 \times N} \times 100 \quad (4)$$

where  $i$  represents the severity of symptoms (0–5),  $n_i$  is the number of fruit with the severity of  $i$ , and  $N$  is the total number of evaluated fruit. For each cultivar and replication, the standardized area under the disease progress curve (SAUDPC) was calculated by trapezoidal integration of MKI values over time (80 days) expressed as a percentage of a maximum theoretical curve. The SAUDPC was transformed to  $\arcsin\sqrt{\text{SAUDPC}/100}$  when necessary for homogeneity of variance. ANOVA was performed on the SAUDPC data, and treatment means were compared using LSD test at  $P = 0.05$ . Eta squared ( $\eta^2$ ) and omega squared ( $\omega^2$ ) were computed in MS Excel (Microsoft, Redmond, WA). Data from all experiments were analyzed using Statistix 9 (Analytical Software, Tallahassee, FL) except K-Means Cluster Analysis that was conducted using the SPSS 16 software.

## RESULTS

### Susceptibility of Cultivars in the Field

The incidence of fruit affected by *C. acutatum* s. lat. on the trees of the WOGBC was evaluated during three seasons (1997–1998, 2005–2006, and 2006–2007) due to the low fruit-rot incidence during the remaining years. Among these epidemic seasons, the most severe outbreak occurred in 1997–1998, whereas the



lowest levels of the disease occurred in 2005–2006. For example, about 50% of the evaluated olive trees had a rating value scale  $\geq 8$  (i.e., median  $\approx 8$ ) in 1997–1998 and 2006–2007, while the media value was around 5 in 2005–2006 indicating a higher degree of dispersion (Figure 1). In the studied seasons, the fruit-rot incidence greatly varied among cultivars and season. According to  $\eta^2$  and  $\omega^2$ , the cultivar was the most important factor explaining the total variance (approximately 60%) followed by the season and the interaction genotype-season. In other words, the differences in severity of symptoms among olive trees during the epidemic seasons were mainly due to the genotype.

Because cultivar reaction to the pathogen ranged continuously from highly susceptible to highly resistant, we selected five cultivars across the susceptibility range. Each of the selected cultivars was represented by at least five repetitions (trees) in the experimental plot. These cultivars were: “Frantoio” ( $n = 8$  trees) for the Highly Resistant cultivars (HR, 0–2 in the rating scale), “Picual” ( $n = 8$  trees) for the Resistant cultivars (R, 2.1–4.3 in the rating scale), “Arbequina” ( $n = 6$  trees) for the Moderately Susceptible cultivars (MS, 4.4–6.2 in the rating scale), “Lechín de Sevilla” ( $n = 5$  trees) for the Susceptible cultivars (S, 6.3–7.9 in the rating scale), and “Ocal” ( $n = 21$  trees) as representative of the Highly Susceptible cultivars (HS,  $> 8$  on the rating scale). The fruit-rot incidence caused by *C. acutatum* s. lat. on the five selected cultivars varied significantly ( $P < 0.05$ ) among them albeit these differences depended on the year (Table 3). Concurrent with the greatest anthracnose epidemic in 1997–1998, we detected main differences among these cultivars according to their anthracnose resistance. The cultivars Picual (R) and Arbequina (MS) did not differ significantly from each other during the three epidemic seasons according to the fruit-rot incidence, although the difference was significant when the analysis was applied to the average fruit-rot incidence of the three seasons. Likewise, the branch-dieback severity showed

**TABLE 3 |** Incidence of anthracnose in five olive cultivars in the World Olive Germplasm Bank of Córdoba during 1997–1998, 2005–2006, and 2006–2007.

Cultivar	Fruit-rot incidence (0–10) <sup>a</sup>				Branch <sup>b</sup> dieback (0–5)
	1997–1998	2005–2006	2006–2007	Average	2007
Frantoio	0.1d <sup>c</sup>	0.0d	0.1d	0.1e	0.0c
Picual	3.5c	1.0c	4.7c	3.4d	0.0c
Arbequina	4.5c	3.7bc	5.5c	4.9c	0.14b
Lechín de Sevilla	8.3b	5.3b	7.1b	7.0b	0.17b
Ocal	9.9a	7.7a	9.6a	9.2a	1.38a
Mean	6.5	5.1	6.1	5.7	0.34

<sup>a</sup>Fruit-rot incidence was estimated using a 0–10 rating scale in which binary data (proportion of affected fruit) are normalized by applying the logit transformation of proportion. Scale values were directly subjected to analysis of variance and mean comparison tests.

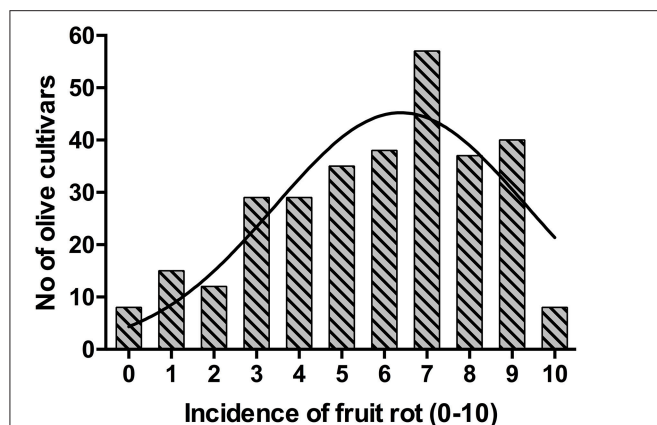
<sup>b</sup>Volume of olive tree canopy affected dieback syndrome (chlorosis and wilting of leaves and dieback of shoots and branches) according to the following rating scale: 0 < 10%, 1 = 10–24%, 2 = 25–49%, 3 = 50–74%, 4 = 75–90%, and 5 = > 90% (Moral and Trapero, 2009).

<sup>c</sup>Means with the same letter are not significantly different according to the Fisher's protected LSD test at  $P = 0.05$ .

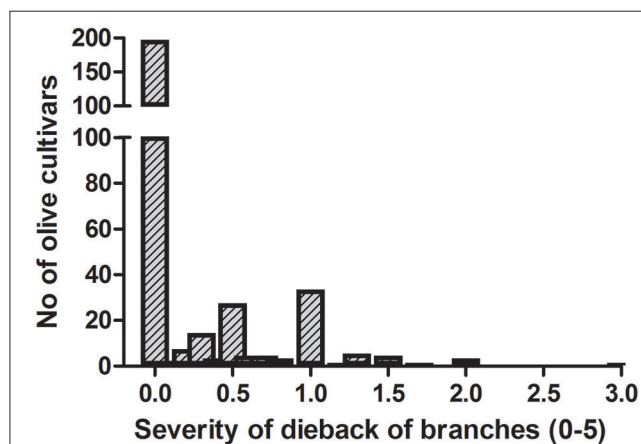
by reference cultivars differed significantly, being particularly severe in “Ocal” (Table 3). There was a significant and positive correlation (Spearman's correlation;  $r = 0.691$ ,  $P < 0.0001$ ) between the fruit-rot incidence and the volume (%) of tree canopy affected by the pathogen for these five cultivars.

Due to the large number of cultivars evaluated, the results showed a continuous gradation in the susceptibility of cultivars ranging from completely resistant to extremely susceptible. In other words, the resistance ranged from cultivars with none or very few affected fruit (e.g., “Dolce Agogia,” “Frantoio,” “Grappolo,” or “Mavreya”), to cultivars with all affected fruit (e.g., “Acebuchera,” “Picudo Blanco de Estepa,” or “Uovo Piccione”). By applying the cluster analysis to the 308 olive cultivars, these were classified as follows: 32 cvs. HR (10.4%), 61 cvs. R (19.8%), 66 cvs. MS (21.4%), 83 cvs. S (26.9%), and 66 cvs. HS (21.4%) (Table 1).

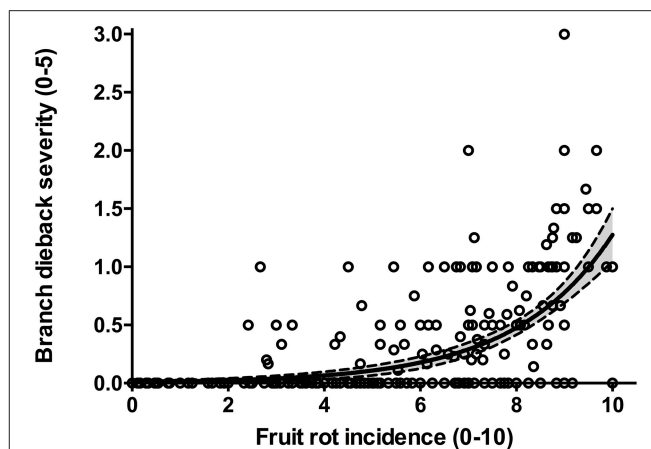
The frequency distribution of cultivars according to the categories of fruit-rot incidence is skewed to the left (skew parameter  $k = -0.436$ ) and had a negative kurtosis ( $k_u = -0.694$ ), which highlighted the prevalence of S or HS cultivars in the collection (Figure 2). Instead, the frequency distribution of cultivars according to the categories of branch-dieback severity showed a positive skew ( $k = 2.01$ ) and a positive kurtosis ( $k_u = 4.907$ ) due to the fact that most of cultivars did not show this second disease syndrome since it only occurred in some cultivars that had a high fruit-rot incidence (Figure 3). There was also a significant correlation (Spearman rank correlation:  $r = 0.530$ ,  $P < 0.001$ ) between the fruit-rot incidence and the volume (%) of tree canopy affected by the pathogen for all of the cultivars. This relationship was weak albeit it improved when we only considered the cultivars that showed a fruit-rot incidence higher than 2.5 (0.71% of affected fruit). A simple exponential growth curve was well fitted to the values of volume



**FIGURE 2 |** Frequency distribution of fruit-rot incidence caused by *Colletotrichum acutatum* in 308 olive cultivars in the World Olive Germplasm Bank of Córdoba (Spain).



**FIGURE 3 |** Frequency distribution of branch dieback caused by *Colletotrichum acutatum* in 308 olive cultivars in the World Olive Germplasm Bank of Córdoba (Spain).



**FIGURE 4 |** Relationship between fruit-rot incidence and branch-dieback severity caused by *Colletotrichum acutatum* in 308 olive cultivars in the World Olive Germplasm Bank of Córdoba (Spain). Lines represent the fitted exponential growth equation  $Y = 0.009 \times e^{0.492}$  and the confidential intervals at 95%.

dominance of susceptible cultivar for all of the cultivars. Conversely, when the comparisons were individually conducted among cultivars from the same country, there was no prevalence of resistant or susceptible cultivars in most cases, probably due to the low number of cultivars available. However, the susceptible cultivars were dominant in Spain and Syria while resistant ones were prevalent in Italy (Table 4).

The pathogen *C. acutatum* s. lat. was isolated from sampled leaves, shoots, and branches with dieback symptoms albeit with a relatively low frequency (<2.8%).

## Susceptibility of Cultivars in Artificial Inoculation

All inoculated cultivars developed typical anthracnose (soapy-rot) with significant differences among them. None of the non-inoculated fruit showed disease symptoms during the 80 days of incubation in humid chambers. The first symptoms were observed in the fruit of the cv. Ocal at 5 days after inoculation, whereas the fruit of the cv. Frantoio showed the first symptoms 23 days after inoculation. Likewise, the pathogen caused the complete rot (severity = 5) of all the fruit of the cvs. Lechín de Sevilla and Ocal on 21 days, while it needed more than 80 days to cause the completely rot of all of the fruit of the cv. Frantoio. In this latter cultivar, only the 17% of inoculated fruit had been shown anthracnose symptoms 40 days after inoculation. The SAUDPC analysis significantly separated the five inoculated cultivars (Figure 5). Finally, there was a significant correlation ( $r = 0.989$ ,  $P = 0.0015$ ) between the SAUDPC of inoculated fruit and the fruit-rot incidence in the field.

## DISCUSSION

Here, we present the largest evaluation of olive cultivars for their resistance to anthracnose, which is considered the most

(%) of affected canopy over fruit-rot incidence (Figure 4). Even so, some cultivars that showed high values of fruit-rot incidence did not show branch dieback, such as “Bouteillan,” “Grosal de Cieza,” “Imperial,” “Manzanilla de Almería,” and “Picudo Blanco de Estepa,” which had a fruit-rot incidence >9 but had no dieback symptoms. On the other hand, the cv. Salonenque olive trees showed most of their fruit affected by the pathogen (severity value = 9) and, also, a 50% branch-dieback (severity value = 3) of their canopy affected by dieback of branches.

Comparisons for each country between the frequency of resistant cultivars (HR and R categories) and the frequency of susceptible ones (MS, S, and HS categories) were conducted to determine whether the origin of the cultivars had any effect on their susceptibility. The null hypothesis in these tests was that there was no prevalence of any category of susceptibility while the alternative hypothesis was, that there was a prevalence of resistant or susceptible cultivars. Overall, there was a significant



**TABLE 4** | Origin and susceptibility to anthracnosis of the olive cultivars evaluated in the World Olive Germplasm Bank of Córdoba (Spain).

ORIGIN	Cultivars (No)	HR <sup>a</sup>	R	MS	S	HS	P-value <sup>b</sup>	R/S <sup>c</sup>
ALBANIA	7	2	2	3	0	0	0.3103	-
ALGERIA	1	0	1	0	0	0	0.2632	-
CHILE	1	0	0	0	0	1	0.4449	-
CROATIA	7	2	1	3	1	0	0.5480	-
CYPRUS	2	1	1	0	0	0	0.1136	-
EGYPT	3	1	1	1	0	0	0.3539	-
FRANCE	7	0	1	2	1	3	0.1693	-
GREECE	16	3	2	4	5	2	0.2752	-
IRAN	4	0	3	1	0	0	0.1758	-
ISRAEL	2	0	0	1	0	1	0.2800	-
ITALY	20	9	5	1	3	2	0.0082	R
LEBANON	2	1	1	0	0	0	0.1136	-
MEXICO	2	0	1	0	1	0	0.6850	-
MOROCCO	1	0	1	0	0	0	0.2632	-
PAKISTAN	1	0	0	1	0	0	0.4449	-
PORTUGAL	6	0	0	3	3	0	0.0613	-
SPAIN	174	9	28	34	55	48	0.0000	S
SYRIA	29	1	5	6	11	6	0.0249	S
TUNISIA	6	2	2	0	0	2	0.1898	-
TURKEY	15	1	6	4	3	1	0.3284	-
USA	2	0	0	2	0	0	0.2800	-
TOTAL	308	32	61	66	83	66	0.0000	S

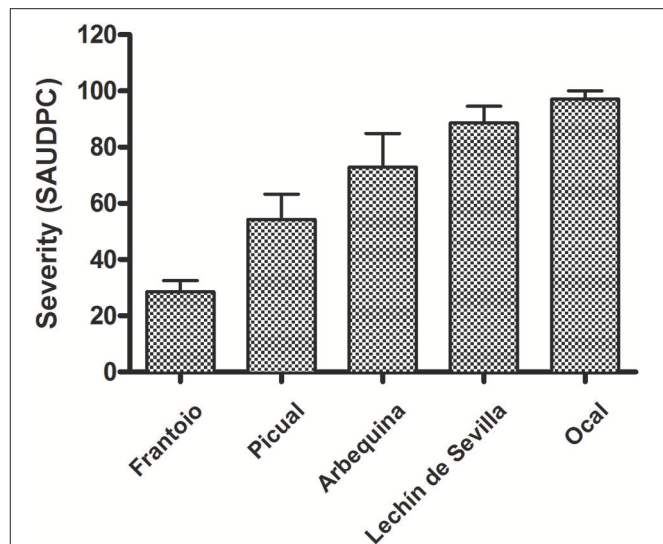
<sup>a</sup>HR = highly resistant (0–2 in the 0–0 rating scale), R = resistant (2.1–4.3), MS = intermediate (4.4–6.2), S = susceptible (6.3–7.9), and HS = highly susceptible (8–10).

<sup>b</sup>P-value of the chi-square test used to determine the dominance or non-dominance of resistant or susceptible cultivars.

<sup>c</sup>R = dominance of resistant cultivars, S = dominance of susceptible cultivars. - = no dominance of resistant or susceptible cultivars.

important fruit disease of this crop (Cacciola et al., 2012; Moral et al., 2014). We evaluated the reaction of 308 well-identified cultivars growing in the WOGBC during three epidemic seasons. Among others, two important reasons for which the WOGBC is an excellent experimental plot to evaluate the resistance to *C. acutatum* s. lat. are: (i) the plot is located in an endemic area for anthracnose due to the proximity of the biggest river of Andalusia (Guadalquivir river); (ii) the plot has a high number of HS olive cultivars that act as an inoculum source of *Colletotrichum* (Moral et al., 2015). In a previous evaluation, we presented a limited number of olive cultivars based on a single observation (Moral et al., 2005) and without the correct identification of the trees using molecular markers (Trujillo et al., 2013). In the latter study, Trujillo et al. (2013) clarified numerous problems of homonymy, synonymy, and wrong denominations in the WOGBC. For example, the original cultivars Arauco, Cañivano Blanco, Carrasqueño de Lucena, and Razzola, which were planted in the WOGBC, have been identified as synonymous of the cultivars Azapa, Picholine Marocaine, Picual, and Frantoio, respectively (Trujillo et al., 2013).

Olive fruit infection occurs at all stages of its development, from flower bud emergence to ripening (Moral et al., 2009). These infections occur mainly as a result of water-splashed

**FIGURE 5** | Standard Area Under Disease Progress Curve (SAUDPC) of five reference olive cultivars inoculated with *Colletotrichum acutatum*. All the cultivars were significantly different according to the Fisher's protected least significant difference (LSD) test at  $P = 0.05$ .

conidia. Also, the anthracnose disease cycle can be influenced by the activity of olive fruit fly (*B. oleae*), which may increase the fruit's susceptibility by causing wounds or directly act as a spores carrier (Malacrino et al., 2015). Fruit ripeness increases fruit susceptibility to anthracnose (Moral et al., 2008), with the unripen (or developing) fruit being very resistant to the pathogen, regardless of cultivar (Moral et al., 2008; Cacciola et al., 2012; Moral and Trapero, 2012). Therefore, cultivar resistance is evaluated by inoculating yellowish-green fruit with a spore suspension of the pathogen (Moral et al., 2008; Talhinhos et al., 2015). In any case, the use of this rating scale under field conditions to evaluate the cultivar reaction to the pathogen provides an important (>20-fold) economic saving respect to the artificial inoculation method (Moral, 2009). However, the correct evaluation of the olive cultivars under field conditions can only be conducted during epidemic seasons (Moral and Trapero, 2009), which are sporadic in Mediterranean conditions (Cacciola et al., 2012; Moral and Trapero, 2012). In our case, there were only three epidemic seasons during a period of 10 years. The epidemic intensity during the studied years (Figure 1) was associated with conducive weather conditions for anthracnose, mainly the annual rainfall. Thereafter, the greatest epidemic was in 1997, a wet year (888.4 mm) in Cordoba, which had been preceded by a very wet 1996 (951.1 mm). Furthermore, the epidemic years 2005 and 2006 were moderately wet with an annual rainfall of 548.8 and 560.7 mm, respectively.

In this study, the phenotypic resistance of the cultivars was very stable, as shown by the fact that the experimental error explained only around 10% of the total variance ( $\eta^2$ ), while the cultivar genotype explained over a 65% of this. This fact implies a high stability of the cultivar response to the pathogen (resistance/susceptibility) during the epidemic years. The season (which includes its weather conditions) and the

interaction cultivar-season had also a similar effect as that of the experimental error. We did not conduct anthracnose evaluations during non-epidemic seasons due to the almost complete absence of the disease during these years, even in the HS cultivars (Moral, 2009). In other words, anthracnose is a highly weather-dependent disease (Moral and Trapero, 2012).

As it was previously described, the cultivar reaction to *Colletotrichum* is a continuous variable ranging from HR to HS cultivars (Moral et al., 2015). Nevertheless, the susceptibility/resistance of the cultivars is more useful and easily understood by the farmers and agronomist if the cultivars are placed into distinct ordinal classes (Pataky et al., 2011). In this study, as in previous studies of olive diseases (Moral et al., 2005; Trapero and López-Doncel, 2005), we classified the cultivars in five categorical groups using a K-means analysis that minimized the variances. The intervals (the ranging values) of each categorical group can shift slightly depending on the number of evaluated cultivars since K-analysis uses a centroid value (average of the data of each group) for each given categorical group (Jain et al., 1999).

In general, susceptible cultivars (MS, S, and HS, total 215 cultivars) were more prevalent than resistant cultivars (R and HR, total 93 cultivars) regardless of country of origin, except for Italy. For Italian cultivars, resistance to the pathogen (*C. godetiae*) in the WOGB was prevalent. Since *Colletotrichum* spp. is endemic in north of Italy, most of these Italian cultivars could be selected by the farmers according their resistance to the pathogen (Barranco et al., 2000; Bartolini and Cerreti, 2017). The studied Italian cultivars also show a high degree of genetic homogeneity, for example, they all belong to the chlorotype group E1.1, except the S cultivar Carolea that belongs to the group E1.2 (Besnard et al., 2013). Likewise, all the Italian cultivars in the WOGBC, which belong to the genetic cluster 2 described by Trujillo et al. (2013), are resistant to anthracnose except “Cipressino.”

In the extremes of the resistance/susceptibility, we found the cultivar Dolce Agogia, which showed a complete resistance to the pathogen (i.e., absence of disease symptoms), and the cultivars Acebuchera, Picudo Blanco de Estepa, and Uovo Piccione, in which the fruit-rot incidence was 100%. In this study, we described most of the evaluated cultivars as susceptible (MS, S, or HS) to the pathogen, including the species *Olea ferruginea*, which was moderately susceptible. Results of present study agree with previous observations for susceptible cultivars: Ascolana tenera, Barnea, Galega vulgar, Gordal sevillana, Hojiblanca, Manzanilla de Sevilla, Morisca, Ocal, Picudo, Sant Agostino, and Verdial de Badajoz; and for the resistant or moderately resistant cultivars: Bical de Castelo Branco, Coratina, Frantoio, Empeltre, Leccino, Manzanilla cacereña, Mixani, and Picual (Barranco et al., 2000; Rallo et al., 2005; Moral et al., 2015; Talhinas et al., 2015; Bartolini and Cerreti, 2017). Conversely, the published response to anthracnose of some cultivars does not match with our current results. In our study, four of these cultivars [“Abou-Salt Mohazan,” “Azapa” (syn. “Arauco”), “Blanqueta,” and “Cordovil de Castelo Branco”] were more susceptible and seven of them (“Arbequina,” “Cobrançosa,” “Itrana,” “Moraiolo,” “Morrut,” and “Picholine”) were somewhat less susceptible than their respective previous classification (Bartolini and Cerreti, 2017).

Errors in the classification of olive cultivars according to their resistance to anthracnose have been extensively discussed by us and they are usually associated with: cultivar misidentification, effect of the ripeness time during the evaluation moment, low inoculum pressure, unfavorable environmental conditions for disease development, and confusion with other fruit-rot diseases caused by species of *Alternaria*, *Botryosphaeria*, *Fusarium*, or *Neofabraea* (Moral et al., 2008, 2014, 2015; Moral and Trapero, 2009). Furthermore, the potential interaction between *Colletotrichum* species (or isolate) and the olive cultivar needs a special mention. In general, this type of interaction has been described with MS and S cultivars, such as “Galega vulgar,” “Cobrançosa,” or “Hojiblanca.” Fortunately, R or HR cultivars, such as “Blanqueta,” “Picual,” or “Frantoio” show a high degree of resistance to the different isolates (Xaviér, 2009; Talhinas et al., 2015). Concomitantly, some *Colletotrichum* species (or isolates) tend to be weakly (e.g., *C. acutatum* s. str. or *C. rhombiforme*) or highly virulent (e.g. *C. godetiae* or *C. nymphaeae*) against a broad range of olive cultivars (Xaviér, 2009; Schena et al., 2014; Talhinas et al., 2015).

The histogram of resistance/susceptibility of the olive cultivars according to fruit-rot incidence showed that data fit a normal distribution but with a dominance of the susceptible cultivars. Similarly, a substantial deviation from normal distribution among cultivars for different agronomic characteristics has been described, such as yield, ripening data, and oil content (Caballero et al., 2006; León et al., 2008, 2011). Although our study was not intended to address the resistance mechanisms against the pathogen, it suggests a complex and polygenic control of the resistance to *Colletotrichum* species. In any case, other types of genetic control, including combinations of minor and major genes, could be involved in the resistance of the olive tree to *Colletotrichum* species (Geffroy et al., 2000). However, the case of major genes mediating the defense against non-biotrophic pathogens is rare (Poland et al., 2009). In addition, phenolic compounds have an important role in the defense of the fruit against fungal pathogen on different crops (Prusky, 1996), including olive (Moral et al., 2015); in this case, and due to phenolic acids derivate from different metabolic pathways in the plant (Boudet, 2007), there is not a clear relationship between genes and phenolic compound. Fortunately for the olive breeding programs, crosses between resistant cultivars produce a high frequency of resistant descendants (Moral et al., 2015).

Likewise, important differences have been observed according to the severity of branch-dieback among olive cultivars. This second syndrome of olive anthracnose is associated with the production of phytotoxins by the pathogen on rotten fruit (Moral et al., 2009, 2014). In our study, both fruit-rot incidence and branch-dieback severity appear correlated albeit some of the cultivars, which showed high values of fruit-rot incidence, did not show dieback symptoms. These results suggest differences in the resistance mechanisms for both syndromes. In the experimental plot, the percentage of isolation of *Colletotrichum* in semi-selective medium from symptomatic tissues of the olive trees were relatively low (<3%) during spring-summer and it was associated with the rainfall events. Schena et al. (2017), using a duplex qPCR, have described a high colonization of these

vegetative tissues from May to October. In our conditions, we have never observed fruiting bodies (acervuli) of the pathogen on leaves or shoots under field conditions (Moral et al., 2009), although they can be induced in humid chamber after 1 month of incubation (Moral et al., 2014). For this reason, we think that these tissues have a limited role as inoculum sources in comparison with affected fruit in southern Spain (Moral and Trapero, 2012; Moral et al., 2014). In Southern Italy, the pathogen is able to infect directly leaves and shoots beside, these latter tissues can be also colonized by the pathogen through the peduncles of rotten fruit (Martelli, 1960). Furthermore, acervuli of the pathogen have been described on olive leaves in other countries, such as Australia and Italy (Martelli, 1961; Sergeeva et al., 2008). These differences could be due to variation among pathogen populations, weather conditions, or cultivar resistance.

It is worthy of note that the evaluation of the severity of branch-dieback caused by *Colletotrichum* sp. may mislead the inexperienced evaluator since other pathogens can cause similar symptoms. In the event of doubt, the evaluator should conduct isolation for other candidate pathogens. For example, we diagnosed Verticillium wilt in more than 20 olive trees belonging to different cultivars in the WOGBC (Morello et al., 2016).

Current trends in planting high-density orchards (which are very conducive for olive anthracnose) and reducing the use of copper-based fungicides are contrary to the high-quality oils demanded by consumers (Moral et al., 2012, 2014; Díez et al., 2016). Thereafter, the selection of less susceptible cultivars to anthracnose is essential for new plantations. The information about the resistance to olive anthracnose that is presented in this study is fundamental for farmers, technicians, and breeders.

## CONCLUSIONS

In this paper, we evaluated the resistance of 308 olive cultivars to *C. acutatum* s. lat. during three epidemic seasons under field conditions. However, there is a clear predominance (69.7%) of

susceptible cultivars (MS, S, and HS), we have also identified 32 cultivars (10.4%) highly resistant to the pathogen. The most notable cultivar was “Dolce Agogia,” which did not show any anthracnose symptom during the three seasons. This work constitutes the largest evaluation of olive cultivars according to their resistance to *C. acutatum* to date, albeit the response of other many cultivars to the pathogen is not well-known and, thus, it should be evaluated. For future work, the methodology described in this paper should be also used to evaluate the cultivar response to other aerial fungal pathogen affecting leaves and fruit, such as *V. oleaginea* and *Pseudocercospora cladosporioides*, causal agents of peacock spot and cercosporiosis, respectively.

## AUTHOR CONTRIBUTIONS

Conceived and designed the study: AT and JM. Performed the evaluation under field conditions: JM, JRV, CX, and LFR. Curator of the WOGBC: JC. Performed the evaluations in controlled conditions: JM. Analyzed the data: JM and AT. Wrote the paper: JM and AT.

## ACKNOWLEDGMENTS

This research was funded by the Spanish Ministry of Education and Science (project AGL2004-7495 co-financed by the European Union FEDER Funds) and by the Andalusia Regional Government (Project P08-AGR-03635). JM holds a Marie Skłodowska Curie fellowship launched by the European Union's H2020 (contract number 658579). CX was the holder of a fellowship from the Agencia Española de Cooperación Internacional (AECI). Thanks are due to the IFAPA for the use of experimental orchards. We are especially grateful to C. del Río and A. Belaj, who have been in charge of the WOGBC, and to Francisca Luque for helping under controlled conditions. They also thank T. J. Michailides, W. J. Kaiser, and A. Gordon for critical review of the manuscript.

## REFERENCES

- Ballio, A., Bottalico, A., Bounocore, V., Carilli, A., Di Vittorio, V., and Graniti, A. (1969). Production and isolation of aspergillomarasmine B (lycomarasmine acid) from cultures of *Colletotrichum gloeosporioides* (*Gloeosporium olivarum*). *Phytopathol. Mediterr.* 8, 187–196.
- Barranco, D., Cimato, A., Fiorino, P., Rallo, L., Touzani, A., Castañeda, C., et al. (2000). *World Catalogue of Olive Varieties*. Madrid: International Olive Oil Council and Mundi-Prensa.
- Bartolini, G., and Cerreti, S. (2017). *Olive germplasm (Olea europaea L.)* Available online at: <http://www.oleadb.it/olivodb.html>.
- Bartolini, G., Prevost, G., Messeri, C., Carignani, G., and Menini, U. G. (1998). *Olive Germplasm. Cultivars and World-Wide Collections*. Rome: FAO.
- Beauchamp, G. K., Keast, R. S. J., Morel, D., Lin, J., Pika, J., Han, Q., et al. (2005). Phytochemistry: ibuprofen-like activity in extra-virgin olive oil. *Nature* 437, 45–46. doi: 10.1038/437045a
- Besnard, G., Khadari, B., Navascués, M., Fernández-Mazuecos, M., El Bakkali, A., Arrigo, N., et al. (2013). The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. *Proc. R. Soc. B.* 280:20122833. doi: 10.1098/rspb.2012.2833
- Boudet, A. M. (2007). Evolution and current status of research in phenolic compounds. *Phytochemistry* 68, 2722–2735. doi: 10.1016/j.phytochem.2007.06.012
- Caballero, J. M., del Río, C., Barranco, D., and Trujillo, I. (2006). The olive world germplasm bank of Córdoba, Spain. *Olea* 25, 14–19. doi: 10.17660/ActaHortic.2008.791.1
- Cacciola, S. O., Faedda, R., Sinatra, F., Agosteo, G. E., Schena, S., Frisullo, S., et al. (2012). Olive anthracnose. *J. Plant Pathol.* 94, 29–44.
- Chattaoui, M., Raya, M. C., Bourri, M., Moral, J., Perez-Rodriguez, M., Trapero, A., et al. (2016). Characterization of a *Colletotrichum* population causing anthracnose disease on Olive in northern Tunisia. *J. Appl. Microbiol.* 120, 1368–1381. doi: 10.1111/jam.13096
- Damm, U., Cannon, P. F., Woudenberg, J. H. C., and Crous, P. W. (2012). The *Colletotrichum acutatum* species complex. *Stud. Mycol.* 73, 37–113. doi: 10.3114/sim0010
- Díez, C. M., Moral, J., Cabello, D., Morello, P., Rallo, L., and Barranco, D. (2016). Cultivar and tree density as key factors in the long-term performance of super high-density olive orchards. *Front. Plant Sci.* 7:1226. doi: 10.3389/fpls.2016.01226
- Díez, C. M., Trujillo, I., MartínezUrdiroz, N., Barranco, D., Rallo, L., Marfil, P., et al. (2015). Olive domestication and diversification in the Mediterranean Basin. *New Phytol.* 206, 436–447. doi: 10.1111/nph.13181
- Espósito, K., Marfella, R., Ciotola, M., Di Palo, C., Giugliano, F., Giugliano, G., et al. (2004). Effect of a Mediterranean-style diet on endothelial dysfunction and markers of vascular inflammation in the metabolic syndrome: a randomized trial. *JAMA* 292, 1440–1446. doi: 10.1001/jama.292.12.1440



- FAO (2015). *The FAO Statistical Database (FAOSTAT): Food and Agriculture Organization of the United Nations*. Available online at: <http://faostat.fao.org>.
- Faieda, R., Agosteo, G. E., Schena, L., Mosca, S., Frisullo, S., Magnano di San Lio, G., et al. (2011). *Colletotrichum clavatum* sp. nov. identified as the causal agent of olive Anthracnose in Italy. *Phytopathol. Medit.* 50, 283–302. doi: 10.14601/Phytopathol\_Mediterr-9547
- Fritz, C. O., Morris, P. E., and Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *J. Exp. Psychol.* 141, 2–18. doi: 10.1037/a0024338
- Ganino, T., Bartolini, G., and Fabbri, A. (2006). The classification of olive germplasm. *J. Hort. Sci. Biotechnol.* 81, 319–334. doi: 10.1080/14620316.2006.11512069
- García-López, J., Lorite, I. J., García-Ruiz, R., Ordoñez, R., and Dominguez, J. (2016). Yield response of sunflower to irrigation and fertilization under semi-arid conditions. *Agric. Water Manag.* 176, 151–162. doi: 10.1016/j.agwat.2016.05.020
- Geffroy, V., Sévignac, M., De Oliveira, J. C., Fouilloux, G., Skroch, P., Thoquet, P., et al. (2000). Inheritance of partial resistance against *Colletotrichum lindemuthianum* in *Phaseolus vulgaris* and co-localization of quantitative trait loci with genes involved in specific resistance. *Mol. Plant Microbe Interact.* 13, 287–296. doi: 10.1094/MPMI.2000.13.3.287
- Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *CSUR* 31, 264–323. doi: 10.1145/331499.331504
- León, L., Beltrán, G., Aguilera, M. P., Rallo, L., Barranco, D., and De la Rosa, R. (2011). Oil composition of advanced selections from an olive breeding program. *Eur. J. Lipid Sci. Technol.* 113, 870–875. doi: 10.1002/ejlt.201000535
- León, L., De la Rosa, R., Gracia, A., Barranco, D., and Rallo, L. (2008). Fatty acid composition of advanced olive selections obtained by crossbreeding. *J. Sci. Food Agric.* 88, 1921–1926. doi: 10.1002/jsfa.3296
- López-Doncel, L. M. (2003). *Evaluación de la Resistencia del olivo a Spilocaea oleagina, Agente del Repilo*. Ph.D. Thesis, Universidad de Córdoba, Córdoba.
- Malacrino, A., Schena, L., Campolo, O., Laudani, F., and Palmeri, V. (2015). Molecular analysis of the fungal microbiome associated with the olive fruit fly *Bactrocera oleae*. *Fungal Ecol.* 18, 67–74. doi: 10.1016/j.funeco.2015.08.006
- Martelli, G. P. (1960). Primo contributo alla conoscenza della biologia di *Gloeosporium olivarum* Alm. *Phytopathol. Medit.* 1, 31–43.
- Martelli, G. P. (1961). Acervuli di *Gloeosporium olivarum* Alm. su foglie di olivo. *Phytopathol. Medit.* 1, 43–51.
- Martín, M. P., and García-Figueres, F. (1999). *Colletotrichum acutatum* and *C. gloeosporioides* cause anthracnose on olives. *Eur. J. Plant Pathol.* 105, 733–741. doi: 10.1023/A:1008785703330
- McKinney, H. H. (1923). Influence of soil temperature and moisture on infection of wheat seedlings by *Helminthosporium sativum*. *J. Agricul. Res.* 26, 195–218.
- Moral, J. (2009). *Etiología, Epidemiología, y Resistencia Varietal en la Antracnosis del olivo Causada por Colletotrichum spp.* Ph.D. Thesis, Universidad de Córdoba, Córdoba.
- Moral, J., Ávila, A., López-Doncel, L. M., Alsalmiya, M., Oliveira, R., Gutiérrez, F., et al. (2005). Resistencia a los Repilos de distintas variedades de olivo. *Vida Rural* 208, 34–40.
- Moral, J., Alsalmiya, M., Roca, L. F., Díez, C., León, L., De la Rosa, R., et al. (2015). Relative susceptibility of new olive cultivars to *Spilocaea oleagina*, *Colletotrichum acutatum*, and *Pseudocercospora cladosporioides*. *Plant Dis.* 99, 58–64. doi: 10.1094/PDIS-04-14-0355-RE
- Moral, J., Bouhmid, K., and Trapero, A. (2008). Influence of fruit maturity, cultivar susceptibility, and inoculation method on infection of olive fruit by *Colletotrichum acutatum*. *Plant Dis.* 92, 1421–1426. doi: 10.1094/PDIS-92-10-1421
- Moral, J., Jurado-Bello, J., Sánchez, M. I., De Oliveira, R., and Trapero, A. (2012). Effect of temperature, wetness duration, and planting density on olive anthracnose caused by *Colletotrichum* spp. *Phytopathology* 102, 974–981. doi: 10.1094/PHYTO-12-11-0343
- Moral, J., Oliveira, R., and Trapero, A. (2009). Elucidation of the disease cycle of olive anthracnose caused by *Colletotrichum acutatum*. *Phytopathology* 99, 548–556. doi: 10.1094/PHYTO-99-5-0548
- Moral, J., and Trapero, A. (2009). Assessing the susceptibility of olive cultivars to anthracnose caused by *Colletotrichum acutatum*. *Plant Dis.* 93, 1028–1036. doi: 10.1094/PDIS-93-10-1028
- Moral, J., and Trapero, A. (2012). Mummified fruit as a source of inoculum and disease dynamics of olive anthracnose caused by *Colletotrichum* spp. *Phytopathology* 102, 982–989. doi: 10.1094/PHYTO-12-11-0344
- Moral, J., Xaviér, C., Roca, L. F., Romero, J., Moreda, W., and Trapero, A. (2014). La Antracnosis del olivo y su efecto en la calidad del aceite. *Grasas y Aceites* 65, e028. doi: 10.3989/gya.110913
- Morello, P., Díez, C. M., Codes, M., Rallo, L., Barranco, D., Trapero, A., et al. (2016). Sanitation of olive plants infected by *Verticillium dahliae* using heat treatments. *Plant Pathol.* 65, 412–421. doi: 10.1111/ppa.12432
- Pataky, J. K., Williams, M. M. II., Headrick, J. M., Nankam, C., du Toit, L. J., and Michener, P. M. (2011). Observations from a quarter century of evaluating reactions of sweet corn hybrids in disease nurseries. *Plant Dis.* 95, 1402–1506. doi: 10.1094/PDIS-03-11-0236
- Poland, J. A., Balint-Kurti, P. J., Wisser, R. J., Pratt, R. C., and Nelson, R. J. (2009). Shades of gray: the world of quantitative disease resistance. *Trends Plant Sci.* 14, 21–29. doi: 10.1016/j.tplants.2008.10.006
- Preto, G., Martins, F., Pereira, J. A., and Baptista, P. (2017). Fungal community in olive fruits of cultivars with different susceptibilities to anthracnose and selection of isolates to be used as biocontrol agents. *Biol. Control* 110, 1–9. doi: 10.1016/j.biocontrol.2017.03.011
- Prusky, D. (1996). Pathogen quiescence in postharvest diseases. *Annu. Rev. Phytopathol.* 34, 413–434. doi: 10.1146/annurev.phyto.34.1.413
- Rallo, L., Barranco, D., Caballero, J. M., del Río, C., Martín, A., Tous, J., et al. (2005). *Variedades de Olivo en España*. Madrid: Junta de Andalucía, M.A.P.A. and Mundi-Prensa.
- Roca, L. F., Moral, J., Viruega, J. R., Ávila, A., Oliveira, R., and Trapero, A. (2007). Copper fungicides in the control of olive diseases. *Olea* 26, 48–50.
- Schena, L., Abdelfattah, A., Mosca, S., Li Destri Nicosia, M. G., Agosteo, G. E., and Cacciola, S. O. (2017). Quantitative detection of *Colletotrichum godetiae* and *C. acutatum sensu stricto* in the phyllosphere and carposphere of olive during four phenological phases. *Eur. J. Plant Pathol.* 149, 337–347. doi: 10.1007/s10658-017-1185-x
- Schena, L., Mosca, S., Cacciola, S. O., Faieda, R., Sanzani, S. M., Agosteo, G. E., et al. (2014). Species of the *Colletotrichum gloeosporioides* and *C. boninense* complexes associated with olive anthracnose. *Plant Pathol.* 63, 437–446. doi: 10.1111/ppa.12110
- Sergeeva, V., Spooner-Hart, R., and Nair, N. G. (2008). First report of *Colletotrichum acutatum* and *C. gloeosporioides* causing leaf spots of olives (*Olea europaea*) in Australia. *Aust. Plant Dis. Not.* 3, 143–144. doi: 10.1071/DN08055
- Talhinhas, P., Gonçalves, E., Sreenivasaprasad, S., and Oliveira, H. (2015). Virulence diversity of anthracnose pathogens (*Colletotrichum acutatum* and *C. gloeosporioides* species complexes) on eight olive cultivars commonly grown in Portugal. *Eur. J. Plant Pathol.* 142, 73–83. doi: 10.1007/s10658-014-0590-7
- Talhinhas, P., Sreenivasaprasad, S., Neves-Martins, J., and Oliveira, H. (2005). Molecular and phenotypic analyses reveal association of diverse *Colletotrichum acutatum* groups and a low level of *C. gloeosporioides* with olive anthracnose. *Appl. Environ. Microbiol.* 71, 2987–2998. doi: 10.1128/AEM.71.6.2987-2998.2005
- Trapero, A., and López-Doncel, L. (2005). “Resistencia y susceptibilidad al Repilo,” in *Variedades de Olivo en España*, eds L. Rallo, D. Barranco, J. M. Caballero, C. del Río, A. Martín, J. Tous, and I. Trujillo (Madrid: Junta de Andalucía and Mundi-Prensa), 323–328.
- Trujillo, I., Ojeda, M. A., Urdirroz, N. M., Potter, D., Barranco, D., Rallo, L., et al. (2013). Identification of the Worldwide Olive Germplasm Bank of Córdoba (Spain) using SSR and morphological markers. *Tree Genet. Genomes* 10, 141–155. doi: 10.1007/s11295-013-0671-3
- Xaviér, C. (2009). *Resistencia y Grupos de Virulencia en la Antracnosis del Olivo Causada por Colletotrichum spp.* Master Thesis, Universidad de Córdoba, Córdoba.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Moral, Xaviér, Viruega, Roca, Caballero and Trapero. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Olive Cultivation in the Southern Hemisphere: Flowering, Water Requirements and Oil Quality Responses to New Crop Environments

**Mariela Torres<sup>1</sup>, Pierluigi Pierantozzi<sup>1</sup>, Peter Searles<sup>2</sup>, M. Cecilia Rousseaux<sup>2</sup>, Georgina García-Inza<sup>2</sup>, Andrea Miserere<sup>2</sup>, Romina Bodoira<sup>3</sup>, Cibeles Contreras<sup>1</sup> and Damián Maestri<sup>3\*</sup>**

## OPEN ACCESS

### Edited by:

Marcello Mastrorilli,  
Consiglio per la Ricerca in Agricoltura  
e l'Analisi dell'Economia Agraria  
(CREA), Italy

### Reviewed by:

Sergio Tombesi,  
Università Cattolica del Sacro Cuore,  
Italy

George A Manganaris,

Cyprus University of Technology,  
Cyprus

### \*Correspondence:

Damián Maestri  
dmaestri@unc.edu.ar

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 12 July 2017

**Accepted:** 10 October 2017

**Published:** 27 October 2017

### Citation:

Torres M, Pierantozzi P, Searles P,  
Rousseaux MC, García-Inza G,  
Miserere A, Bodoira R, Contreras C  
and Maestri D (2017) Olive Cultivation  
in the Southern Hemisphere:  
Flowering, Water Requirements and  
Oil Quality Responses to New Crop  
Environments.  
Front. Plant Sci. 8:1830.  
doi: 10.3389/fpls.2017.01830

<sup>1</sup> Estación Experimental Agropecuaria San Juan, Instituto Nacional de Tecnología Agropecuaria (Inta), CONICET, San Juan, Argentina, <sup>2</sup> Centro Regional de Investigaciones Científicas y Transferencia Tecnológica de La Rioja (CRILAR, Provincia de La Rioja, UNLaR, SEGEMAR, UNCa, CONICET), La Rioja, Argentina, <sup>3</sup> Instituto Multidisciplinario de Biología Vegetal, Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Nacional de Córdoba, Córdoba, Argentina

Olive (*Olea europaea* L.) is a crop well adapted to the environmental conditions prevailing in the Mediterranean Basin. Nevertheless, the increasing international demand for olive oil and table olives in the last two decades has led to expansion of olive cultivation in some countries of the southern hemisphere, notably in Argentina, Chile, Perú and Australia. While the percentage of world production represented by these countries is still low, many of the new production regions do not have typical Mediterranean climates, and some are located at subtropical latitudes where there is relatively little information about crop function. Thus, the primary objective of this review was to assess recently published scientific literature on olive cultivation in these new crop environments. The review focuses on three main aspects: (a) chilling requirements for flowering, (b) water requirements and irrigation management, and (c) environmental effects on fruit oil concentration and quality. In many arid and semiarid regions of South America, temperatures are high and rainfall is low in the winter and early spring months compared to conditions in much of the Mediterranean Basin. High temperatures have often been found to have detrimental effects on olive flowering in many olive cultivars that have been introduced to South America, and a better understanding of chilling requirements is needed. Lack of rainfall in the winter and spring also has resulted in an urgent need to evaluate water requirements from the flower differentiation period in the winter to early fruit bearing. Additionally, in some olive growing areas of South America and Australia, high early season temperatures affect the timing of phenological events such that the onset of oil synthesis occurs sooner than in the Mediterranean Basin with most oil accumulation taking place in the summer when temperatures are very high. Increasing mean daily temperatures have been demonstrated to decrease fruit oil concentration (%) and negatively affect some aspects

of oil quality based on both correlative field studies and manipulative experiments. From a practical standpoint, current findings could be used as approximate tools to determine whether the temperature conditions in a proposed new growing region are appropriate for achieving sustainable oil productivity and quality.

**Keywords:** chilling requirements, fatty acids, irrigation, oil concentration, oil yield, water requirements, *Olea europaea* L.

## INTRODUCTION

The geographic origin of cultivated olive (*Olea europaea* L.) can be traced to areas along the eastern Mediterranean coast where Turkey, Syria, Lebanon, Palestine, and Israel are currently located. Some records indicate that olive trees have been cultivated in those areas since at least 3000 BC (Connor, 2005). Olive then spread widely around southern Europe, northern Africa, and the Iberian Peninsula. Today, approximately 98% of olives are cultivated in Mediterranean Basin countries. Spain, Italy, and Greece together produce about 77% of the world's olive oil. Portugal, Tunisia, Turkey, Morocco, Syria, and Egypt also have an important amount of production, but oil yields are low per hectare in many instances and modern processing technology is underutilized (El-Kholy, 2012).

In the last 20–30 years, interest in olive oil production and consumption has expanded olive cultivation to regions and countries outside the Mediterranean Basin such as Australia, China, India, and South America. In the southern hemisphere, the biggest hectareage for olive cultivation is located in Argentina. Until 1990, olive cultivation covered a total area of approximately 30,000 ha, most of which corresponded to small orchards (<10 ha) with traditional management (i.e., low planting density and flood irrigation). Subsequent tax exemption laws brought in large investments that included large commercial orchards (>100 ha) with higher planting densities and drip irrigation. Currently, there are about 110,000 ha under cultivation mainly in the central-western and north-western regions bordering the Andes mountain range (27–33°S latitude). Spanish and Italian cultivars including “Arbequina,” “Manzanilla,” “Picual,” and “Frantoio” have been extensively planted with about 70% of production being devoted to olive oil. “Arauco” is the only cultivar recognized from Argentina in the World Catalog of Olive Varieties (IOOC, 2000).

In South America, Chile is ranked second in area planted with about 24,000 ha of olive orchards. Production is almost exclusively dedicated to olive oil, and similarly to Argentina, the most important cultivars are Spanish and Italian cultivars with “Arbequina” comprising 50% of the total production area. Other cultivars include “Frantoio,” “Arbosana,” “Picual,” and “Leccino.” “Azapa” is a local table olive cultivar with a close resemblance to “Arauco.” Perú (approximately 28,000 ha of olive orchards) is not a large producer of olive oil, but table olive production has increased considerably over the last decade. Being located near the equator, the climate conditions in Perú are very different from those found in traditional olive growing regions (Ayerza and Sibbett, 2001). Uruguay and Brazil both have minor, but increasing olive production areas (10,000 and 1,300 ha,

respectively), as a result of ongoing expansion projects that are mainly for oil production.

Currently, Australia has about 11 million olive trees spread across approximately 35,000 ha. Although early orchards included a large number of cultivars, about 90% of Australian olive oil is produced from common European cultivars (“Arbequina,” “Frantoio,” “Coratina,” “Corregiola,” “Manzanilla,” “Picual,” and “Koroneiki”) and more recently from the Israeli cv. Barnea. Australian areas under olive cultivation include a wide natural diversity of environments from the most southern point of Western Australia to the northern tropical areas of Queensland. Olive production has expanded rapidly in recent years due to the adaptation of intensive and super-high density planting systems in new commercial orchards. As a result of the low rainfall and the unpredictable nature of Australian olive crop environments, almost all Australian olive orchards are irrigated (Mailer, 2012). This is also the case in Argentina because annual rainfall is most often between 100 and 400 mm (Searles et al., 2011).

At this point, it is important to bear in mind that many of the olive growing areas in the southern hemisphere have temperature and precipitation regimes that are very different from those of the Mediterranean Basin where olive trees are traditionally cultivated (Table 1). This reality has encouraged, or even forced, both growers and academics to seek new approaches to crop management. Although scientific studies conducted outside the Mediterranean Basin are still limited, it is important to review and synthesize the knowledge currently available on several critical topics. Lavee (2014) provided some general guidelines on olive adaptation to new environments based mostly on knowledge from the Mediterranean Basin, and concluded that there is a strong need for local research in new production areas. For these reasons, the primary objective of this review was to analyze recently published scientific literature on olive cultivation in non-Mediterranean environments in the southern hemisphere. The review focuses on three main aspects: (a) chilling requirements for flowering, (b) water requirements and irrigation management, and (c) environmental effects on fruit oil concentration and quality. The revision also contributes to identifying areas where knowledge is insufficient and to set priorities for further research.

## Chilling Requirements for Flowering

Olive is a crop that flowers profusely and produces high olive oil yields under the prevailing climatic and agro-ecological conditions of the Mediterranean Basin with most production being confined traditionally to latitudes between 30° and 45° North. Yet, olive trees have the ability to adjust to a wide range

**TABLE 1 |** Temperature, rainfall, and evapotranspiration (ETo) values from different olive growing areas in South America and Australia compared with those of typical Mediterranean regions in Spain, Italy, and Tunisia.

Location	Parameter	Spring	Summer	Autumn	Winter	Annual
Argentina (North-western region)	Tmax	29.1	33.5	26.3	20.1	27.2
Latitude 27°-29°S	Tmin	13.8	19.5	13	4.2	12.6
Altitude (masl) 420–1,200	Rainfall (mm)	48.5	166	73.0	10.0	298
	ETo (mm)	483	555	339	243	1,620
Argentina (Central región)	Tmax	27.5	33.8	25.3	20.9	26.1
Latitude 30.5°S	Tmin	10.6	18.1	10.9	6.35	10.3
Altitude (masl) 450	Rainfall (mm)	92.4	330	120	10.0	556
	ETo (mm)	405	556	230	178	1,369
Argentina (Central-western region)	Tmax	28.5	32.0	21.2	18.2	25.3
Latitude 31°-33°S	Tmin	13.8	18.5	7.5	2.8	11.0
Altitude (masl) 590–920	Rainfall (mm)	38.5	73.9	18.6	9.8	141
	ETo (mm)	484.	538.	216.	196	1,435
Peru (South-western Coast, e.g., Tacna and Ilo valleys)	Tmax	23.4	28.1	22.7	17.8	23.0
Latitude 17°-18°S	Tmin	13.5	18.1	13.7	10.0	13.8
Altitude (masl) 895	Rainfall (mm)	3	7.3	1.4	14.5	26.2
	ETo (mm)	434	491	278	226	1,429
Brazil (South-eastern region, Minas Gerais state)	Tmax	27.7	27.6	23.2	23.8	25.6
Latitude 22°-23°S	Tmin	14.8	14.9	10.6	9.2	12.4
Altitude (masl) 1,200–1,300	Rainfall (mm)	376	492	140	247	1,255
	ETo (mm)	343	373	233	212	1,159
Uruguay (South-eastern region, e.g., Rocha)	Tmax	23.3	27.9	20.8	17.1	22.3
Latitude 33°-33.6 S	Tmin	11.8	15.7	9.6	7.0	11.0
Altitude (masl) 4–63	Rainfall (mm)	279	316	271	342	1,208
	ETo (mm)	NA	NA	NA	NA	NA
Australia (South-eastern region, e.g., Victoria, New South Wales)	Tmax	25.7	30.5	20.6	16.1	23.3
Latitude 31°-36°S	Tmin	11.0	15.5	7.7	4.1	9.6
Altitude (masl) 150–420	Rainfall (mm)	179	170	163	183	695
	ETo (mm)	426	501	219	192	1,338
Australia (South-western region, e.g., Perth)	Tmax	23.4	29.2	22.3	17.8	23.2
Latitude 30°-33°S	Tmin	11.3	15.1	10.4	7.7	11.1
Altitude (masl) 5	Rainfall (mm)	93.6	33.0	238	358	723
	ETo (mm)	392	467	228	189	1,276
Spain (Southern region, e.g., Sevilla)	Tmax	23.2	34.0	26.0	17.1	25.1
Latitude 37.2°-37.6°N	Tmin	10.6	18.3	13.5	6.60	12.2
Altitude (masl) 8–358	Rainfall (mm)	134	20	167	233	554
	ETo (mm)	372	600	288	147	1,408
Spain (Central region, e.g., Toledo)	Tmax	19.7	31.9	21.7	12.1	21.3
Latitude 39°-39.5° N	Tmin	7.5	17.3	10.0	2.5	9.3
Altitude (masl) 510	Rainfall (mm)	110	49	100	100	359
	ETo (mm)	324	556	238	107	1,225
Italy (Southern region, e.g., Benevento)	Tmax	19.7	22.5	14.4	8.0	16.2
Latitude 41.1° N	Tmin	10.8	18.2	6.1	5.0	10.0
Altitude (masl) 250	Rainfall (mm)	153	123	263	189	728
	ETo (mm)	432	546	125	129	1,232
Tunisia (Eastern coast, e.g., Sousse)	Tmax	19.7	31.9	21.7	12.1	21.3
Latitude 39°-39.5° N	Tmin	7.5	17.3	10.0	2.5	9.3
Altitude (masl) 516	Rainfall (mm)	110	49	100	100	359
	ETo (mm)	324	556	238	107	1,225

*Tmax* and *Tmin* are the average seasonal maximum and minimum temperatures (°C), respectively. NA, not available. ETo values were calculated using the FAO Penman-Monteith method (Allen et al., 1998).

of different environments due to a number of specific biological and anatomical characteristics (Gucci and Caruso, 2011). This adaptation often leads to significant effects on several aspects of reproductive performance such as flowering, oil yield, and oil quality that vary depending on the environmental conditions (e.g., Tura et al., 2007; Lazzez et al., 2008; Temine et al., 2008; Torres et al., 2009; Di Vaio et al., 2012; Rondanini et al., 2014). Rapoport (2014) has reviewed many of the reproductive biology responses to drought and temperature in olive under extreme conditions. Thus, our aim in this section is to focus on the specific issue of chilling hours for flowering, which is of great importance in many production areas in the southern hemisphere.

Flowering is one of the major yield determinants in olive, and although olive trees are capable of producing a large number of flowers, the percentage of flowers that set fruit is usually very low with values of about 2% (Lavee et al., 1996). Flowering occurs once buds induced the previous growing season receive sufficient chilling during the winter dormancy period to end dormancy, differentiate anatomically, and accumulate warmer temperatures adequate for budburst (Rallo and Cuevas, 2008). The accumulation of chilling requirements for flowering during winter dormancy is most often referred to as vernalization, and high temperatures during the winter may adversely affect the number of chilling hours accrued (Malik and Perez, 2011). Thus, flowering and therefore fruiting may be reduced due to insufficient chilling temperatures at low latitudes ( $<30^{\circ}$ ).

**Figure 1** shows the chronological sequence of the main phenological stages of olive cultivation in the Mediterranean Basin compared with the main growing regions in Argentina (i.e., central-western and north-western Argentina). In NW Argentina, it has been observed that fairly high winter and spring temperatures lead to earlier flowering, and eventually to earlier oil accumulation, relative to the Mediterranean Basin (Gómez del Campo et al., 2010). Early flowering was also reported in other low latitude South American production areas such as Perú (Lavee, 2014). High temperatures have been shown to result in a lack of chilling hours for flowering in some cultivars growing in NW Argentina (Aybar et al., 2015) and in Tacna, Perú (Castillo-Llanque et al., 2014). In addition, it has been observed that trees exposed to insufficient chilling temperatures and high temperature events can flower, but the flowers are of low quality and have a low set percentage. This phenomenon has been documented in olive growing areas at low latitudes where some olive varieties produce deformed floral buds and fruit (**Figure 2**). This is in accordance with previous suggestions that winter chilling is necessary not only for floral differentiation, but also for proper formation of floral buds (Rallo and Martin, 1991; De Melo-Abreu et al., 2004).

The optimum temperature regime for reproductive development of olive buds has been considered to include fluctuating temperatures from 2 to  $19^{\circ}\text{C}$  (Denney and McEachern, 1983). However, in greenhouse experiments under controlled temperature conditions, Malik and Bradford (2009) observed maximal flowering in “Arbequina” at minimum temperatures ranging from  $4.4$  to  $7.8^{\circ}\text{C}$  and flowering reductions when night temperature was maintained at  $2.2^{\circ}\text{C}$ . Hence, these authors suggest an inhibitory effect on flowering even at

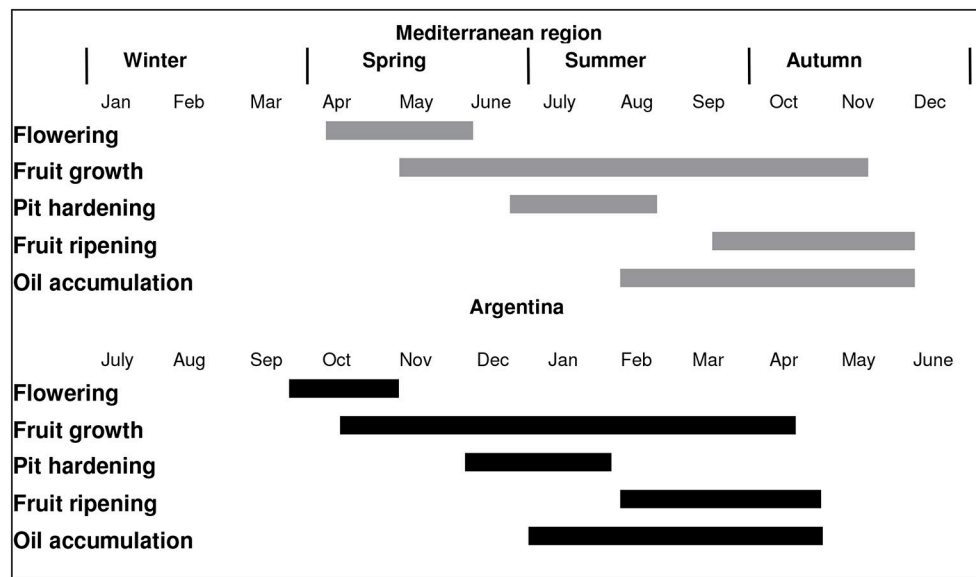
temperatures that were previously considered satisfactory for reaching chilling requirements.

At a large geographical scale, the potential of new sites for olive cultivation in the Arid Chaco Region in northern Argentina was assessed through temperature regime comparisons with more established sites in central Argentina, Italy, Spain, and the USA (Ayerza and Sibbett, 2001). At these sites, the frequencies of minimum ( $0.0$ – $12.5^{\circ}\text{C}$ ) and maximum ( $12.5$ – $21.1^{\circ}\text{C}$ ) temperatures during the winter were determined along with those of extreme cold ( $<0^{\circ}\text{C}$ ) and heat ( $>37^{\circ}\text{C}$ ) during the flowering period. Following these criteria, all Italian and Spanish olive growing sites had at least 150 days per year that were adequate for chilling hour accumulation, while some established Argentinean sites (San Juan,  $31^{\circ} 34' \text{ S}$ ; Mendoza,  $32^{\circ} 50' \text{ S}$ ) did not exceed 110 days, and all proposed new Arid Chaco sites had less than 60 days. Another distinctive feature of the climate in the Arid Chaco region during the winter season when chilling accumulation occurs is alternating periods (i.e., several days) of high and low temperatures (Ayerza and Sibbett, 2001). This reality may add an additional drawback for flower development and fruit set because temperatures above  $21^{\circ}\text{C}$  partially reverse chilling accumulation when they occur before chilling requirements are completed (De Melo-Abreu et al., 2004; Malik and Perez, 2011).

Lavee (2014) observed that the cold Humboldt Current along the Pacific coast of Perú lowers the expected temperatures for this latitude ( $<20^{\circ}\text{S}$ ). Interestingly, both an introduced Italian cultivar (“Frantoio”) and a local Peruvian cultivar (“Criolla”) clearly flower, in spite of the latitude. Castillo-Llanque et al. (2014) reported that this latter cultivar shows greater and more rapid floral bud development suggesting that it may be better adapted than the Italian cultivar to the local climatic conditions. Reductions in light intensity during the winter months due to continuous cloudiness in coastal Perú have also been suggested anecdotally to compensate for insufficient chilling (Lavee, 2014).

Even from casual observations, it seems clear that olive cultivars differ in their chilling requirement. An analytical approach for assessing the potential for flowering occurrence and date in different cultivars involves simulation models based on cultivar-specific thermal requirements (De Melo-Abreu et al., 2004). One of the models proposed by De Melo-Abreu et al. (2004) predicts the date when the chilling requirement is reached as well as the full flowering date after the accumulation of warm temperatures. Recently, this model was validated over several years at eight low latitude sites in north-western Argentina (Aybar et al., 2015). In “Arbequina,” normal flowering was observed at almost all sites and in all years, while normal flowering events in “Frantoio” and “Leccino” were uncommon. The results confirmed that these two latter cultivars require a very high number of chilling units in accordance with values from the World Olive Germplasm Bank of Córdoba (Spain), and winter temperatures in NW Argentina do not meet their chilling requirements for normal flowering in most years. In several tropical areas, expansion of fruit trees has been made possible by the use of growth regulators, which replace chilling requirements to overcome endodormancy. For example, hydrogen cyanamide (HC) has been used successfully in apple trees (Mohamed,





**FIGURE 1 |** The chronological sequence of the main phenological stages of olive cultivation in the Mediterranean Basin compared with the main olive growing regions in Argentina. The Mediterranean and Argentine sequences are adapted from Sanz-Cortés et al. (2002) and Gómez del Campo et al. (2010), respectively.

2008), grapes (Or et al., 2000), and peaches (Marodin et al., 2002). Benzyladenine (BA), a synthetic cytokinin that acts during dormancy release in apple (Bubán, 2000), is another growth regulator often used. However, BA did not affect flower initiation in the olive cv. Sevillano (Badr and Hartmann, 1972), and a more recent field trial in NW Argentina did not successfully lead to normal flowering in “Frantoio” when BA or HC were applied (Aybar, 2010).

From this section, we conclude that insufficient chilling hours during winter dormancy in many areas of South America and potentially other parts of the southern hemisphere often lead to reductions in flowering in some cultivars. Thus, cultivar-specific simulation models are recommended as approximate tools to predict whether individual cultivars will likely flower in proposed new growing regions.

## Water Requirements and Irrigation Management

Olive has been cultivated traditionally under rainfed conditions in the Mediterranean Basin without supplemental irrigation (Connor and Fereres, 2005). Good olive oil yields can be obtained there without irrigation in growing areas where annual rainfall is greater than 600 mm (Gucci and Fereres, 2012). Although such a value may serve as a benchmark for olive cultivation in many Mediterranean countries, additional factors such as rainfall seasonality, soil water-holding capacity, and crop evapotranspiration should be carefully considered for specific growing regions.

Since the early pioneering study of Hartmann and Panetsos (1961) in California, several dozen studies have addressed the irrigation needs of olive trees growing under Mediterranean climate conditions based on the sensibility of the main olive

phenological stages to water deficit (e.g., Tognetti et al., 2005; Lavee et al., 2007; Gucci et al., 2009; Rapoport et al., 2012; Gómez-del-Campo et al., 2014). These studies have mostly focused on evaluating irrigation needs under the hot dry summer conditions when rainfall is limited and provide the physiological and agronomic knowledge necessary for the application of irrigation strategies under many commercial growing conditions (Fernández, 2014). However, some recent advances in our understanding of olive tree response to irrigation have been possible due to the different climate regimes found in South America and elsewhere in the southern hemisphere.

In the Mediterranean Basin, irrigation is normally suspended during the winter months because rainfall is more than sufficient to satisfy crop evapotranspiration (ETc) under the fairly cold and cloudy conditions. The soil moisture stored during the winter also may preclude the need to irrigate in the spring during flowering and subsequent fruit set. By contrast, in many southern hemisphere climates where olive is cultivated (e.g., the subtropics of Australia and Argentina), rainfall events occur mostly in the summer with little or no winter rainfall. Somewhat greater temperatures during the winter and spring months at these latitudes compared to those of the Mediterranean also suggest that ETc should be higher. Thus, there are not irrigation experiences from Mediterranean countries that are applicable to these regions for this time of the year.

In the last several years, some studies evaluating irrigation needs during the winter have been conducted in the arid and semi-arid regions of central and North-western Argentina. A preliminary study by Rousseaux et al. (2008) examined some physiological and yield responses to the suspension of irrigation for 6–7 weeks during the winter season (July–August) in La Rioja province (NW Argentina). Soil water content decreased after



**FIGURE 2 |** Deformed floral buds and fruit from olive trees exposed to insufficient chilling temperatures and high temperature events. Photograph (A) is from the olive collection at INTA-San Juan (31°S, Argentina); photographs (B,C) are from semi-tropical regions of Brazil, 22°–23°S (courtesy of Dr. Shimon Lavee).

15 days in non-irrigated plots, and both pre-dawn and midday leaf water potential showed mild reductions during most of the experiment. However, this change in leaf water status was not enough to significantly decrease net photosynthetic rate, and fruit yield at harvest the following growing season only showed a modest percentage reduction (–20%). Estimates from this study suggested that little irrigation would be needed to satisfy crop water demand during the winter. Later sap flow measurements in whole trees combined with soil micro-lysimeter data during the winter also confirmed that ETc was only about 40% of reference transpiration (*sensu* Allen et al., 1998). Although maximum daily temperature was often around 20°C, transpiration per unit leaf area was minimal when daily average temperature was below 13°C (Rousseaux et al., 2009), a condition frequently observed during July–August due to low night temperatures.

In another study (Pierantozzi et al., 2013), when the deficit irrigation period was extended from early winter until mid-spring (i.e., mid-June through October), significant reductions in both photosynthetic pigment levels and net photosynthetic rates were observed in trees grown under moderate (50% ETc) and severe (25 and 0% ETc) deficit irrigation. Vegetative shoot growth started toward the end of the winter, and was significantly reduced by deficit irrigation (Pierantozzi et al., 2014). Flowering was also delayed in the trees receiving no irrigation. This delay, as well as reductions in flowering intensity, ultimately decreased oil yield in the treatments under moderate or severe water stress compared to the well-irrigated treatments (100 or 75% ETc).

The greater responses to deficit irrigation in Pierantozzi et al. (2013, 2014) than in Rousseaux et al. (2008) are most likely explained by the duration of the deficit period. When the influence of water availability was partitioned in four periods, water deficit during winter dormancy did not affect either flowering or fruiting parameters, but deficit during inflorescence development reduced many different flowering traits and ovule development (Rapoport et al., 2012). Thus, it appears that a short and mild water deficit during the colder, winter dormancy period (Rousseaux et al., 2008) may not greatly affect reproductive responses, but intensifying water deficit from late fall through mid-spring including winter dormancy, flower differentiation, and flower opening leads to detrimental effects (Pierantozzi et al., 2013, 2014). According to these two latter studies, little irrigation (50% ETc) may be sufficient to maintain adequate plant water

potentials for the coldest winter months, but high (75% ETc) or full (100% ETc) irrigation rates could be needed by mid-August, which is approximately 2 months before full flowering, to guarantee adequate oil yields in regions with dry winter and spring seasons.

Many South American olive cultivation regions in Argentina, Chile and Perú are located at subtropical latitudes (<30°S) where high temperatures will likely affect annual water requirements. In a field experiment in NW Argentina (28° 33' S, province of La Rioja, Argentina), the warm climate facilitated excessive shoot growth when very high irrigation levels were applied (Correa-Tedesco et al., 2010). Based on both vegetative growth and fruit yield, an optimal crop coefficient (Kc) value of approximately 0.70 was determined for the entire growing season in this drip-irrigated, intensively-planted orchard (“Manzanilla fina”). According to these authors, such a Kc value may be sufficient to maintain the vegetative growth at an appropriate level without adversely affecting fruit yield. Based on transpiration and soil evaporation values in the same orchard, Rousseaux et al. (2009) calculated Kc values of about 0.65–0.70 and 0.85–0.90 for either moderately or excessively irrigated olive plots, respectively. The proposed Kc value of 0.7 is similar to reported values for many areas with Mediterranean climates such as California (Goldhamer et al., 1994) and Spain (Girona et al., 2002). However, the reference ET values in NW Argentina (1,600 mm) are higher than in most of the Mediterranean Basin (1,000–1,400 mm). These high ET values require a greater amount of total annual water in order to satisfy crop demand. Additionally, due to the low annual rainfall (100–400 mm) in this region, a very large proportion of the total water applied must be in the form of irrigation, rather than relying on rainfall.

Irrigation studies from Australia also provide partial information about olive water requirements and irrigation management (Yunusa et al., 2008; Zeleke et al., 2012; Zeleke, 2014). Estimates of soil water use from neutron probe measurements and canopy transpiration from porometer readings were obtained in four olive orchards in southern Australia (34°S) by Yunusa et al. (2008). Similar to NW Argentina, these orchards experienced high annual potential ET conditions (1,600 mm) with little rainfall during much of the growing season. It was observed that crop evapotranspiration in these orchards was fairly low (600 mm; Kc = 0.4) primarily

because they were not irrigated during the spring when rainfall was scarce. Based on the results of Pierantozzi et al. (2013, 2014) from Argentina, yield in these orchards might be significantly increased by irrigating during the spring when flowering, fruit set, and most shoot growth occur. In an area with a more evenly distributed rainfall regime, a water balance approach indicated that annual ET<sub>c</sub> was about 700 mm, but a yearly value for reference evapotranspiration was not given (Zeleeke, 2014).

As basic information concerning water requirements has started to accrue in the southern hemisphere, more sophisticated approaches are now being examined. For example, it has been suggested from studies in New Zealand and Argentina that plant-based indicators such as fluctuations in trunk diameter and stem water potential have considerable potential for programming deficit irrigation (Greven et al., 2009; Trentacoste et al., 2015; Agüero Alcaras et al., 2016). In this regard, oil yield was only slightly reduced over three seasons when irrigation was applied below a stem water potential threshold of  $-2.5$  MPa in cv. Arbequina in central-western Argentina (Trentacoste et al., 2015). Thus, this technique appears to be promising for reducing irrigation as competition between sectors for water resources increases. In super-high density olive orchards in Chile, there are also significant advances in determining energy balance components using ground-based measurements and unmanned aerial vehicles that have direct applications for irrigation management (López-Olivari et al., 2016; Ortega-Farías et al., 2016).

To summarize this section, the differences in rainfall distribution between some southern hemisphere sites and the Mediterranean Basin have provided new knowledge about water requirements in olive trees. Unlike the Mediterranean, little rainfall during the winter and early spring occurs in many of the main olive growing areas in Argentina and Australia. Assessments of water requirements from flower differentiation to early fruit growth indicate that irrigation during the pre-flowering—flowering period is essential to enhance reproductive performance and oil yields in areas with a dry winter-spring season.

## Oil Concentration and Composition

The various stages of olive fruit growth and oil synthesis occur over a prolonged total period of 5–6 months (Figure 1). Under the environmental conditions in the Mediterranean Basin, most oil accumulation coincides with the late summer and fall months when temperatures are decreasing from maximum summer values. In contrast, onset of the olive oil biogenesis period takes place somewhat earlier in southern hemisphere growing regions located at relatively low latitudes, and most of the oil accumulation occurs during the summer when temperatures are higher. For this reason, this section will explore the potential importance of temperature on oil concentration and composition in non-Mediterranean climate growing regions.

In the Mediterranean Basin, the dynamics of oil accumulation are considered to have sigmoidal-type curves, irrespective of cultivar, although the rates and duration of the oil synthesis period may vary for a given cultivar according to local environmental conditions (Allalout et al., 2011; Camposeo et al.,

2013). A sigmoidal, S-shaped curve includes little increase in oil concentration early in fruit growth before pit hardening, followed by an extended linear period of oil concentration increase, and then a leveling off of oil concentration late in the season. Consistent with this pattern, recent studies in Argentina focused on modeling fruit oil concentration (%) in a number of cultivars from the beginning of pit hardening until harvest described bi-linear relationships where oil concentration increases linearly from pit hardening until it reaches a threshold value, above which oil concentration does not increase further (Trentacoste et al., 2012; Rondanini et al., 2014). In a fairly similar manner, oil synthesis in the local cv. Arauco was low until pit hardening and then followed a saturation curve, first increasing almost linearly, and then appearing to reach a plateau 170 days after full flowering (Bodoira et al., 2015). These results tend to indicate that the overall, basic pattern of oil accumulation is not altered by the climatic conditions in the southern hemisphere.

In contrast, the amount of oil accumulated does seem to be greatly affected by high temperatures. Based on correlative field studies using data from different cultivars, years, and locations in the warm desert region of NW Argentina, Rondanini et al. (2014) found that final fruit oil concentration on a dry weight basis decreased over a mean temperature range of  $23^{\circ}$ – $27^{\circ}$ C. Additionally, a comparison of oil concentration values reported from cooler central-western Argentina ( $45.5$ – $57.4\%$ ;  $33^{\circ}$ S latitude; Trentacoste et al., 2012) with values from NW Argentina ( $36.5$ – $48.5\%$ ;  $28$ – $29^{\circ}$ S; Rondanini et al., 2014) shows that final oil concentrations were lower in the warm NW region. While such apparent differences are fairly clear when considering new crop environments over significant latitudinal gradients, decreases in oil concentration have been reported even over short geographical distances within the Mediterranean Basin. In this regard, final oil concentrations averaged  $18\%$  on a fresh weight basis for the cooler coastal plain and only  $14\%$  for a warmer interior valley in Israel (Lavee et al., 2012). Maximum daily temperatures were about  $5^{\circ}$ C greater during oil accumulation at the interior valley site.

An important approach to directly evaluate the fruit oil concentration response to high temperature has been carried out by heating or cooling fruiting branches (cv. Arauco) in transparent plastic chambers under field conditions in NW Argentina (García-Inza et al., 2014). After 4 months of treatment, this study found a negative linear relationship between oil concentration and mean daily temperature during the oil synthesis period with oil concentration decreasing  $1.1\%$  per  $^{\circ}$ C across the range of average seasonal temperatures tested ( $16$ – $32^{\circ}$ C). When temperature was manipulated for shorter periods (i.e., 1 month), temperatures  $7^{\circ}$ C above ambient resulted in a negative effect on oil concentration even at final harvest, particularly when the exposure to high temperature took place at the beginning of oil accumulation. Such controlled, experimental results provide further evidence that oil concentration is strongly influenced by temperature.

To better understand oil concentration responses to temperature, the influence of temperature on the duration of the fruit-oil filling period and on the rate of oil accumulation needs to be considered. In central-western Argentina, data from several

cultivars indicated that the fruit oil-filling period was shortened by about 40 days with increasing maximum daily temperature and solar radiation (Trentacoste et al., 2012). On the other hand, fruit oil concentration was linearly related to the rate of oil accumulation in NW Argentina, but not to the duration of the oil-filling period (Rondanini et al., 2014). Thus, further studies are needed to assess the underlying mechanisms of temperature on oil accumulation.

Regarding olive oil composition, increasing evidence shows that some European olive cultivars grown in many regions of South America and Australia produce oils with different fatty acid compositions compared with those obtained from the same cultivars in their original Mediterranean Basin growing areas (Torres and Maestri, 2006; Ceci and Carelli, 2007; Torres et al., 2009; Mailer et al., 2010; Rondanini et al., 2011, 2014; Bodoira et al., 2016). In some cases, the percentages of fatty acids such as oleic acid do not even meet current International Olive Oil Council (IOOC) trade standards. This occurs despite proper agronomic practices, harvesting, and fruit processing standards.

While genotype is considered to be the major source of variability for VOO fatty acid composition (Ripa et al., 2008), the effect of environment and the genotype  $\times$  environment interaction may also be significant. Many Spanish ("Arbequina," "Manzanilla") and Italian ("Frantoio," "Coratina") cultivars show consistently lower oleic acid content and higher palmitic and linoleic acid contents when grown in Argentina vs. the Mediterranean (Ceci and Carelli, 2007; Torres et al., 2009; Rondanini et al., 2011, 2014). In contrast, tree cluster analysis indicated that the fatty acid composition was fairly similar in "Picual" when grown in Argentina or in the Mediterranean region (Rondanini et al., 2011). Additionally, oleic acid (%) appears to decrease to a much greater degree in some cultivars such as "Arbequina" than in others including "Coratina."

The cv. Arbequina was first introduced to Argentina from Spain about 70 years ago. A study by Torres et al. (2009) using AFLP DNA markers has showed high genetic homogeneity in this cultivar in central Argentina compared to its original Spanish growing region. In central Argentina, oleic acid content in "Arbequina" VOOs is 10–15% lower than in Spanish oils (Table 2), but this difference is unlikely to be explained by a founder effect associated with a relatively low number of "Arbequina" individuals being introduced originally to Argentina. Much newer "Arbequina" orchards established in NW Argentina, which originated from cuttings introduced from Europe in the 1990s, also produce oils with much lower oleic acid contents than oils from Spain.

Recently, some studies have evaluated the dynamics of fatty acid accumulation during fruit ontogeny in olive cultivars growing in several environments in Argentina (Rondanini et al., 2014; Bodoira et al., 2015, 2016). In all cultivars tested, the oleic acid content had similar maximum values (about 70% of the total fatty acid content) at early fruit growth stages and then it generally decreased, albeit at different times and rates depending on both the olive cultivar and the environment. This could indicate that the enzymatic activities of fatty acid desaturation metabolism are influenced by genotype  $\times$  environment interactions. To assess this possibility, enzymatic

studies are currently being conducted using fruit collected over a wide latitudinal gradient in western Argentina to obtain relationships between enzymatic activities and the levels of oleic and linoleic acids in two cultivars ("Arbequina" and "Coratina") showing different fatty acid profiles.

Correlation studies of olive oils from the warm valleys of NW Argentina suggest that temperature during the oil synthesis period could be the main environmental factor affecting fatty acid composition of VOOs. In this region, negative relationships between oleic acid concentrations at final harvest and seasonal mean temperatures during oil synthesis have been found for the cv. Arbequina (Rondanini et al., 2011). When the dynamics of fatty acid accumulation were modeled as a function of thermal time, oils from "Arbequina" showed a significant reduction in oleic acid content with thermal time (approximately 0.8% per 100°Cd), resulting in concentrations of less than 50% at final harvest, which coincided with a thermal time of about 3,500°Cd (Rondanini et al., 2014).

More direct evidence of the response of oil fatty acid composition to temperature was found by enclosing fruiting branches in transparent plastic chambers during the fruit filling period for 4 months under field conditions (García-Inza et al., 2014). These authors found that fruits (cv. Arauco) developed under temperatures 5 or 10°C warmer than the seasonal mean ambient temperature (20.6°C) produced oils with lower oleic acid contents. Across the whole range of temperature explored, the oleic acid concentration decreased linearly 0.7% per °C, while palmitic, linoleic and linolenic acids percentages increased with increasing temperature. Interestingly, further study also found that oleic acid content in oils extracted from the seed and the mesocarp showed opposite responses to ambient temperature (García-Inza et al., 2016). In other words, oleic acid content increased with temperature in the seed, but it decreased in the mesocarp. The increase of oleic acid in the seed of olive fruit is consistent with the response observed in oil-seed crops, such as sunflower and soybean (Rondanini et al., 2003; Zuñil et al., 2012). Detailed biochemical and molecular research is needed to understand why oleic acid content in the mesocarp decreases with temperature under field conditions. The response of fatty acid composition to temperature has not been a subject of major concern in the Mediterranean Basin. However, it may become of interest with global warming. At the very least, this issue presents a drawback for commercial olive oil production in areas that already have high temperatures.

Similar to Argentina, the wide variations in Australian olive crop environments sometimes result in oils with chemical and sensory attributes being more variable than those observed in oils produced in Mediterranean countries (Ayton et al., 2007; Mailer et al., 2010). Particularly, the fatty acid composition of "Arbequina" is notably influenced by the environmental conditions of the different Australian growing regions (Table 2). The oleic acid content of "Arbequina" VOOs is markedly lower in the northern warmer areas from New South Wales and Queensland (54.5% in average) than in the southern cooler region (81% in VOOs from Tasmania). Variations in oleic acid content in other olive cultivars grown in Australia are also significant and follow the same tendency, i.e., decreased oleic acid content in oils



**TABLE 2 |** Fatty acid composition of virgin olive oils from cv. Arbequina cultivated at different growing areas in Spain (Tous et al., 1997; Pardo et al., 2007), Argentina (Ceci and Carelli, 2007; Torres et al., 2009; Rondanini et al., 2011), and Australia (Mailer et al., 2010).

Parameters	Spain (37°–42° N) <sup>a</sup>	North-western Argentina (27°–29° S) <sup>b</sup>	Central Argentina (30° 5'S) <sup>b</sup>	Central-western Argentina (31°–33° S) <sup>b</sup>	Australia, including Tasmania (29°–42° S) <sup>a</sup>
Maturity index	Veraison—ripe	3.1 ± 0.6	3.4 ± 0.3	4.5	Veraison—ripe
<b>FATTY ACIDS (%)</b>					
Palmitic	11.3–13.9	19.2 ± 1.3	17.5 ± 1.4	15.8 ± 1.2	10.4–19.7
Palmitoleic	1.1–1.2	3.3 ± 0.7	2.5 ± 0.3	1.6 ± 0.3	0.8–3.5
Stearic	1.7–2.4	1.6 ± 0.1	1.4 ± 0.4	1.9 ± 0.2	1.2–1.9
Oleic	69.8–74.6	51.8 ± 4.1	61.3 ± 3.9	63.3 ± 3.1	54.5–81.0
Linoleic	8.3–11.4	21.9 ± 2.8	16.0 ± 3.7	15.5 ± 2.3	4.4–19.4
Linolenic	0.5–0.6	1.0 ± 0.1	0.8 ± 0.1	0.6 ± 0.2	0.6–0.7
MUFAs/PUFAs <sup>c</sup>	5.5–8.4	2.4 ± 0.5	3.9 ± 0.3	4.1 ± 0.4	3.0–8.0

<sup>a</sup>Data from Spain and Australia are presented as a range of mean values. <sup>b</sup>Data from each olive growing region in Argentina were averaged and reported as mean values (± standard deviation). <sup>c</sup>MUFAs, monounsaturated fatty acids; PUFAs, polyunsaturated fatty acids.

from warmer climates (Mailer et al., 2010). The concentrations of the different fatty acids generally fall within the IOOC acceptable limits, but this is not always the case.

Differences in fatty acid composition attributable to geographic variations can be also found in “Arbequina” growing in Chile where expansion has led to olive oil production in regions varying widely in latitude from 18°S (Azapa Valley) to 36°S (Central Valley). Although there is no evidence of a direct effect of temperature on composition of Chilean “Arbequina” olive oils, Portilla et al. (2014) reported lower oleic acid contents in oils from the most northern latitudes compared with those from the most southern ones (57 and 76% in average, respectively). Interestingly, Brazil and Uruguay are beginning to cultivate olives and to produce VOOs. Brazilian olive production is carried out mainly in the State of Minas Gerais in southeastern Brazil at subtropical latitudes (22–23°S), in an area where the mean annual temperature is about 19°C and fluctuates between 12°C (minimum) and 26°C (maximum), with mean annual rainfall being approximately 1,300 mm. Under these mild temperature conditions, the oleic acid content of olive oils from 11 cultivars with different origins was found to be in the range 70.8–84.3%, and concentrations of all individual fatty acids were within the IOOC standards for VOOs (Ballus et al., 2014).

Overall, this section indicates that olive oil from regions with warmer temperatures often has lower fruit oil concentration (%) and oleic acid content than from regions with more moderate temperatures. In addition, the reductions in oleic acid appear to be cultivar-specific. This suggests that a genotype × environment interaction is likely important in olive oil quality responses to temperature. Attempts to modify oleic acid content with agricultural practices such as irrigation management have so far been unsuccessful (e.g., Berenguer et al., 2006; Vita Serman et al., 2011; Caruso et al., 2014). Only very subtle responses to irrigation level have been observed. This suggests that cultivar selection and potentially breeding will be of significant importance in obtaining olive oil with high oleic acid content in warm areas.

## CONCLUSIONS AND FURTHER RESEARCH

Increasing global demand for olive oil has expanded olive cultivation to new growing areas in the southern hemisphere. These new crop environments often do not have typical Mediterranean climates, and some of them are in the subtropics where the response of the crop is relatively unknown. Based on the results of recently published studies, this review has highlighted: (1) the occurrence of insufficient chilling hours for flowering during winter dormancy in some high chilling requirement cultivars, such as “Frantoio” and “Leccino” in specific areas; (2) the lack of winter and spring rainfall in parts of Argentina and Australia illustrate the importance of rainfall distribution and indicate that some amount of irrigation is likely needed throughout the entire year to avoid declines in oil yield in some areas; and (3) reductions in oil concentration and oleic acid content in warm areas emphasize what may be expected for cooler regions such as the Mediterranean with global warming.

With respect to selecting specific cultivars for new southern hemisphere environments, the cv. Arbequina, which is the most common cultivar worldwide in modern super-high density orchards, has been shown to flower consistently even in warm subtropical regions, but its oil concentration and oleic acid content are often much lower than when grown in the Mediterranean. Other cultivars have also been shown to have positive and negative attributes in these new environments. Nevertheless, cultivar-specific simulation models are recommended as approximate tools to predict whether individual cultivars will likely flower in proposed new growing areas.

Temperature has emerged as a key variable considering the geographic variability found in the southern hemisphere. A critical aspect of future research may be the response of olive trees to temperature from the biochemical-molecular level to the whole-plant level. It will be important to take into account the considerable genetic variability in olive trees and the apparent genotype × environment interactions that exist for some aspects of olive quality. Thus, the use of many cultivars in studies would

be desirable when practical. Lastly, basic studies are not yet available for many growing regions in the southern hemisphere. Such information would enhance our overall understanding of olive cultivation, and reduce the necessity to extrapolate from only a few regions.

## AUTHOR CONTRIBUTIONS

MT, PP, PS, MR, GG, and DM contributed substantially to the conception and design of the review; MT, PP, PS, MR, and DM

drafted the text; MT, PP, PS, MR, GG, AM, RB, CC, and DM approved the version to be published; MT, PP, PS, MR, and DM agreed to be accountable for all aspects of the work.

## FUNDING

This research was supported by grants from the Ministerio de Ciencia, Tecnología e Innovación Productiva de Argentina (ANPCyT, PICT2015 0195) and CONICET (PUE 2016 22920160100125 and PIP 2014-16 N°542).

## REFERENCES

- Agüero Alcaras, L. M., Rousseaux, M. C., and Searles, P. S. (2016). Responses of several soil and plant indicators to post-harvest regulated deficit irrigation in olive trees and their potential for irrigation scheduling. *Agric. Water Manag.* 171, 10–20. doi: 10.1016/j.agwat.2016.03.006
- Allalout, A., Krichene, D., Methenni, K., Taamalli, A., Daoud, D., and Zarrouk, M. (2011). Behaviour of super-intensive Spanish and Greek olive cultivars grown in Northern Tunisia. *J. Food Biochem.* 35, 27–43. doi: 10.1111/j.1745-4514.2010.00364.x
- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*. FAO Irrigation and Drainage Paper No. 56. Rome: FAO.
- Aybar, V. E. (2010). *Floración en olivo (Olea europaea L.): Evaluación del Ajuste de un Modelo Predictivo para las Condiciones del Chaco Árido Argentino y Utilización de Hormonas Exógenas*. Master's thesis, Buenos Aires, University of Buenos Aires.
- Aybar, V. E., De Melo-Abreu, J. P., Searles, P. S., Matias, A. C., Del Río, C., Caballero, J. M., et al. (2015). Evaluation of olive flowering at low latitude sites in Argentina using a chilling requirement model. *Span. J. Agric. Res.* 13, 1–10. doi: 10.5424/sjar/2015131-6375
- Ayerza, R., and Sibbett, S. (2001). Thermal adaptability of olive (*Olea europaea* L.) to the Arid Chaco of Argentina. *Agric. Ecosys. Environ.* 84, 277–285. doi: 10.1016/S0167-8809(00)00260-7
- Ayton, J., Mailer, R. J., Haigh, A., Tronson, D., and Conlan, D. (2007). Quality and oxidative stability of Australian olive oil according to harvest date and irrigation. *J. Food Lip.* 14, 138–156. doi: 10.1111/j.1745-4522.2007.00076.x
- Badr, S. A., and Hartmann, H. T. (1972). Flowering response of the olive (*Olea europaea* L.) to certain growth regulators applied under inductive and noninductive environments. *Bot. Gaz.* 133, 387–392. doi: 10.1086/336659
- Ballus, C. A., Dillenburger Meinhardt, A., De Souza Campos, F. A. Jr., De Oliveira da Silva, F. L., De Oliveira, A. F., and Teixeira Godoy, H. (2014). A quantitative study on the phenolic compound, tocopherol and fatty acid contents of monovarietal virgin olive oils produced in the southeast region of Brazil. *Food Res. Int.* 62, 74–83. doi: 10.1016/j.foodres.2014.02.040
- Berenguer, M. J., Vossen, P. M., Grattan, S. R., Connell, J. H., and Polito, V. S. (2006). Tree irrigation levels for optimum chemical and sensory properties of olive oil. *HortScience* 41, 427–432.
- Bodoira, R., Torres, M., Pierantozzi, P., Aguade, F., Taticchi, A., Servili, M., et al. (2016). Dynamics of fatty acids, tocopherols and phenolic compounds biogenesis during olive (*Olea europaea* L.) fruit ontogeny. *J. Am. Oil Chem. Soc.* 93, 1289–1299. doi: 10.1007/s11746-016-2877-7
- Bodoira, R., Torres, M., Pierantozzi, P., Taticchi, A., Servili, M., and Maestri, D. (2015). Oil biogenesis and antioxidant compounds from Arauco olive (*Olea europaea* L.) cultivar during fruit development and ripening. *Eur. J. Lip. Sci. Technol.* 117, 377–388. doi: 10.1002/ejlt.201400234
- Bubán, T. (2000). The use of benzyladenine in orchard fruit growing: a mini review. *Plant Growth Regul.* 32, 381–390. doi: 10.1023/A:1010785604339
- Camposeo, S., Vivaldi, G. A., and Gattullo, C. E. (2013). Ripening indices and harvesting times of different olive cultivars for continuous harvest. *Sci. Hortic.* 151, 1–10. doi: 10.1016/j.scienta.2012.12.019
- Caruso, G., Gucci, R., Urbani, S., Esposto, S., Taticchi, A., Di Maio, I., et al. (2014). Effect of different irrigation volumes during fruit development on quality of virgin olive oil of cv. Frantoio. *Agric. Water Manag.* 134, 94–103. doi: 10.1016/j.agwat.2013.12.003
- Castillo-Llanque, F. J., Rapoport, H. F., and Baumann Samanez, H. (2014). Irrigation withholding effects on olive reproductive bud development for conditions with insufficient winter chilling. *Acta Hortic.* 1057, 113–119. doi: 10.17660/ActaHortic.2014.1057.12
- Ceci, L. N., and Carelli, A. A. (2007). Characterization of monovarietal Argentinean olive oils from new productive zones. *J. Am. Oil Chem. Soc.* 84, 1125–1136. doi: 10.1007/s11746-007-1140-7
- Connor, D. J. (2005). Adaptation of olive (*Olea europaea* L.) to water-limited environments. *Aust. J. Agric. Res.* 56, 1181–1189. doi: 10.1071/AR05169
- Connor, D. J., and Fereres, E. (2005). “The physiology of adaptation and yield expression in olive,” in *Horticultural Reviews*, Vol. 31, ed J. Janick (Hoboken, NJ: John Wiley and Sons, Inc.), 155–229.
- Correa-Tedesco, G., Rousseaux, M. C., and Searles, P. S. (2010). Plant growth and yield responses in olive (*Olea europaea*) to different irrigation levels in an arid region of Argentina. *Agric. Water Manag.* 97, 1829–1837. doi: 10.1016/j.agwat.2010.06.020
- De Melo-Abreu, J. P., Barranco, D., Cordeiro, A. M., Tous, J., Rogado, B. M., and Villalobos, F. J. (2004). Modelling olive flowering date using chilling for dormancy release and thermal time. *Agric. Forest Meteorol.* 125, 121–127. doi: 10.1016/j.agrformet.2004.02.009
- Denney, J. O., and McEachern, G. R. (1983). An analysis of several climatic temperature variables dealing with olive reproduction. *J. Amer. Soc. Hortic. Sci.* 108, 578–581.
- Di Vaio, C., Nocerino, S., Paduano, A., and Sacchi, R. (2012). Influence of some environmental factors on drupe maturation and olive oil composition. *J. Sci. Food Agric.* 93, 1134–1139. doi: 10.1002/jsfa.5863
- El-Kholi, M. (2012). *Following Olive Footprints (Olea europaea L.). Cultivation and Culture, Folklore and History, Traditions and Uses*. Córdoba: International Society for Horticultural Science, Series Scripta Horticulturae. Imprenta Luque.
- Fernández, J. E. (2014). Understanding olive adaptation to abiotic stresses as a tool to increase crop performance. *Environ. Exp. Bot.* 103, 158–179. doi: 10.1016/j.envexpbot.2013.12.003
- García-Inza, G. P., Castro, D. N., Hall, A. J., and Rousseaux, M. C. (2014). Responses to temperature of fruit dry weight, oil concentration, and fatty acid composition in olive (*Olea europaea* L. var. ‘Arauco’). *Eur. J. Agron.* 54, 107–115. doi: 10.1016/j.eja.2013.12.005
- García-Inza, G. P., Castro, D. N., Hall, A. J., and Rousseaux, M. C. (2016). Opposite oleic acid responses to temperature in oils from the seed and mesocarp of the olive fruit. *Eur. J. Agron.* 76, 138–147. doi: 10.1016/j.eja.2016.03.003
- Girona, J., Luna, M., Arbones, M., Mata, J., Rufat, J., and Marsal, J. (2002). Young olive tree responses (*Olea europea*, cv “Arbequina”) to different water supplies. Water function determination. *Acta Hortic.* 568, 277–280. doi: 10.17660/ActaHortic.2002.568.53
- Goldhamer, D. A., Dunai, J., and Ferguson, L. F. (1994). Irrigation requirements of olive trees and responses to sustained deficit irrigation. *Acta Hortic.* 365, 172–175. doi: 10.17660/ActaHortic.1994.356.36

- Gómez del Campo, M., Morales-Sillero, A., Vita Serman, F., Rousseaux, M. C., and Searles, P. S. (2010). Olive growing in the arid valleys of Northwest Argentina (provinces of Catamarca, La Rioja and San Juan). *Oliva* 114, 23–45.
- Gómez-del-Campo, M., Pérez-Expósito, M. A., Hammami, S. B. M., Centeno, A., and Rapoport, H. F. (2014). Effect of varied summer deficit irrigation on components of olive fruit growth and development. *Agric. Water Manage.* 137, 84–91. doi: 10.1016/j.agwat.2014.02.009
- Greven, M., Neal, S., Green, S., Dichio, B., and Clothier, B. (2009). The effects of drought on the water use, fruit development and oil yield from young olive trees. *Agric. Water Manage.* 96, 1525–1531. doi: 10.1016/j.agwat.2009.06.002
- Gucci, R., and Caruso, G. (2011). Environmental stresses and sustainable olive growing. *Acta Hortic.* 924, 19–30. doi: 10.17660/ActaHortic.2011.924.1
- Gucci, R., and Fereres, E. (2012). “Fruit trees and vines. Olive,” in *Crop Yield Response to Water*. FAO Irrigation and drainage paper 66 (Rome: Food and Agriculture Organization of the United Nations), 300–313.
- Gucci, R., Lodolini, E. M., and Rapoport, H. F. (2009). Water deficit induced changes in mesocarp cellular processes and the relationship between mesocarp and endocarp during olive fruit development. *Tree Physiol.* 29, 1575–1585. doi: 10.1093/treephys/tpp086
- Hartmann, H. T., and Panetos, C. (1961). Effect of soil moisture deficiency during floral development on fruitfulness in the olive. *Proc. Am. Soc. Horti. Sci.* 78, 209–217.
- IOOC (2000). *World Catalogue of Olive Varieties*. Madrid: International Olive Oil Council.
- Lavee, S. (2014). Adaptation of commercial olive cultivation to new production zones and environments including in the Southern Hemisphere – possibilities, considerations and the problems involved. *Acta Hortic.* 1057, 27–40. doi: 10.17660/ActaHortic.2014.1057.1
- Lavee, S., Hanoch, E., Wodner, M., and Abramovich, H. (2007). The effect of predetermined deficit irrigation on the performance of cv. Muhasan olives (*Olea europaea* L.) in the eastern coastal plain of Israel. *Sci. Hortic.* 112, 156–163. doi: 10.1016/j.scienta.2006.12.017
- Lavee, S., Haskal, A., and Avidan, B. (2012). The effect of planting distances and tree shape on yield and harvest efficiency of cv. Manzanillo table olives. *Sci. Hortic.* 142, 166–173. doi: 10.1016/j.scienta.2012.05.010
- Lavee, S., Rallo, L., Rapoport, H. F., and Troncoso, A. (1996). The floral biology of the olive: effect of flower number, type and distribution on fruitset. *Sci. Hortic.* 66, 149–158. doi: 10.1016/S0304-4238(96)00941-7
- Lazzez, A., Perri, E., Caravita, M. A., Khlif, M., and Cossentini, M. (2008). Influence of olive maturity stage and geographical origin on some minor components in virgin olive oil of the Chemlali variety. *J. Agric. Food Chem.* 56, 982–988. doi: 10.1021/jf0722147
- López-Olivari, R., Ortega-Farías, S., and Poblete-Echeverría, C. (2016). Partitioning of net radiation and evapotranspiration over a superintensive drip irrigated olive orchard. *Irrig. Sci.* 34, 17–31. doi: 10.1007/s00271-015-0484-2
- Mailer, R. J. (2012). “Cultivation of olives in Australia,” in *Olive Oil - Constituents, Quality, Health Properties and Bioconversions*, ed D. Boskou (Rijeka: InTech), 211–232. doi: 10.5772/36235
- Mailer, R. J., Ayton, J., and Graham, K. (2010). The influence of growing region, cultivar and harvest timing on the diversity of Australian olive oil. *J. Am. Oil Chem. Soc.* 87, 877–884. doi: 10.1007/s11746-010-1608-8
- Malik, N. S. A., and Bradford, J. M. (2009). Inhibition of flowering in ‘Arbequina’ olives from chilling at lower temperatures. *J. Food Agric. Environ.* 7, 429–431.
- Malik, N. S. A., and Perez, J. L. (2011). The effect of high temperature interruptions during inductive period on the extent of flowering and on metabolic responses in olives (*Olea europaea* L.). *Sci. Hortic.* 129, 207–212. doi: 10.1016/j.scienta.2011.03.028
- Marodin, G. A. B., Sartori, I. A., and Guerra, D. S. (2002). Effect of the application of hydrogen cyanamide and mineral oil on dormancy breaking and peach production, cv. Flamecrest. *Rev. Bras. Frutic.* 24, 426–430. doi: 10.1590/S0100-29452002000200029
- Mohamed, A. K. A. (2008). The effect of chilling, defoliation and hydrogen cyanamide on dormancy release, bud break and fruiting of Anna apple cultivar. *Sci. Hortic.* 118, 25–32. doi: 10.1016/j.scienta.2008.05.015
- Or, E., Vilozny, I., Eyal, Y., and Ogrdovitch, A. (2000). The transduction of the signal for grape bud dormancy breaking induced by hydrogen cyanamide may involve the SNF-like protein kinase GDBRPK. *Plant Molec. Biol.* 43, 483–494. doi: 10.1023/A:1006450516982
- Ortega-Farías, S., Ortega-Salazar, S., Poblete, T., Kilic, A., Allen, R., Poblete-Echeverría, C., et al. (2016). Estimation of energy balance components over a drip-irrigated olive orchard using thermal and multispectral cameras placed on a helicopter-based unmanned aerial vehicle (UAV). *Remote Sens.* 8, 638–656. doi: 10.3390/rs8080638
- Pardo, J. E., Cuesta, M. A., and Alvarruiz, A. (2007). Evaluation of potential and real quality of virgin olive oil from the designation of origin “Aceite Campo de Montiel” (Ciudad Real, España). *Food Chem.* 100, 977–984. doi: 10.1016/j.foodchem.2005.10.059
- Pierantozzi, P., Torres, M., Bodoira, R., and Maestri, D. (2013). Water relations, biochemical-physiological and yield responses of olive trees (*Olea europaea* L. cvs. Arbequina and Manzanilla) under drought stress during the pre-flowering and flowering period. *Agric. Water Manage.* 125, 13–25. doi: 10.1016/j.agwat.2013.04.003
- Pierantozzi, P., Torres, M., Lavee, S., and Maestri, D. (2014). Vegetative and reproductive responses, oil yield and composition from olive trees (*Olea europaea* L.) under contrasting water availability during the dry winter-spring period in central Argentina. *An. Appl. Biol.* 164, 116–127. doi: 10.1111/aab.12086
- Portilla, G., Robert, P., Sepúlveda, B., and Romero, N. (2014). Carotenoid pigments, polyphenols, tocopherols and fat composition of extra virgin olive oils produced in Chile. *Acta Hortic.* 1057, 619–626. doi: 10.17660/ActaHortic.2014.1057.78
- Rallo, L., and Cuevas, J. (2008). “Fructificación y producción,” in *El cultivo del olivo*, eds D. Barranco, R. Fernández-Escobar, and L. Rallo (Madrid: Ediciones Mundi Prensas), 127–162.
- Rallo, L., and Martin, G. C. (1991). The role of chilling in releasing olive floral buds from dormancy. *J. Am. Soc. Horti. Sci.* 116, 1058–1062.
- Rapoport, H. F. (2014). The reproductive biology of the olive tree and its relationship to extreme environmental conditions. *Acta Hortic.* 1057, 41–50. doi: 10.17660/ActaHortic.2014.1057.2
- Rapoport, H. F., Hammami, S. B. M., Martins, P., Pérez-Priego, O., and Orgaz, F. (2012). Influence of water deficits at different times during olive tree inflorescence and flower development. *Environ. Exp. Bot.* 77, 227–233. doi: 10.1016/j.envexpbot.2011.11.021
- Ripa, V., De Rose, F., Caravita, M. A., Parise, M. R., Perri, E., Rosati, A., et al. (2008). Qualitative evaluation of olive oils from new olive selections and effects of genotype and environment on oil quality. *Adv. Hortic. Sci.* 22, 95–103.
- Rondanini, D. P., Castro, D. N., Searles, P. S., and Rousseaux, M. C. (2011). Fatty acid profiles of varietal virgin olive oils (*Olea europaea* L.) from mature orchards in warm arid valleys of Northwestern Argentina (La Rioja). *Grasas y Aceites* 62, 399–409. doi: 10.3989/gya.125110
- Rondanini, D. P., Castro, D. N., Searles, P. S., and Rousseaux, M. C. (2014). Contrasting patterns of fatty acid composition and oil accumulation during fruit growth in several olive varieties and locations in a non-mediterranean region. *Eur. J. Agron.* 52, 237–246. doi: 10.1016/j.eja.2013.09.002
- Rondanini, D., Savin, R., and Hall, A. (2003). Dynamics of fruit growth and oil quality of sunflower (*Helianthus annuus* L.) exposed to brief intervals of high temperature during grain filling. *Field Crops Res.* 83, 79–90. doi: 10.1016/S0378-4290(03)00064-9
- Rousseaux, M. C., Benedetti, J. P., and Searles, P. S. (2008). Leaf-level responses of olive trees (*Olea europaea*) to the suspension of irrigation during the winter in an arid region of Argentina. *Sci. Hortic.* 115, 135–141. doi: 10.1016/j.scienta.2007.08.005
- Rousseaux, M. C., Figuerola, P. I., Correa-Tedesco, G., and Searles, P. S. (2009). Seasonal variations in sap flow and soil evaporation in an olive (*Olea europaea* L.) grove under two irrigation regimes in an arid region of Argentina. *Agric. Water Manag.* 96, 1037–1044. doi: 10.1016/j.agwat.2009.02.003
- Sanz-Cortés, F., Martínez-Calvo, J., Badenes, M. L., Bleiholder, H., Hack, H., Llacer, G., et al. (2002). Phenological growth stages of olive trees (*Olea europaea*). *An. Appl. Biol.* 140, 151–157. doi: 10.1111/j.1744-7348.2002.tb00167.x
- Searles, P. S., Agüero Alcarás, M., and Rousseaux, M. C. (2011). El consumo de agua por el cultivo de olivo (*Olea europaea* L.) en el noroeste de Argentina: una comparación con la Cuenca Mediterránea. *Ecología Austral* 21, 15–28.
- Temine, S. B., Manai, H., Methenni, K., Baccouri, B., Abaza, L., Daoud, D., et al. (2008). Sterolic composition of Chétoui virgin

- olive oil: influence of geographical origin. *Food Chem.* 110, 368–374. doi: 10.1016/j.foodchem.2008.02.012
- Tognetti, R., D'Andria, R., Morelli, G., and Alvino, A. (2005). The effect of deficit irrigation on seasonal variations of plant water use in *Olea europaea* L. *Plant Soil* 273, 139–155. doi: 10.1007/s11104-004-7244-z
- Torres, M. M., and Maestri, D. M. (2006). The effects of genotype and extraction methods on chemical composition of virgin olive oils from Traslasierra Valley (Córdoba, Argentina). *Food Chem.* 96, 507–511. doi: 10.1016/j.foodchem.2005.03.003
- Torres, M. M., Pierantozzi, P., Cáceres, M. E., Labombarda, P., Fontanazza, G., and Maestri, D. M. (2009). Genetic and chemical assessment in Arbequina olive cultivar grown in Córdoba province (Argentina). *J. Sci. Food Agric.* 89, 523–530. doi: 10.1002/jsfa.3483
- Tous, J., Romero, A., Plana, J., Guerrero, L., Díaz, I., and Hermoso, J. F. (1997). Características químico-sensoriales de los aceites de oliva “Arbequina” obtenidos en distintas zonas de España. *Grasas y Aceites* 48, 415–424. doi: 10.3989/gya.1997.v48.i6.814
- Trentacoste, E. R., Connor, D. J., and Gómez del Campo, M. (2015). Effect of row spacing on vegetative structure, fruit characteristics and oil productivity of N-S and E-W oriented olive hedgerows. *Sci. Hortic.* 193, 240–248. doi: 10.1016/j.scienta.2015.07.013
- Trentacoste, E. R., Puertas, C. M., and Sadras, V. O. (2012). Modelling the intraspecific variation in the dynamics of fruit growth, oil and water concentration in olive (*Olea europaea* L.). *Eur. J. Agron.* 38, 83–93. doi: 10.1016/j.eja.2012.01.001
- Tura, D., Gigliotti, C., Pedo, S., Failla, O., Bassi, D., and Serraiocco, A. (2007). Influence of cultivar and site of cultivation on levels of lipophilic and hydrophilic antioxidants in virgin olive oils (*Olea europaea* L.) and correlations with oxidative stability. *Sci. Hortic.* 112, 108–119. doi: 10.1016/j.scienta.2006.12.036
- Vita Serman, F., Pacheco, D., Olguín Pringles, A., Bueno, L., and Capraro, F. (2011). Effect of regulated deficit irrigation strategies on productivity, quality and water use efficiency in a high-density “Arbequina” olive orchard located in an arid region of Argentina. *Acta Hortic.* 888, 81–88. doi: 10.17660/ActaHortic.2011.888.8
- Yunusa, I. A. M., Zeppel, M. J. B., and Nuberg, I. K. (2008). Water-use efficiency reflects management practices in Australian olive groves. *J. Hortic. Sci. Biotechnol.* 83, 232–238. doi: 10.1080/14620316.2008.11512374
- Zeleeke, K. (2014). Water use and root zone water dynamics of drip-irrigated olive (*Olea europaea* L.) under different soil water regimes. *New Zeal. J. Crop Hortic. Sci.* 42, 217–232. doi: 10.1080/01140671.2014.891527
- Zeleeke, K., Mailer, R., Eberbach, P., and Wünsche, J. (2012). Oil content and fruit quality of nine olive (*Olea europaea* L.) varieties affected by irrigation and harvest times. *New Zeal. J. Crop Hortic. Sci.* 40, 241–252. doi: 10.1080/01140671.2012.662159
- Zuil, S., Izquierdo, N., Cantarero, L., and Aguirrezábal, L. (2012). Oil quality of maize and soybean genotypes with increased oleic acid percentage as affected by intercepted solar radiation and temperature. *Field Crops Res.* 127, 203–214. doi: 10.1016/j.fcr.2011.11.019

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Torres, Pierantozzi, Searles, Rousseaux, García-Inza, Miserere, Bodoira, Contreras and Maestri. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Genomic Prediction of Sunflower Hybrids Oil Content

Brigitte Mangin<sup>1\*</sup>, Fanny Bonnafous<sup>1</sup>, Nicolas Blanchet<sup>1</sup>, Marie-Claude Boniface<sup>1</sup>, Emmanuelle Bret-Mestries<sup>2</sup>, Sébastien Carrère<sup>1</sup>, Ludovic Cottret<sup>1</sup>, Ludovic Legrand<sup>1</sup>, Gwenola Marage<sup>1</sup>, Prune Pegot-Espagnet<sup>1</sup>, Stéphane Munos<sup>1</sup>, Nicolas Pouilly<sup>1</sup>, Felicity Vear<sup>3</sup>, Patrick Vincourt<sup>1</sup> and Nicolas B. Langlade<sup>1</sup>

<sup>1</sup> LIPM, Université de Toulouse, INRA, Centre National de la Recherche Scientifique, Castanet-Tolosan, France, <sup>2</sup> Terres Inovia, AGIR, Castanet-Tolosan, France, <sup>3</sup> GDEC, INRA, Université Clermont II Blaise Pascal, Clermont-Ferrand, France

## OPEN ACCESS

### Edited by:

Leire Molinero-Ruiz,  
Instituto de Agricultura Sostenible  
(CSIC), Spain

### Reviewed by:

Miguel Perez-Enciso,  
Universitat Autònoma de Barcelona,  
Spain  
Delin Hong,  
Nanjing Agricultural University, China

### \*Correspondence:

Brigitte Mangin  
brigitte.mangin@inra.fr

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 28 June 2017

**Accepted:** 06 September 2017

**Published:** 21 September 2017

### Citation:

Mangin B, Bonnafous F, Blanchet N, Boniface M-C, Bret-Mestries E, Carrère S, Cottret L, Legrand L, Marage G, Pegot-Espagnet P, Munos S, Pouilly N, Vear F, Vincourt P and Langlade NB (2017) Genomic Prediction of Sunflower Hybrids Oil Content. *Front. Plant Sci.* 8:1633. doi: 10.3389/fpls.2017.01633

Prediction of hybrid performance using incomplete factorial mating designs is widely used in breeding programs including different heterotic groups. Based on the general combining ability (GCA) of the parents, predictions are accurate only if the genetic variance resulting from the specific combining ability is small and both parents have phenotyped descendants. Genomic selection (GS) can predict performance using a model trained on both phenotyped and genotyped hybrids that do not necessarily include all hybrid parents. Therefore, GS could overcome the issue of unknown parent GCA. Here, we compared the accuracy of classical GCA-based and genomic predictions for oil content of sunflower seeds using several GS models. Our study involved 452 sunflower hybrids from an incomplete factorial design of 36 female and 36 male lines. Re-sequencing of parental lines allowed to identify 468,194 non-redundant SNPs and to infer the hybrid genotypes. Oil content was observed in a multi-environment trial (MET) over 3 years, leading to nine different environments. We compared GCA-based model to different GS models including female and male genomic kinships with the addition of the female-by-male interaction genomic kinship, the use of functional knowledge as SNPs in genes of oil metabolic pathways, and with epistasis modeling. When both parents have descendants in the training set, the predictive ability was high even for GCA-based prediction, with an average MET value of 0.782. GS performed slightly better (+0.2%). Neither the inclusion of the female-by-male interaction, nor functional knowledge of oil metabolism, nor epistasis modeling improved the GS accuracy. GS greatly improved predictive ability when one or both parents were untested in the training set, increasing GCA-based predictive ability by 10.4% from 0.575 to 0.635 in the MET. In this scenario, performing GS only considering SNPs in oil metabolic pathways did not improve whole genome GS prediction but increased GCA-based prediction ability by 6.4%. Our results show that GS is a major improvement to breeding efficiency compared to the classical GCA modeling when either one or both parents are not well-characterized. This finding could therefore accelerate breeding through reducing phenotyping efforts and more effectively targeting for the most promising crosses.

**Keywords:** genomic selection, factorial design, sunflower, oil content, hybrid, GBS

# 1. INTRODUCTION

Sunflower is one of the main oilseed crops worldwide. Although this crop was domesticated in North America, the sunflower was developed as a major crop in Russia in the first half of the twentieth century, when the breeding programs of V.S. Pustovoi increased the seed oil content from 25–30 to 45–50%. This success largely reflected the high heritability of this key breeding trait. Based on two segregating populations, involving wild-type and improved germplasms, Fick (1975) provided the first estimation of the narrow-sense heritability of seed oil content as 0.52–0.61, suggesting that the contribution of genetic additive variance is prominent. As the seed oil content can now be rapidly and inexpensively measured using nuclear magnetic resonance (NMR), selection can be performed in segregating progenies and on a single plant basis, from the F<sub>2</sub> generation onwards. Vear et al. (2010) indicated that hybrids generally show heterosis for oil content, which is not typically the case when the parents contain approximately 50% oil.

Mapping quantitative trait locus (QTL) for oil content was initiated more than 20 years ago (Leon et al., 1995), providing congruent results across the segregating populations involved (Mestries et al., 1998; Bert et al., 2002; Bachlava et al., 2010; Merah et al., 2012) and confirming both the quantitative nature of the trait (several loci) and its high heritability (mapped QTLs accounted for 10 to 51% of the phenotypic variability). The high level of heritability for oil content suggests that this trait is an easy character to breed for, and the absence of important interactions with environmental conditions makes it feasible to obtain valid general conclusions as to the interest of a genotype for this character based on a small number of measurements under different conditions. Thus, there is no direct requirement for genomic studies to replace phenotypic measurements. However, because robust oil content data are easy to obtain, oil content is a good model trait to test the power of genomic selection models prior to applying these models to explore more complex characters, such as seed yield or quantitative resistance to fungal diseases.

Genomic prediction refers to the prediction of genetic value based on markers spread throughout the entire genome. In this framework, a mathematical model is trained on past genotyped and phenotyped resources, and new unobserved individuals who are genotyped but not phenotyped are predicted with this learned model. Among the different models since the work of Meuwissen et al. (2001), the mixed model and the genome-wide best linear prediction (GBLUP) of unobserved individuals proposed by VanRaden (2008) is the most popular model. Originally, the mixed model of Meuwissen et al. (2001) assumes that the haplotype effects of all genomic regions follow the same Gaussian distribution. When limiting each genomic region to a single marker, this model is known as the ridge regression BLUP (RR-BLUP). The RR-BLUP and GBLUP models are equivalent models (Endelman, 2011) and Goddard (2009) showed that they are similar to the classical pedigree mixed model when relatedness between individuals are estimated with markers. Mixed models and BLUP have been comprehensively compared to other methods of genomic prediction as penalized

regressions (Li and Sillanpää, 2012, for a review), Bayesian modeling (Kärkkäinen and Sillanpää, 2012, for a review), semi-parametric learners as the reproducing kernel Hilbert space (RKHS) (Gianola et al., 2006) and non-parametric methods, such as random forest (Chen and Ishwaran, 2012). Depending on the trait studied, one or the other of these methods was demonstrated as more reliable, but the best performers provided comparable accuracies (Heslot et al., 2012; Haws et al., 2015). The mixed model framework has consistently produced comparative results to those obtained with more complicated models. The simplicity, efficient computer implementation and flexibility of this model have meant that most novel modeling ideas have been based on this framework.

As previously described, in GBLUP modeling, effects of genetic markers are assumed to follow the same Gaussian distribution. This unrealistic assumption does not consider the biological mechanisms underlying phenotypic variation. Speed and Balding (2014) proposed an extension of GBLUP, called MultiBLUP, to include multiple random effects allocated to different sets of SNP markers. The close variants are grouped and the relatedness of random genetic effects is determined for each set using a similarity matrix calculated using the SNPs of the region of interest, thus modeling a different effect-size distribution for each set. Using MultiBLUP, Wolfe et al. (2016) predicted disease resistance in cassava. Delimiting the genome to a region representing between 30 and 66% of the genetic resistance and using the remaining SNPs facilitated an increase in the precision of prediction from 0.53 to 0.58 compared to GBLUP. Similarly, Sarup et al. (2016) using genomic feature BLUP, which is equivalent to MultiBLUP, asserted using several porcine traits that MultiBLUP prediction accuracy is better than GBLUP when the set of SNPs linked to previously known QTLs explained more than 10% of trait variability. Other methods for integrating information a priori have also been tested. Zhang et al. (2014) proposed the consideration of QTLs based on assigning a predefined weight to the region of interest in the relatedness. These QTLs could also be included as fixed effects (Bernardo, 2014; Spindel et al., 2016). Prediction accuracy increases up to 30% depending on the trait when SNPs are derived from a GWAS performed with the data used to train the genomic model (Spindel et al., 2016). However, the integration of SNPs from the literature does not show the same ability to improve on the model accuracy. MultiBLUP is currently included in the framework of multi-kernel mixed models (Weissbrod et al., 2016), as are included linear mixed models for complex trait architecture (dominance and epistasis) (de los Campos et al., 2009).

Mixed models for hybrid predictions based on GCA and/or specific combining ability (SCA) have long been applied prior to the use of genetic markers. In maize, Bernardo (1996) enhanced this old model by proposing the pedigree BLUP model, which uses co-ancestry coefficients between parents of hybrids. First attempt to estimate these co-ancestry coefficients using molecular markers was proposed in Schrag et al. (2006) and further generalized by Technow et al. (2014) using whole genomic data. In sunflower, Reif et al. (2013) did not observe any improvement of genomic BLUP compared to the pedigree BLUP. Equality between the two approaches was consistent with the work of

Goddard (2009), as Reif et al. (2013) estimated co-ancestry coefficients using the same markers included in the genomic BLUP, so the two predictions are equivalent. In contrast to Reif et al. (2013), we want to compare prediction accuracy of hybrid genetic values using a classical mixed model that makes use of only pedigree information, if available, to other mixed models that use genomic data to compute relatedness between hybrids and parents. We make this comparison using seed oil content phenotypes observed in an incomplete factorial design produced in the course of the SUNRISE project.

## 2. MATERIALS AND METHODS

### 2.1. Plant Materials

Hybrids were obtained as an incomplete factorial design by crossing 36 maintainer lines with 36 restorer lines. The complete hybrid panel contained 492 hybrids. These plants were sown in 11 different environments (5 different environments in 2013, 3 different environments in 2014, and 3 different environments in 2015) (Bonnafeous et al., 2017), but for the present study of oil content, we discarded 2 environments due to imperfect randomizations and inaccurate phenotypic observations.

The parents were genotyped by sequencing using the XRQ genome as the reference parent, and their genotypes were imputed by chromosome using Beagle (Browning and Browning, 2009) as described in Badouin et al. (2017). SNPs that were not polymorphic in either the maintainer or the restorer panels were discarded, and a single referent SNP was maintained, representing each set of redundant SNPs (i.e., SNPs in complete linkage disequilibrium in the 72 parent panel). Finally, the genomic data comprised 468,194 non-redundant SNPs, and hybrid genotypes were deduced from the parent genotypes.

Measurement of oil seed content was observed by NMR using a minispec (MQ10H, mq Series, version 1.2, January 2000, Bruker, Germany). Each 20-ml seed sample was first dried for 24 h at 80°C and subsequently analyzed at room temperature.

Genes related to oil metabolism have been identified through the metabolic network reconstruction of the genome annotation of the sunflower (Badouin et al., 2017). The oil metabolism super-pathway has been manually constructed from several inferred metabolic pathways. Relations between genes and reactions were automatically inferred and curated based on the literature. Further details are provided in the on-line materials Badouin et al. (2017), and the examined genes are listed in the Supplementary material (Data sheet S1). An interactive view of the pathway showing with the gene/reaction links is available at <https://pathway-tools.toulouse.inra.fr/HANXRQ/NEW-IMAGE?type=PATHWAY&object=PWY198A-2> We considered all SNPs identified in the genes listed above, and we added all the SNPs located 1,000 bases upstream and downstream of these genes.

### 2.2. Predictions of Hybrid Performances

The phenotypes were initially adjusted using a spatial model, including the line and column numbers in the field, the repetition when necessary and the genotype status (check variety or hybrid) as fixed factors, and a random independent effect modeling the

genotypic value of observed individuals completed the model as described in Bonnafeous et al. (2017).

Predictions of hybrid performance were computed based on BLUP using several linear mixed models within each environment. Variance components of linear models were estimated using restricted maximum likelihood (REML) with the ASReml-R package (Butler et al., 2007). The models are similar to the progeny models described in Bouvet et al. (2016).

#### 2.2.1. GCA-Based Prediction

The hybrid genetic value of the  $fm$  hybrid was predicted using  $\widehat{GCA}_f + \widehat{GCA}_m$ , where  $f$  denoted the female line and  $m$  the male line. GCA BLUPs were obtained using the following model:

$$y_{fm} = \mu + GCA_f + GCA_m + \epsilon_{fm} \quad (\text{GCA model}) \quad (1)$$

where  $y_{fm}$  is the adjusted phenotype in an environment,  $\mu$  is the mean,  $GCA_f$  and  $GCA_m$  are the random GCA effects of female  $f$  and male  $m$ , respectively, and  $\epsilon_{fm}$  denotes error. All random effect are assumed Gaussian and independent with  $\sigma_{GCA_f}^2$ ,  $\sigma_{GCA_m}^2$ , and  $\sigma_{\epsilon}^2$  for the GCA female, GCA male, and residual variances, respectively. When the parent pedigree is known, the relatedness of parents can be included in the variances of GCA random effects using a coancestry coefficient matrix in this model. However, the pedigree of the parental lines was considered to have too much uncertainty to account for using this analysis. Moreover, parents of the factorial design were chosen to be as unrelated to provide a good representation of the core collection studied in Cadic et al. (2013). Therefore, these parental lines are assumed independent.

#### 2.2.2. FM and FMI Model Predictions

The hybrid genetic value of the  $fm$  hybrid was predicted using BLUPs of  $\widehat{F}_f + \widehat{M}_m$  in the FM model and  $\widehat{F}_f + \widehat{M}_m + \widehat{I}_{fm}$  in the FMI model.

$$y_{fm} = \mu + F_f + M_m + \epsilon_{fm} \quad (\text{FM model}) \quad (2)$$

$$y_{fm} = \mu + F_f + M_m + I_{fm} + \epsilon_{fm} \quad (\text{FMI model}) \quad (3)$$

where  $F_f$ ,  $M_m$ , and  $I_{fm}$  are the random effects of female  $f$  and male  $m$  lines and their interactions, respectively, and  $\epsilon_{fm}$  denotes error. Let  $F$ ,  $M$ ,  $I$ , and  $\epsilon$  denote vectors of female, male, interaction and error residual effects, respectively.  $F \sim \mathcal{N}(0, \sigma_f^2 K_f)$ ,  $M \sim \mathcal{N}(0, \sigma_m^2 K_m)$ ,  $I \sim \mathcal{N}(0, \sigma_{fm}^2 K_{fm})$ ,  $\epsilon \sim \mathcal{N}(0, \sigma_{\epsilon}^2 Id)$  where  $K_f$  is the kinship matrix for females;  $K_m$  is the kinship matrix for males;  $K_{fm}$  is the kinship matrix for the interaction between males and females; and  $\sigma_f^2$ ,  $\sigma_m^2$ ,  $\sigma_{fm}^2$  and  $\sigma_{\epsilon}^2$  are female, male, female by male interaction and residual variance, respectively.

$K_f = X_f X_f'$ ,  $K_m = Z_m Z_m'$ , and  $K_{fm} = W_{fm} W_{fm}'$  with  $W_{fm}$  as the Hadamard product between  $X_f$  and  $Z_m$ , where  $X_f$  is the vector of  $x_f^l$ , the centered (0 or 1) allele transmitted by female  $f$  at the  $l$ th marker locus, and  $Z_m$  is the vector of  $z_m^l$ , the centered (0 or 1) allele transmitted by male  $m$  at the  $l$ th marker locus.

Note that the GCA model and the FM model differ only by the assumptions made on the variance-covariance of the random parental effects. The variance-covariance matrix of these parental

effects was proportional to the identity matrix in the GCA model when it was computed using markers in the FM model. Both models predict the parental GCA and the hybrid prediction is the sum of predicted parental GCA.

We performed two FM models: (i) in one model, the parental design matrices ( $X_f$  and  $X_m$ ) were computed including 468,194 genome SNPs, (ii) in the other model, these matrices included only a pre-selected set of SNPs in genes previously demonstrated as involved in the oil content metabolism network.

### 2.2.3. Multi-Kernel Model Predictions

MultiBLUP was proposed by Speed and Balding (2014) in trait additive modeling. This model was further extended to consider more complex trait architecture, such as epistasis, and this model was included in the general and highly flexible framework of the multi-kernel model (Weissbrod et al., 2016). In the simplest linear additive form of Speed and Balding (2014), these models comprise several additive random effects, each with its own kinship (linear kernel) and variance. These models can easily be generalized to FM or FMI models by modeling several groups of parental random factors, each group having its own kinship and variance. The hybrid genetic value is subsequently predicted based on the sum of the BLUP values for the female and male effects in different groups in the FM model, as an example.

We performed two multi-kernel BLUP models using female and male SNP allelic effects. One prediction is the generalization of MultiBLUP to FM model using two SNP groups, with the SNPs in genes or close to genes involved in the oil content metabolism network in one group, and all remaining SNPs in the other group.

The other multi-kernel model adds to female and male kinships, two epistasis parental kinships computed using the Hadamard product  $K_f * K_f$  and  $K_m * K_m$  for the femalefemale and malemale epistasis kinship, respectively. This model is a generalization of additivexadditive epistasis modeling proposed by Su et al. (2012) to FM model. The Su et al. (2012) epistasis modeling was demonstrated to explicitly model all pairwise additivexadditive SNP interactions by Jiang and Reif (2015) and is similar to the model of Bouvet et al. (2016).

### 2.2.4. Predictive Ability of Hybrid Performances

Predictive ability or phenotypic accuracy of predictions was based on the Pearson's correlation between the observed phenotypes and their predicted values for hybrids that were not used to train the models, the so-called test individuals or out-of-population hybrids. This accuracy was computed as the mean of 100 test sets. We used two sampling schemes: a random draw of 10% of the hybrids or a random draw of 10% of the parent lines for which all observed descendants were included in the test set. This latter sampling enables the generation of test sets comprising only T1 or T0 hybrids, consistent with Technow et al. (2014), i.e., out-of-population samples with parents never observed through hybrid progeny. As for the SUNRISE incomplete factorial design, all parents had a nearly equal number of descendants, this sampling scheme generated approximately 10% of hybrids.

## 3. RESULTS

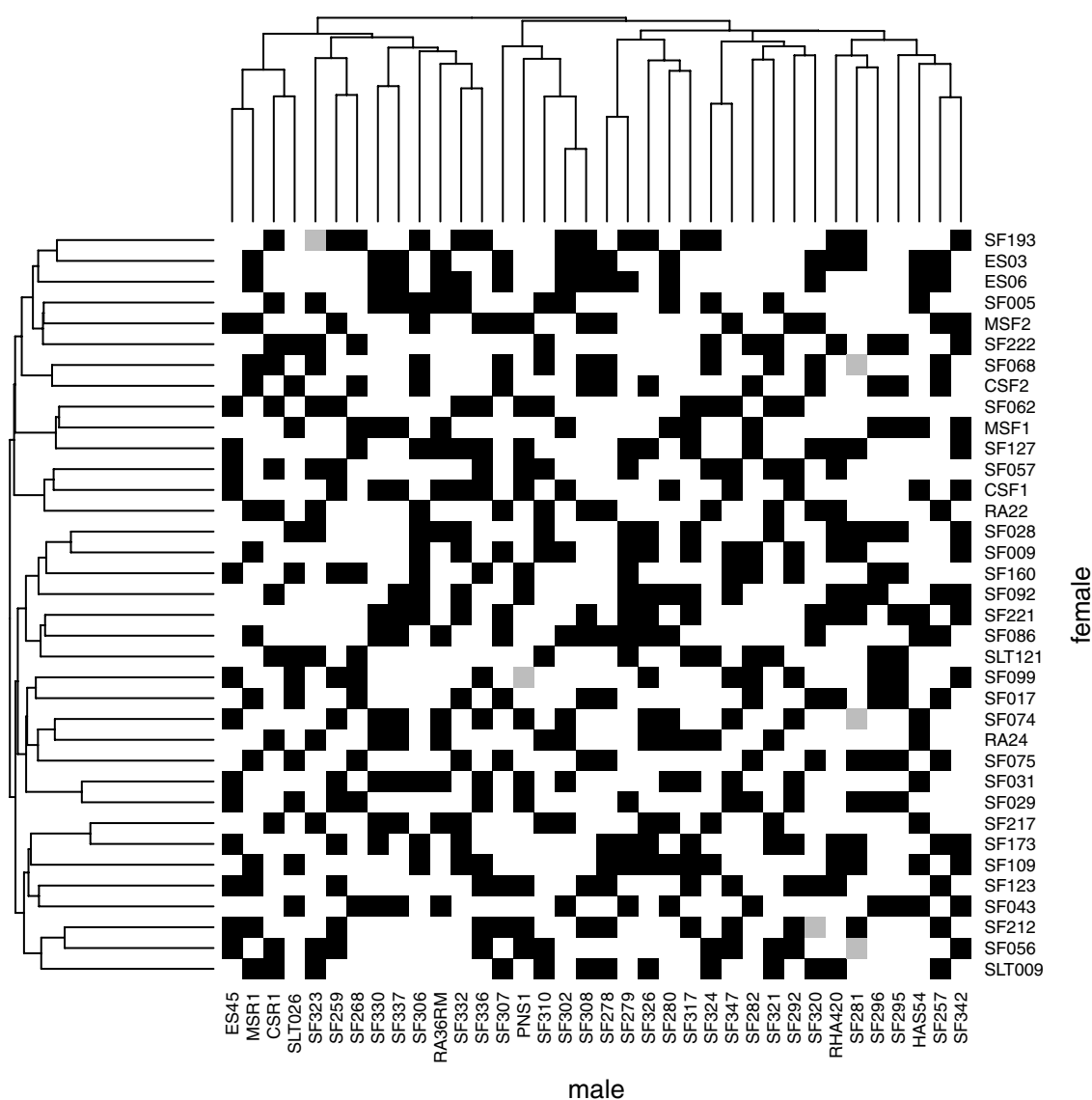
**Figure 1** presents the SUNRISE incomplete factorial design with females and males arranged according to a hierarchical clustering based on VanRaden's kinship matrices (VanRaden, 2008). As part of the sunflower elite collection studied by Cadic et al. (2013), male parents are restorers of the CMS PET1 cytoplasmic male sterility, [R-lines] for which female parents are maintainers, [B-lines]. Male parents seem slightly more structured and related than female parents, consistent with the findings of Cadic et al. (2013), who distinguished two main subgroups in the B-germplasm among the core collection. The factorial design was completely connected and almost balanced, as the parents were involved in nearly an equal number of crosses. Among the 492 hybrids generated, 486 hybrids were observed in the MET at least once for the oil content phenotype; thus, for this trait, parents were observed for a minimum of 12 to a maximum of 15 descendants (number of observed hybrids per parents in the MET are detailed in Supplementary Material, Tables S1, S2).

The oil content-adjusted phenotype of hybrids varied from 31.7 to 59.0% on the MET (see histograms in Supplementary Material, Figure S1). Hybrid-adjusted phenotypes were positively and significantly correlated between environments (**Figure 2**) with a minimum of 0.47, a maximum of 0.77 and average of 0.64. Intra-year correlations were slightly higher than between-year correlations, and environments observed in 2014 (14EX04, 14RV01) were less correlated with the two other year environments (13EX01, 13EX03, 13EX04 and 13EX05 sown in 2013, and 15EX05, 15EX06, and 15EX07 sown in 2015).

Using the three principal models of prediction (GCA, FM and FMI with all SNPs), we compared the REML variance components and their part of variance (**Table 1**). Female and male parts of variance were stable in the MET, despite visible differences in variance component values, particularly the environment 15EX07, implanted in Romany with a wider inter-row spacing (0.7 m) than the other environments (0.5 to 0.6 m). A decrease in the female part of variance and an increase in the male part of variance were observed in all environments using the correction based on the genomic relatedness of parents performed in the FM and FMI models. A significant female  $\times$  male interaction (z-ratio equal to 2.71) was observed in a single environment (13EX05), and when included in the model, residual error was divided by 2. In all environments and for all models, the female part of genetic variance was superior to that of the male counterpart, showing roughly a ratio of (3/2) in favor of the female parent in the inheritance of hybrid genetic value for oil content.

The three models described above were compared for their ability to predict unobserved hybrid genetic values on the same test sets (**Table 2**). Two sampling processes were experimented to estimate the reliability of GS either to complete the factorial design by predicting missing hybrids or to predict hybrids for which one or both parents were never observed by a descendant in the factorial design (the so-called T0 and T1 hybrids, (Technow et al., 2014)). The predictive ability of GS is high for oil content on the MET (0.783 in average for

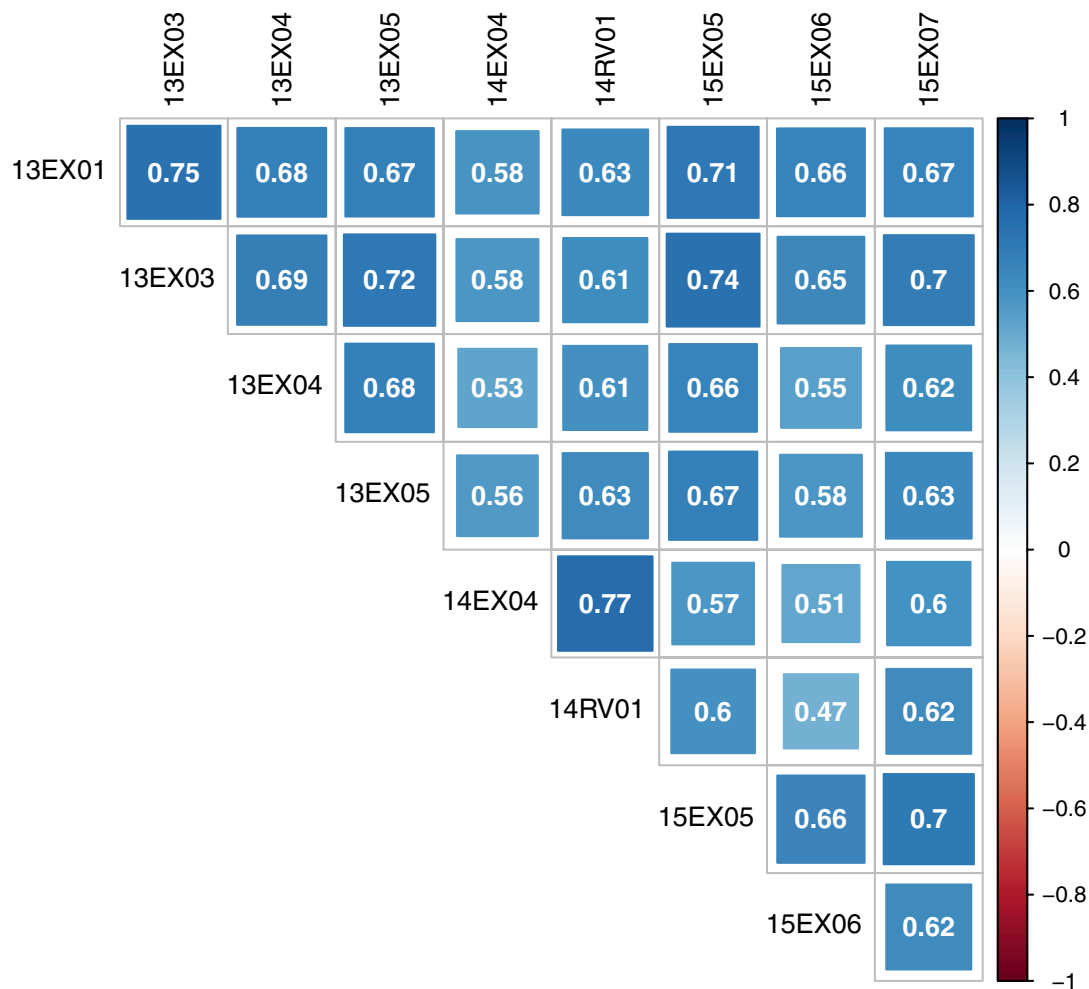




**FIGURE 1** | Incomplete factorial design of SUNRISE from 36 maintainer lines (females, in column) and 36 restorer lines (males, in row). Black squares indicate the 486 hybrids with phenotypic observations on at least one environment and gray squares indicate hybrids that were planned but were never observed for oil content in the MET. Male and female dendrograms are based on their kinship matrices.

*FM* model) when the goal is to predict missing hybrids. The three models were nearly equally accurate, with only a 0.2% increase between the *GCA* model (the worse) and *FM* model (the best). The *FMI* model performed slightly better than the *FM* model in two environments (13EX05 and 14RV01), and these two environments had the greatest estimates of female  $\times$  male interaction variances. The predictive ability is lower when the goal is to predict T0 or T1 hybrids with an average of 0.635 as the best performer (*FM* model). Once again, the *GCA* model was the least accurate model (0.575 in average), showing a 10% decrease in predictive ability compared to the *FM* model. The ranking between the *FM* and *FMI* models was similar to the previous sampling schema.

Having observed that all methods are equally accurate to predict the missing hybrids of the factorial design, we focused on the prediction of T0 and T1 hybrids. Moreover, we made a prediction without considering the female  $\times$  male interaction, as this interaction did not improve the accuracy and was CPU-time consuming. We attempted to improve the *FM* model by considering the genes involved in the oil metabolic pathway. Three hundred and seventy-two genes located throughout all chromosomes, having 3,746 non-redundant SNPs inside or 1,000 bp upstream and downstream, were considered (see details in Supplementary Material, Table S3). Our first attempt was to compute the female and male kinships involved in the *FM* model by considering only the 3,746 pre-selected SNPs. We named this



**FIGURE 2 |** Correlation of oil content adjusted phenotype of hybrids between environments of the MET.

model the *FM\_oil* model. Boxplots of *GCA*, *FM* and *FM\_oil* prediction accuracies for 100 random test sets of T0 and T1 hybrids are presented in **Figure 3**. *FM\_oil* predictions were more accurate than *GCA* predictions, but the *FM* model was still the best, not only in mean, but it also showed a less variability in test set accuracies (**Table 2**).

By limiting the computation of parent relatedness to pre-selected oil SNPs, the *FM\_oil* model is simplified and assumes that all important causal genes explaining oil content variability are already included in the considered metabolic pathway. To avoid an over-simplified assumption, we performed a multi-kernel model with two kinships for each parental effect, generated using pre-selected oil SNPs for one group and all remaining SNPs for the other group. This model assumes a different variance for each group of SNPs and each parental effect, leading to a more flexible model. With an average predictive ability in the MET of 0.628, this multi-kernel model slightly improved the average

predictive ability of 0.612 for the *FM\_oil* model but did not reach that of 0.635 for the *FM* model (**Table 3**). The *FM* model assumes that no interaction occurs between SNPs, neglecting the epistasis phenomena. We performed a multi-kernel BLUP model considering both the femalexfemale and the malexmale parts of the epistasis as a generalization of the additiveadditive epistasis modeling proposed by Su et al. (2012). With an average predictive ability of 0.623, this model did not improve the *FM* BLUPs (**Table 3**).

Having access to genetic value prediction of all hybrids in each environment of the MET with a high level of accuracy (0.783 in average for the *FM* BLUPs) facilitates selection of the best hybrids on average and affords an opportunity to examine their stability across environments. The distribution of hybrid mean predicted performance on the MET is shown in **Figure 4**. The least productive hybrid was predicted with a mean performance of 38.8%, and the most productive hybrid was predicted with a

**TABLE 1** | Number of observed hybrids (n.obs), mean oil content (in %), variance components and parts of variance (in %) [female, male, interaction female × male (inter.) and residual (resi.)] estimated using REML in GCA, FM, and FMI models, per environment (Env.).

Model	Env.	13EX01	13EX03	13EX04	13EX05	14EX04	14RV01	15EX05	15EX06	15EX07
	n.obs	272	423	407	411	418	411	459	461	458
	mean	44.81	45.26	48.49	41.43	43.46	40.72	47.40	45.45	49.33
<b>VARIANCE COMPONENTS</b>										
GCA	Female	2.74	1.68	1.44	2.00	4.19	3.70	2.60	1.58	6.12
	Male	1.44	0.67	0.75	0.61	1.66	1.74	0.88	0.73	2.91
	Resi.	2.09	1.03	1.03	1.40	1.79	1.70	1.40	1.40	4.00
FM	Female	2.13	1.21	1.07	1.55	3.22	2.90	1.84	1.22	4.88
	Male	1.50	0.77	0.78	0.69	1.79	1.83	1.09	0.83	3.23
	Resi.	2.08	1.03	1.03	1.39	1.79	1.70	1.40	1.40	4.00
FMI	Female	2.13	1.21	1.07	1.51	3.22	2.91	1.84	1.22	4.88
	Male	1.50	0.78	0.78	0.72	1.80	1.84	1.09	0.83	3.23
	Inter.	0.00	0.10	0.00	0.72	0.14	0.47	0.00	0.00	0.00
	Resi.	2.08	0.93	1.03	0.65	1.64	1.21	1.40	1.40	4.00
<b>PARTS OF VARIANCE</b>										
GCA	Female	0.44	0.50	0.45	0.50	0.55	0.52	0.53	0.43	0.47
	Male	0.23	0.20	0.23	0.15	0.22	0.24	0.18	0.20	0.22
FM	Female	0.37	0.40	0.37	0.43	0.47	0.45	0.42	0.35	0.40
	Male	0.26	0.26	0.27	0.19	0.26	0.28	0.25	0.24	0.27
FMI	Female	0.37	0.40	0.37	0.42	0.47	0.45	0.42	0.35	0.40
	Male	0.26	0.26	0.27	0.20	0.26	0.29	0.25	0.24	0.27
	Inter.	0.00	0.03	0.00	0.20	0.02	0.07	0.00	0.00	0.00

**TABLE 2** | Predictive ability of hybrid performances per environment (Env.) and average on the MET with GCA, FM, and FMI model BLUPs as the mean over the same 100 test sets (TS) using two sampling processes.

Env.	TS: any hybrids			TS: T1 or T0 hybrids		
	GCA	FM	FMI	GCA	FM	FMI
13EX01	0.756	0.762	0.761	0.580	0.653	0.651
13EX03	0.780	0.780	0.776	0.588	0.652	0.648
13EX04	0.767	0.768	0.766	0.572	0.641	0.639
13EX05	0.739	0.739	0.744	0.537	0.599	0.604
14EX04	0.835	0.836	0.835	0.589	0.665	0.665
14RV01	0.824	0.825	0.827	0.587	0.658	0.659
15EX05	0.800	0.800	0.799	0.596	0.634	0.633
15EX06	0.738	0.738	0.736	0.533	0.580	0.578
15EX07	0.796	0.797	0.796	0.590	0.635	0.633
Average	0.782	0.783	0.782	0.575	0.635	0.634

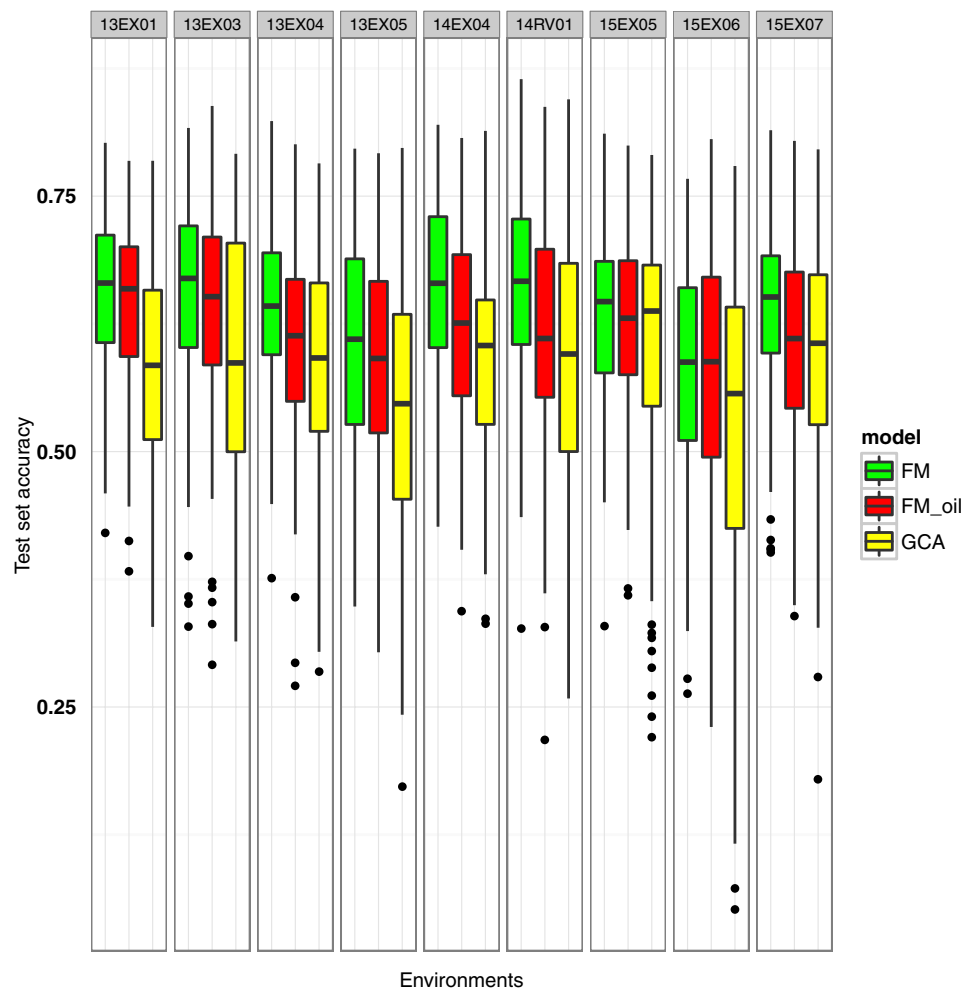
T1 and T0 hybrids are hybrids for which one or both parents have no observed descendant in the training set.

mean performance of 48.8% of seed oil content. Approximately 10% of hybrids had a predicted mean performance greater than 47%. To examine the stability of hybrids across environments, we computed the Wricke's ecovalence stability index (Wricke, 1962) using the hybrid predicted performances. This stability index measures how the hybrid predicted performances vary from an environment to another. **Figure 5** is a heat map representation

of the mean predicted performance of hybrid with hybrids having a Wricke's ecovalence stability index (Wricke, 1962) less than 5, highlighted as a blank square. Hybrids predicted as producing a high oil content on average are generally not stable, only a single hybrid is predicted as stable in the right top corner of the heat map, its predicted mean performance and its Wricke's ecovalence were 48.3% and 4.83, respectively.

## 4. DISCUSSION

As a starting point to evaluate the benefits of GS, in the present study, we compared the accuracy of hybrid performance predictions for seed oil content, a highly heritable breeding trait in sunflower. The simplest GCA-based model was compared with different genomic multi-kernel linear mixed models. We showed that the GCA-based model, ignoring parental pedigrees, is globally as accurate as more complex models to predict the oil content of unobserved sunflower hybrids in an incomplete factorial design where 36 maintainer lines (CMS form) were crossed with 36 restorer lines. This result reflects three main factors: (i) the accurate knowledge of the parental GCAs estimated in each environment from an average of at least 7 hybrid combinations, (ii) the strong additive effect of oil content in the MET, and (iii) the genetic distance between parents selected as unrelated to provide a good representation of the core collection studied in Cadic et al. (2013). However, there is an advantage to GS prediction (10% increase in accuracy) for hybrids of untested parents.



**FIGURE 3** | Boxplot of test set accuracy per environment for GCA, FM and FM\_oil (FM modeling using knowledge of oil metabolic network) BLUPs. The model BLUPs were computed on the same 100 test sets. The test sets contained only T1 or T0 hybrids with untested parents.

Hybrids from untested parents are more distant from those observed than random missing combinations in the incomplete factorial design. Indeed, Hayes et al. (2009); Clark et al. (2012) indicated that it is more challenging to predict the values of unrelated genotypes and suggested that, in such situations, genomic predictions are more accurate than classical pedigree predictions.

GCA-based or GS predictions of missing hybrid performances is accurate in the MET (predictive ability of 0.78 on average), but with much less accuracy compared with Reif et al. (2013) (predictive ability of 0.97 by a leave-one-out hybrid cross validation). These two values are, in fact, not comparable as Reif et al. (2013) predicted the hybrid mean performances on the MET, whereas we predicted intra-environment hybrid performances. It is simpler to predict the mean performance compared with intra-environment performance, as the latter depends on the genetic by environment interaction, and therefore

is more variable and less heritable. The lower a trait is heritable, the lower the GS predictive ability. However, intra-environment predictions are essential to access hybrid stability. Heffner et al. (2009) highlighted that GS is an important tool to address the challenge of genetic by environment interaction. Moreover, the lower the trait is heritable, the greater the prediction improvement expected from GS.

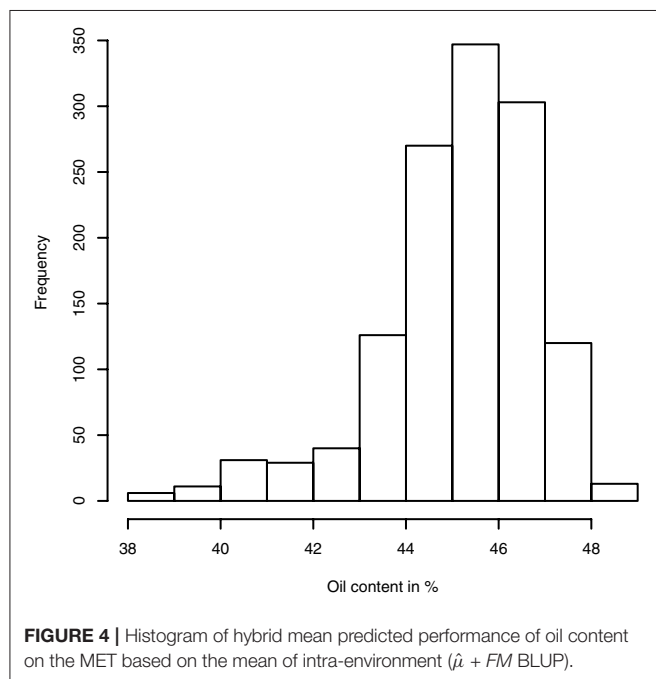
FM and FMI models differ by an interaction term that models parental allelic interaction or dominance. These models generally showed similar levels of accuracy in predicting untested hybrids or hybrids between untested parents. When their accuracies differed in one environment, a significant variance of the parental allelic interaction was observed, suggesting that only factors with sufficient variability could increase the accuracy of models including dominance compared to additive models. Moreover, a systematic small decrease of the FMI model accuracy compared to the FM model was observed when the variance component of



**TABLE 3 |** Predictive ability of hybrid performances (mean over of 100 test sets and its variance) per environment (Env.) and in average on the MET with GCA, *FM*, *FM\_oil*, *mk\_oil* (multi-kernel *FM* model with two groups of SNPs) and *mk\_epi* (multi-kernel model with female, male, female  $\times$  female epistasis and male  $\times$  male epistasis kernels) model BLUPs.

	Mean					Variance				
	GCA	FM	FM_oil	mk_oil	mk_epi	GCA	FM	FM_oil	mk_oil	mk_epi
13EX01	0.580	0.653	0.646	0.641	0.650	$8.58 \cdot 10^{-3}$	$6.46 \cdot 10^{-3}$	$6.53 \cdot 10^{-3}$	$6.69 \cdot 10^{-3}$	$6.34 \cdot 10^{-3}$
13EX03	0.588	0.653	0.642	0.645	0.645	$1.55 \cdot 10^{-2}$	$1.01 \cdot 10^{-2}$	$1.17 \cdot 10^{-2}$	$1.02 \cdot 10^{-2}$	$9.91 \cdot 10^{-3}$
13EX04	0.572	0.641	0.601	0.640	0.628	$1.42 \cdot 10^{-2}$	$6.87 \cdot 10^{-3}$	$9.75 \cdot 10^{-3}$	$6.97 \cdot 10^{-3}$	$8.07 \cdot 10^{-3}$
13EX05	0.537	0.599	0.579	0.594	0.575	$1.71 \cdot 10^{-2}$	$1.19 \cdot 10^{-2}$	$1.25 \cdot 10^{-2}$	$1.17 \cdot 10^{-2}$	$1.43 \cdot 10^{-2}$
14EX04	0.589	0.666	0.619	0.666	0.662	$1.01 \cdot 10^{-2}$	$6.53 \cdot 10^{-3}$	$9.81 \cdot 10^{-3}$	$6.52 \cdot 10^{-3}$	$6.41 \cdot 10^{-3}$
14RV01	0.587	0.659	0.616	0.656	0.632	$1.61 \cdot 10^{-2}$	$1.04 \cdot 10^{-2}$	$1.34 \cdot 10^{-2}$	$1.08 \cdot 10^{-2}$	$1.09 \cdot 10^{-2}$
15EX05	0.596	0.634	0.622	0.623	0.637	$1.77 \cdot 10^{-2}$	$7.13 \cdot 10^{-3}$	$7.09 \cdot 10^{-3}$	$7.51 \cdot 10^{-3}$	$7.22 \cdot 10^{-3}$
15EX06	0.533	0.580	0.574	0.564	0.555	$2.61 \cdot 10^{-2}$	$1.20 \cdot 10^{-2}$	$1.48 \cdot 10^{-2}$	$1.38 \cdot 10^{-2}$	$1.88 \cdot 10^{-2}$
15EX07	0.590	0.635	0.610	0.625	0.625	$1.35 \cdot 10^{-2}$	$8.34 \cdot 10^{-3}$	$9.70 \cdot 10^{-3}$	$9.05 \cdot 10^{-3}$	$9.11 \cdot 10^{-3}$
Average	0.575	0.635	0.612	0.628	0.623	$1.54 \cdot 10^{-2}$	$8.86 \cdot 10^{-3}$	$1.06 \cdot 10^{-2}$	$9.25 \cdot 10^{-3}$	$1.01 \cdot 10^{-2}$

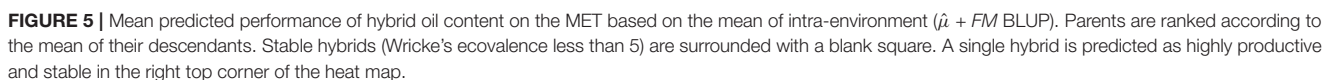
The model BLUPs were computed on the same 100 test sets. The test sets contained only T1 or T0 hybrids with parents never observed by their descendants.



this interaction was estimated as zero. The benefit of the inclusion of non-additive effects in an additive GS model is still subject to debate. Using simulations, Toro and Varona (2010) observed that inclusion of the dominance effect never decreased genetic gain in first generation selection in animal breeding programs whatever the ratio between additive and non-additive parts. Similarly, in a pig population, although Heidaritabar et al. (2016) showed no impact of dominance modeling in GS model accuracy for traits with a small ratio between additive and dominance, these authors did not observe any drawback. In contrast to these studies, the results of the present study are consistent with those of Reif et al. (2013) who observed small decreases in accuracy when

dominance effects were included, depending on the traits and the intra or inter [B/R] group crosses. The significance of this decrease is important, but the lack of independence between the sampled test sets made it impossible to obtain a correct estimate of the variance of the mean accuracy necessary to build a test of significance. Neither the division by the square root of the number of sampled test sets nor the bootstrapped variance is correct with dependent results. Both methods provide a too small variance of the mean accuracy and thus conclude significance where there is no significance. Altogether, it might be assumed that the narrow-sense heritability of the trait plays an important role regarding the introduction of dominance effects in prediction models. As seed oil content is highly heritable (both narrow sense and broad sense), it is difficult to make a general conclusion. However, the high predictability of either GCA or *FM* models can explain why, without dense molecular scan and GS model, breeders have rapidly succeeded in transforming sunflower into a high valuable oil crop in the first half of the 20th century.

The use of biological information to enhance the accuracy of GS predictions was studied using simulations published by Pérez-Enciso et al. (2015). These authors showed that imprecision on QTL locations and non-exhaustive knowledge of all causal QTLs result in the rapid decline of the nearly perfect accuracy obtained when causal QTLs are all perfectly known. However, even with imperfect knowledge of 50% of genes, including causal QTLs, these authors showed a better accuracy compared to GS predictions with all SNPs. This encouraging result shows the interest of including functional knowledge in GS models. We tested the incorporation of biological knowledge on the oil metabolic network but we did not observe any improvement of the *FM* model predictions despite an improvement of the GCA model predictions. This finding is not surprising and is consistent with results of Spindel et al. (2016), who did not observe any improvement in accuracy with inclusion of historical GWAS results. Nevertheless,



We showed that in such a design, classical GCA and GS predictions of hybrid performance had equal accuracy, as the GCA of each parent is well estimated for oil content, a highly

heritable and mostly additive trait. However, predictions of hybrid performances of at least one untested parent are more accurate using GS models, showing that GS can accelerate the genetic gain by enabling better selection in hybrid panel of poorly known parent lines.

## AUTHOR CONTRIBUTIONS

BM and NL designed the study. BM performed the data analyses. FB, PP participated in the data analyses. NB, MB, GM, and EB provided genetic resources and phenotypic data. SC, LL, SM, and NP provided genomic data. LC provided metabolic data. PV designed the hybrid factorial design. BM, FB, LC, FV, PV, and NL drafted the manuscript. All authors have read and approved the manuscript.

## REFERENCES

- Bachlava, E., Tang, S., Pizarro, G., Schuppert, G. F., Brunick, R. K., Draeger, D., et al. (2010). Pleiotropy of the branching locus (b) masks linked and unlinked quantitative trait loci affecting seed traits in sunflower. *Theor. Appl. Genet.* 120, 829–842. doi: 10.1007/s00122-009-1212-1
- Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., et al. (2017). The sunflower genome provides insights into oil metabolism, flowering and asterid evolution. *Nature* 546, 148–152. doi: 10.1038/nature22380
- Bernardo, R. (1996). Best linear unbiased prediction of the performance of crosses between untested maize inbreds. *Crop Sci.* 36, 872–876. doi: 10.2135/cropsci1996.0011183X003600040009x
- Bernardo, R. (2014). Genomewide selection when major genes are known. *Crop Sci.* 54, 68–75. doi: 10.2135/cropsci2013.05.0315
- Bert, P.-F., Jouan, I., de Labrouhe, T. D., Serre, F., Nicolas, P., and Vear, F. (2002). Comparative genetic analysis of quantitative traits in sunflower (*Helianthus annuus* L.) 1. qtl involved in resistance to sclerotinia sclerotiorum and diaporthe helianthi. *Theor. Appl. Genet.* 105, 985–993. doi: 10.1007/s00122-002-1004-3
- Bonnaïfous, F., Fievet, G., Blanchet, N., Boniface, M.-C., Carrère, S., Gouzy, J., et al. (2017). Comparison of GWAS models to identify non-additive genetic control of flowering time in sunflower hybrids. *BioRxiv*. doi: 10.1101/188235
- Bouvet, J.-M., Makouanzi, G., Cros, D., and Vigneron, P. (2016). Modeling additive and non-additive effects in a hybrid population using genome-wide genotyping: prediction accuracy implications. *Heredity* 116, 146–157. doi: 10.1038/hdy.2015.78
- Browning, B. L., and Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* 84, 210–223. doi: 10.1016/j.ajhg.2009.01.005
- Butler, D., Cullis, B., Gilmour, A., and Gogel, B. (2007). *AS-Reml-R reference manual*. Brisbane: Department of Primary Industries and Fisheries. Available online at: [www.vsn-intl.com/products/asreml/](http://www.vsn-intl.com/products/asreml/)
- Cadic, E., Coque, M., Vear, F., Grezes-Basset, B., Pauquet, J., Piquemal, J., et al. (2013). Combined linkage and association mapping of flowering time in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 126, 1337–1356. doi: 10.1007/s00122-013-2056-2
- Chen, X., and Ishwaran, H. (2012). Random forests for genomic data analysis. *Genomics* 99, 323–329. doi: 10.1016/j.ygeno.2012.04.003
- Clark, S. A., Hickey, J. M., Daetwyler, H. D., and van der Werf, J. H. (2012). The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet. Select. Evol.* 44:4. doi: 10.1186/1297-9686-44-4
- de los Campos, G., Gianola, D., and Rosa, G. J. (2009). Reproducing kernel hilbert spaces regression: a general framework for genetic evaluation. *J. Anim. Sci.* 87, 1883–1887. doi: 10.2527/jas.2008-1259

## ACKNOWLEDGMENTS

This work was part of the SUNRISE project of the French National Research Agency (ANR-11-BTBR-0005, 2012–2019). We thank our partners: Biogemma, Caussades semences, Maisadour semences, RAGT 2n, Soltis, Syngenta, and Terres Inovia for providing experimental data. The authors thank Thierry André who suggested the comparison between GCA- and GS-based predictions.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.01633/full#supplementary-material>

- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024
- Fick, G. (1975). Heritability of oil content in sunflowers. *Crop Sci.* 15, 77–78. doi: 10.2135/cropsci1975.0011183X001500010022x
- Gianola, D., Fernando, R. L., and Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173, 1761–1776. doi: 10.1534/genetics.105.049510
- Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257. doi: 10.1007/s10709-008-9308-0
- Haws, D. C., Rish, I., Teyssedre, S., He, D., Lozano, A. C., Kambadur, P., et al. (2015). Variable-selection emerges on top in empirical comparison of whole-genome complex-trait prediction methods. *PLoS ONE* 10:e0138903. doi: 10.1371/journal.pone.0138903
- Hayes, B. J., Visscher, P. M., and Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* 91, 47–60. doi: 10.1017/S0016672308009981
- Heffner, E. L., Sorrells, M. E., and Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Sci.* 49, 1–12. doi: 10.2135/cropsci2008.08.0512
- Heidaritabar, M., Wolc, A., Arango, J., Zeng, J., Settari, P., Fulton, J. E., et al. (2016). Impact of fitting dominance and additive effects on accuracy of genomic prediction of breeding values in layers. *J. Anim. Breed. Genet.* 133, 334–346. doi: 10.1111/jbg.12225
- Heslot, N., Yang, H. P., Sorrells, M. E., and Jannink, J. L. (2012). Genomic selection in plant breeding: a comparison of models. *Crop Sci.* 52, 146–160. doi: 10.2135/cropsci2011.06.0297
- Jiang, Y., and Reif, J. C. (2015). Modeling epistasis in genomic selection. *Genetics* 201, 759–768. doi: 10.1534/genetics.115.177907
- Kärkkäinen, H. P., and Sillanpää, M. J. (2012). Back to basics for bayesian model building in genomic selection. *Genetics* 191, 969–987. doi: 10.1534/genetics.112.139014
- Leon, A., Lee, M., Rufener, G., Berry, S., and Mowers, R. (1995). Use of rflp markers for genetic linkage analysis of oil percentage in sunflower seed. *Crop Sci.* 35, 558–564. doi: 10.2135/cropsci1995.0011183X003500020046x
- Li, Z., and Sillanpää, M. J. (2012). Overview of lasso-related penalized regression methods for quantitative trait mapping and genomic selection. *Theor. Appl. Genet.* 125, 419–435. doi: 10.1007/s00122-012-1892-9
- Merah, O., Langlade, N., Alignan, M., Roche, J., Pouilly, N., Lippi, Y., et al. (2012). Genetic analysis of phytosterol content in sunflower seeds. *Theor. Appl. Genet.* 125, 1589–1601. doi: 10.1007/s00122-012-1937-0
- Mestries, E., Gentzbittel, L., Tourvielle de Labrouhe, D., Nicolas, P., and Vear, F. (1998). Analyses of quantitative trait loci associated with resistance to shape sclerotinia sclerotiorum in sunflowers (shape *Helianthus annuus* L.) using molecular markers. *Mol. Breed.* 4, 215–226. doi: 10.1023/A:1009694626992
- Meuwissen, T., Hayes, B., and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

- Pérez-Enciso, M., Rincón, J. C., and Legarra, A. (2015). Sequence-vs. chip-assisted genomic selection: accurate biological information is advised. *Genet. Select. Evol.* 47:43. doi: 10.1186/s12711-015-0117-5
- Reif, J. C., Zhao, Y., Würschum, T., Gowda, M., and Hahn, V. (2013). Genomic prediction of sunflower hybrid performance. *Plant Breed.* 132, 107–114. doi: 10.1111/pbr.12007
- Sarup, P., Jensen, J., Ostensen, T., Henryon, M., and Sørensen, P. (2016). Increased prediction accuracy using a genomic feature model including prior information on quantitative trait locus regions in purebred danish duroc pigs. *BMC Genet.* 17:11. doi: 10.1186/s12863-015-0322-9
- Schrag, T., Melchinger, A., Sørensen, A., and Frisch, M. (2006). Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using aflp markers associated with qtl. *Theor. Appl. Genet.* 113, 1037–1047. doi: 10.1007/s00122-006-0363-6
- Speed, D., and Balding, D. J. (2014). Multiblup: improved snp-based prediction for complex traits. *Genome Res.* 24, 1550–1557. doi: 10.1101/gr.169375.113
- Spindel, J., Begum, H., Akdemir, D., Collard, B., Redoña, E., Jannink, J., et al. (2016). Genome-wide prediction models that incorporate *de novo* GWAS are a powerful new tool for tropical rice improvement. *Heredity* 116, 395–408. doi: 10.1038/hdy.2015.113
- Su, G., Christensen, O. F., Ostensen, T., Henryon, M., and Lund, M. S. (2012). Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. *PLoS ONE* 7:e45293. doi: 10.1371/journal.pone.0045293
- Technow, F., Schrag, T. A., Schipprack, W., Bauer, E., Simianer, H., and Melchinger, A. E. (2014). Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics* 197, 1343–1355. doi: 10.1534/genetics.114.165860
- Toro, M. A., and Varona, L. (2010). A note on mate allocation for dominance handling in genomic selection. *Genet. Select. Evol.* 42:33. doi: 10.1186/1297-9686-42-33
- VanRaden, P. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Vear, F. (2010). “Classic genetics and breeding,” in *Genetics, Genomics and Breeding of Sunflower*, eds J. Hu, G. Seiler, and C. Kole (Lebanon: Science Publishers, Inc.) 51–77.
- Weissbrod, O., Geiger, D., and Rosset, S. (2016). Multikernel linear mixed models for complex phenotype prediction. *Genome Res.* 26, 969–979. doi: 10.1101/gr.201996.115
- Wolfe, M. D., Rabbi, I. Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., et al. (2016). Genome-wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. *Plant Genome* 9, 1–13. doi: 10.3835/plantgenome2015.11.0118
- Wricke, G. (1962). Über eine methode zur erfassung der ökologischen streubreite in feldversuchen. *Z. Pflanzenzüchtg* 47, 92–96.
- Zhang, Z., Ober, U., Erbe, M., Zhang, H., Gao, N., He, J., et al. (2014). Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. *PLoS ONE* 9:e93017. doi: 10.1371/journal.pone.0093017

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Mangin, Bonnafous, Blanchet, Boniface, Bret-Mestries, Carrère, Cottret, Legrand, Marage, Pegot-Espagnet, Munos, Pouilly, Vear, Vincourt and Langlade. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Identification and Functional Annotation of Genes Differentially Expressed in the Reproductive Tissues of the Olive Tree (*Olea europaea* L.) through the Generation of Subtractive Libraries

## OPEN ACCESS

### Edited by:

Mariela Torres,  
Instituto Nacional de Tecnología  
Agropecuaria (INTA), Argentina

### Reviewed by:

Luciana Baldoni,  
Istituto di Bioscienze e Biorisorse  
(IBBR), Italy  
Rosario Muleo,  
Università degli Studi della Tuscia, Italy

### \*Correspondence:

Juan D. Alche  
juandedios.alche@eez.csic.es

### † Present Address:

José A. Traverso,  
Department of Cell Biology, University  
of Granada, Granada, Spain  
Simon J. Hiscock,  
Oxford Botanic Garden and Harcourt  
Arboretum, Department of Plant  
Sciences, University of Oxford,  
Oxford, United Kingdom

‡ These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 28 April 2017

**Accepted:** 28 August 2017

**Published:** 13 September 2017

### Citation:

Zafra A, Carmona R, Traverso JA,  
Hancock JT, Goldman MHS,  
Claros MG, Hiscock SJ and Alche JD  
(2017) Identification and Functional  
Annotation of Genes Differentially  
Expressed in the Reproductive Tissues  
of the Olive Tree (*Olea europaea* L.)  
through the Generation of Subtractive  
Libraries. *Front. Plant Sci.* 8:1576.  
doi: 10.3389/fpls.2017.01576

Adoración Zafra<sup>1†</sup>, Rosario Carmona<sup>1†</sup>, José A. Traverso<sup>1†</sup>, John T. Hancock<sup>2</sup>,  
Maria H. S. Goldman<sup>3</sup>, M. Gonzalo Claros<sup>4</sup>, Simon J. Hiscock<sup>5†</sup> and Juan D. Alche<sup>1\*</sup>

<sup>1</sup> Plant Reproductive Biology Laboratory, Department of Biochemistry, Cellular and Molecular Biology of Plants, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, Granada, Spain, <sup>2</sup> Faculty of Health and Life Sciences, University of the West of England, Bristol, United Kingdom, <sup>3</sup> Departamento de Biología, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, São Paulo, Brazil, <sup>4</sup> Departamento de Biología Molecular y Bioquímica, Universidad de Málaga, Málaga, Spain, <sup>5</sup> School of Biological Sciences, University of Bristol, Bristol, United Kingdom

The olive tree is a crop of high socio-economical importance in the Mediterranean area. Sexual reproduction in this plant is an essential process, which determines the yield. Successful fertilization is mainly favored and sometimes needed of the presence of pollen grains from a different cultivar as the olive seizes a self-incompatibility system allegedly determined of the sporophytic type. The purpose of the present study was to identify key gene products involved in the function of olive pollen and pistil, in order to help elucidate the events and signaling processes, which happen during the courtship, pollen grain germination, and fertilization in olive. The use of subtractive SSH libraries constructed using, on the one hand one specific stage of the pistil development with germinating pollen grains, and on the other hand mature pollen grains may help to reveal the specific transcripts involved in the cited events. Such libraries have also been created by subtracting vegetative mRNAs (from leaves), in order to identify reproductive sequences only. A variety of transcripts have been identified in the mature pollen grains and in the pistil at the receptive stage. Among them, those related to defense, transport and oxidative metabolism are highlighted mainly in the pistil libraries where transcripts related to stress, and response to biotic and abiotic stimulus have a prominent position. Extensive lists containing information as regard to the specific transcripts determined for each stage and tissue are provided, as well as functional classifications of these gene products. Such lists were faced up to two recent datasets obtained in olive after transcriptomic and genomic approaches. The sequences and the differential expression level of the SSH-transcripts identified here, highly matched the transcriptomic information. Moreover, the unique presence of a representative number of these transcripts has been validated by means of qPCR approaches. The construction

of SSH libraries using pistil and pollen, considering the high interaction between male-female counterparts, allowed the identification of transcripts with important roles in stigma physiology. The functions of many of the transcripts obtained are intimately related, and most of them are of pivotal importance in defense, pollen-stigma interaction and signaling.

**Keywords:** gynoeceum, leaf, olive, pollen, self-incompatibility, SSH, transcripts

## INTRODUCTION

The olive (*Olea europaea* L.) is an important crop in Mediterranean countries. The fruit is used for the production of olive oil. Olive oil yield, organoleptic properties, quality, fatty acid content and many other parameters are highly dependent on the procedures used for olive oil production, including which olive cultivars are used. Asexual propagation of this tree, achieved by different methods (Böhm, 2013), is the usual practice since its domestication. This practice results in very high heteroplasmy, as assessed by the accumulation of mutations in a non-coding sequence of the mitochondrial genome when vegetative propagation is maintained for a long period of time (García-Díaz et al., 2003). However, olive production relies on the successful achievement of sexual reproduction. This plant has been suggested to harbor a self-incompatibility system of the gametophytic type (Cuevas and Polito, 1997; Ateyyeh et al., 2000; Wu et al., 2002), as described for the Oleaceae family (Igic and Kohn, 2001). However, most recent and abundant literature on the issue demonstrates that the self-incompatibility in olive is sporophytic (Kusaba et al., 2001; Allen et al., 2010a; Breton et al., 2014, 2016, 2017; Farinelli et al., 2015; Saumitou-Laprade et al., 2017). The fine mechanisms governing this system are currently being deciphered, and are likely to explain the divergence of incompatibility mechanisms which can occur among members of the same family, as seen in *Arabidopsis* (Kusaba et al., 2001). The SI mechanism described in the olive involves the preferential presence of pollen grains from a different cultivar for successful fertilization (allogamy). The main worry of the growers is the yield, which is affected by the pollinisers–pollinator relationship (Breton and Bervillé, 2012). In the case of the olive, wind is the main factor affecting the yield, as the dispersion of the pollen in olive is mainly anemophilous.

The use of high-throughput analytic methods based in next-generation sequencing (NGS) is rapidly reaching the study of the olive tree. A number of recent studies have described the generation of several olive transcriptomes (reviewed by Muleo et al., 2016), which have been generated from different organs and adaptive responses, sometimes discriminating their varietal origin and built preferentially by pyrosequencing/Illumina sequencing and array technologies. Thus, transcriptomes have been generated to approach flower and fruit development (Alagna et al., 2009, 2016; Galla et al., 2009; Ozgenturk et al., 2010; Muñoz-Mérida et al., 2013; Carmona et al., 2015; Iraia et al., 2016), fruit abscission (GilAmado and Gomez-Jimenez, 2012; Parra et al., 2013), abiotic stress responses (Bazakos et al., 2015; Guerra et al., 2015; Leyva-Pérez et al., 2015), miRNA (Donaire et al., 2011; Yanik et al., 2013), plant architecture (González-Plaza

et al., 2016), and even comparative transcriptomics (Sarah et al., 2017). Genome sequence of the olive tree, corresponding to 95–99% of the estimated genome length was recently obtained and annotated by Cruz et al. (2016). Such annotation was assisted by RNAseq from different tissues and stages, and represents an important resource for future research on olive tree, as well as for breeding purposes.

The present study was based on the construction of several cDNA libraries that were subtracted using the SSH method. The aim was to study the reproductive biology of the olive and particularly to obtain clues regarding pollen and stigma physiology, including the presence of differentially expressed enzymes, allergens and other relevant gene products. For that purpose we used reproductive tissues (pollen and pistil) as well as vegetative tissues (leaf as the subtractive item). For each pair combination, the forward and reverse libraries were constructed.

## MATERIALS AND METHODS

### Plant Material

The different tissues were obtained from adult olive trees (*Olea europaea*, cv. Picual) growing at the Estación Experimental del Zaidín (Granada, Spain). Pistils were excised from the complete flower at the stage of development 4, dehiscent anthers, as defined by Zafra et al. (2010). These pistils normally include a relatively high number of mature (dehiscent) pollen, hydrated pollen grains, and even germinating pollen grains and pollen tubes, either over the stigma surface or through the transmitting tissues of the style or the ovary. The mature pollen grains were collected during the anthesis period using large paper bags by vigorously shaking the inflorescences. Pollen was sequentially sieved through a mesh in order to separate the grains from the debris. Young leaves were also selected. In all the three cases the different tissues were quickly frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . Samples from three consecutive years were used for the present analysis.

### Construction of the Suppression Subtractive Hybridization (SSH) Libraries

Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen) from samples of the different years, and the contaminating genomic DNA was removed by DNAase I (Qiagen) treatment followed by a clean-up with the RNeasy MinElute Cleanup kit (Qiagen). cDNA was then synthesized from pistil, leaf, and mature pollen total RNA using the SMART PCR cDNA Synthesis kit (Clontech). The subtracted libraries were constructed with the PCR-Select cDNA Subtraction Kit (Clontech). A total of 6 libraries were constructed: 1. Pistil subtracted with pollen

[P(Po)]; 2. Pollen subtracted with pistil [Po(P)]; 3. Pistil subtracted with leaf [P(L)]; 4. Leaf subtracted with pistil [L(P)]; 5. Pollen subtracted with leaf [Po(L)]; 6. Leaf subtracted with pollen [L(Po)], according to the manufacturer's instructions. Two rounds of PCR amplifications were also performed according to the manufacturer's protocol in order to enrich differentially regulated genes, by using the PCR Primer 1 and the Nested PCR primer 1 and 2R as indicated in the manufacturer's instructions and provided by the kit.

## Cloning and Differential Screening

The secondary PCR products were cloned into the T/A cloning vector pGEM-T Easy (Promega) according to the manufacturer's instructions and transformed into DH5 $\alpha$  *E. coli* cells. The colonies containing inserts were picked and used as template for PCR. The primers used in this case were SP6 and T7. Sanger sequencing of PCR products was carried out at the Estación Experimental del Zaidín DNA Sequencing Service (CSIC, Granada, Spain), the Laboratório de Biologia Molecular de Plantas (Universidade de São Paulo, Brazil), and other commercially available facilities. With the aim to perform the differential screenings, a number of membrane replicates were prepared, each one containing 1  $\mu$ l of the PCR product per dot, which were spotted onto nylon membranes and fixed with a brief wash in 2x SSC followed by baking at 120°C during 30 min. The membrane replicates were probed with the forward-subtracted probe, the reverse-subtracted probe, the unsubtracted tester probe, and the unsubtracted driver probe in each case. The labeled probes were generated from the secondary PCR products described in the (SSH) library construction section, which were purified using the MinElute PCR Purification Kit (Qiagen). DIG-DNA labeling, determination of labeling efficiency, hybridization, and immunological detection were carried out as described in the DIG High Prime DNA Labeling and Detection Starter Kit II (Roche) instruction manual. The membranes were revealed with the CSPD ready-to-use chemiluminescent substrate (Roche), exposed to ChemiDoc XRS system (Bio-Rad). Images were gathered with a supersensitive 12-bit CCD after 30 min of exposition (Supplementary Figure 1). All hybridizations and image captures were repeated twice.

## Sequencing and Data Analysis

Transcripts were compared (using BLASTn) against non-redundant protein databases at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>; Altschul et al., 1990) ( $E$ -value  $10^{-4}$ ) and also against the non-redundant proteins unique transcripts Olea EST database (Alagna et al., 2009). The Blast2Go (<http://www.blast2go.com/b2ghome>) software was used (Conesa et al., 2005) to carry out the statistical analysis of GO (Gene Ontology) terms. For the analysis of the contigs obtained from the singletons, the Codon Code Aligner software was used (<http://www.codoncode.com/aligner/>).

The Venn diagrams were constructed using the transcripts of the 6 SSH libraries analyzed. Three groups were considered, corresponding to pollen, pistil and leaf transcripts. The VENNY software (<http://bioinfo.gp.cnb.csic.es/tools/venny/>) was used for

this purpose. With the aim to compare the output results after the comparison against the NCBI and Olea EST databases, two diagrams were performed separately.

To retrieve the putative Arabidopsis homologs of the olive clones obtained, the sequences from the transcripts of two selected libraries [Po(P) and P(Po)] were compared against the Arabidopsis Information Resource (TAIR) webpage (<http://www.arabidopsis.org/Blast/>). A BLASTn against the TAIR10 Transcripts (–introns, +UTRs) (DNA) was carried out. The matrix weight was Blosum45, the nucleic mismatch –3, gapped alignments ON. The output results were used as input data in the plant biology resource from Genevestigator (<https://www.genevestigator.com/gv/plant.jsp>). The anatomy tool from this webpage was used to construct the heatmap representing the level of expression of the transcripts in olive corresponding to defense, oxidative metabolism and transport. These categories were chosen attending to two criteria: firstly a highly represented number of transcripts and also due to the implication in the reproductive process.

## qPCR Validation of Subtractive Transcripts

Total RNA from pollen, pistil and leaf from olive cv. Picual were extracted using the RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions from samples obtained after three consecutive years as described above. Two  $\mu$ g of total RNA was reversed transcribed using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems, Thermo Scientific). Nine independent reverse transcriptase reactions were carried out. cDNA was stored at –20°C until use for qPCR analyses. Primers were designed for transcripts putatively specific of each tissue as determined by the differential screening. The Primer BLAST software (NCBI) was used for primer design with modifications in the default settings [PCR product size: 70–150 bp; Primer Tm 58–62°C; Organism: green plants; Primer size: 18–23 bp; Primer GC content: 30–80%; Hyb Oligo Size: 18–30 bp; Hyb Oligo Tm: 68–72°C; Hyb Oligo GC content: 30–80%]. Putative amplicons were blasted against Reproline database aimed to confirm the specificity. For details of the primer sequence and expected sizes of the amplicons, see Supplementary Table 1. LightCycler FastStart DNA Master Syber Green (Roche) was used in a Light Cyclor 480 Instrument II (Roche Diagnostic, Mannheim, Germany) in a 20  $\mu$ l reaction volume. Samples were run in duplicate for each experiment. Expression levels of target genes were normalized as regard to the expression level of two housekeeping genes (Zinc-finger and katanin p60) and their relative expression levels were calculated with the  $\Delta\Delta C_t$  rule (Taylor et al., 2016). Housekeeping genes were gathered based on an automatic screening of the Reproline database and were previously tested in pollen, pistil and leaf tissues (Carmona et al., 2016, 2017). Housekeeping sequences and amplicon sizes are detailed in Supplementary Table 1.

## Comparative Transcriptomics Pre-processing of Subtractive Sequences

All clones sequenced were pre-processed using SeqTrimNext (Falgueras et al., 2010) as described in Carmona et al. (2015)

to remove linkers, adaptors, vector fragments and contaminated sequences among others, while keeping the longest informative part of the sequence. Sequences below 100 bp were discarded.

### Annotation and Correspondence with Olive Transcriptomes

Useful sequences were annotated using Full-LengtherNext (Seoane et al., in preparation). Additionally, the correspondence of the subtractive sequences with the transcriptomes reported in ReprOlive database (<http://reprolive.eez.csic.es>; Carmona et al., 2015), as well as with the transcripts deduced from the first olive tree genome draft (89,982 transcripts) (Cruz et al., 2016) were determined. This correspondence was estimated by comparing (using BLASTn,  $E$ -value  $10^{-6}$ ) the subtractive sequences against the mentioned transcriptomes. With regard to ReprOlive transcriptomes, sequences from pollen libraries [Po(L) and Po(P)] were compared against the pollen transcriptome (27,823 transcripts), sequences from pistil libraries [P(Po) and P(L)] were compared against the pistil transcriptome (60,400 transcripts), whereas sequences from leaf libraries [L(Po) and L(P)] were compared against the vegetative transcriptome (38,919 transcripts).

### Presence/Absence of Subtractive Sequences in Different Stages/Tissues

In order to estimate the validity of the subtractive SSH libraries constructed, an *in silico* approximation was performed, taking advantage of the availability of Roche/454 reads used in ReprOlive (Carmona et al., 2015), which belong to seven different tissues and/or developmental stages: leaf, mature pollen, germinated pollen at two different times after hydration (1 and 5 h) and pistil at developmental stages 2, 3, and 4 as defined by Zafra et al. (2010). Each one of the subtractive sequences was compared by BLASTn ( $e$ -value  $1e^{-6}$ ) against the reads of each of the seven tissues/stages separately. In the event of at least one read matching with the subtractive sequence (considering a 90% or greater of identity and an alignment of 85% or greater of the subtractive sequence length), the subtractive sequence was considered to be present in this particular tissue/stage (and reported as “YES” in the corresponding output information).

### Availability of Data and Materials

The datasets supporting the conclusion of this article are available in the European Nucleotide Archive (<http://www.ebi.ac.uk/ena/data/view/PRJEB13716>) in the fastq.gz format with the title “SSH libraries of the olive tree (*Olea europaea* L.) reproductive tissues.” Fastq identifiers in Supplementary Tables 2–7 allow location of every single sequence in its corresponding fastq file deposited at ENA.

## RESULTS AND DISCUSSION

Six libraries were generated by using the combination of tester/driver tissues indicated in **Table 1**. Special attention was given to the P(Po), P(L), and Po(L) libraries. **Table 1** includes information as regarded to the number of clones identified and sequenced from each one of the six SSH libraries generated.

The P(Po) library provided information about those transcripts that are expressed during the pollen tube germination in comparison with the mature pollen grains, within the context of the whole pistil as in this stage, the pistil is full of germinating pollen grains. This library also offered information of the pistil transcripts. The P(L) library reveals the presence of transcripts in a tissue which is a distinct form from the leaf, being in addition a reproductive tissue. Lastly, the Po(L) shows the transcripts of a reproductive dormant tissue (pollen) from which transcripts from vegetative tissue have been subtracted.

A total of 1,344 clones were sequenced. From those, 790 resulted in ESTs and 171 in contigs. The mean length of the contigs compared to the ESTs showed an increase of 25% for Po(P), 27% for P(Po), 15% for P(L), 22% for L(P), 8% for Po(L), and 7% for L(Po). The redundancy levels were relatively low, ranging between 3.17 and 2.15%. BLAST analysis was carried out by using two alternative databases. The percentage of BLAST hits averaged 70% when the alignment was carried out with the NCBI database (<http://www.ncbi.nlm.nih.gov>; Altschul et al., 1990) ( $e$ -value  $e^{-4}$ ), and averaged 95% when the alignment was made against the OLEA EST database (Alagna et al., 2009) ( $e$ -values  $e^{-1}$ ).

In order to assess the subtractive efficiency of the libraries, PCR-amplified samples of the DNA inserts of each clone were transferred to membranes and subjected to multiple hybridizations with: (a) unsubtracted tester, (b) unsubtracted driver, (c) forward-subtracted, and (d) reverse-subtracted probes. An example of the procedure is displayed in Supplementary Figure 1.

The criteria followed in order to define clones as regard to their tissue-specificity were as follows: clones hybridizing exclusively with the forward-subtracted probe were considered to be differentially expressed. The clones that hybridize to the forward-subtracted probe and the unsubtracted tester probe also correspond to differentially expressed genes with a 95% of probability. Clones that hybridize to all the four probes correspond to non-differentially expressed clones. The results from the screening (Supplementary Tables 2–7) revealed that 33.3% of the clones were differentially expressed in the Po(P) library, 47.9% in the P(Po) library, 29.1% in the P(L) library, 41.6% in the L(P) library, 73% in the Po(L), and 68.2% in the L(Po).

The use of subtractive libraries allowed us to obtain a relatively large number of tissue-specific transcripts. The number of such putative tissue-specific transcripts was highly dependent on the database searched by means of BLAST. Thus, the number of specific reproductive sequences (from pollen and pistil) was larger after using the OLEA EST db than after using NCBI db. The opposite tendency occurred with vegetative (leaf) transcripts (**Figure 1**).

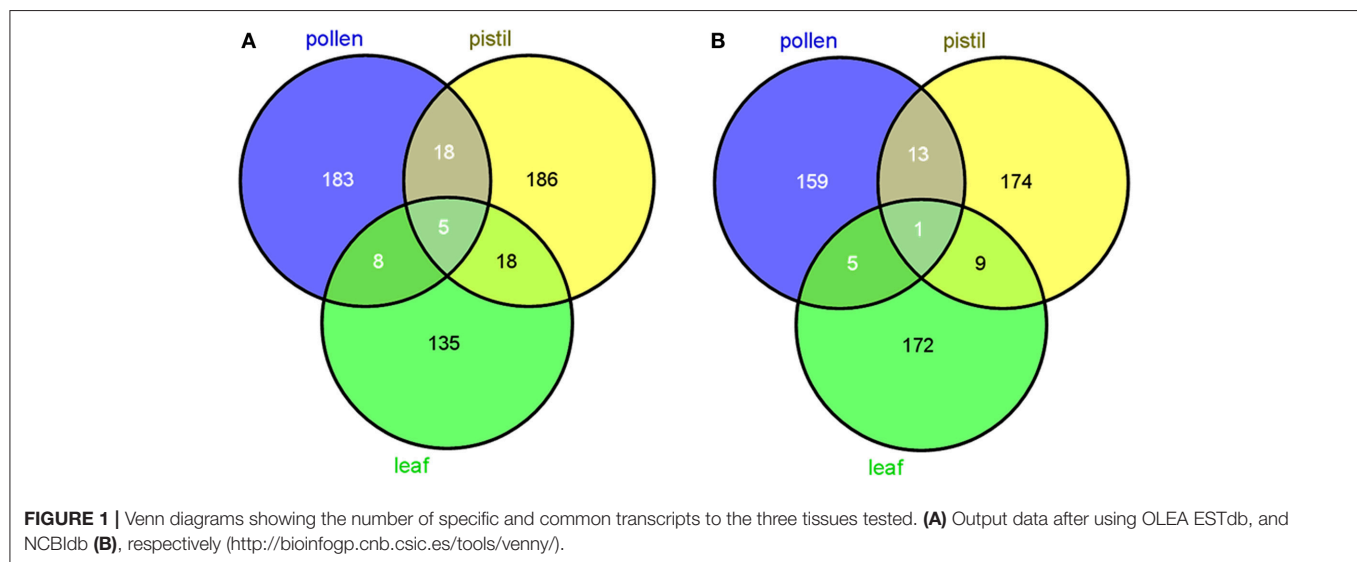
In order to validate the results obtained after the construction and screening of the SSH libraries, a total of 12 genes were selected for further assessment of their expression profiles by means of qPCR on the basis of representing a panel of different expression situations (**Table 2**). Four transcripts (Expansin 11 precursor, Pectinesterase inhibitor 21, Pectin methyl esterase 2.1 and Pathogenesis related protein-1) were selected according to



**TABLE 1** | SSH libraries constructed and descriptive parameters about the clones sequenced.

Name of the library	Tester tissue	Driver tissue	No. Sequenced clones	Total ESTs	Total contigs	Mean contig(bp)	Mean ESTs (bp)	% Redundancy	BLAST Hits(%) (NCBIdb)	BLAST Hits(%) (OLEA ESTdb)
Po(P)	Pollen	Pistil	288	127	46	588	440	3.17	76	90
P(Po)	Pistil	Pollen	288	200	30	589	431	3.10	57	97
P(L)	Pistil	Leaf	192	116	17	615	523	2.51	64	97
L(P)	Leaf	Pistil	192	60	34	576	447	2.15	74	98
Po(L)	Pollen	Leaf	192	129	28	543	500	2.40	70	92
L(Po)	Leaf	Pollen	192	158	16	565	525	2.52	77	95
Total			1344	790	171	579	478	2.64	70	95

P, Pistil; Po, pollen; L, leaf.



their putatively unique or prevailing presence in pollen. The qPCR assays performed indicated that these transcripts were preferentially present in pollen (**Figure 2A**). Same approaches carried out with the remaining transcripts selected on the basis of their putative uniqueness or majoritarian expression in pistil and leaf (**Table 2**) yielded the qPCR results shown in **Figures 2B,C**, respectively.

Quantitative-PCR amplifications of all 12 transcripts showed the predicted patterns of tissue and abundance distribution predicted on the basis of their presence in the generated libraries. These 12 gene products exemplify a broad panel of functions and molecular characteristics, ranging from allergens, cell wall modification or loosening, secondary metabolism, hydrolase and catalytic activity, photosynthesis, oxidative stress, light reception and defense to pathogens. The importance of some of these functions in the reproductive process is discussed next.

The concentration of high- and low-abundance sequences was equalized in the different libraries, which allowed us to identify low abundance transcripts but with the drawback of missing details about their real abundance. However, analysis of the distribution of Gene Ontology terms provided a first approach to define the implication of these transcripts (**Figures 3–5**). The two

pollen subtractive libraries [Po(L) and Po(P)] (**Figure 3**) included exclusive transcripts involved in biological processes related to the categories of pollination, responses to extracellular stimulus and post-embryonic development, and a high abundance of transcripts connected with cell differentiation, cell growth, and cellular component organization, in comparison to the rest of the libraries. The large presence of transcripts involved in cellular component organization may suggest that although the mature pollen grains have not yet started to germinate, many transcripts needed for pollen tube formation are already accumulated inside the pollen grains. In the same way, transcripts connected to transport activity also occupy a relevant role in both pollen libraries [Po(P) and Po(L)]. Therefore, the presence of such “preparative” transcripts in the mature pollen grain may be a determinant for the correct development of the pollen tube once the pollen grain arrives on the stigma and starts germinating through the style.

As regard to the distribution of transcripts among the different cellular components, the presence of transcripts apparently targeted to plastids is surprising, as olive pollen grains contain only poorly-differentiated plastids, probably lacking a highly structured biochemical machinery (Rodríguez-García and

**TABLE 2 |** Selected gene products for pPCR validation.

Name of the transcript	Sample	Po(P)	P(Po)	P(L)	L(P)	Po(L)	L(Po)
Expansin 11 precursor	P1C2-PoP	+	—	—	—	—	—
Pectinesterase inhibitor 21	P2C45-PoP	+	—	+	—	+	—
Pectin metil esterase 2.1	P1C13-PoP	+	—	—	—	—	—
Pathogenesis related protein-1	P2C35-PoP	+	—	—	—	+	—
Pathogenesis related protein-5	P1C20-PPo	—	+	+	—	—	—
Disease resistance response protein-206	P1C40-PPo	—	+	—	—	—	—
Esterase PIR7A	P2C34-PPo	—	+	+	—	—	—
14-3-3 protein 4	P3C82-PPo	—	+	—	—	+	—
Photosystem II 10 kda polypeptide	P2C61-LPo	—	—	—	+	—	+
Mitogen-activated protein kinase 3	P1C13-LPo	—	—	—	—	—	+
Fructose-biphosphate aldolase	P2C28-LPo	—	—	—	+	—	+
Chlorophyll a-b binding protein	P2C39-LPo	—	+	—	+	—	+

García, 1978; Rodríguez-García et al., 1995). In both the Po(P) and Po(L) libraries, transcripts likely assigned to intracellular membrane-bounded organelle represent approximately one-third of the total transcripts.

Regarding the pie charts showing molecular function, considerable differences between the two pollen subtractive libraries were observed. These differences mainly result from the massive presence of Pectin Methyl Esterases (PMEs) in pollen. PMEs are enzymes present in higher plants, fungi and bacteria. They catalyze the demethylesterification of homogalacturonan residues of pectin, releasing methanol as the reaction product. Such modification is responsible for changes in the pectin molecule, which can then be cross-linked by calcium, and this further results in changes in the mechanical properties of the plant cell wall, altering its plasticity. This particularly affects the ability for growth and guidance of pollen tubes (Castro et al., 2013). Pollen specific PMEs have been described in other species (Tian et al., 2006; Gómez et al., 2013), with key roles during pollen germination (Leroux et al., 2015), during pollen tube elongation along the transmitting tract and when the pollen tube reaches the embryo sac in the ovule (Gómez et al., 2013). The olive pollen PME is considered a highly prevalent allergen present in the olive pollen (Salamanca et al., 2010).

Approximately half of the transcripts in the Po(P) library were involved in hydrolase activity. On the other hand, the Po(L) library did not show a high abundance of these transcripts. The subtraction carried out to create the Po(P) library possibly removed most of the PMEs of the pollen, and allowing the identification a wide variety of PME isoforms, as well as one pollen-specific PME. Therefore, the information provided by the

pie charts for molecular functions delivers particular evidence on the processes happening within the pistil at stage 4.

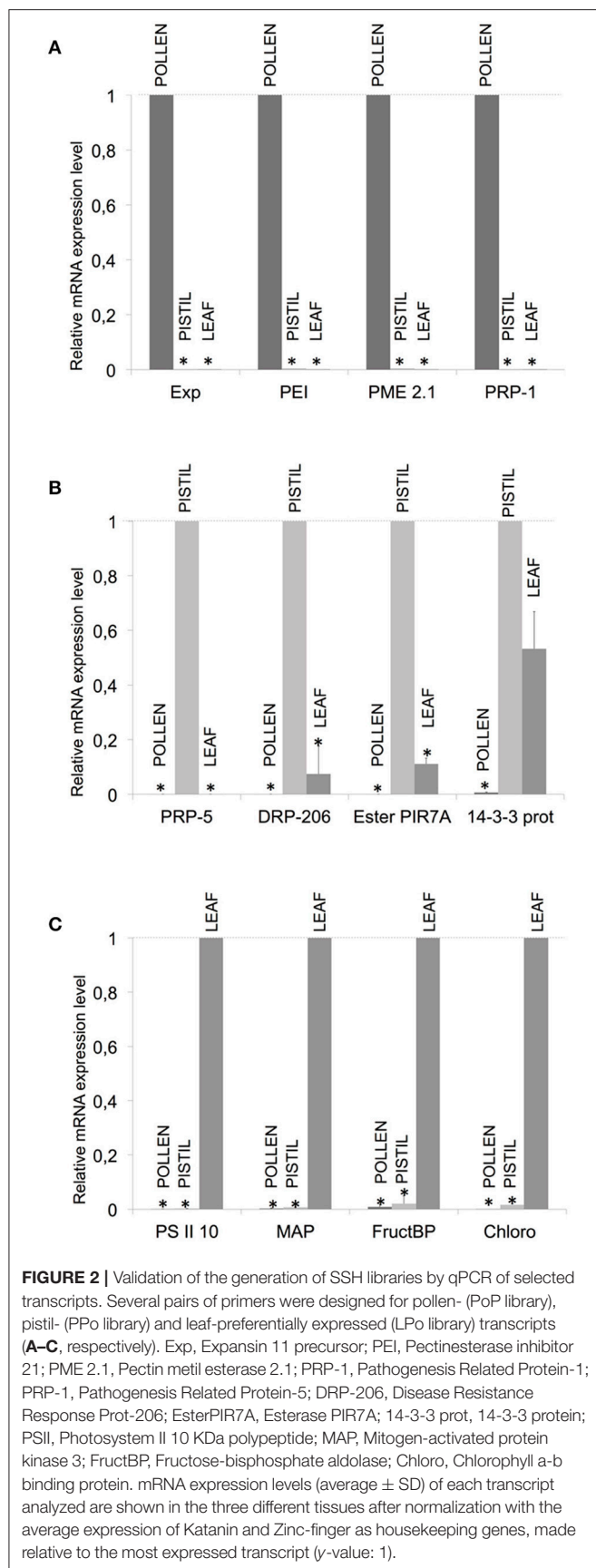
In the pistil subtractive libraries P(Po) and P(L) (Figure 4), the presence of an exclusive transcript related to symbiosis (encompassing mutualism through parasitism) was detected. Further to this, both the pistil and the leaf tissues contain wide pools of transcripts related to stress and defense. A highlight here is the presence of transcripts linked to responses to biotic stimulus in the P(L) library as it has been proposed that certain self-incompatibility processes may have evolved from pathogen-defense mechanisms (Hodgkin et al., 1988; de Nettancourt, 1997; Elleman and Dickinson, 1999; Hiscock and Allen, 2008). Within these transcripts, we found Pathogenesis related proteins-1, 5, and 10, Beta-glucosidases and PME Inhibitors.

On the other hand, the pistil harbored numerous PME inhibitor transcripts, but they did not appear in the P(Po) library as the consequence of the subtraction carried out. This is likely due to the homology between PME and the PME inhibitors (PMEI) in the N-terminal pro-region, present in the stigmatic exudate of the olive (Rejón et al., 2013). This similarity among many plant species was previously described by Jolie et al. (2010) suggesting the inhibitory role of the PME pro-peptide region. PMEI may be originated from a rearrangement of plant genome which could have the starting point of the PME inhibitors (Giovane et al., 2004) dare consistent with the localization of the PME inhibitors described at the pollen tube apex (Röckel et al., 2008), which were detected in the pistil library, probably as a result of the presence of pollen tubes growing through the stigma-style.

In a similar way to both pollen libraries, there was a large presence of transcripts for proteins associated to intracellular membrane bounded organelles in the P(Po) and P(L) libraries. Among these, transcripts associated to plastids scored as an important proportion, and were more numerous in the P(Po) library than in the P(L) one.

The analysis of the molecular functions of the transcripts present in pistils revealed the noticeable presence of transcripts for proteins with electron carrier activity in both cases, which did not appeared in the pollen libraries. The number of transcripts for proteins with DNA-binding was almost six-fold higher in the P(Po) library compared to the P(L) library. Even though transcripts within the nucleotide-binding category were abundant in both libraries, they were almost three times more abundant in the P(Po) library than in the P(L) one.

DNA-binding proteins are key players in the process of expression and regulation of new proteins as such interactions are considered to be central for many basic biological processes, including transcription regulation, DNA replication and DNA repair (Bonocora and Wade, 2015). The high levels of transcripts encoding DNA-binding proteins in the pistil could be indicative of the ability to have quick responses, where finely tuned regulation is needed. On the other hand, the presence of transcripts for nucleotide binding proteins could also be related to the changes happening outside the gynoeceum cells, as the plant disease resistance genes have been described to frequently encode nucleotide binding proteins (Meyers et al., 1999).



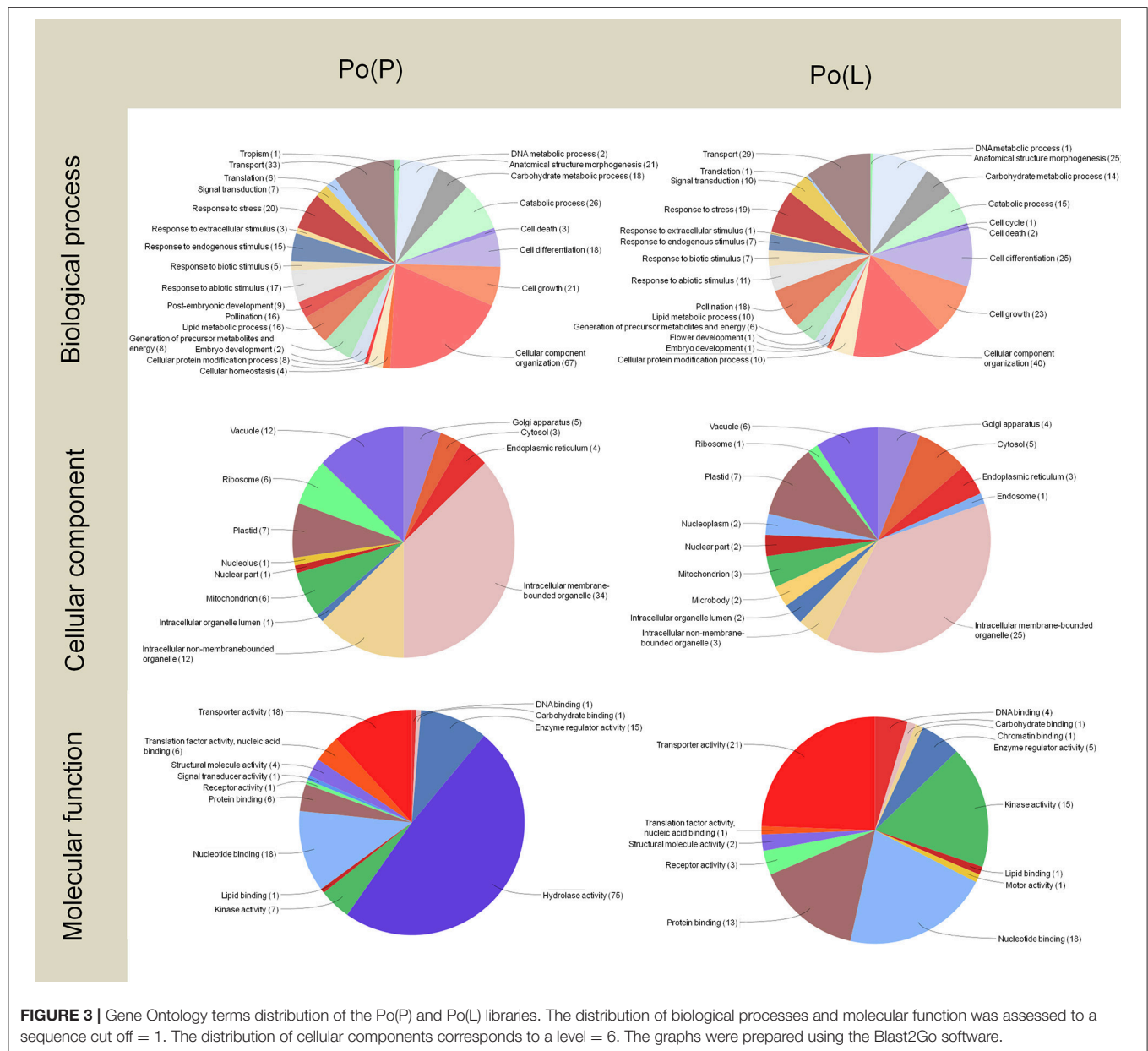
Regarding the leaf subtractive libraries L(Po) and L(P) (Figure 5), most of the transcripts corresponded, not surprisingly, to proteins involved in photosynthetic metabolism, with the detection of a large proportion of transcripts for proteins located at plastids as well as a majority of transcripts implicated in biological processes such as carbohydrate metabolism and the generation of precursor metabolites and energy. This is also consistent with the large presence of transcripts for proteins with electron carrier function. As expected, these transcripts are also present in the pistil, although less abundantly, and absent in pollen.

To sum up and as discussed above, the putative origins, biological processes, cellular localizations and molecular functions of the transcripts identified in the different subtractive libraries analyzed are in good agreement with the predicted nature of such transcripts derived from the methods and tissues used for their construction, as displayed in Table 3. It is necessary to take into account that pistil tissues are expected to include dehiscent (mature) olive pollen, as well as hydrated pollen grains and even germinating pollen grains and pollen tubes, as described in the methods section.

Therefore, taking into account the origins for the described transcripts, the SSH libraries which best describe the pollen-pistil interactions in olive and the pollen hydration and pollen tube growth are Po(P) and P(Po). For both of these SSH libraries, the corresponding GO terms have been identified, and been represented together for comparison purposes (Figure 6).

Within the pool of transcripts present in the pistil from which pollen has been subtracted [P(Po)], those involved in regulation, response to stress/stimulus, and signaling/cell communication are more abundantly represented in terms of relative percentage. On the other hand, the library of pollen from which pistil has been subtracted [Po(P)], is mainly rich in transcripts involved in cellular organization, localization, developmental processes, pollination and growth. Detailed lists of the transcripts detected for each pollen SSH library [Po(P) and Po(L)], together with BLAST relevant scores for each one are listed in Supplementary Tables 2, 6. Amongst these pollen transcripts it was found LAT52, which is known to play a role in pollen hydration and germination (Muschiatti et al., 1998; Tang et al., 2002). The presence of SF21, with a putative function in pollen-pistil interaction (Allen et al., 2010a) confirms the importance of these transcripts in reproduction. However, no transcripts of SF21 were found in the pistil [see Supplementary Tables 3, 4, corresponding to P(Po) and P(L)], despite a putative function in pollen tube guidance (Allen et al., 2010a). The Soluble NSF Attachment Protein Receptor (SNARE) proteins within the mature pollen grains are also present in the spore, with a role in the pollen tube movements (Bushart and Roux, 2007).

In the pistil (Supplementary Tables 3, 4), the response to stress is mainly represented by the Pathogenesis related proteins (PRPs); the signaling processes occurring in the pistil is emphasized by the presence of auxin responsive factors. Regulation is carried out by the interaction with the pollen specific auxin induced/repressed proteins (present in the mature pollen; they were found in both pollen libraries). The output results from the pistil libraries showed a similarity to that seen

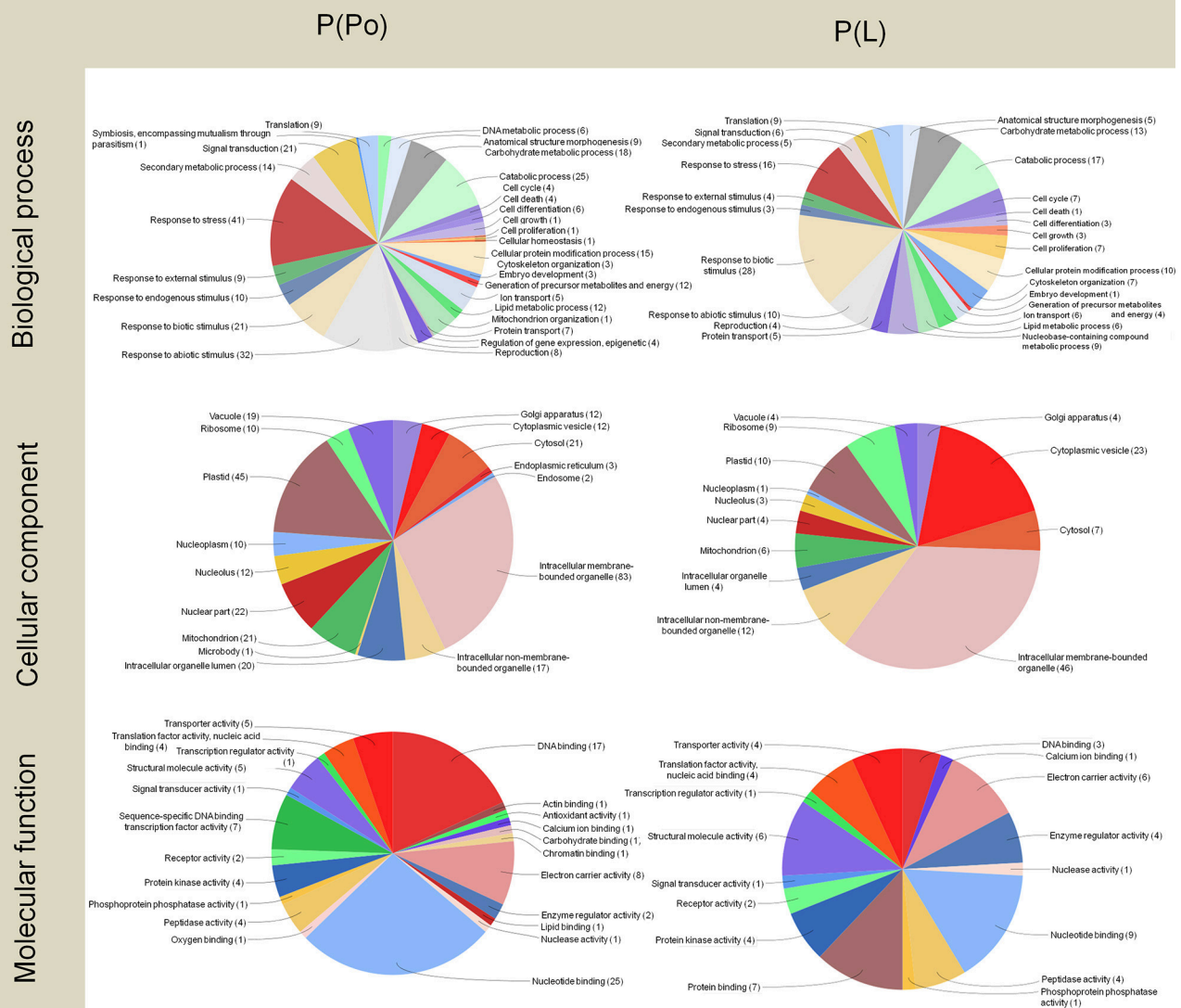


with auxin-induced root cultures (Neuteboom et al., 1999), but with an unknown function in the mature pollen grains. The auxin responsive proteins present in both pistil libraries are important for pollen tube formation (Yang et al., 2013), which indicates the presence of growing pollen tubes. However, they are not present in the mature pollen grains (Supplementary Tables 2, 6). The 14-3-3 protein is also important in pollen germination as it is involved in the regulation of turgor pressure of the pollen tube (Pertl et al., 2010).

The use of BLAST analysis of the SSH-retrieved sequences against the OLEA ESTdb, specifically containing sequences from olive mesocarp only, provided in many cases further information through the annotation of our sequences. For example, the described pistil-specific thaumatin/PRP-5 (Kuboyama,

1998; Sassa et al., 2002) is identified in the OLEA ESTdb as “Thaumatococcus-like protein, Pathogenesis-related protein 5,” whereas the NCBIdb identify them with the general term “Thaumatococcus-like protein.” We were able to discriminate between two different thaumatin-like proteins: Pathogenesis-related protein 5, which was only found in the pistil, and the Pathogenesis-related protein 1, which was pollen specific. Moreover, three isolated transcripts from different thaumatin-like proteins were found in the pistil, the first one identified as “STS14 protein/ Pathogenesis-related protein 1C,” and the others as “Osmotin-like protein” (OSML13 and OSM34, respectively). The STS14 protein is proposed to be involved in the protection of the outer tissues of the pistil from pathogen attack or guidance of the pollen tubes through the pistil. It is highly expressed





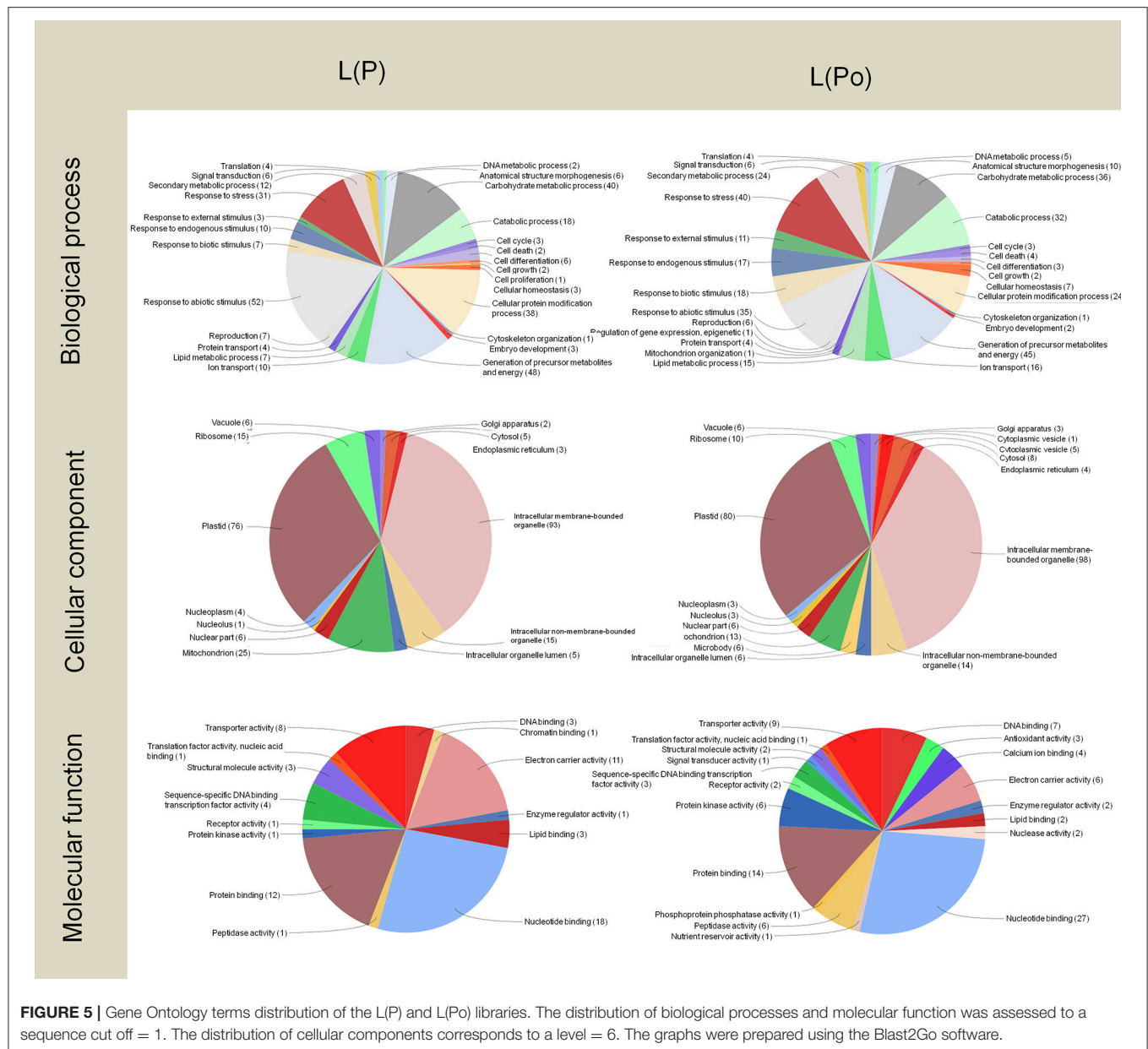
**FIGURE 4 |** Gene Ontology terms distribution of the P(Po) and P(L) libraries. The distribution of biological processes and molecular function was assessed to a sequence cut off = 1. The distribution of cellular components corresponds to a level = 6. The graphs were prepared using the Blast2Go software.

in the stigma and stylar cortex around 120 h before anthesis and increases toward the end of flower development, with a maximum at anthesis (Van Eldik et al., 1996).

As an example, three groups of transcripts were selected (defense, oxidative metabolism and transport: **Figure 7**) with the aim to further analyze and discuss the expression of transcripts with high abundance as well as their key putative roles in the pollen-pistil interaction, pollen tube germination and growth. The genes considered putative homologs to *Arabidopsis* from each group were analyzed throughout the anatomy tool of Geninvestigator. The specificity of the transcripts and the biological implications of their differential expression are described and discussed below.

## Different Transcripts Putatively Involved in Defense Are Present in the Pistil

One of the stigma-specific transcripts detected in olive was that corresponding to the Pathogenesis-related protein 5 (PRP-5). Members of this protein group have been associated with resistance to fungal infection and to responses to biotic/abiotic stresses, disease resistance or hormonal responses by inducing transcripts such as DOR, MYB, AP2, and WRKY (El-kereamy et al., 2011). A pistil-specific thaumatocin/PRP-5 has been described in Japanese pear (Sassa et al., 2002) and in tobacco (Kuboyama, 1998), where the maximum levels of the transcript were reached at anthesis. This gene product has been proposed to play a role in pollen recognition and pollen tube pathways. Among the



stigma-specific PRP-5 sequences obtained in olive, we observed high homology with the SE39B specific-stigma thaumatin from tobacco (Kuboyama, 1998) and also with a specific thaumatin from the fruit of *Olea europaea* highly expressed in response to phytophagous larvae (Corrado et al., 2012). Another olive stigma-specific transcript of interest, involved in defense, and also considered an allergen, is Mal d 1 from apple (Vanek-krebitz et al., 1995). This belongs to the PRP-10 group. These gene products have been described throughout several developmental stages and plant tissues, with a dual role associated with defense functions and regulation/signaling (Zubini et al., 2009; Choi et al., 2012). Pectin Methyl Esterases (PMEs) and their inhibitors (PMEIs) have been considered to be involved in defense function in vegetative tissues (McMillan et al., 1993; Boudart et al., 1998;

Lionetti et al., 2007, 2012; Wydra and Beri, 2007; Ann et al., 2008; Körner et al., 2009; Volpi et al., 2011, 2013). Thus, the PMEIs have been reported to enhance RNAi action, acting in gene-regulatory mechanisms (Dorokhov et al., 2006), which include virus-induced gene silencing (VIGS) and the fight against other pathogens (Collmer and Keen, 1986). Interestingly, it has been suggested that PMEIs might be internalized by endocytosis at the flanks of the pollen tube tip, regulating pollen-tube wall stability by locally inhibiting pollen PME activity (Röckel et al., 2008). It has also been suggested that PMEIs are able to reduce the activity of cell wall PMEs, leading to a drop-in pollen tube stability (Paynel et al., 2014). A pollen specific PME was described in broccoli triggering partial male sterility and decreased seed set by inhibition of pollen tube growth (Zhang et al., 2010).

**TABLE 3 |** Putative composition and nature of the transcripts present in the generated SSH libraries.

Name of the library	Tester tissue	Driver tissue	Expected biological significance
Po(P)	Pollen	Pistil	Transcripts expressed exclusively in mature pollen grains, not expressed in pistil nor in hydrated, or germinated pollen grains.
P(Po)	Pistil	Pollen	Transcripts expressed exclusively in pistil and hydrated and/or germinated pollen grains but not expressed in mature pollen grains.
P(L)	Pistil	Leaf	Transcripts expressed exclusively in pistils, but not in vegetative tissues (leaf) despite their common evaluative origin.
L(P)	Leaf	Pistil	Transcripts expressed exclusively in vegetative tissues (leaf) and hydrated or germinated pollen grains, but not in pistil.
Po(L)	Pollen	Leaf	Transcripts expressed exclusively in male reproductive tissues (pollen grains), but not in vegetative tissues (leaf).
L(Po)	Leaf	Pollen	Transcripts expressed exclusively in vegetative tissues (leaf), but not in mature pollen grains.

The large presence of cysteine proteinase in the pistil may be attributed to defense mechanisms, similar to that already described by Grudkowska and Zagdańska (2004). Another defense mechanism which seems to work actively in the olive pistil is the “disease resistance response protein 206,” that has been described to be induced in pea in response to the infection by *F. solani* f. sp. *phaseoli* (Culley et al., 1995). Within the Late Embryogenesis Abundant proteins, the pistil possesses transcripts for the dehydrin Rab 18. The Late Embryogenesis Abundant-18 transcript decreases during the germination process in pea, though it is present again in the emerging hypocotyls. Therefore, this transcript might be related to the elongation process under optimal growing conditions (Colmenero-Flores et al., 1999). In the case of the olive pistil, the presence of these transcripts could be related to the elongation process occurring in the growing pollen tubes within the stigma/style. Other transcripts found in the pistil libraries were: beta-glucosidases, late blight resistance proteins, WRKY genes, mitogen-activated kinase proteins, and the MYB genes expressed in the olive pistil are also involved in defense (Pandey and Somssich, 2009; Engelhardt et al., 2012). The MYB transcription factor itself has been described to be involved in pollen development (Niwa et al., 1993; Katiyar et al., 2012). A pistil specific nodulin has been also described in the pistil of several species (Allen et al., 2010b), being involved in a successful fertilization (Shi et al., 2012).

## Different Transcripts Putatively Involved in Defense Are Present in the Pollen Grain

Defense genes highly expressed in the olive pollen also comprise PME again, the PME inhibitor U1, and a panel of eight pathogenesis-related proteins.

PRP-1 was detected exclusively in olive pollen subtracted Po(P) and Po(L) libraries. To date, PRP-1 has only been described

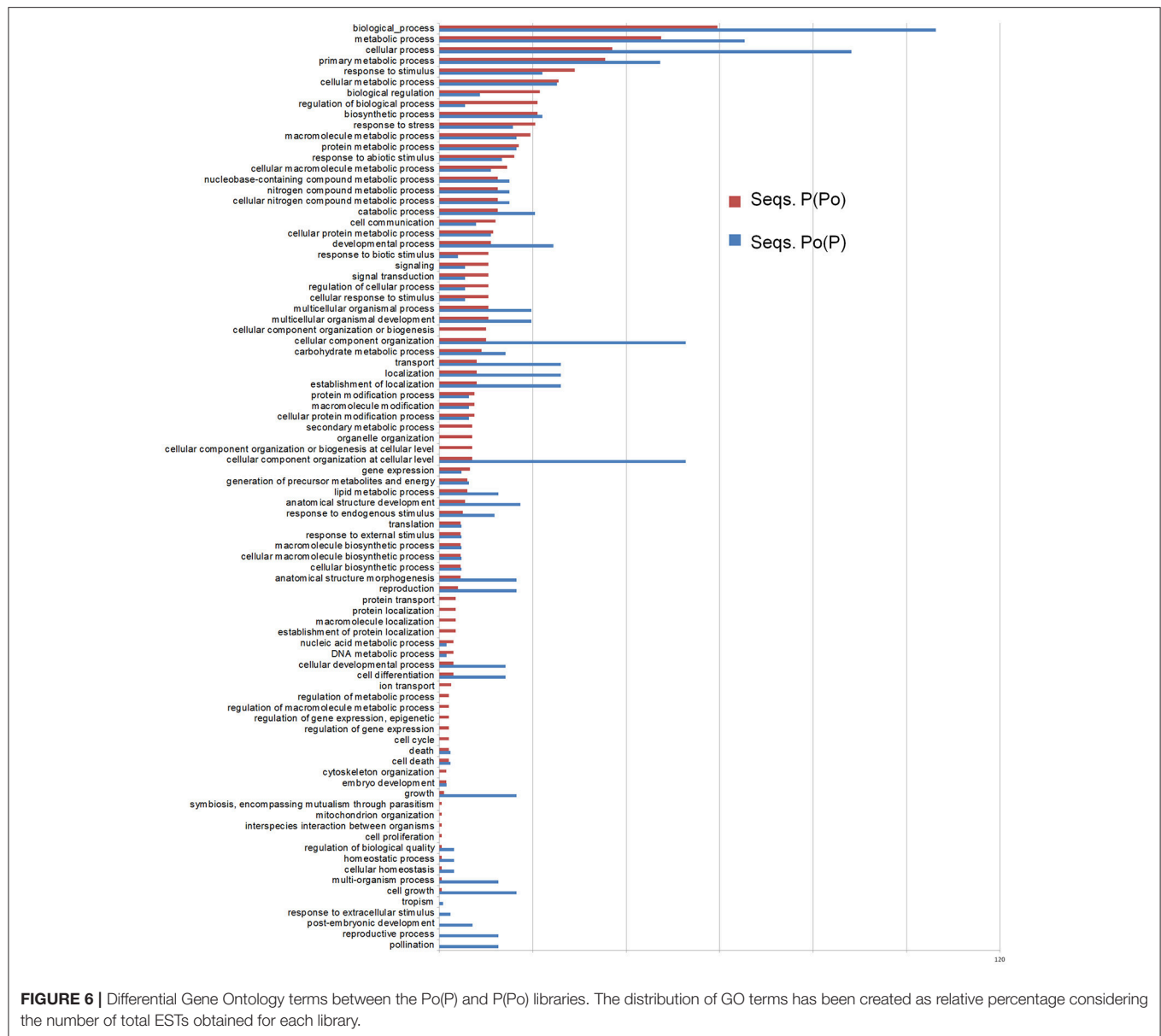
to be involved in food allergy (Asensio et al., 2004), as the precise function of these proteins is not in the pollen itself known. The specific expression of the heat stress transcription factor (HsfA2) was also detected. HsfA2, together with chaperones, are important protectors of the pollen maturation, viability and pollen tube germination from heat damage (Frank et al., 2009; Giorno et al., 2010; Zinn et al., 2010).

## Oxidative Metabolism in the Pistil

Closely related to defense mechanisms, oxidative metabolism interplays a dual role, keeping the balance between defense and signaling. In the case of the pistil, these two functions are even more finely tuned as the signaling processes are very important for a successful reproduction. Therefore, it is important to highlight the presence of transcripts corresponding to Glutathione S-transferases (GSTs), Ferredoxin-1, NAD(P)H-dependent oxidoreductase, Peroxidase 72 and Quinone oxidoreductases. Most of these transcripts have not been described as pistil-specific in *Arabidopsis* (Figure 7). Among these, stigma-specific peroxidases have been previously studied in several species (McInnis et al., 2005; Swanson et al., 2005; Beltramo et al., 2012), with the implication in the pollen-pistil interaction, pollination process, and signaling. The glutathione S-transferase has been classified as an allergenic protein in animal species (Yu and Huang, 2000; Huang et al., 2006). Later it was identified in birch pollen (Deifl et al., 2014). However, when compared to other birch pollen allergens such as Bet v 1, the release kinetics of Glutathione S-transferase from pollen grains upon contact with water and different physiologic solutions was much slower. It was suggested that the amount of glutathione S-transferases released during this time period was too low to induce allergic sensitization (Deifl et al., 2014).

## Oxidative Metabolism in the Pollen Grain

The presence of transcripts from Tpr repeat-containing thioredoxin ttl1-like was observed. Such gene products have been described to accumulate in response to osmotic stress and abscisic acid (ABA), and also may be involved in pollen compatibility (Haffani et al., 2004). Using analysis to look for members of the oxidoreductase family of proteins we could find transcripts for galactose oxidase, glyoxal oxidase and a specific L-ascorbate oxidase homolog (Pollen-specific protein NTP303). To our knowledge, the presence of galactose oxidase has not been connected to any particular characteristic of the plant reproductive tissues. Interestingly, the enzyme glyoxal oxidase has been described to be involved in male sterility, jointly to other enzymes implicated in cell wall expansion (Chen et al., 2012; Suzuki et al., 2013). The presence of L-ascorbate oxidase transcripts has been described in *in vitro* germinating pollen grains (Weterings et al., 1992), although we failed to find these mRNAs in the olive pistil, which also contains *in vivo* growing pollen tubes. It is interesting to highlight the presence of the olive pollen allergenic protein Cu, Zn Superoxide Dismutase which is involved in the protection against oxidative stress during pollen development. Its dual role, i.e., as an allergen and as part of the antioxidant/signaling metabolism, makes its study particularly interesting (Butteroni et al., 2005). Moreover, it has



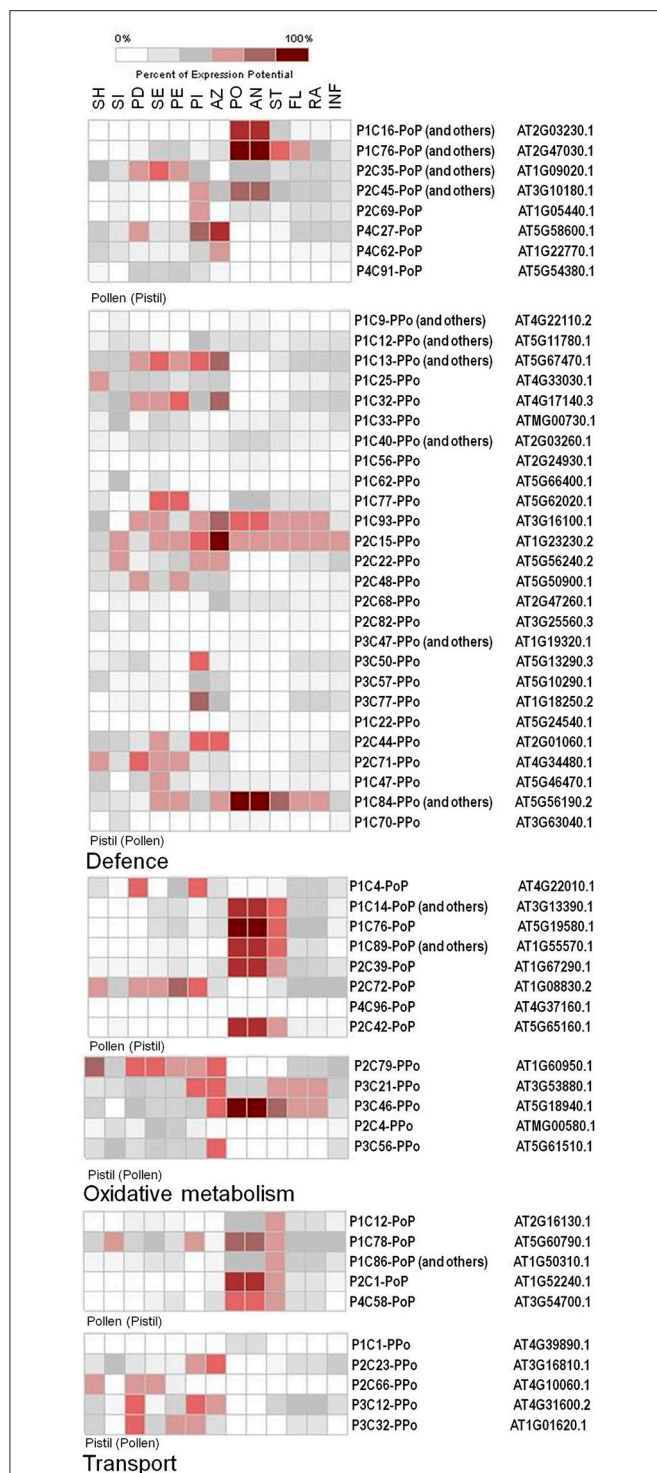
been described to be implicated in the development of the male reproductive tissues of the olive tree (Zafra et al., 2012).

## Transcripts Connected with Transport of Molecules in the Pistil

Pollen-stigma interactions and the growth of the pollen tube throughout the pistil tissues encompass a large exchange of molecules among these tissues, either positively or negatively regulating and/or permitting such growth, throughout providing energy, ions or structural molecules. Among the pistil preferential transcripts detected in this work, several have been attributed with functions facilitating transport of such molecules. This is the case for the Ras-related transport protein, which facilitates proteins movement through membranes, and the mitochondrial import inner membrane translocase

subunit Tim13. Transcripts from a member of the solute carrier family 35 (B1) are also present in the olive gynoecium. Other transporters that have been described also in primary roots (with a growing processes comparable to that of pollen tubes within the style of receptive flowers) are the specific lipid-transfer protein (LTP) AKCS9 (present in membranes) and aquaporins, both specifically present within the olive pistil transcripts and with described vegetative/reproductive differential meanings: lipid-transfer proteins were correlated with root hair deformation and pistil abortion (Krause et al., 1994; Shi et al., 2012) whereas specific aquaporins were found in the region adjacent to the root tip and have been demonstrated to be required for the self-incompatibility process displayed for members of the family *Cruciferae* (Ikeda et al., 1997; Sakurai et al., 2008).





**FIGURE 7 |** Level of expression across several reproductive/vegetative tissues of genes considered putative homologs to *Arabidopsis* using the anatomy tool Genvestigator. Three categories were considered for analysis of the levels of expression: oxidative metabolism (upper part), defense (middle part), and transport (lower part). The transcripts of two selected libraries were considered only: Po(P) and P(Po). SH, shoot; SI, silique; PD, pedicel; SE, sepal; PE, petal; PI, pistil; AZ, abscission zone; PO, pollen; AN, anther; ST, stamen; FL, flower; RA, raceme; INF, inflorescence. Identities of both the olive SSH transcripts and their corresponding Arabidopsis homologous are shown for reference purposes.

## Transcripts Connected with Transport of Molecules in the Pollen Grain

Transcripts for several transporters were found in the mature olive pollen grain. The sugar transport protein must represent a key transcript in pollen and pollen germination as it has been described in tobacco (Lemoine et al., 1999). Also, the polyol transporter present in the olive pollen could share similar functions to the polyol/monosaccharide transporter 2 expressed in mature pollen grains, growing pollen tubes, hydathodes, and young xylem cells (Klepek et al., 2010). Moreover, boron transporters expression reveals the regulatory role of boron in pollen germination and pollen tube growth (Qinli et al., 2003). Nitrate transporters also act as a nitrate sensor that triggers a specific signaling pathway stimulating lateral root growth (Guo et al., 2001), which may have a similar significance in pollen tube growth. The presence of the cation proton exchanger is critical for maintaining polarity, directing pollen growth toward the ovule, and to allow cell expansion and flower development (Bassil et al., 2011; Lu et al., 2011). The transcripts encoding ABC transporters, also found in the olive pollen, could be related to the transport of sporopollenin precursors for exine formation in developing pollen (Choi et al., 2011). Rho guanine nucleotide exchange factors are crucial in polar growth of pollen tubes (Zhang and McCormick, 2007). Finally, phosphate transporters have also been described as central for gametophyte development (Niewiadomski et al., 2005).

## Other Transcripts

The present analysis also has reported some unexpected results. As an example, anther-specific proline-rich protein APG transcripts have been found in the pistil, when they have been considered to be confined to the anther during the period of microspore development, with a dramatic decline during pollen maturation (Roberts et al., 1993). This result could be explained by the implication of the proline-rich protein APG in the pollen tube during the germination process, through a process yet to be determined.

Even though our data still do not reveal substantial information as regards some key aspects of the olive reproductive biology which are still open, such as the demonstration of the presence of a self-incompatibility system of the gametophytic type (largely suspected). Many of the transcripts detected here (either tissue-specific or not) are of great interest for the further characterization of the species, and in some cases for important issues like olive pollen and stigma physiology, as discussed above. Current knowledge of olive pollen allergenicity can also be improved, as several of the identified transcripts correspond to potential allergenic molecules already described in other species, but as such not yet described in olive. This is the case, for example, with glutathione S-transferase, considered a minor allergen in birch pollen (Zwicker, 2013; Deifl et al., 2014). Gene products corresponding to transcripts detected in the resulting SSH libraries P(Po) and P(L) described here are also consistent with proteins characterized in the olive stigma exudate by means of proteomic approaches (Rejón et al., 2014), which may act as Additional positive controls for the present methodology, because the presence of olive pollen originated peptides among those detected by was almost completely avoided.

**TABLE 4 |** Summary stats about the pre-processing of the libraries.

Name of the library	Number of sequences before cleaning	Number of sequences after SeqTrimNext treatment
Po(P)	288	191
P(Po)	288	137
P(L)	192	119
L(P)	192	91
Po(L)	192	141
L(Po)	192	141
Total	1344	820

**TABLE 5 |** Percentages of “pre-cleaned” SSH transcripts mapping to the two recently developed olive transcriptome databases (ReprOlive: Carmona et al., 2015; Cruz et al., 2016).

Name of the library	Number (%) of sequences mapping to ReprOlive	% of sequences mapping to the olive genome-derived transcriptome	Average (%)
Po(P)	189 (99)	169 (88.5)	93.8
P(Po)	135 (98.5)	127 (92.7)	95.6
P(L)	115 (96.6)	110 (92.4)	94.5
L(P)	91 (100)	89 (97.8)	98.9
Po(L)	140 (99.3)	129 (91.5)	95.4
L(Po)	136 (96.5)	136 (96.5)	96.5
Average	98.3	93.2	95.8

## COMPARATIVE TRANSCRIPTOMICS

Comparative analysis of the SSH-derived transcripts with transcriptomic information present in two recently developed databases (ReprOlive: Carmona et al., 2015; Cruz et al., 2016) was preceded by a bioinformatics cleaning by SeqTrimNext which resulted in 820 useful sequences retained from 1,344 clones sequenced with the distribution displayed in **Table 4**.

Further annotation of the cleaned sequences from pollen libraries [Po(P) and Po(L)] and screening for correspondence with the two transcriptomic databases yielded the information displayed in Supplementary Tables 8, 9. Similarly, Supplementary Tables 10–13 include the correspondence of pistil libraries [P(Po) and P(L)] and leaf libraries [L(Po) and L(P)] with both transcriptomic databases.

Overall, a high proportion (88.5–100%, with an average of 95.8%) of the SSH sequences identified here mapped to both transcriptome databases (**Table 5**). Tissue specificity of the SSH sequences described here, assessed by BLAST between the different ReprOlive datasets with Roche/454 reads (Additional files 8–13), also yielded a high proportion of “YES” matches at the appropriate tissue, thus validating *in silico* the usefulness of the SSH approach which was carried out experimentally here.

## CONCLUSIONS

The generation and analysis of different SSH subtractive libraries has provided a dataset of sequences, consisting in

about a thousand entries of great value for the understanding of the physiological processes taking place in olive pollen and pistil during their development and interaction. They are particularly important as many of these inputs have been demonstrated to be exclusively or preferentially expressed in the reproductive tissues, and not in the leaf tissues, as this material was used to build the subtractive strategy.

The subtractive transcripts have been annotated according to their homology as regard to four main databases: a general plant database provided by the NCBI, and three olive-specific databases constructed from mesocarp, reproductive tissues and a final one derived from the lately published olive genome draft. Moreover, they have been extensively classified and their presence discussed as regard to their putative biological function, cellular localization, and the molecular functions expected to exert.

Such information will be used in the near future as the basis to examine further aspects of the olive reproductive biology through the specific analysis of the expression of these products. These aspects may include compatibility, cell-to-cell communication, pollen tube growth and guidance, and pollen allergenicity among others.

## AUTHOR CONTRIBUTIONS

AZ and JA designed the experiments and redacted the manuscript. AZ performed the experiments and analyzed the results. JT was particularly involved in the work with the databases and tools on the web servers. SH and JH took part in the lab hosting and supervision during the AZ stay in their respective laboratories. MHG participated in the sequencing and interpretation of results. RC and MGC performed database searches and other bioinformatics analyses and organized and deposited all sequences at ENA.

## FUNDING

This work was supported by ERDF co-funded projects BFU2011-22779, BFU2016-77243-P, RTC-2015-4181-2 and RTC-2016-4824-2 (MINECO), P2011-CVI-7487 (Junta de Andalucía) and 201540E065 (CSIC). AZ thanks ceiA3 grant for Ph.D. in enterprises.

## ACKNOWLEDGMENTS

We thank T. Batstone and A. Allen for their help with laboratory work at the University of Bristol, A. Chueca for preparing samples for sequencing and E. Lima for helping with qPCR.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.01576/full#supplementary-material>

## REFERENCES

- Alagna, F., Cirilli, M., Galla, G., Carbone, F., Daddiego, L., Facella, P., et al. (2016). Transcript analysis and regulative events during flower development in olive (*Olea europaea* L.). *PLoS ONE* 11:E0152943. doi: 10.1371/journal.pone.0152943
- Alagna, F., D'Agostino, N., Torchia, L., Servili, M., Rao, R., Pietrella, M., et al. (2009). Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics* 10:399. doi: 10.1186/1471-2164-10-399
- Allen, A. M., Lexer, C., and Hiscock, S. J. (2010a). Characterisation of sunflower-21 (SF21) genes expressed in pollen and pistil of *Senecio squalidus* (Asteraceae) and their relationship with other members of the SF21 gene family. *Sex. Plant Reprod.* 23, 173–186. doi: 10.1007/s00497-010-0137-9
- Allen, A. M., Lexer, C., and Hiscock, S. J. (2010b). Comparative analysis of pistil transcriptomes reveals conserved and novel genes expressed in dry, wet, and semidry stigmas. *Plant Physiol.* 154, 1347–1360. doi: 10.1104/pp.110.162172
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Ann, H. S., Sohn, K. H., Choi, H. W., Hwang, I. S., Lee, S. C., and Hwang, B. K. (2008). Pepper pectin methylesterase inhibitor protein CaPMEI1 is required for antifungal activity, basal disease resistance and abiotic stress tolerance. *Planta* 228, 61–78. doi: 10.1007/s00425-008-0719-z
- Asensio, T., Crespo, J. F., Sanchez-Monge, R., Lopez-Torrejon, G., Somoza, M. L., Rodriguez, J., et al. (2004). Novel plant pathogenesis-related protein family involved in food allergy. *J. Allergy Clin. Immunol.* 114, 896–899. doi: 10.1016/j.jaci.2004.06.014
- Ateyyeh, A. F., Stosser, R., and Qrunfle, M. (2000). Reproductive biology of the olive (*Olea europaea* L.) cultivar “Nabali Baladi” *J. Appl. Bot.* 74, 255–270.
- Bassil, E., Tajima, H., Liang, Y.-C., Ohto, M.-A., Ushijima, K., Nakano, R., et al. (2011). The Arabidopsis Na<sup>+</sup>/H<sup>+</sup> Antiporters NHX1 and NHX2 control vacuolar pH and K<sup>+</sup> homeostasis to regulate growth, flower development, and reproduction. *Plant Cell* 23, 3482–3497. doi: 10.1105/tpc.111.089581
- Bazakos, C., Manioudaki, M. E., Sarropoulou, E., Spano, T., and Kalaitzis, P. (2015). 454 pyrosequencing of olive (*Olea europaea* L.) transcriptome in response to salinity. *PLoS ONE* 10:e0143000. doi: 10.1371/journal.pone.0143000
- Beltramo, C., Torello Marinoni, D., Perrone, I., and Botta, R. (2012). Isolation of a gene encoding for a class III peroxidase in female flower of *Corylus avellana* L. *Mol. Biol. Rep.* 39, 4997–5008. doi: 10.1007/s11033-011-1296-y
- Böhmer, J. (2013). *O Grande Livro da Oliveira e do Azeite. Portugal Oleícola*. Lisbon: Dinalvivo Editora.
- Bonocora, R. P., and Wade, J. T. (2015). “ChIP-seq for genome-scale analysis of bacterial DNA-binding proteins,” in *Bacterial Transcriptional Control Methods in Molecular Biology*, eds I. Artsimovitch and T. J. Santangelo (New York, NY: Springer), 327–340.
- Boudart, G., Lafitte, C., Barthe, J. P., Frasez, D., and Esquerré-Tugayé, M. T. (1998). Differential elicitation of defense responses by pectic fragments in bean seedlings. *Planta* 206, 86–94. doi: 10.1007/s004250050377
- Breton, C., Koubouris, G., Villemur, P., and Bervillé, A. J. (2017). Comment on Saumitou et al. (2017): Elucidation of the genetic architecture of self-incompatibility in olive: evolutionary consequences and perspectives for orchard management. *Evol. Appl.* doi: 10.1111/eva.12494. [Epub ahead of print].
- Breton, C. M., and Bervillé, A. (2012). New hypothesis elucidates self-incompatibility in the olive tree regarding S-alleles dominance relationships as in the sporophytic model. *C. R. Biol.* 335, 563–572. doi: 10.1016/j.crv.2012.07.006
- Breton, C. M., Farinelli, D., Shafiq, S., Heslop-Harrison, J. S., Sedgley, M., and Bervillé, A. J. (2014). The self-incompatibility mating system of the olive (*Olea europaea* L.) functions with dominance between S-alleles. *Tree Genet. Genomes* 10, 1055–1067. doi: 10.1007/s11295-014-0742-0
- Breton, C. M., Farinelli, D., Koubouris, G., and Bervillé, A. J. (2016). A model based on S-allele dominance relationships to explain pseudo self-fertility of varieties in the olive tree. *Euphytica* 210, 105–117. doi: 10.1007/s10681-016-1708-0
- Bushart, T. J., and Roux, S. J. (2007). Conserved features of germination and polarized cell growth: a few insights from a pollen-fern spore comparison. *Ann. Bot.* 99, 9–17. doi: 10.1093/aob/mcl159
- Butteroni, C., Afferni, C., Barletta, B., Iacovacci, P., Corinti, S., Brunetto, B., et al. (2005). Cloning and expression of the *Olea europaea* allergen Ole e 5, the pollen Cu/Zn superoxide dismutase. *Int. Arch. Allergy Immunol.* 137, 9–17. doi: 10.1159/000084608
- Carmona, R., Arroyo, M., Jiménez-Quesada, M. J., Seoane, P., Zafra, A., Larrosa, R., et al. (2017). Automated identification of reference genes based on RNA-seq data. *Biomed. Eng. Online* 16(Suppl. 1):65. doi: 10.1186/s12938-017-0356-5
- Carmona, R. M., Zafra, A., Seoane, P., Castro, A. J., Guerrero-Fernández, D., Castillo-Castillo, T., et al. (2015). ReprOlive: a database with linked-data for the olive tree (*Olea europaea* L.) reproductive transcriptome. *Front. Plant Sci.* 6:625. doi: 10.3389/fpls.2015.00625
- Carmona, R., Seoane, P., Zafra, A., Alché, J. D., and Claros, G. (2016). “Automatic workflow for the identification of constitutively-expressed genes based on NGS reads mapping,” in *Bioinformatics and Biomedical Engineering, IWBBIO 2016. Lecture Notes in Computer Science*, eds F. Ortuño and I. Rojas (Cham: Springer), 403–414.
- Castro, A. J., Suárez, C., Zienkiewicz, K., Alché, J. D. D., Zienkiewicz, A., and Rodríguez-García, M. (2013). Electrophoretic profiling and immunocytochemical detection of pectins and arabinogalactan proteins in olive pollen during germination and pollen tube growth. *Ann. Bot.* 112, 503–513. doi: 10.1093/aob/mct118
- Chen, C. M., Hao, X. F., Chen, G. J., Cao, B. H., Chen, Q. H., Liu, S. Q., et al. (2012). Characterization of a new male sterility-related gene Camf 1 in *Capsicum annuum* L. *Mol. Biol. Rep.* 39, 737–744. doi: 10.1007/s11033-011-0793-3
- Choi, D. S., Hwang, I. S., and Hwang, B. K. (2012). Requirement of the cytosolic interaction between PATHOGENESIS-RELATED PROTEIN10 and LEUCINE-RICH REPEAT PROTEIN1 for cell death and defense signaling in pepper. *Plant Cell* 24, 1675–1690. doi: 10.1105/tpc.112.095869
- Choi, S., Jin, J., Hwang, J., Kim, Y., Suh, M., and Lee, Y. (2011). An ABCG/WBC-type ABC transporter is essential for transport of sporopollenin precursors for exine formation in developing pollen. *Plant J.* 65, 181–193. doi: 10.1111/j.1365-3113.2010.04412.x
- Collmer, A., and Keen, N. T. (1986). The role of pectic enzymes in plant pathogenesis. *Annu. Rev. Phytopathol.* 24, 383–409. doi: 10.1146/annurev.py.24.090186.002123
- Colmenero-Flores, J. M., Moreno, L. P., Smith, C. E., and Covarrubias, A. A. (1999). Pvlea-18, a member of a new late-embryogenesis-abundant protein family that accumulates during water stress and in the growing regions of well-irrigated bean seedlings. *Plant Physiol.* 120, 93–103. doi: 10.1104/pp.120.1.93
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Corrado, G., Alagna, F., Rocco, M., Renzone, G., Varricchio, P., Coppola, V., et al. (2012). Molecular interactions between the olive and the fruit fly *Bactrocera oleae*. *BMC Plant Biol.* 12:86. doi: 10.1186/1471-2229-12-86
- Cruz, F., Julca, I., Gómez-Garrido, J., Loska, D., Marcet-Houben, M., Cano, E., et al. (2016). Genome sequence of the olive tree, *Olea europaea*. *Gigascience* 5, 29. doi: 10.1186/s13742-016-0134-5
- Cuevas, J., and Polito, V. S. (1997). Compatibility relationships in “Manzanillo” olive. *HortScience* 32, 1056–1058.
- Culley, D. E., Horovitz, D., and Hadwiger, L. A. (1995). Molecular characterization of disease-resistance response gene DRR206-d from *Pisum sativum* (L.). *Plant Physiol.* 107, 301–302. doi: 10.1104/pp.107.1.301
- Deifl, S., Zwicker, C., Vejvar, E., Kitzmu, C., Gadermaier, G., Nagl, B., et al. (2014). Glutathione-S-transferase: a minor allergen in birch pollen due to limited release from hydrated pollen. *PLoS ONE* 9:e109075. doi: 10.1371/journal.pone.0109075
- de Nettancourt, D. (1997). Incompatibility in angiosperms. *Sex. Plant Reprod.* 10, 185–199. doi: 10.1007/s004970050087
- Donaire, L., Pedrola, L., de la Rosa, R., and Llave, C. (2011). High-throughput sequencing of RNA silencing-associated small RNAs in olive (*Olea europaea* L.). *PLoS ONE* 6:e27916. doi: 10.1371/journal.pone.0027916
- Dorokhov, Y. L., Frolova, O. Y., Skurat, E. V., Ivanov, P. A., Gasanova, T. V., Sheveleva, A. A., et al. (2006). A novel function for a ubiquitous plant enzyme pectin methylesterase: the enhancer of RNA silencing. *FEBS Lett.* 580, 3872–3878. doi: 10.1016/j.febslet.2006.06.013



- El-kereamy, A., El-sharkawy, I., Ramamoorthy, R., Taheri, A., Errampalli, D., Kumar, P., et al. (2011). *Prunus domestica* pathogenesis-related protein-5 activates the defense response pathway and enhances the resistance to fungal infection. *PLoS ONE* 6:e17973. doi: 10.1371/journal.pone.0017973
- Elleman, C., and Dickinson, H. G. (1999). Commonalities between pollen/stigma and host/pathogen interactions: calcium accumulation during stigmatic penetration by *Brassica oleracea* pollen tubes. *Sex. Plant Reprod.* 12, 194–202. doi: 10.1007/s004970050192
- Engelhardt, S., Boevink, P. C., Armstrong, M. R., Ramos, M. B., Hein, I., and Birch, P. R. J. (2012). Relocalization of late blight resistance protein R3a to endosomal compartments is associated with effector recognition and required for the immune response. *Plant Cell* 24, 5142–5158. doi: 10.1105/tpc.112.104992
- Falgueras, J., Lara, A. J., Fernández-Pozo, N., Cantón, F. R., Pérez-Trabajo, G., and Claros, M. G. (2010). SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics* 11:38. doi: 10.1186/1471-2105-11-38
- Farinelli, D., Breton, C. M., Famiani, F., and Bervillé, A. (2015). Specific features in the model of olive self-incompatibility system: method to decipher S-allele pairs for varieties based on fruit setting. *Sci. Horticult.* 181, 62–75. doi: 10.1016/j.scienta.2014.10.056
- Frank, G., Pressman, E., Ophir, R., Althan, L., Shaked, R., Freedman, M., et al. (2009). Transcriptional profiling of maturing tomato (*Solanum lycopersicum* L.) microspores reveals the involvement of heat shock proteins, ROS scavengers, hormones, and sugars in the heat stress response. *J. Exp. Bot.* 60, 3891–3908. doi: 10.1093/jxb/erp234
- García-Díaz, A., Oya, R., Sónchez, A., and Luque, F. (2003). Effect of prolonged vegetative reproduction of olive tree cultivars (*Olea europaea* L.) in mitochondrial homoplasmies and heteroplasmies. *Genome* 46, 377–381. doi: 10.1139/G03-017
- Galla, G., Barcaccia, G., Ramina, A., Collani, S., Alagna, F., Baldoni, L., et al. (2009). Computational annotation of genes differentially expressed along olive fruit development. *BMC Plant Biol.* 9:128. doi: 10.1186/1471-2229-9-128
- GilAmado, J. A., and Gomez-Jimenez, M. C. (2012). Transcriptome analysis of mature fruit abscission control in olive. *Plant Cell Physiol.* 54, 244–269. doi: 10.1093/pcp/pcs179
- Giorno, F., Wolters-Arts, M., Grillo, S., Scharf, K.-D., Vriezen, W. H., and Mariani, C. (2010). Developmental and heat stress-regulated expression of HsfA2 and small heat shock proteins in tomato anthers. *J. Exp. Bot.* 61, 453–462. doi: 10.1093/jxb/erp316
- Giovane, A., Servillo, L., Balestrieri, C., Raiola, A., Avino, R. D., and Tamburrini, M. (2004). Pectin methylesterase inhibitor. *Biochim. Biophys. Acta* 169, 245–252. doi: 10.1016/j.bbapap.2003.08.011
- Gómez, M., Ranau-Morata, B., Roque, E., Polaina, J., Beltran, J., and Ca-as, L. (2013). PpPMEP, a pollen-specific pectin methylesterase of pea (*Pisum sativum* L.). *Plant Reprod.* 26, 245–254. doi: 10.1007/s00497-013-0220-0
- González-Plaza, J. J., Ortiz-Martín, I., Mu-oz-Mérida, A., García-López, C., Sánchez-Sevilla, J. F., Luque, F., et al. (2016). Transcriptomic analysis using olive varieties and breeding progenies identifies candidate genes involved in plant architecture. *Front. Plant Sci.* 7:240. doi: 10.3389/fpls.2016.00240
- Grudkowska, M., and Zagdańska, B. (2004). Multifunctional role of plant cysteine proteinases. *Acta Biochim. Pol.* 51, 609–624.
- Guerra, D., Lamontanara, A., Bagnaresi, P., Orrù, L., Rizza, F., Zelasco, S., et al. (2015). Transcriptome changes associated with cold acclimation in leaves of olive tree (*Olea europaea* L.). *Tree Genet. Genomics* 11, 113. doi: 10.1007/s11295-015-0939-x
- Guo, F. Q., Wang, R., Chen, M., and Crawford, N. M. (2001). The Arabidopsis dual-affinity nitrate transporter gene AtNRT1.1 (CHL1) is activated and functions in nascent organ development during vegetative and reproductive growth. *Plant Cell* 13, 1761–1777. doi: 10.1105/tpc.13.8.1761
- Haffani, Y. Z., Gaude, T., Cock, J. M., and Goring, D. R. (2004). Antisense suppression of thioredoxin h mRNA in *Brassica napus* cv. westar pistils causes a low level constitutive pollen rejection response. *Plant Mol. Biol.* 55, 619–630. doi: 10.1007/s11103-004-1126-x
- Hiscock, S. J., and Allen, A. M. (2008). Diverse cell signalling pathways regulate pollen-stigma interactions: the search for consensus. *New Phytol.* 179, 286–317. doi: 10.1111/j.1469-8137.2008.02457.x
- Hodgkin, T., Lyon, G. D., and Dickinson, H. G. (1988). Recognition in flowering plants: a comparison of the Brassica self-incompatibility system and plant pathogen interactions. *New Phytol.* 110, 557–569. doi: 10.1111/j.1469-8137.1988.tb00296.x
- Huang, C. H., Liew, L. M., Mah, K. W., Kuo, I. C., Lee, B. W., and Chua, K. Y. (2006). Characterization of glutathione S-transferase from dust mite, Der p 8 and its immunoglobulin E cross-reactivity with cockroach glutathione S-transferase. *Clin. Exp. Allergy* 36, 369–376. doi: 10.1111/j.1365-2222.2006.02447.x
- Igic, B., and Kohn, J. R. (2001). Evolutionary relationships among self-incompatibility RNases. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13167–13171. doi: 10.1073/pnas.231386798
- Ikedo, S., Nasrallah, J. B., Dixit, R., Preiss, S., and Nasrallah, M. E. (1997). An aquaporin-like gene required for the brassica self-incompatibility response. *Science* 276, 1564–1566. doi: 10.1126/science.276.5318.1564
- Iraia, D. L., Chiappetta, A., and Muzzalupo, I. (2016). A *de novo* transcriptomic approach to identify flavonoids and anthocyanins “Schich-Off” in olive (*Olea europaea* L.) drupes at different stages of maturation. *Front. Plant Sci.* 6:1246. doi: 10.3389/fpls.2015.01246
- Jolie, R. P., Duvetter, T., Van Loey, A. M., and Hendrickx, M. E. (2010). Pectin methylesterase and its proteinaceous inhibitor: a review. *Carbohydr. Res.* 345, 2583–2595. doi: 10.1016/j.carres.2010.10.002
- Katiyar, A., Smita, S., Lenka, S. K., Rajwanshi, R., Chinnusamy, V., and Bansal, K. C. (2012). Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. *BMC Genomics* 13:544. doi: 10.1186/1471-2164-13-544
- Klepek, Y.-S., Volke, M., Konrad, K. R., Wippel, K., Hoth, S., Hedrich, R., et al. (2010). *Arabidopsis thaliana* POLYOL/MONOSACCHARIDE TRANSPORTERS 1 and 2: fructose and xylitol/H<sup>+</sup> symporters in pollen and young xylem cells. *J. Exp. Bot.* 61, 537–550. doi: 10.1093/jxb/erp322
- Körner, E., von Dahl, C. C., Bonaventure, G., and Baldwin, I. T. (2009). Pectin methylesterase NapME1 contributes to the emission of methanol during insect herbivory and to the elicitation of defence responses in *Nicotiana attenuata*. *J. Exp. Bot.* 60, 2631–2640. doi: 10.1093/jxb/erp106
- Krause, A., Sigrist, C., Dehning, I., Sommer, H., and Broughton, W. (1994). Accumulation of transcripts encoding a lipid transfer-like protein during deformation of nodulation-competent vigna unguiculata root hairs. *Mol. Plant Microbe Interact.* 7, 411–418. doi: 10.1094/MPMI-7-0411
- Kuboyama, T. (1998). A novel thaumatin-like protein gene of tobacco is specifically expressed in the transmitting tissue of stigma and style. *Sex. Plant Reprod.* 11, 251–256. doi: 10.1007/s004970050150
- Kusaba, M., Dwyer, K., Hendershot, J., Vrebalov, J., Nasrallah, J. B., and Nasrallah, M. E. (2001). Self-Incompatibility in the genus *Arabidopsis*: characterization of the S locus in the outcrossing, *A. lyrata* and its autogamous relative, *A. thaliana*. *Plant Cell* 13, 627–643. doi: 10.1105/tpc.13.3.627
- Lemoine, R., Bürkle, L., Barker, L., Sakr, S., Kühn, C., Regnacq, M., et al. (1999). Identification of a pollen-specific sucrose transporter-like protein NtSUT3 from tobacco. *FEBS Lett.* 454, 325–330. doi: 10.1016/S0014-5793(99)00843-1
- Leroux, C., Bouton, S., Fabrice, T. N., Mareck, A., Guénin, S., Fournet, F., et al. (2015). PECTIN METHYLESTERASE48 Is involved in Arabidopsis pollen grain germination. *Plant Physiol.* 167, 367–380. doi: 10.1104/pp.114.250928
- Leyva-Pérez, M. O., Valverde-Corredor, A., Valderrama, R., Jiménez-Ruiz, J., Mu-oz-Merida, A., Trelles, O., et al. (2015). Early and delayed long-term transcriptional changes and short-term transient responses during cold acclimation in olive leaves. *DNA Res.* 22, 1–11. doi: 10.1093/dnares/dsu033
- Lionetti, V., Cervone, F., and Bellincampi, D. (2012). Methyl esterification of pectin plays a role during plant-pathogen interactions and affects plant resistance to diseases. *J. Plant Physiol.* 169, 1623–1630. doi: 10.1016/j.jplph.2012.05.006
- Lionetti, V., Raiola, A., Camardella, L., Giovane, A., Obel, N., Pauly, M., et al. (2007). Overexpression of pectin methylesterase inhibitors in arabidopsis restricts fungal infection by *Botrytis cinerea*. *Plant Physiol.* 143, 1871–1880. doi: 10.1104/pp.106.090803
- Lu, Y., Chanroj, S., Zulkifli, L., Johnson, M. A., Uozumi, N., Cheung, A., et al. (2011). Pollen tubes lacking a pair of K<sup>+</sup> transporters fail to target ovules in *Arabidopsis*. *Plant Cell* 23, 81–93. doi: 10.1105/tpc.110.080499



- McInnis, S. M., Costa, L. M., Gutiérrez-Marcos, J. F., Henderson, C. A., and Hiscock, S. J. (2005). Isolation and characterization of a polymorphic stigma-specific class III peroxidase gene from *Senecio squalidus* L. (Asteraceae). *Plant Mol. Biol.* 57, 659–677. doi: 10.1007/s11103-005-1426-9
- McMillan, G., Hedley, D., Fyfe, L., and Pérombelon, M. (1993). Potato resistance to soft-rot erwinias is related to cell wall pectin esterification. *Physiol. Mol. Plant Pathol.* 42, 279–289. doi: 10.1006/pmpp.1993.1026
- Meyers, B. C., Dickerman, A. W., Micheltore, R. W., Sivaramakrishnan, S., Sobral, B. W., and Young, N. D. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* 20, 317–332. doi: 10.1046/j.1365-313X.1999.t01-1-00606.x
- Muleo, R., Morgante, M., Cattonaro, F., Scalabrini, S., Cavallini, A., Natali, L., et al. (2016). “Genome sequencing, transcriptomics, and proteomics”, in *The Olive Tree Genome*, eds E. Rugini, L. Baldoni, R. R. Muleo, and L. Sebastiani (Cham: Compendium of Plant Genomes. Springer), 141–161.
- Muñoz-Mérida, A., González-Plaza, J. J., Ca-ada, A., Blanco, A. M., García-López, M. D. C., Rodríguez, J. M., et al. (2013). *De novo* assembly and functional annotation of the olive (*Olea europaea*) transcriptome. *DNA Res.* 20, 93–108. doi: 10.1093/dnares/dss036
- Muschietti, J., Eyal, Y., and McCormick, S. (1998). Pollen Tube localization implies a role in pollen-pistil interactions for the tomato receptor-like protein kinases LePRK1 and LePRK2. *Plant Cell* 10, 319–330. doi: 10.2307/3870591
- Neuteboom, L. W., Ng, J. M. Y., Kuyper, M., Cljdesdale, O. R., Hooykaas, P. J. J., and Zaai Van Der, B. J. (1999). Isolation and characterization of cDNA clones corresponding with mRNAs that accumulate during auxin-induced lateral root formation. *Plant Mol. Biol.* 39, 273–287. doi: 10.1023/A:1006104205959
- Niewiadomski, P., Knappe, S., Geimer, S., Fischer, K., Schulz, B., Unte, U. S., et al. (2005). The Arabidopsis plastidic glucose 6-Phosphate/Phosphate translocator GPT1 Is essential for pollen maturation and embryo sac development. *Plant Cell* 17, 760–775. doi: 10.1105/tpc.104.029124
- Niwa, Y., Izawa, O., Ishimoto, K., Akamatsu, H., and Kanoh, T. (1993). Age-dependent basal level and induction capacity of copper-zinc and manganese superoxide dismutase and other scavenging enzyme activities in leukocytes from young and elderly adults. *Am. J. Pathol.* 143, 312–320.
- Ozgenturk, N. O., Oruç, F., Sezerman, U., Kuçukural, A., Vural Korkut, S., Toksoz, F., et al. (2010). Generation and analysis of expressed sequence tags from *Olea europaea* L. *Comp. Funct. Genomics* 2010, 1–9. doi: 10.1155/2010/757512
- Pandey, S. P., and Somssich, I. E. (2009). The role of WRKY transcription factors in plant immunity. *Plant Physiol.* 150, 1648–1655. doi: 10.1104/pp.109.138990
- Parra, R., Paredes, M. A., Sanchez-Calle, I. M., and Gomez-Jimenez, M. C. (2013). Comparative transcriptional profiling analysis of olive ripe-fruit pericarp and abscission zone tissues shows expression differences and distinct patterns of transcriptional regulation. *BMC Genomics* 14:866. doi: 10.1186/1471-2164-14-866
- Paynel, F., Leroux, C., Surcouf, O., Schaumann, A., Pelloux, J., Driouich, A., et al. (2014). Kiwi fruit PME1 inhibits PME activity, modulates root elongation and induces pollen tube burst in *Arabidopsis thaliana*. *Plant Growth Regul.* 74, 285–297. doi: 10.1007/s10725-014-9919-7
- Pertl, H., Pöckl, M., Blaschke, C., and Obermeyer, G. (2010). Osmoregulation in *Lilium* pollen grains occurs via modulation of the plasma membrane H<sup>+</sup> ATPase activity by 14-3-3 proteins. *Plant Physiol.* 154, 1921–1928. doi: 10.1104/pp.110.165696
- Qinli, W., Lu, L., Wu, X., Li, Y., and Lin, J. (2003). Boron influences pollen germination and pollen tube growth in *Picea meyeri*. *Tree Physiol.* 23, 345–351. doi: 10.1093/treephys/23.5.345
- Rejón, J. D., Delalande, F., Schaeffer-reiss, C., Carapito, C., Zienkiewicz, K., Alché, J. D. D., et al. (2013). Proteomics profiling reveals novel proteins and functions of the plant stigma exudate. *J. Exp. Bot.* 64, 5695–5705. doi: 10.1093/jxb/ert345
- Rejón, J., Delalande, F., Schaeffer-Reiss, C., Carapito, C., Zienkiewicz, K., Alché, J., et al. (2014). The plant stigma exudate. *Plant Signal. Behav.* 9:e28274. doi: 10.4161/psb.28274
- Roberts, M. R., Foster, G. D., Blundell, R. P., Robinson, S. W., Kumar, A., Draper, J., et al. (1993). Gametophytic and sporophytic expression of an anther-specific Arabidopsis thaliana gene. *Plant J.* 3, 111–120. doi: 10.1046/j.1365-313X.1993.t01-5-00999.x
- Röckel, N., Wolf, S., Kost, B., Rausch, T., and Greiner, S. (2008). Elaborate spatial patterning of cell-wall PME and PME1 at the pollen tube tip involves PME1 endocytosis, and reflects the distribution of esterified and de-esterified pectins. *Plant J.* 53, 133–143. doi: 10.1111/j.1365-313X.2007.03325.x
- Rodríguez-García, M., and García, A. (1978). Differentiation of the plastid population during microsporogenesis and the development of the pollen grain in the liliaceae. *Biol. Cell.* 33, 63–70.
- Rodríguez-García, M. I., Fernandez, M. C., Alche, J. D., and Olmedilla, A. (1995). Endoplasmic reticulum as a storage site for allergenic proteins in pollen grains of several Oleaceae. *Protoplasma* 187, 111–116. doi: 10.1007/BF01280238
- Sakurai, J., Ahamed, A., Murai, M., Maeshima, M., and Uemura, M. (2008). Tissue and cell-specific localization of rice aquaporins and their water transport activities. *Plant Cell Physiol.* 49, 30–39. doi: 10.1093/pcp/pcm162
- Salamanca, G., Rodríguez, R., Quiralte, J., Moreno, C., Pascual, C. Y., Barber, D., et al. (2010). Pectin methylesterases of pollen tissue, a major allergen in olive tree. *FEBS J.* 277, 2729–2739. doi: 10.1111/j.1742-4658.2010.07689.x
- Sarah, G., Homa, F., Pointet, S., Contreras, S., Sabot, F., Nabholz, B., et al. (2017). A large set of 26 new reference transcriptomes dedicated to comparative population genomics in crops and wild relatives. *Mol. Ecol. Resour.* 17, 565–580. doi: 10.1111/1755-0998.12587
- Sassa, H., Ushijima, K., and Hirano, H. (2002). A pistil-specific thaumatin/PR5-like protein gene of Japanese pear (*Pyrus serotina*): sequence and promoter activity of the 5' region in transgenic tobacco. *Plant Mol. Biol.* 50, 371–377. doi: 10.1023/A:1019866029629
- Saumitou-Laprade, P., Vernet, P., Vekemans, X., Castric, V., Barcaccia, G., Khadari, B., et al. (2017). Controlling for genetic identity of varieties, pollen contamination and stigma receptivity is essential to characterize the self-incompatibility system of *Olea europaea* L. *Evol. Appl.* doi: 10.1111/eva.12498. [Epub ahead of print].
- Shi, T., Gao, Z., Wang, L., Zhang, Z., Zhuang, W., Sun, H., et al. (2012). Identification of differentially-expressed genes associated with pistil abortion in Japanese apricot by genome-wide transcriptional analysis. *PLoS ONE* 7:e47810. doi: 10.1371/journal.pone.0047810
- Suzuki, H., Rodríguez-Urbe, L., Xu, J., and Zhang, J. (2013). Transcriptome analysis of cytoplasmic male sterility and restoration in CMS-D8 cotton. *Plant Cell Rep.* 32, 1531–1542. doi: 10.1007/s00299-013-1465-7
- Swanson, R., Clark, T., and Preuss, D. (2005). Expression profiling of Arabidopsis stigma tissue identifies stigma-specific genes. *Sex. Plant Reprod.* 18, 163–171. doi: 10.1007/s00497-005-0009-x
- Tang, W., Ezcurra, I., Muschietti, J., and McCormick, S. (2002). A cysteine-rich extracellular protein, LAT52, interacts with the extracellular domain of the pollen receptor kinase LePRK2. *Plant Cell* 14, 2277–2287. doi: 10.1105/tpc.003103
- Taylor, C. M., Jost, R., Erskine, W., and Nelson, M. N. (2016). Identifying stable reference genes for qRT-PCR normalisation in gene expression studies of narrow-leaved Lupin (*Lupinus angustifolius* L.). *PLoS ONE* 11:e0148300. doi: 10.1371/journal.pone.0148300
- Tian, G., Chen, M., Zaltsman, A., and Citovsky, V. (2006). Pollen-specific pectin methylesterase involved in pollen tube growth. *Dev. Biol.* 294, 83–91. doi: 10.1016/j.ydbio.2006.02.026
- Vanek-krebitz, M., Hoffmann-Sommergruber, K., Laimer da Machado, M., Susani, M., Ebner, C., Kraft, D., et al. (1995). Cloning and sequencing of Mad d 1, the major allergen from apple (*Malus domestica*) and its immunological relationship to Bet v 1, the major birch pollen allergen. *Biochem. Biophys. Res. Commun.* 214, 538–551. doi: 10.1006/bbrc.1995.2320
- Van Eldik, G. J., Wiggins, M., Ruiter, R. K., Van Herpen, M. M., Schrauwen, J. A., and Wullems, G. J. (1996). Molecular analysis of a pistil-specific gene expressed in the stigma and cortex of *Solanum tuberosum*. *Plant Mol. Biol.* 30, 171–176. doi: 10.1007/BF00017811
- Volpi, C., Janni, M., Lionetti, V., Bellincampi, D., Favaron, F., and D'Ovidio, R. (2011). The ectopic expression of a pectin methyl esterase inhibitor increases pectin methyl esterification and limits fungal diseases in wheat. *Mol. Plant Microbe Interact.* 24, 1012–1019. doi: 10.1094/MPMI-01-11-0021
- Volpi, C., Raiola, A., Janni, M., Gordon, A., Sullivan, D. M. O., Favaron, F., et al. (2013). Plant physiology and biochemistry *Claviceps purpurea* expressing polygalacturonases escaping PGIP inhibition fully infects PvPGIP2 wheat transgenic plants but its infection is delayed in wheat transgenic plants with increased level of pectin methyl esterif. *Plant Physiol. Biochem.* 73, 294–301. doi: 10.1016/j.plaphy.2013.10.011

- Weterings, K., Reijnen, W., van Aarssen, R., Kortstee, A., Spijkers, J., van Herpen, M., et al. (1992). Characterization of a pollen-specific cDNA clone from *Nicotiana tabacum* expressed during microgametogenesis and germination. *Plant Mol. Biol.* 18, 1101–1111. doi: 10.1007/BF00047713
- Wu, S.-B., Collins, G., and Sedgley, M. (2002). Sexual compatibility within and between olive cultivars. *J. Hortic. Sci. Biotechnol.* 77, 665–673. doi: 10.1080/14620316.2002.11511554
- Wydra, K., and Beri, H. (2007). Immunohistochemical changes in methyl-ester distribution of homogalacturonan and side chain composition of rhamnogalacturonan I as possible components of basal resistance in tomato inoculated with *Ralstonia solanacearum*. *Physiol. Mol. Plant Pathol.* 70, 13–24. doi: 10.1016/j.pmp.2007.05.006
- Yang, J., Tian, L., Sun, M.-X., Huang, X.-Y., Zhu, J., Guan, Y.-Y., et al. (2013). Auxin response factor17 is essential for pollen wall pattern formation in *Arabidopsis*. *Plant Physiol.* 162, 720–731. doi: 10.1104/pp.113.214940
- Yanik, H., Turktas, M., Dundar, E., Hernandez, P., Dorado, G., and Unver, T. (2013). Genome-wide identification of alternate bearing-associated microRNAs (miRNAs) in olive (*Olea europaea* L.). *BMC Plant Biol.* 13:10. doi: 10.1186/1471-2229-13-10
- Yu, S. J., and Huang, S. W. (2000). Purification and characterization of glutathione S-transferases from the German Cockroach, *Blattella germanica* (L.). *Pestic. Biochem. Physiol.* 67, 36–45. doi: 10.1006/pest.1999.2472
- Zafra, A., Jiménez-Quesada, M. J., Traverso, J. A., Corpas, F. J., and Alché, J. D. (2012). Peroxisomal localization of CuZn superoxide dismutase in the male reproductive tissues of the olive tree. *Microsc. Microanal.* 18, 33–34. doi: 10.1017/S1431927612012822
- Zafra, A., Rodríguez-García, M. I., and Alche, J. D. (2010). Cellular localization of ROS and NO in olive reproductive tissues during flower development. *BMC Plant Biol.* 19:36. doi: 10.1186/1471-2229-10-36
- Zhang, Y. G., Jian, F., Wu, J., and Wang, X. W. (2010). BoPMEI1, a pollen-specific pectin methylesterase inhibitor, has an essential role in pollen tube growth. *Planta* 231, 1323–1334. doi: 10.1007/s00425-010-1136-7
- Zhang, Y., and McCormick, S. (2007). A distinct mechanism regulating a pollen-specific guanine nucleotide exchange factor for the small GTPase Rop in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18830–18835. doi: 10.1073/pnas.0705874104
- Zinn, K. E., Tunc-Ozdemir, M., and Harper, J. F. (2010). Temperature stress and plant sexual reproduction: uncovering the weakest links. *J. Exp. Bot.* 61, 1959–1968. doi: 10.1093/jxb/erq053
- Zubini, P., Zambelli, B., Musiani, F., Ciurli, S., Bertolini, P., and Baraldi, E. (2009). The RNA hydrolysis and the cytokinin binding activities of PR-10 proteins are differently performed by two isoforms of the Pru p 1 peach major allergen and are possibly functionally related. *Plant Physiol.* 150, 1235–1247. doi: 10.1104/pp.109.139543
- Zwicker, C. (2013). *Biochemical and Immunological Characterization of Glutathione-S-transferase in Birch Pollen*. Masterarbeit, University of Vienna Zentrum für Molekulare Biologie BetreuerIn: Kovarik, Pavel. Available online at: <http://othes.univie.ac.at/26851/>

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Zafra, Carmona, Traverso, Hancock, Goldman, Claros, Hiscock and Alche. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Genetic Tracing of *Jatropha curcas* L. from Its Mesoamerican Origin to the World

Haiyan Li<sup>1</sup>, Suguru Tsuchimoto<sup>2</sup>, Kyuya Harada<sup>2</sup>, Masanori Yamasaki<sup>3</sup>, Hiroe Sakai<sup>2</sup>, Naoki Wada<sup>2</sup>, Atefeh Alipour<sup>1</sup>, Tomohiro Sasai<sup>1</sup>, Atsushi Tsunekawa<sup>4</sup>, Hisashi Tsujimoto<sup>4</sup>, Takayuki Ando<sup>5</sup>, Hisashi Tomemori<sup>4</sup>, Shusei Sato<sup>6</sup>, Hideki Hirakawa<sup>7</sup>, Victor P. Quintero<sup>8</sup>, Alfredo Zamarripa<sup>9</sup>, Primitivo Santos<sup>10</sup>, Adel Hegazy<sup>11</sup>, Abdalla M. Ali<sup>12</sup> and Kiichi Fukui<sup>13\*</sup>

## OPEN ACCESS

### Edited by:

Johann Vollmann,  
University of Natural Resources  
and Life Sciences, Vienna, Austria

### Reviewed by:

Nobuko Ohmido,  
Kobe University, Japan  
Veronique Sylvie Decroocq,  
Institut National de la Recherche  
Agronomique (INRA), France

### \*Correspondence:

Kiichi Fukui  
kfukui@bio.eng.osaka-u.ac.jp

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 04 April 2017

**Accepted:** 22 August 2017

**Published:** 07 September 2017

### Citation:

Li H, Tsuchimoto S, Harada K, Yamasaki M, Sakai H, Wada N, Alipour A, Sasai T, Tsunekawa A, Tsujimoto H, Ando T, Tomemori H, Sato S, Hirakawa H, Quintero VP, Zamarripa A, Santos P, Hegazy A, Ali AM and Fukui K (2017) Genetic Tracing of *Jatropha curcas* L. from Its Mesoamerican Origin to the World. *Front. Plant Sci.* 8:1539. doi: 10.3389/fpls.2017.01539

<sup>1</sup> Department of Biotechnology, Graduate School of Engineering, Osaka University, Osaka, Japan, <sup>2</sup> Plant Bioengineering for Bioenergy Laboratory, Graduate School of Engineering, Osaka University, Osaka, Japan, <sup>3</sup> Kobe Food Resources Education and Research Center, Graduate School of Agricultural Science, Kobe University, Hyogo, Japan, <sup>4</sup> Arid Land Research Center, Tottori University, Tottori, Japan, <sup>5</sup> The Center for International Affairs, Tottori University, Tottori, Japan, <sup>6</sup> Graduate School of Life Sciences, Tohoku University, Miyagi, Japan, <sup>7</sup> Kazusa DNA Research Institute, Chiba, Japan, <sup>8</sup> INIFAP-Campo Experimental Bajío, Guanajuato, Mexico, <sup>9</sup> INIFAP-Campo Experimental Rosario Izapa, Chiapas, Mexico, <sup>10</sup> College of Agriculture, University of the Philippines Los Banos, Laguna, Philippines, <sup>11</sup> Genetic Engineering and Biotechnology Research Institute, University of Sadat City, Sadat City, Egypt, <sup>12</sup> Faculty of Agriculture, Shambat, University of Khartoum, Khartoum, Sudan, <sup>13</sup> Graduate School of Pharmaceutical Science, Osaka University, Osaka, Japan

*Jatropha curcas* L. (*Jatropha*), a shrub species of the family Euphorbiaceae, has been recognized as a promising biofuel plant for reducing greenhouse gas emissions. However, recent attempts at commercial cultivation in Africa and Asia have failed because of low productivity. It is important to elucidate genetic diversity and relationship in worldwide *Jatropha* genetic resources for breeding of better commercial cultivars. Here, genetic diversity was analyzed by using 246 accessions from Mesoamerica, Africa and Asia, based on 59 simple sequence repeat markers and eight retrotransposon-based insertion polymorphism markers. We found that central Chiapas of Mexico possesses the most diverse genetic resources, and the Chiapas Central Depression could be the center of origin. We identified three genetic groups in Mesoamerica, whose distribution revealed a distinct geographic cline. One of them consists mainly of accessions from central Chiapas. This suggests that it represents the original genetic group. We found two Veracruz accessions in another group, whose ancestors might be shipped from Port of Veracruz to the Old World, to be the source of all African and Asian *Jatropha*. Our results suggest the human selection that caused low productivity in Africa and Asia, and also breeding strategies to improve African and Asian *Jatropha*. Cultivars improved in the productivity will contribute to expand mass commercial cultivation of *Jatropha* in Africa and Asia to increase biofuel production, and finally will support in the battle against the climate change.

**Keywords:** *Jatropha*, biofuel plant, genetic diversity, center of origin, genetic resources

## INTRODUCTION

*Jatropha curcas* L. (*Jatropha*) is a shrub that produces non-edible seed oil suitable for biodiesel fuel. The oil content ranges from 30 to 50% of its seed weight (Heller, 1996; Kumari et al., 2009). *Jatropha* has a potential to reduce the consumption of fossil fuel and carbon dioxide emissions (Bailis and Baka, 2010; Bahadur et al., 2013). *Jatropha* is drought tolerant, thereby its biofuel production could avoid competition with food crops (Fairless, 2007; Bahadur et al., 2013). Saving of carbon emissions depends on how biofuels are produced. Growing perennial plants on degraded and/or fallow fields for biofuel production has an advantage of sustained greenhouse gas reduction compared with converting forests or grasslands to produce annual crop-based biofuels (Fargione et al., 2008; Searchinger et al., 2008).

Large *Jatropha* plantations have been planned worldwide, particularly in the semiarid and arid areas of African and Asian countries by using seeds derived from local plants (Fairless, 2007; Von Maltitz and Setzkorn, 2013). Many commercial plantations, however, have not resulted in as high yields as were anticipated because of the limited productivity (Sanderson, 2009; Singh et al., 2014; Van Eijck et al., 2014; Von Maltitz et al., 2014). Breeding high-yielding *Jatropha* varieties is in its infancy so far (Divakara et al., 2010). Most of previous studies showed that genetic diversity of *Jatropha* is low in Africa, Asia and South America, but higher in Mesoamerica, including Mexico and Central America (Basha and Sujatha, 2007; Basha et al., 2009; Pamidimarri et al., 2010; Rosado et al., 2010; He et al., 2011; Montes et al., 2014; Montes Osorio et al., 2014; Siju et al., 2016). The low genetic diversity of African and Asian accessions has limited the potential for successful breeding by using local resources. The challenge now is to develop well-adapted, high-yielding varieties that are suitable for a wide range of climate conditions in African and Asian countries, since only wide plantation with high oil-production can guarantee a good supply for biofuel. Characterization and preservation of a diverse collection of *Jatropha* accessions, including those from Mesoamerica, is the key step to develop them.

Mesoamerica, especially Mexico, has been shown to possess high genetic variations of *Jatropha* and was assumed to be its place of origin (Makkar and Becker, 2009; Dias et al., 2012). Mexico has not only toxic *Jatropha* varieties, but also non-toxic ones, suggesting that Mexico might be also the domestication center of *Jatropha* (Dias et al., 2012). Recent studies described that Chiapas, the southernmost state of Mexico bordering Guatemala, might be the center of *Jatropha* biodiversity (Ovando Medina et al., 2011; Pecina Quintero et al., 2011, 2014; Salvador Figueroa et al., 2015). However, a comprehensive study is required to obtain more evidences for this conclusion.

It has been widely approved that *Jatropha* was brought by Portuguese from Mesoamerica to Cape Verde Islands, and then brought to Africa and Asia (Heller, 1996), while the transmission route in Mesoamerica is unknown and the genetic background of African and Asian *Jatropha* in the world population remains unclear. To improve the traits in the breeding goal of *Jatropha* in Asia and Africa, it is of great significance not only to identify diverse genetic resources but also to unveil the ancestral

genotype of African and Asian *Jatropha* when shipped from Mesoamerica.

In this study, 59 simple sequence repeat (SSR) markers and eight retrotransposon-based insertion polymorphism (RBIP) markers were applied to assess the genetic diversity of 246 *Jatropha* accessions from eight Mexican states (Chiapas, Guerrero, Michoacan, Morelos, Oaxaca, Tabasco, Veracruz, and Yucatan), Guatemala and countries of Africa and Asia. We also evaluated the genetic variability of accessions from each of nine regions in Chiapas, to identify the center of origin. We further identified the voyage from Chiapas to the Old World by tracing the progenies in Mesoamerica sharing the same ancestors with nowadays African and Asian accessions. Human selection causing a low genetic basis in Africa and Asia is discussed. Finally, we propose strategies to improve African and Asian *Jatropha*, in order to increase the phenotypic performance including productivity, which will lead us to relieve from the climate change by mass production of the biofuel.

## MATERIALS AND METHODS

### Plant Materials

A worldwide collection of 246 *Jatropha* accessions in this study consisted of 207 accessions from Mesoamerica (198 from Mexico and nine from Guatemala), seven from Africa, and 32 from Asia (Supplementary Table 1; see **Figure 3** for the locations of Mexican states and regions in the state of Chiapas). The Mexican accessions and seven Guatemalan accessions were obtained from the collection of the Instituto Nacional de Investigaciones Forestales, Agropecuarias (INIFAP, Mexico). Seven Philippine accessions were obtained from the collection of UPLB originating in four provinces of Luzon and Mindanao islands. Twenty-one Vietnamese accessions were obtained from plants grown in Quang Tri Province. Sudanese and Egyptian accessions were obtained from the University of Khartoum and University of Sadat City, respectively. Other accessions were the same as those used in our previous study (Sato et al., 2011). *Jatropha* sampling and transferring to Japan were performed in compliance with the Nagoya agreement.

### DNA Markers

Over 500 SSR markers were developed by designing primer pairs surrounding SSR sequences identified by the genomic database of Asian accessions based on our whole genome sequencing project (Sato et al., 2011; Hirakawa et al., 2012). Effective markers were selected from them based on clear polymorphisms and low number of null alleles. Eight RBIP markers employed in this study were developed from members of the copia-type families identified in the genomic database (Alipour et al., 2013). All of them were expected to have retrotransposed more recently than other members. Primers were designed for both flanking (FLK) sequences and long terminal repeats (LTRs). Retrotransposon insertion at each locus was shown by combining primers designed from the FLK sequences at both sides and primers designed from LTR sequences. Linkages between the markers were tested by TASSEL2.1 (Bradbury et al., 2007).



## DNA Extraction

Genomic DNA was extracted from *Jatropha* leaves with a DNeasy Plant Mini Kit (Qiagen, United States) according to the manufacturer's instructions. The DNA samples were diluted to a final concentration of 0.35 ng/ $\mu$ l and stored at  $-20^{\circ}\text{C}$  until use.

## Genotyping and Scoring

Polymerase chain reaction (PCR) for SSR markers was performed in a 10  $\mu$ l total volume solution containing 1.4 ng of DNA template, 0.16  $\mu$ l of SSR primer mix (25  $\mu\text{M}$  each), 1 $\times$  PCR buffer, 0.03  $\mu\text{M}$   $\text{MgCl}_2$ , 0.8  $\mu$ l dNTPs (2.5 mM), 0.08 U Bio-taq (5 units/ $\mu\text{L}$ ), and Milli Q water. Touch-down amplification was performed as follows: 3 min hold at  $94^{\circ}\text{C}$ , followed by three cycles at  $94^{\circ}\text{C}$  for 30 s, and  $68^{\circ}\text{C}$  for 30 s, with the annealing temperature reduced by  $2^{\circ}\text{C}$  until  $64^{\circ}\text{C}$  every three cycles. Continuing with three cycles at  $94^{\circ}\text{C}$  for 30 s,  $62^{\circ}\text{C}$  for 30 s, and  $72^{\circ}\text{C}$  for 30 s, the annealing temperature was reduced by  $2^{\circ}\text{C}$  until  $58^{\circ}\text{C}$  every three cycles. A further 30 cycles of amplification was performed at  $94^{\circ}\text{C}$  for 30 s,  $55^{\circ}\text{C}$  for 30 s, and  $72^{\circ}\text{C}$  for 30 s. The PCR products were stored at  $-20^{\circ}\text{C}$  until separation by polyacrylamide gel electrophoresis (PAGE). Amplified bands were stained with ethidium bromide. For markers that showed ambiguous bands, experiments were repeated until clear bands were observed by increasing annealing temperatures.

For RBIP markers, PCR was performed in a 5  $\mu$ l total volume solution containing 0.7 ng of DNA template, 0.08  $\mu$ l of primer mix (50  $\mu\text{M}$  each), 1 $\times$  PCR buffer, 0.03  $\mu\text{M}$   $\text{MgCl}_2$ , 0.4  $\mu$ l dNTPs (2.5 mM), and 0.08 U Bio-taq (5 units/ $\mu\text{L}$ ), made to the volume with Milli Q water. PCR reaction was started with denaturation at  $94^{\circ}\text{C}$  for 2 min, followed by 35 cycles of 45 s at  $94^{\circ}\text{C}$ , 45 s at  $55^{\circ}\text{C}$  and 2 min at  $72^{\circ}\text{C}$ , continued with a final elongation step at  $72^{\circ}\text{C}$  for 10 min. Amplified bands were stained with ethidium bromide. The PCR products were stored and analyzed like the SSR markers.

For all the SSR markers and RBIP markers, bands were individually scored (presence 1, absence 0). All data analyses were performed based on the genotype data matrix.

## Genetic Heterozygosity

Expected heterozygosity ( $H_E$ ) and observed heterozygosity ( $H_O$ ) show the portion of heterozygotes in populations. They are measures of the extent of genetic variation in a population.  $H_E$  is heterozygosity under Hardy–Weinberg equilibrium (HWE) by random mating of individuals (Nei, 1987).  $H_O$  is heterozygosity at the observed level. Differences between  $H_E$  and  $H_O$  are caused by inbreeding often as results of selection or the small population size. Inbreeding coefficient ( $F_{IS}$ ) is the value to estimate the level of inbreeding of the population.  $F_{IS}$  was calculated from  $H_E$  and  $H_O$  by the equation:  $F_{IS} = 1 - H_O/H_E$ . The larger  $F_{IS}$  shows the larger extent of inbreeding in a population.  $H_E$  and  $H_O$  were calculated with GenALEx 6.5 (Peakall and Smouse, 2006). Comparisons of the significances of  $H_E$  and  $H_O$  among populations were performed using Kruskal–Wallis one-way analysis of variance. Calculation of the  $F_{IS}$  and  $P$ -values, which show the significance of inbreeding, was performed with FSTAT version 2.9.3.2 (Goudet, 2001).

## Population Structure

Model-based clustering analysis was performed using Mesoamerican accessions or all accessions including African and Asian accessions with STRUCTURE 2.3.4 (Pritchard et al., 2000). The analysis was performed using the Markov chain Monte Carlo method based on  $1 \times 10^5$  iterations followed by a burn-in period of  $5 \times 10^4$  iterations. To determine the optimal number of populations ( $K$ ),  $\Delta K$  method was applied.  $K$  varied from 1 to 10, after 10 independent runs for each  $K$ ,  $\Delta K$  was calculated based on the rate of change of the log likelihood for  $K$  values between successive  $K$  values following Evanno et al. (2005).  $\Delta K$  should have a clear peak at true value of  $K$ .

Principal coordinates analysis (PCoA) of all individuals was performed with GenALEx 6.5. The final results were shown as a three-dimensional plot ( $X$ ,  $Y$ , and  $Z$  represent the three principal coordinates) of all the accessions.

## Phylogenetic Relationship

A neighbor-joining (NJ) tree of African and Asian accessions and non-admixed Mesoamerican accessions with more than 80% of the group membership in Mesoamerican structure analysis was constructed on the basis of genetic distances (Nei et al., 1983) with POPULATION 1.2.32 (Langella, 2002). The tree was drawn with Treeview 1.6.6. (Page, 1996) and Figtree 1.4.2. (Rambaut and Drummond, 2014). Genetic distances (Nei's  $D_A$  distance) between groups (admixed accessions with less than 80% of the group membership in Mesoamerican structure analysis were excluded in the analysis) were also calculated with POPULATION 1.2.32.

## Number of Retrotransposons

Presence of each of the eight retrotransposons was judged by presence of the amplified band using a FLK primer and a LTR primer, and in the case of full-length members, also by absence of the band using FLK primers at both sides of the retrotransposon. According to the results, number of retrotransposons present in each accession was counted.

## Approximate Bayesian Computation (ABC) Analysis

To characterize the Mesoamerican ancestral genetic group of African and Asian accessions, we assumed that African and Asian *Jatropha* originated from different Mesoamerican genetic groups and then constructed scenarios. To decide which scenario is the best supported by data, scenarios were compared in DIYABC v2.1.0 (Cornuet et al., 2008, 2010).

## Phenotypic Variation

Number of inflorescences per plant, number of female flowers per plant, ratio of female to male flowers and seed yield (g) per plant of 100 representative accessions from seven Mexican states [Chiapas (57), Guerrero (4), Michoacan (4), Morelos (3), Oaxaca (7), Veracruz (20), and Yucatan (5)] were measured individually at the Rosario Izapa Experimental Farm, INIFAP (Chiapas, Mexico). Chiapas accessions were further sub-classified into those from six regions [Centro (14), Frailesca (8), Fronteriza

(3), Sierra (4), Selva (2), and Soconusco (26)]. To compare the variation among accessions from central Chiapas (Centro, Fronteriza, Frailesca, and Sierra), peripheral areas of Chiapas (Selva and Soconusco), and other Mexican states, the mean value and standard deviation (SD) of each area were calculated.

## RESULTS

### Genetic Diversity in Mesoamerica

To survey the worldwide genetic variation of *Jatropha*, 59 effective SSR markers that showed the intraspecific polymorphism were selected from more than 500 *Jatropha* SSRs. The mean percentage of the missing data of SSR markers was 0.79%. We also identified six full-length members with two LTRs showing 100% sequence identity with each other, as well as two solo LTRs displaying 100% sequence identity with the consensus sequence, from *Jatropha* copia-type retrotransposon families (Alipour et al., 2013) to use as RBIP markers. All of the eight RBIP markers showed the intraspecific polymorphism. The mean percentage of the missing data of RBIP markers was 1.37%. Primers for the SSR markers and RBIP markers used in our study were listed in Supplementary Tables 2, 3. There were no or low linkage between all the markers (data not shown).

The SSR markers yielded 221 polymorphic bands ranging from two to seven with an average of 3.75 alleles per marker in all accessions. Among them, 161 alleles were specific to Mesoamerica, whereas no allele was specific to Africa or Asia (Supplementary Table 1). The state of Chiapas had 47 specific alleles, which was the highest number among Mexican states and Guatemala. In Chiapas, 30 alleles were specific to the central areas (Centro, Altos, Fronteriza, Frailesca, and Sierra), while only one allele was specific to the peripheral areas (Norte, Selva, Soconusco, and Istomo-Costa) (Supplementary Table 1).

Remarkably, 39 Asian and African accessions were almost monomorphic and homozygote in both kinds of markers, except for five Vietnamese accessions, each of which was heterozygote in a single marker. Negligible expected heterozygosity ( $H_E$ ) or genetic variation was observed in African and Asian accessions ( $H_E = 0.000$  and  $H_E = 0.002$ , respectively) (Figure 1A and Supplementary Table 1). On the contrary, a significantly higher  $H_E$  value was obtained for Mesoamerican accessions (Figure 1A and Supplementary Table 1).

In Mesoamerica, the highest genetic variation was found in Chiapas, with a clear decreasing cline of  $H_E$  from Chiapas and its bordering states to non-bordering states (Figure 1B). These results showed that the highest genetic diversity of *Jatropha* exists in Chiapas.

To further narrow down the center of genetic variation,  $H_E$  was examined among the regions within Chiapas. Five regions in central Chiapas had significantly higher  $H_E$  values than four regions in the peripheral areas ( $P < 0.001$ , Figure 1C). Slight differences between expected and observed heterozygosity ( $H_E$  versus  $H_O$ ) were observed among the accessions from central Chiapas, whereas large differences between these statistics were detected among the accessions from the peripheral areas of Chiapas, other Mexican states, and Guatemala (Figures 1B,C).

These resulted in lower inbreeding coefficients ( $F_{IS}$ ) in central Chiapas than in most of the peripheral areas of Chiapas, other Mexican states, and Guatemala (Supplementary Table 1). This may be due to human selection and/or decrease of effective population size in the areas other than central Chiapas. Interestingly, in Soconusco, the southernmost peripheral region of Chiapas,  $F_{IS}$  was the lowest, and in Sierra, one of the central regions, locating near to Soconusco,  $F_{IS}$  was the second lowest. It is likely that effect of the human selection was little in these regions, but further research would be required to prove it.

### Phenotypic Diversity in Mesoamerica

The phenotypic variation of four yield-related traits was evaluated for 100 Mexican accessions. The mean values and SDs of the accessions from central Chiapas were higher than those from the peripheral areas of Chiapas and other Mexican states (Supplementary Table 4). These suggest that *Jatropha* population in central Chiapas has the highest phenotypic variation in Mexico.

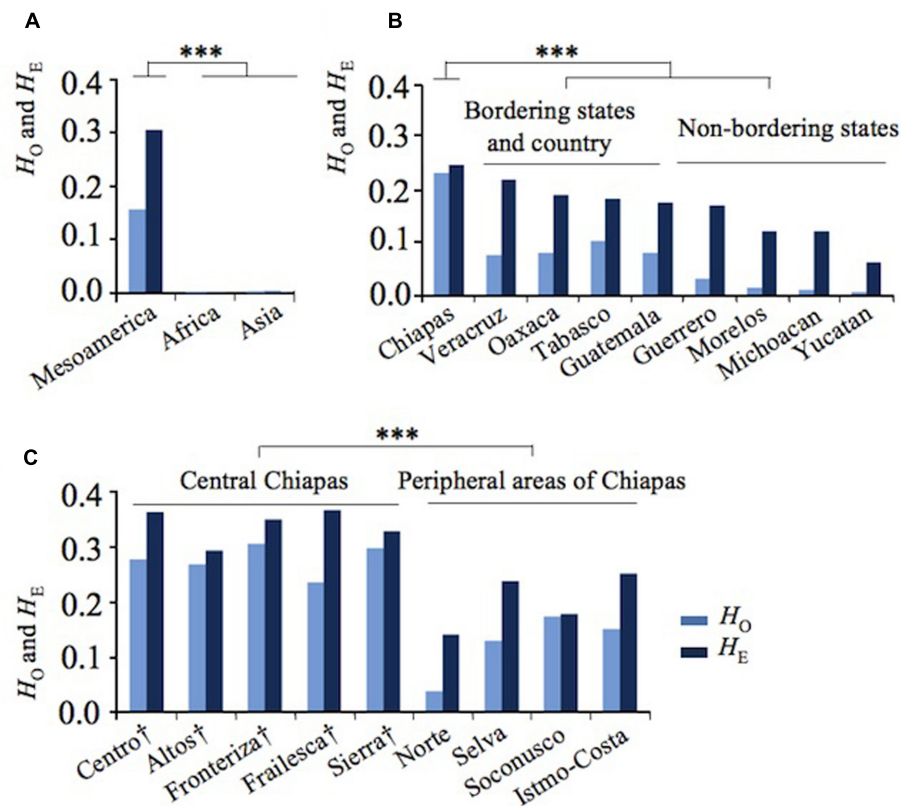
### Genetic Groups in Mesoamerica

The genetic constitutions of all the Mesoamerican accessions were examined by structure analysis. The  $\Delta K$  value with respect to  $K$  (number of groups in the population) showed a peak when  $K$  equaled to 3 (Supplementary Figure 1A; see Population Structure). This indicates that the optimal group number of all Mesoamerican accessions is 3, and then accessions from each region were classified into three clear genetic groups: A, B and C, by the model-based clustering analysis (Figure 2). The numbers and ratios of accessions assigned to each group in each Chiapas region, other Mexican states and Guatemala are presented in Figure 3. The distribution of the groups revealed a distinct geographic cline. Accessions from central Chiapas were mostly in Group A (orange). Accessions in Group B (purple) were mainly distributed in the peripheral areas of Chiapas and in neighboring states and countries, whereas accessions in Group C (blue) were mainly distributed in states distant from Chiapas (Guerrero, Michoacan, and Morelos). This geographical distribution of *Jatropha* genetic groups in Mesoamerica has not been reported before. The distribution of Group B is especially interesting, because it seems to surround the area of Group A.

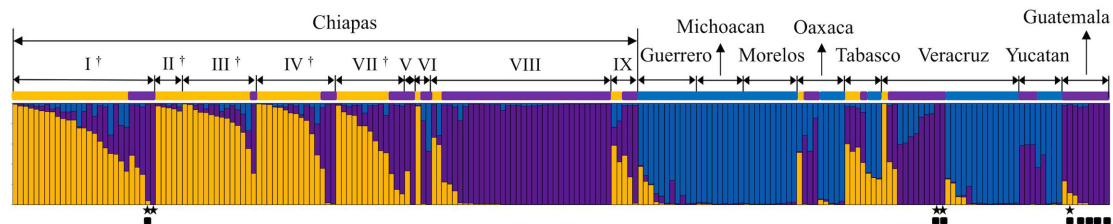
The highest genetic variation and the lowest  $F_{IS}$  among the three groups were observed in Group A (Supplementary Table 5), which was as expected from the results of the previous section. Mean values and SDs of phenotypic variations in Groups A, B, and C were also calculated and compared. Accessions from Group A had higher mean values and variations of the four traits than those from Groups B and C (Supplementary Table 4).

### Mesoamerican Accessions Genetically Closest to African and Asian *Jatropha*

To estimate the ancestral genotypes of African and Asian *Jatropha* in their origin, we searched Mesoamerican accessions that are genetically close to them. We performed population structure and phylogenetic analyses on *Jatropha* including



**FIGURE 1 |** Genetic divergence of *Jatropha* in Mesoamerica. **(A)** Observed ( $H_O$ ) and expected ( $H_E$ ) heterozygosity in Mesoamerica, Africa, and Asia. **(B)**  $H_O$  and  $H_E$  in Mexican states and Guatemala. **(C)**  $H_O$  and  $H_E$  of regions in Chiapas states. Regions in central Chiapas covering the Chiapas Central Depression are shown by †. \*\*\* $P < 0.001$ .

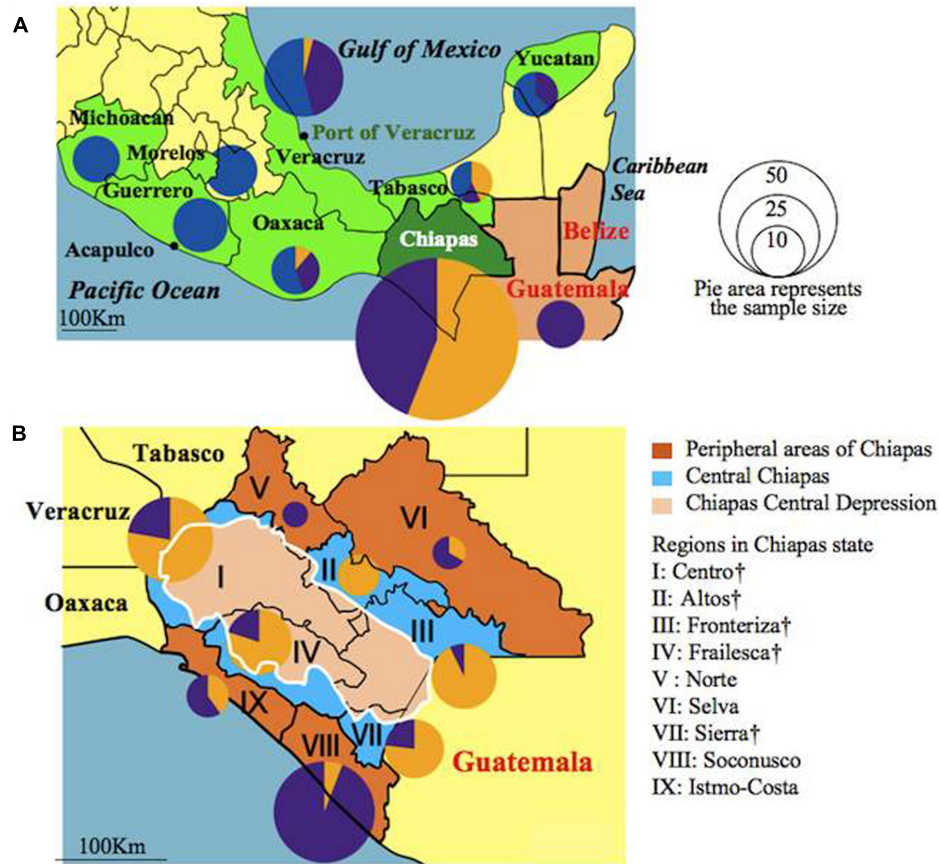


**FIGURE 2 |** Model-based clustering ( $K = 3$ ) of Mesoamerican accessions. Three genetic populations, Groups A, B and C, are indicated in orange, purple, and blue, respectively. Mesoamerican group of each accession is indicated with different group color. The origin of each accession is shown above (see Supplementary Table 1 for origins). I ~ IX show regions in the state of Chiapas (I: Centro, II: Altos, III: Fronteriza, IV: Frailesca, V: Norte, IV: Selva, VII: Sierra, VIII: Soconusco, and IX: Istmo-Costa), and regions in central Chiapas covering the Chiapas Central Depression are shown by †. ■ Eight accessions that classified into the African and Asian group (Supplementary Figure 2). ★ Accessions carrying all the eight retrotransposons.

African and Asian accessions. The  $\Delta K$  value with respect to  $K$  showed the maximum when  $K$  equaled to 3 (Supplementary Figure 1B), meaning that the optimal group number of all the accessions is 3, and then they were classified into three genetic populations: Groups I, II, and III, by the model-based clustering analysis (Supplementary Figure 2). Group II mainly corresponds to Mesoamerican Group A, whereas Group III is composed of most of Mesoamerican accessions of Groups B and C, further indicating the distinct differences between Group

A and other accessions in Mesoamerica. Group I comprises all the African and Asian accessions and eight Mesoamerican accessions: one from central Chiapas; two from Veracruz; and five from Guatemala. The genetic constitution of African and Asian accessions proved to be almost identical, indicating that they are the progenies of inbreeding through successive generations within limited ancestral genotypes. Interestingly, the eight Mesoamerican accessions in Group I were those classified into Group B in the previous section (Figure 2), and African





**FIGURE 3 |** Distribution of accessions assigned to Groups A, B, and C in Mexican states, Guatemala (A) and Chiapas regions (B) illustrated as pie graphs. Sizes of the circles correspond to the sample sizes. Mexican states were colored green. Central Chiapas is colored blue and peripheral areas of Chiapas is colored brown. The Chiapas Central Depression is colored pink. Regions in central Chiapas covering the Chiapas Central Depression are shown by †.

and Asian accessions had the closest genetic and phylogenetic relationships with Group B (Supplementary Table 6). Similarly, accessions from the peripheral areas of Chiapas, Veracruz and Guatemala had the closest relationships with African and Asian accessions (Supplementary Table 7). Results of Approximate Bayesian Computation (ABC) analysis supported our conclusion that African and Asian accessions were derived from Group B accessions (Supplementary Figure 3).

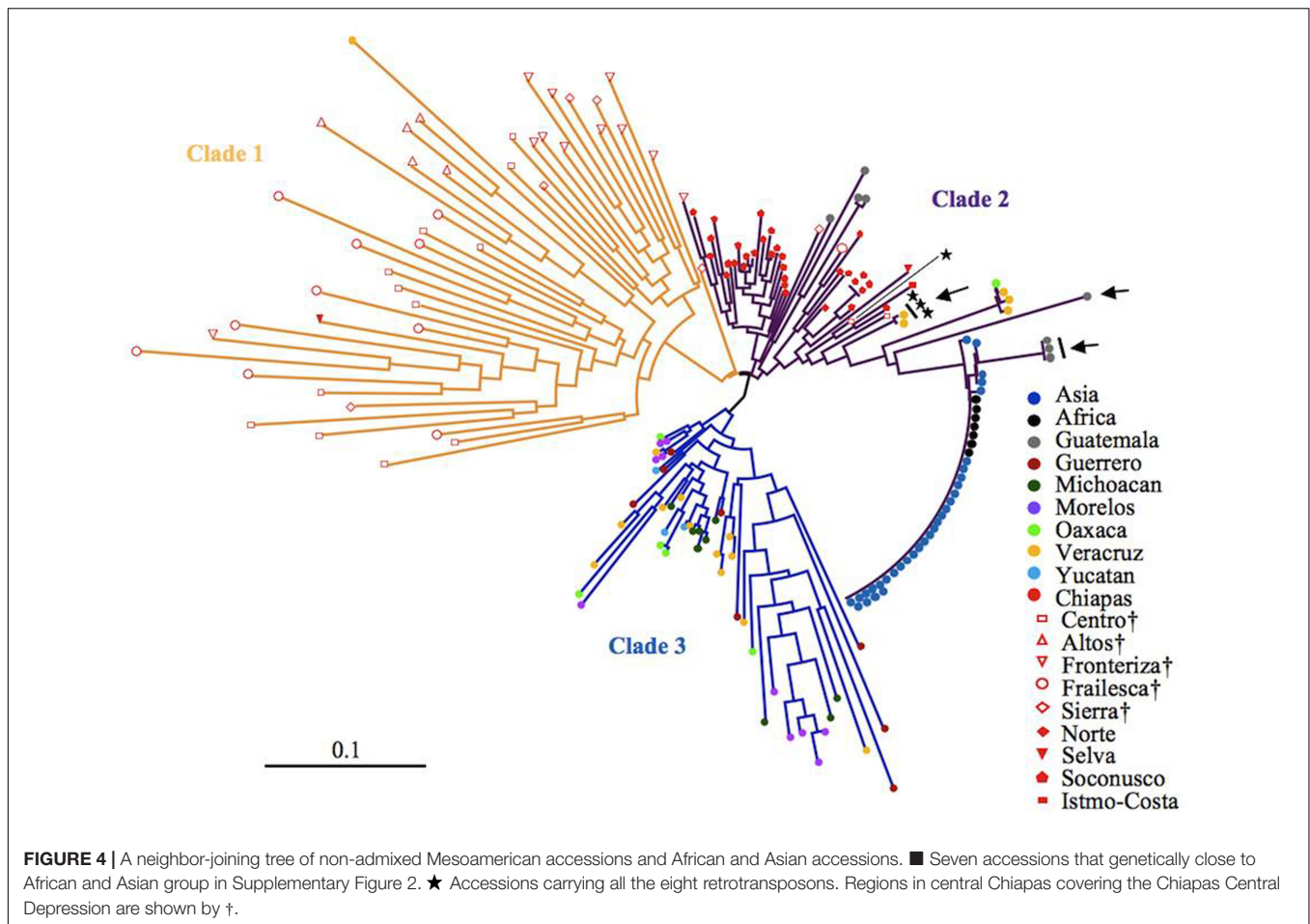
After eliminating admixed accessions which showed < 80% of the membership of each Mesoamerican Groups (see Figure 2), a phylogenetic tree of African and Asian accessions and non-admixed Mesoamerican accessions was constructed. The tree showed three clades, and interestingly, all of Mesoamerican accessions in each of the three clades (clades 1, 2, and 3) belonged to each of the three Mesoamerican Groups (Groups A, B, and C in Figure 2), respectively. One (Guatemalan accession) of the eight Mesoamerican accessions that genetically close to African and Asian groups (see above and Supplementary Figure 2) was not included in the tree because of its admixed membership. Remaining seven Mesoamerican accessions were in the same clade with all African and Asian accessions in the phylogenetic tree (Figure 4). PCoA analysis also revealed a close

relationship between these seven accessions and African and Asian ones (Supplementary Figure 4). These seven accessions are genetically closest to African and Asian *Jatropha*, and most likely share common ancestors with them. Because insertions of copia-type retrotransposons in the genome are irreversible, the presence/absence of them is a reliable indicator for determining parental lineages. Among the seven accessions, one from the Centro region (central Chiapas) and two from Veracruz shared all eight retrotransposons identified in the African and Asian accessions (Figures 2, 4, Supplementary Figure 2 and Table 8). On the other hand, the other four accessions from Guatemala lacked one of the eight retrotransposons. This observation suggests that the genotype of the three accessions [No. 127 (Soledad de Doblado, Veracruz), 210 (Entrada a Independencia, Veracruz), and 354 (Suchiapa, Centro, Chiapas)] (Figure 5) is the closest to the ancestral genotypes of African and Asian *Jatropha*.

## DISCUSSION

Aiming to identify the center of origin possessing the most diverse genetic resources of *Jatropha* that would contribute to





breeding projects in Africa and Asia, we examined genetic diversity of a worldwide collection of *Jatropha* accessions. We found that Mesoamerican populations showed significantly higher genetic diversity than African and Asian populations, and the Chiapas population showed the highest. Low genetic variation of local accessions in Africa and Asia should restrict their phenotypic performance, which would have resulted in the limitation in breeding programs. These results are consistent with previous reports (Basha and Sujatha, 2007; Basha et al., 2009; Pamidimarri et al., 2010; Rosado et al., 2010; He et al., 2011; Montes et al., 2014; Montes Osorio et al., 2014; Ouattara et al., 2014; Yue et al., 2014; Guo et al., 2016; Montes and Melchinger, 2016; Siju et al., 2016). Exceptionally, Maghuly et al. (2015) showed the most diverse population in Cape Verde by using 907 accessions from three continents. It was probably because they used only 19 Mesoamerican (Mexican) accessions from limited collection sites. Note that they found that SNPs of some genes were mostly in Mexican accessions in the same paper.

We further narrowed down the center of diversity to central Chiapas. The main part of central Chiapas is consisted of the Chiapas Central Depression (Pons et al., 2016), which is an ecoregion defined by its vegetative and geographic features. The Depression, which extends over five regions in central Chiapas, is composed of hot and semiarid lowlands dominated

by deciduous shrubs. It is sandwiched by the northern highlands, central plateau and the southern Sierra Madre Mountains. An east-west running Grijalva river divides the Depression. Besides, the northern and southern parts of Chiapas have the humid tropical climate where the rainfall is plenty. Averaged annual rainfall in some areas of these parts exceeds 3,000 mm, and there exists a large scale of the tropical rainforest. On the other hand, the central Chiapas has less humid climate, annual precipitation ranges from 750 to 1500 mm (Newton and Tejedor, 2011). The tropical rainforest in the peripheral areas of Chiapas may have prevented natural diffusion of *Jatropha* from the central part, because the optimal rainfall for *Jatropha* is from 1000 to 1500 mm, and too much rain and humidity will provoke fungus and high rainfall might cause root damage (Ouwens et al., 2007). Isolated geographical and climate situations of the Chiapas Central Depression may have favored the evolution of drought-adapted plant species. The overlap among the *Jatropha* centers of genetic and phenotypic variation, as well as the pronounced geographical and climatic conditions, suggests that the Chiapas Central Depression is most likely the center of origin of *Jatropha*.

Chiapas Central Depression has the most diverse genomic resources, thus unexplored useful natural variants or mutants can be expected there. It would be the most suitable region to



**FIGURE 5 |** *Jatropha*'s voyage from the Chiapas Central Depression to the Old World. The map shows the altitudes of Chiapas and its surrounding areas. The background map is available from [https://commons.wikimedia.org/wiki/File:Mexico\\_topographic\\_map-blank.svg](https://commons.wikimedia.org/wiki/File:Mexico_topographic_map-blank.svg). The stars represent the sampling sites of three accessions that share the same ancestor of African and Asian *Jatropha*. The red line represents a current road between Veracruz and Chiapas.

search *Jatropha* resources for breeding, which will contribute to enriching the pool of *Jatropha* for further genomic study. By using DNA markers derived from the genomic sequencing of an Asian accession (Sato et al., 2011), we could anchor the most diverse and various materials of *Jatropha*. Further genomic sequencing of Mexican accessions will lead to identifying more number of genome-wide effective DNA markers, which will encourage genomic selection or genome wide association studies (GWASs) of *Jatropha* from its center of origin by combing with phenotypic data over years. Diverse genomic resources can also be used for the study on the interactions among genes and/or genomic regions. Our identification of the center of diversity not only pointed out the center of origin, but also will inspire further genomic study of *Jatropha*.

We found genetic grouping (Groups A, B, and C) that correlated with geographic location in Mesoamerica, which has not been reported previously. Our results suggest that Group A, which distributes in central Chiapas represents the original genetic group. Specific genotypes of Group A might have spread from the Chiapas Central Depression to bordering areas, and subsequently formed distinct genotypes that comprise Group B and then C, according to their geographical distribution. We showed that African and Asian accessions were derived from

Group B, rather than Group A or C. Further ABC analysis would be useful to make this conclusion more convincing.

In our study, we found some Group B accessions from Veracruz and Guatemala were genetically close to African and Asian accessions (Supplementary Table 7). Because Veracruz and Guatemala locate in the opposite direction (west and east, respectively) of the Chiapas Central Depression, the *Jatropha* in Group B from the peripheral areas of Chiapas, Veracruz and Guatemala might have experienced some genetic diffusion and exchange before evolving into the ancestral genotype of African and Asian *Jatropha*. This exchange would have involved human selection of the specific *Jatropha* genotype with advantageous phenotypes across these regions.

Identification of the two Veracruz accessions that genetically close to African and Asian *Jatropha* showed that Veracruz was an important junction between Mesoamerica and the Old World for the *Jatropha* transmission. Veracruz has been a major port facing the Gulf of Mexico since 1519, and there was no comparable port on the Caribbean coast of Guatemala. Moreover, the routes from Veracruz to other places in Mexico were gradually developed as the importance of the port increased. During the 16th and 17th centuries, Veracruz was one of the three major American ports for transatlantic trade. There was a fixed route between the

Cape Verde Islands and Veracruz for slave transport from Africa (Thomas, 2006; Bryant, 2014). It is thus reasonable to speculate that *Jatropha* was taken to the Cape Verde Islands and then to Africa via the Port of Veracruz.

Interestingly, the two Veracruz accessions that carry all the eight retrotransposons were collected at Soledad de Doblado and Entrada a Independencia of Veracruz, both of which are within 40 and 100 km from the Port of Veracruz, respectively. A Chiapas accession that is genetically close to the African and Asian group but also carries all the retrotransposons was collected at Suchiapa region in the Chiapas Central Depression (**Figure 5**). It is also noteworthy that this Chiapas accession is in Group B, even though Group A represents the major genetic background there (**Figure 3**). The two Veracruz accessions also belong to Group B. The absence of genetically similar accessions between central Chiapas and the port of Veracruz excludes a gradual natural expansion by seeds and favors the human-mediated spread of *Jatropha* from central Chiapas to the Port of Veracruz. These also suggest that there were selections when people brought *Jatropha* from central Chiapas. One important question is as follows: who first selected the ancestral African and Asian *Jatropha* in central Chiapas? The Spanish conquistadors normally did not enter the Chiapas Central Depression, where no mineral resources were expected to occur (Ochiai, 1991). In addition, there exist various non-Spanish local names of *Jatropha* in Chiapas and Veracruz (unpublished results by Ando). Therefore, it is probable that the indigenous people of Mexico selected and carried a limited genotype of *Jatropha* seeds or cuttings in central Chiapas and transported them to the area close to the Port of Veracruz, to be the ancestor of two Veracruz accessions. Then they would have been transported to Africa and Asia via the Cape Verde Islands by Portuguese from the Port of Veracruz (**Figure 5**) (Heller, 1996).

As a limited number of genotypes would have been brought to the Old World by Portuguese galleons, crosses could have occurred only among small populations in the Old World plantings, yielding very low genetic variation. This breeding bottleneck represents a typical example of the founder effect and is probably the major cause of the failure of *Jatropha* breeding in African and Asian countries. In addition, accumulation of mildly deleterious genes exhibited in the homozygous state by inbreeding might show inbreeding depression. In fact, one of the Veracruz accessions (No. 127), which has been suggested to share a common ancestor with African and Asian accessions as discussed above, had low values for the four yield-related traits (number of inflorescences per plant: 3; number of female flowers per plant: 18; ratio of female to male flowers: 0.06 and seed yield per plant: 15.10 g) comparing with those in each region and group in Mexico (Supplementary Table 4), suggesting that the low seed-yield of African and Asian accessions is genetically determined. Moreover, Mesoamerican people seemed not to have placed any value on the seed-yielding potential of *Jatropha* as they selected and brought *Jatropha* seeds or cuttings from Chiapas to Veracruz for use as live fences and medicines (Duke, 1985; Prasad et al., 2012; Agbogidi et al., 2013), and these traditional applications are still widespread among Mesoamerican inhabitants even nowadays. Fences and medicines are the likely reasons why the

Portuguese shipped the plant to Africa and Asia (Sabandar et al., 2013). Since then, seed productivity did not have a high priority in most part of Africa and Asia until recently. After arrival to Africa and Asia, epigenetic mutations might have occurred to adapt to the local climate within 100s of years, which have caused minor phenotypic variations, as discussed in previous reports (Montes Osorio et al., 2014; Yue et al., 2014).

Based on our findings that the genetic basis of African and Asian accessions is Group B in Mesoamerica and that Group A shows the highest genetic diversity in Mesoamerica, we propose the following strategies for genetic improvement in African and Asian *Jatropha*. First, make wide crosses between elite lines from each country in Africa and Asia, which should have their epigenetic adaptations to the climatic conditions, and accessions of Group A from central Chiapas, which are genetically distant from African and Asian *Jatropha* and have the largest genetic diversity. Not only hybrid vigor, but also some extent of phenotypic variations which are derived from the highest heterozygosity of Group A, would be expected in F<sub>1</sub> progenies. Second, select elite F<sub>1</sub> or F<sub>2</sub> progenies by phenotypic performance. Finally, use vegetative propagation (Behera et al., 2014) of selected progenies to preserve excellent lines in the heterozygous status. This strategy would be better than crossing among African and Asian *Jatropha*, or than just transplanting the Mexican one, because of the higher genetic diversity, heterozygosity and climate adaptability. Production of *Jatropha* cultivars with high productivity in Africa and Asia will help in the mass commercial cultivation of *Jatropha* there. High production of *Jatropha* seeds can ensure a stable and enough supply of biofuel, which will help in the battle against the climate change.

## AUTHOR CONTRIBUTIONS

HL wrote the manuscript; HL, ST, KH, and KF discussed the results, organized and revised the manuscript; MY supported and conducted some statistical analyses; HS and NW commented the manuscript and supported the experiment; HL, AA, and TS performed the genotyping experiment; AT, HTs, TA, and HTo suggested and helped the data collection and analyses. SS and HH designed SSR markers and performed part of the marker analysis; VQ and AZ collected Mexican accessions and provided the phenotype data of Mexican accessions; PS collected and provided Philippines accessions; AH provided one Egyptian accession; AMA provided two Sudanese accessions. KF organized the project.

## ACKNOWLEDGMENTS

We would like to thank Dr. Satoshi Tabata for offering scientific assistance and Dr. Yoshiaki Kitaya for providing *Jatropha* samples from Vietnam. Plant Bioengineering for Bioenergy Laboratory was supported by Corporate Social Responsibility (CSR) Foundation of the Sumitomo Electric Industries (SEI), Ltd. We gratefully acknowledge Mr. Masayori Matsumoto,



Mr. Masanori Yoshikai, Mr. Toshifumi Hosoya, Mr. Yasuhisa Yushio, and Mr. Naoki Ikeguchi of SEI, Ltd., for their support to KF. This study was conducted in part under the Joint Research Program of the Arid Land Research Center, Tottori University, Japan and INIFAP, Mexico.

## REFERENCES

- Agbogidi, O. M., Akparobi, S. O., and Eruotor, P. G. (2013). Health and environmental benefits of *Jatropha curcas* linn. *Appl. Sci. Rep.* 1, 36–39.
- Alipour, A., Tsuchimoto, S., Sakai, H., Ohmido, N., and Fukui, K. (2013). Structural characterization of copia-type retrotransposons leads to insights into the marker development in a biofuel crop, *Jatropha curcas* L. *Biotechnol. Biofuels* 6:129. doi: 10.1186/1754-6834-6-129
- Bahadur, B., Sujatha, M., and Carels, N. (eds) (2013). *Jatropha, Challenges for a New Energy Crop: Genetic Improvement and Biotechnology*, Vol. 2. New York, NY: Springer. doi: 10.1007/978-1-4614-4915-7
- Bailis, R. E., and Baka, J. E. (2010). Greenhouse gas emissions and land use change from *Jatropha curcas*-based jet fuel in Brazil. *Environ. Sci. Technol.* 44, 8684–8691. doi: 10.1021/es1019178
- Basha, S. D., George, F., Makkar, H. P. S., Becker, K., and Sujatha, M. (2009). A comparative study of biochemical traits and molecular markers for assessment of genetic relationships between *Jatropha curcas* L. germplasm from different countries. *Plant Sci.* 176, 812–823. doi: 10.1016/j.plantsci.2009.03.008
- Basha, S. D., and Sujatha, M. (2007). Inter and intra-population variability of *Jatropha curcas* L. characterized by RAPD and ISSR markers and development of population-specific SCAR markers. *Euphytica* 156, 375–386. doi: 10.1007/s10681-007-9387-5
- Behera, L. K., Nayak, M. R., Nayak, D., and Jadeja, D. B. (2014). In vitro mass multiplication of *Jatropha (Jatropha curcas L.)* through axillary bud culture. *J. Appl. Nat. Sci.* 6, 189–192.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Castevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Bryant, S. K. (2014). *Rivers of Gold, Lives of Bondage: Governing Through Slavery in Colonial Quito*. Chapel Hill, NC: University of North Carolina Press. doi: 10.5149/northcarolina/9781469607726.001.0001
- Cornuet, J. M., Ravigné, V., and Estoup, A. (2010). Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* 11:401. doi: 10.1186/1471-2105-11-401
- Cornuet, J. M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J. M., Balding, D. J., et al. (2008). Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24, 2713–2719. doi: 10.1093/bioinformatics/btn514
- Dias, L. A. S., Missio, R. F., and Dias, D. C. F. S. (2012). Antiquity, botany, origin and domestication of *Jatropha curcas* (Euphorbiaceae), a plant species with potential for biodiesel production. *Genet. Mol. Res.* 11, 2719–2728. doi: 10.4238/2012.June.25.6
- Divakara, B. N., Upadhyaya, H. D., Wani, S. P., and Gowda, C. L. L. (2010). Biology and genetic improvement of *Jatropha curcas* L.: a review. *Appl. Energy* 87, 732–742. doi: 10.1016/j.apenergy.2009.07.013
- Duke, J. A. (1985). Medicinal plants. *Science* 229, 1036. doi: 10.1126/science.4035343
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Fairless, D. (2007). Biofuel: the little shrub that could – maybe. *Nature* 449, 652–655. doi: 10.1038/449652a
- Fargione, J., Hill, J., Tilman, D., Polasky, S., and Hawthorne, P. (2008). Land clearing and the biofuel carbon debt. *Science* 319, 1235–1238. doi: 10.1126/science.1152747
- Goudet, J. (2001). *Fstat, Version 2.9.3*. Lausanne: University of Lausanne.
- Guo, G. Y., Chen, F., Shi, X. D., Tian, Y. S., Yu, M. Q., Han, X. Q., et al. (2016). Genetic variation and phylogenetic relationship analysis of *Jatropha curcas* L. inferred from nrDNA ITS sequences. *C R Biol.* 339, 337–346. doi: 10.1016/j.crvi.2016.06.004
- He, W., King, A. J., Khan, M. A., Cuevas, J. A., Ramiamanana, D., and Graham, I. A. (2011). Analysis of seed phorbol-ester and curcumin content together with genetic diversity in multiple provenances of *Jatropha curcas* L. from Madagascar and Mexico. *Plant Physiol. Biochem.* 49, 1183–1190. doi: 10.1016/j.plaphy.2011.07.006
- Heller, J. (1996). *Physic Nut (Jatropha curcas L.): Promoting the Conservation and Use of Underutilized and Neglected Crops*. Rome: Institute of Plant Genetics and Crop Plant Research.
- Hirakawa, H., Tsuchimoto, S., Sakai, H., Nakayama, S., Fujishiro, T., Kishkohara, Y., et al. (2012). Upgraded genomic information of *Jatropha curcas* L. *Plant Biotechnol.* 29, 123–130. doi: 10.5511/plantbiotechnology.12.0515a
- Kumari, A., Mahapatra, P., Garlapati, V. K., and Banerjee, R. (2009). Enzymatic transesterification of *Jatropha* oil. *Biotechnol. Biofuels* 2:1. doi: 10.1186/1754-6834-2-1
- Langella, O. (2002). *POPULATIONS, Version 1.2.29*. Paris: CNRS.
- Maghuly, F., Jankowicz-Cieslak, J., Pabinger, S., Till, B. J., and Laimer, M. (2015). Geographic origin is not supported by the genetic variability found in a large living collection of *Jatropha curcas* with accessions from three continents. *Biotechnol. J.* 10, 536–551. doi: 10.1002/biot.201400196
- Makkar, H. P. S., and Becker, K. (2009). *Jatropha curcas*, a promising crop for the generation of biodiesel and value-added coproducts. *Eur. J. Lipid Sci. Technol.* 111, 773–787. doi: 10.1002/ejlt.200800244
- Montes, J. M., and Melchinger, A. E. (2016). Domestication and breeding of *Jatropha curcas* L. *Trends Plant Sci.* 21, 1045–1057. doi: 10.1016/j.tplants.2016.08.008
- Montes, J. M., Technow, F., Martin, M., and Becker, K. (2014). Genetic diversity in *Jatropha curcas* L. assessed with SSR and SNP markers. *Diversity* 6, 551–566. doi: 10.3390/d6030551
- Montes Osorio, L. R., Torres Salvador, A. F., Jongschaap, R. E., Azurdia Perez, C. A., Berduo Sandoval, J. E., Trindade, L. M., et al. (2014). High level of molecular and phenotypic biodiversity in *Jatropha curcas* from Central America compared to Africa, Asia and South America. *BMC Plant Biol.* 14:77. doi: 10.1186/1471-2229-14-77
- Nei, M. (1987). *Molecular Evolutionary Genetics*. New York City, NY: Columbia University Press.
- Nei, M., Tajima, F., and Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species. *J. Mol. Evol.* 19, 153–170. doi: 10.1007/BF02300753
- Newton, A. C., and Tejedor, N. (eds) (2011). *Principles and Practice of Forest Landscape Restoration: Case Studies from the Drylands of Latin America*. Gland: IUCN.
- Ochiai, K. (1991). *Les Cités perdues des Mayas (in French)*, Japanese trans. C. Baudez and S. Picasso (1987). Paris: Gallimard.
- Ouattara, B., Ndir, K. N., Gueye, M. C., Diedhiou, I., Barnaud, A., Fonckea, D., et al. (2014). Genetic diversity of *Jatropha curcas* L. in Senegal compared with exotic accessions based on microsatellite markers. *Genet. Resour. Crop Evol.* 61, 1039–1045. doi: 10.1007/s10722-014-0106-5
- Ouwens, K. D., Francis, G., Franken, Y. J., Rijssenbeek, W., Riedacker, A., Foidl, N., et al. (2007). *Position Paper on Jatropha curcas. State of the Art, Small and Large Scale Project Development*. Wageningen: FACT Foundation.
- Ovando Medina, I., Sánchez Gutiérrez, A., Adriano Anaya, L., Espinosa García, F., Núñez Farfán, J., and Salvador Figueroa, M. (2011). Genetic diversity of population of *Jatropha curcas* in the State of Chiapas, Mexico. *Diversity* 3, 641–659. doi: 10.3390/d3040641

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.01539/full#supplementary-material>



- Page, R. D. M. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Bioinformatics* 12, 357–358. doi: 10.1093/bioinformatics/12.4.357
- Pamidimarri, D. V., Mastan, S. G., Rahman, H., and Reddy, M. P. (2010). Molecular characterization and genetic diversity analysis of *Jatropha curcas* L. in India using RAPD and AFLP analysis. *Mol. Biol. Rep.* 37, 2249–2257. doi: 10.1007/s11033-009-9712-2
- Peakall, R., and Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6, 288–295. doi: 10.1093/bioinformatics/bts460
- Pecina Quintero, V., Anaya Lopez, J. L., Zamarripa Colmenero, A., Garcia, N. M., Nunez Colin, C. A., Solís Bonilla, J. L., et al. (2011). Molecular characterisation of *Jatropha curcas* L. genetic resources from Chiapas, México through AFLP markers. *Biomass Bioenergy* 35, 1897–1905. doi: 10.1016/j.biombioe.2011.01.027
- Pecina Quintero, V., Anaya López, J. L., Zamarripa Colmenero, A., Nunez Colin, C. A., Montes García, N., Solís Bonilla, J. L., et al. (2014). Genetic structure of *Jatropha curcas* L. in Mexico and probable center of origin. *Biomass Bioenergy* 60, 147–155. doi: 10.1016/j.biombioe.2013.11.005
- Pons, J. M., Vicens, E., Martínez, R., García Barrera, P., Nieto, I. E., Oviedo, A., et al. (2016). The campanian-maastrichtian rudist bivalves succession in the Chiapas Central Depression, Mexico. *Cretac. Res.* 60, 210–220. doi: 10.1016/j.cretres.2015.11.015
- Prasad, D. M. R., Izam, A., and Khan, M. M. R. (2012). *Jatropha curcas*: plant of medical benefits. *J. Med. Plant Res.* 6, 2691–2699. doi: 10.5897/JMPR10.977
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Rambaut, A., and Drummond, A. (2014). *FigTree, Version 1.4.2*. Edinburgh: University of Edinburgh.
- Rosado, T. B., Laviola, B. G., Faria, D. A., Pappas, M. R., Bhering, L. L., Quirino, B., et al. (2010). Molecular markers reveal limited genetic diversity in a large germplasm collection of the biofuel crop *Jatropha curcas* L. in Brazil. *Crop Sci.* 50, 2372–2382. doi: 10.2135/cropsci2010.02.0112
- Sabandar, C. W., Ahmat, N., Jaafar, F. M., and Sahidin, I. (2013). Medicinal property, phytochemistry and pharmacology of several *Jatropha* species (Euphorbiaceae): a review. *Phytochemistry* 85, 7–29. doi: 10.1016/j.phytochem.2012.10.009
- Salvador Figueroa, M., Magaña Ramos, J., Vázquez Ovando, J. A., Adriano Anaya, M. L., and Ovando Medina, I. (2015). Genetic diversity and structure of *Jatropha curcas* L. in its centre of origin. *Plant Genet. Resour.* 13, 9–17. doi: 10.1017/S1479262114000550
- Sanderson, K. (2009). Wonder weed plans fail to flourish. *Nature* 461, 328–329. doi: 10.1038/461328a
- Sato, S., Hirakawa, H., Isobe, S., Fukai, E., Watanabe, A., Kato, M., et al. (2011). Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. *DNA Res.* 18, 65–76. doi: 10.1093/dnares/dsq030
- Searchinger, T., Heimlich, R., Houghton, R. A., Dong, F., Elobeid, A., Fabiosa, J., et al. (2008). Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land use change. *Science* 319, 1238–1240. doi: 10.1126/science.1151861
- Siju, S., Ismanizan, I., and Wickneswari, R. (2016). Genetic homogeneity in *Jatropha curcas* L. individuals as revealed by microsatellite markers: implication to breeding strategies. *Braz. J. Bot.* 39, 861–868. doi: 10.1007/s40415-014-0117-7
- Singh, K., Singh, B., Verma, S. K., and Patra, D. D. (2014). *Jatropha curcas*: a ten year story from hope to despair. *Renew. Sustain. Energy Rev.* 35, 356–360. doi: 10.1016/j.rser.2014.04.033
- Thomas, H. (2006). *The Slave Trade: the Story of the Atlantic Slave Trade*. London: Phoenix.
- Van Eijck, J., Romijn, H., Balkema, A., and Faaij, A. (2014). Global experience with *jatropha* cultivation for bioenergy: an assessment of socio-economic and environmental aspects. *Renew. Sustain. Energy Rev.* 32, 869–889. doi: 10.1016/j.rser.2014.01.028
- Von Maltitz, G., Gasparatos, A., and Fabricius, C. (2014). The rise, fall and potential resilience benefits of *Jatropha* in Southern Africa. *Sustainability* 6, 3615–3643. doi: 10.3390/su6063615
- Von Maltitz, G. P., and Setzkorn, K. A. (2013). A typology of Southern African biofuel feedstock production projects. *Biomass Bioenergy* 59, 33–49. doi: 10.1016/j.biombioe.2012.11.024
- Yue, G. H., Lo, L. C., Sun, F., Cao, S. Y., Yi, C. X., Hong, Y., et al. (2014). No variation at 29 microsatellites in the genome of *Jatropha curcas*. *J. Genomics* 2, 59–63. doi: 10.7150/jgen.7435

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer NO declared a shared affiliation, with no collaboration, with one of the authors MY to the handling Editor.

Copyright © 2017 Li, Tsuchimoto, Harada, Yamasaki, Sakai, Wada, Alipour, Sasai, Tsunekawa, Tsujimoto, Ando, Tomemori, Sato, Hirakawa, Quintero, Zamarripa, Santos, Hegazy, Ali and Fukui. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Development of Highly Informative Genome-Wide Single Sequence Repeat Markers for Breeding Applications in Sesame and Construction of a Web Resource: SisatBase

Komivi Dossa<sup>1,2†</sup>, Jingyin Yu<sup>1†</sup>, Boshou Liao<sup>1</sup>, Ndiaga Cisse<sup>2\*</sup> and Xiurong Zhang<sup>1\*</sup>

<sup>1</sup> Key Laboratory of Biology and Genetic Improvement of Oil Crops, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Ministry of Agriculture, Wuhan, China, <sup>2</sup> Centre d'Etudes Régional pour l'Amélioration de l'Adaptation à la Sécheresse, Thiès, Senegal

## OPEN ACCESS

### Edited by:

Mariela Torres,  
Instituto Nacional de Tecnología  
Agropecuaria (INTA), Argentina

### Reviewed by:

Harsh Raman,  
NSW Department of Primary  
Industries, Australia  
Ioannis Ganopoulos,  
Institute of Plant Breeding  
and Genetic Resources-ELGO  
DEMETER, Greece

### \*Correspondence:

Xiurong Zhang  
zhangxr@oilcrops.cn  
Ndiaga Cisse  
cissendiaga02@hotmail.com

<sup>†</sup> These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 23 June 2017

**Accepted:** 08 August 2017

**Published:** 22 August 2017

### Citation:

Dossa K, Yu J, Liao B, Cisse N and  
Zhang X (2017) Development  
of Highly Informative Genome-Wide  
Single Sequence Repeat Markers  
for Breeding Applications in Sesame  
and Construction of a Web Resource:  
SisatBase. *Front. Plant Sci.* 8:1470.  
doi: 10.3389/fpls.2017.01470

The sequencing of the full nuclear genome of sesame (*Sesamum indicum* L.) provides the platform for functional analyses of genome components and their application in breeding programs. Although the importance of microsatellites markers or simple sequence repeats (SSR) in crop genotyping, genetics, and breeding applications is well established, only a little information exist concerning SSRs at the whole genome level in sesame. In addition, SSRs represent a suitable marker type for sesame molecular breeding in developing countries where it is mainly grown. In this study, we identified 138,194 genome-wide SSRs of which 76.5% were physically mapped onto the 13 pseudo-chromosomes. Among these SSRs, up to three primers pairs were supplied for 101,930 SSRs and used to *in silico* amplify the reference genome together with two newly sequenced sesame accessions. A total of 79,957 SSRs (78%) were polymorphic between the three genomes thereby suggesting their promising use in different genomics-assisted breeding applications. From these polymorphic SSRs, 23 were selected and validated to have high polymorphic potential in 48 sesame accessions from different growing areas of Africa. Furthermore, we have developed an online user-friendly database, SisatBase (<http://www.sesame-bioinfo.org/SisatBase/>), which provides free access to SSRs data as well as an integrated platform for functional analyses. Altogether, the reference SSR and SisatBase would serve as useful resources for genetic assessment, genomic studies, and breeding advancement in sesame, especially in developing countries.

**Keywords:** sesame, microsatellite, web resource, informative markers, molecular breeding

## INTRODUCTION

During the past years, the development in genetic studies and decrease of genotyping costs, have resulted in the rapid growth of the use of molecular markers (Kantartzi, 2013). Different genetic marker systems have been developed including restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism

(AFLP), sequence-related amplified polymorphism (SRAP), Diversity Arrays Technology (DArT), restriction-site associated DNA sequencing (RADseq), single-nucleotide polymorphism (SNP), specific-locus amplified fragment sequencing (SLAFseq), and random selective amplification of microsatellite polymorphic loci (RSAMPL). However, simple sequence repeats (SSR) also known as microsatellite has become the molecular marker of choice because of its versatility, operational flexibility, and low-cost. This has provided the foundation for its successful application in a wide range of fundamental and applicable fields, such as, genetic diversity, linkage/association mapping of gene/QTL, marker-assisted selection (MAS), variety identification, and evolution analysis (Jiao et al., 2012; Zhang Q. et al., 2012; Li et al., 2014; Shi et al., 2014; Dossa et al., 2016c).

SSRs are relatively short tandem repeats (STRs) of DNA that are widely distributed throughout whole genomic sequences (Sharma, 2007). They are present in coding regions but are more abundant in non-coding regions (Hancock, 1995). They are characterized by a high co-dominant inheritance, reproducibility, and multi-allelic variation (Morgante and Olivieri, 1993; Kalia et al., 2011). In addition, SSRs have been demonstrated to have several important biological functions including the regulation of chromatin organization, DNA metabolic processes, gene activity, and RNA structure (Li et al., 2002, 2004).

Sesame (*Sesamum indicum* L.) is an emerging oil crop in the world with one of the highest oil content (up to 64%) and quality (Dossa et al., 2017) among major oilseed crops. It is mainly grown in developing countries, as such, its improvement through modern molecular breeding techniques has lagged behind other oilseed crops. Up to now, different types of molecular markers have been developed and applied to sesame genotyping and breeding efforts, such as RAPD (Bhat et al., 1999; Ercan et al., 2004), inter-SSR (ISSR) (Kim et al., 2002), AFLP (Laurentin and Karlovsky, 2006), but SSR has been the preferential marker (Zhang H. et al., 2012; Zhang Y. et al., 2012; Yepuri et al., 2013; Wei et al., 2014; Dossa et al., 2016c). Although their importance in gene mapping and MAS, only few SSR markers are available for sesame research and the available ones fail to adequately represent the entire genome (Dossa, 2016). More importantly, there is no database to search for sesame SSR information at the whole genome level and to perform functional analyses, as developed in other crops such as chickpea (CicArMiSatDB: Doddamani et al., 2014), (CMsDB: Parida et al., 2015), *Cucumis melo* (CmMDb: Bhawna et al., 2015), tomato (TomSatDB: Iquebal et al., 2013), sugar beet (SBMDb: Iquebal et al., 2015), brassicas (Shi et al., 2014), etc.

The completion of the full nuclear genome sequence (Wang et al., 2014a) recently updated (Wang et al., 2016) and the newly sequenced landraces (Wei et al., 2015, 2016) provide a cardinal framework to identify highly informative SSRs at the whole genome level. In this study, we took advantage of these three genome sequence resources and provided not only a large amount of genome-wide informative SSR markers for large-scale genotyping and

breeding research in sesame, but also a user-friendly online database for convenient search and functional analyses of SSRs.

## MATERIALS AND METHODS

### Data Source

Three genome sequences of the cultivated sesame including the reference genome from the elite variety “Zhongzhi13” (Wang et al., 2014a, 2016) and the genome sequences of the landraces “Baizhima” and “Mishuozhima” (Wei et al., 2015, 2016) were downloaded from Sinbase<sup>1</sup> (Wang et al., 2014b) and SesameFG<sup>2</sup> (Wei et al., 2017), respectively. It is noteworthy that in this study, the latest version (v2) of the reference genome (Wang et al., 2016) with 13 pseudo-chromosomes (309 Mb) was employed for identifying microsatellites while the draft genome sizes of “Baizhima” and “Mishuozhima” are 267 and 254 Mb, respectively.

### Microsatellite Mining and Primer Designing

Perl scripts from MISA (Thiel et al., 2003) were used for identifying SSRs based on the reference genome sequence. Perfect microsatellites, as well as compound microsatellites interrupted by a certain number of bases were searched (Yu et al., 2016). The parameters were set for detecting mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide (nt) motifs with a minimum of 10, 6, 5, 5, 5, and 5 repeats, respectively. The compound ones were defined as  $\geq 2$  repeats interrupted by  $\leq 100$  bp. Primer3 software (Untergasser et al., 2012) was employed to design up to three primer pairs to all the identified SSRs. We named all SSRs from SiSSM1 to SiSSMxx following their order on the pseudo-chromosomes and unanchored sequences. To identify the SSRs within genic regions, the general feature format (GFF) files of genes or transcripts were combined with the positions of the SSRs located on pseudo-chromosomes. The corresponding genes or transcripts linked to each SSR, along with the biological functions were retrieved from “Sinbase.” In addition, Circos (Krzywinski et al., 2009) was used to construct the diagram of the SSR density and their genomic features in sesame.

### Electronic Polymerase Chain Reaction

The primer pairs of 105,879 microsatellites located on the 13 pseudo-chromosomes were used to *in silico* amplify the genomic sequences of “Zhongzhi13,” “Mishuozhima,” and “Baizhima,” employing the software GMATA (Wang and Wang, 2016). The primer nucleotide mismatch allowed was no more than one nucleotide and other parameters were set as default. The polymorphic primers were selected based on difference in number of repeat-units present in the three genomes.

<sup>1</sup>[www.ocri-genomics.org/Sinbase/index.html](http://www.ocri-genomics.org/Sinbase/index.html)

<sup>2</sup><http://ncgr.ac.cn/SesameFG/>

**TABLE 1** | Characteristics of SSRs identified in the whole genome of sesame.

SSR mining		Total	
Total number of sequence scaffolds examined		4,449	
Total number of identified SSRs		138,194	
Number of sequence scaffolds containing SSR		1,279	
Number of sequence scaffolds containing more than 1 SSR		877	
Number of compound SSRs		28,666	
Number of SSRs present in genic regions		20,167	
	Repeat type	Number of SSRs	Percentage
	Mono-nucleotide	67,949	49.17
	Di-nucleotide	59,886	43.33
	Tri-nucleotide	9,116	6.60
	Tetra-nucleotide	933	0.68
	Penta-nucleotide	148	0.11
	Hexa-nucleotide	162	0.12
Total		138,194	100

**TABLE 2** | Chromosome wise distribution of SSR types in the sesame genome.

Chromosomes	Perfect types						Compound types	Total	Percent (%)
	Mono-	Di-	Tri-	Tetra-	Penta-	Hexa-			
chr1	3935	2498	462	57	7	8	1222	8189	7.73
chr2	4057	2438	457	45	10	13	1194	8214	7.76
chr3	5553	3174	588	72	11	16	1722	11136	10.52
chr4	4048	2469	502	55	20	7	1364	8465	7.99
chr5	3523	1945	357	37	3	9	984	6858	6.48
chr6	5164	3022	580	58	14	8	1502	10348	9.77
chr7	3157	1805	354	32	11	5	914	6278	5.93
chr8	4858	2952	589	62	11	8	1531	10011	9.46
chr9	4703	3071	644	60	7	16	1508	10009	9.45
chr10	3895	2216	400	44	5	10	1104	7674	7.25
chr11	2706	1756	309	31	7	5	872	5686	5.37
chr12	3116	1916	365	36	11	10	955	6409	6.05
chr13	3415	1798	360	45	3	5	977	6603	6.24
Total	52130	31060	5967	634	120	120	15849	105880	100.00

## Plant Materials and DNA Extraction

A total of 48 accessions of the cultivated sesame (*S. indicum* L.,  $2n = 26$ ), comprising of landraces and modern cultivars grown in 12 countries of West, Central, and East Africa, were used in this study (Supplementary Table S1).

Leaves from 2 weeks old single seedling per accession were used for DNA isolation using the cetyltrimethylammonium bromide (CTAB) according to method described by Dossa et al. (2016c). DNA quality and quantity were assessed on 1.5% agarose gel and by spectrophotometry (NanoDrop 2000, Thermo Scientific, Wilmington, DE, United States), respectively. DNA samples were stored at  $-20^{\circ}\text{C}$ , for further use.

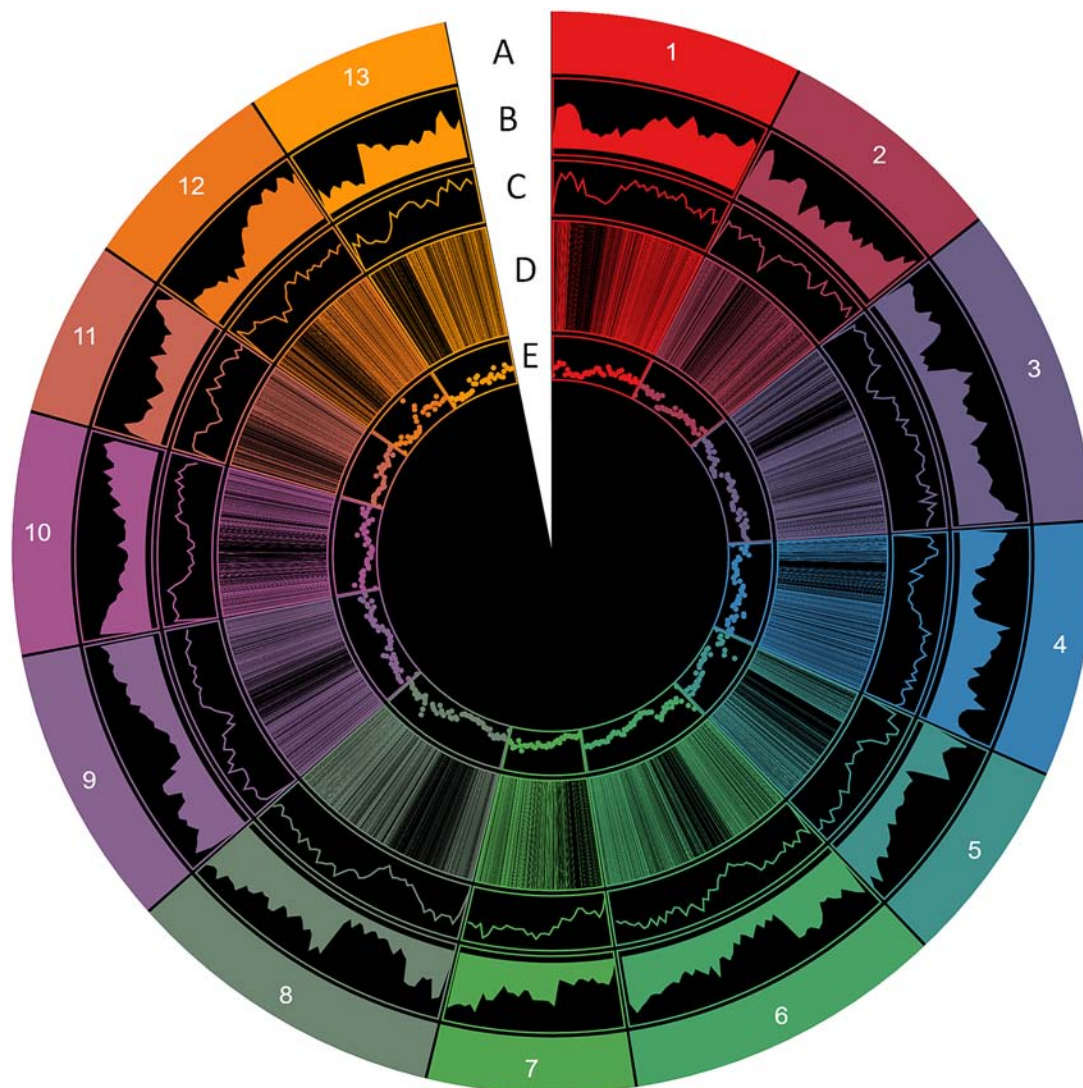
## Polymerase Chain Reaction, Electrophoresis, and Data Analysis

A subset of 23 SSR markers providing coverage across all the 13 pseudo-chromosomes was selected from the entire polymorphic markers identified through electronic polymerase chain reaction

(e-PCR), to validate their polymorphism potential between the 48 sesame accessions. PCR was conducted as described by Dossa et al. (2016c). Briefly, PCR was performed in a total volume of 15  $\mu$ L containing 30 ng of DNA, 1 pmol of each primer, 0.2 U Taq DNA polymerase and 2 $\times$  reaction mix supplied with the dNTPs and MgCl<sub>2</sub>. The PCR cycles were 94°C (5 min), 35 cycles of 94°C (30 s), 55°C (30 s), 72°C (30 s), followed by the extension step for 5 min at 72°C. The PCR amplicon sizes were scored in base pairs (bp) based on migration relative to the internal size standard of 400HD-ROX (Applied Biosystems, Foster, CA, United States) on an ABI 3130xl Genetic Analyzer (Applied Biosystems). Additionally, the amplified products were also electrophoretically separated on 1.5% agarose gel in TAE buffer and stained with ethidium bromide.

The number of alleles (Na), major allele frequency (MAF), and polymorphic information content (PIC) were calculated with the software PowerMarker version 3.25 (Liu and Muse, 2005). Moreover, to identify the pair-wise genetic relationships between





**FIGURE 1 |** Distribution of SSR information in the sesame genome. **(A)** The 13 chromosomes, **(B)** gene density, **(C)** SSR density, **(D)** polymorphic SSRs, **(E)** genic SSRs (window of 500 kb).

the 48 accessions, a neighbor-joining (NJ) tree based on Nei genetic distance (Nei, 1972) was drawn in MEGA version 7 (Kumar et al., 2016).

### Development of SisatBase

The process of SisatBase development can be divided into two steps: (i) integration and consolidation of microsatellites data and (ii) developing SisatBase and embedding useful tools.

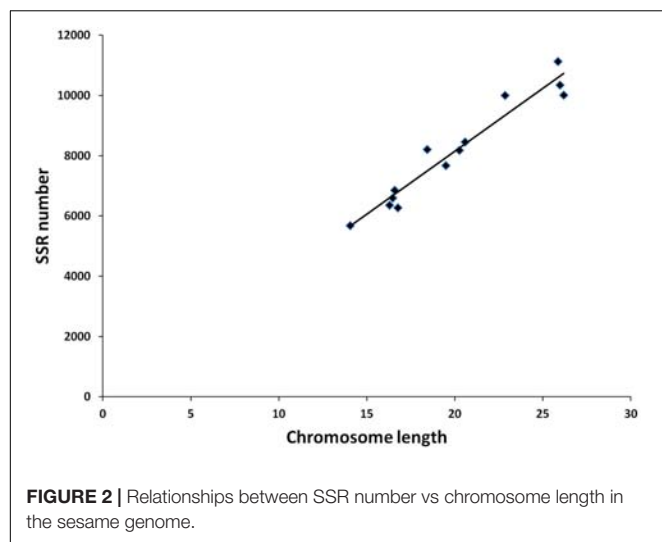
The datasets were curated to create a logic relationship among the different types of microsatellite data for their integration in SisatBase. Thereafter, SisatBase was developed using the LMAP (Linux + Apache + Mysql + Perl/PHP/Python) web application program platform. The HyperText Markup Language (HTML) and JavaScript language were also used to develop a user-friendly web interface. With the aim to enrich the functions of SisatBase, Browse, Search, customized BLAST, and MISAweb

were developed for users to browse, search, and identify SSRs in the sesame genome conveniently (Altschul et al., 1997; Stein, 2013).

## RESULTS

### Identification, Characteristics, and Genomic Distribution of SSRs in the Sesame Genome

A total of 138,194 non-redundant microsatellites were identified from 4,449 sequence scaffolds representing 94.3% of the assembled genome of sesame with an average of 507 microsatellites per Mb (Table 1). Mono-nucleotide and di-nucleotide SSRs were the most represented repeat types

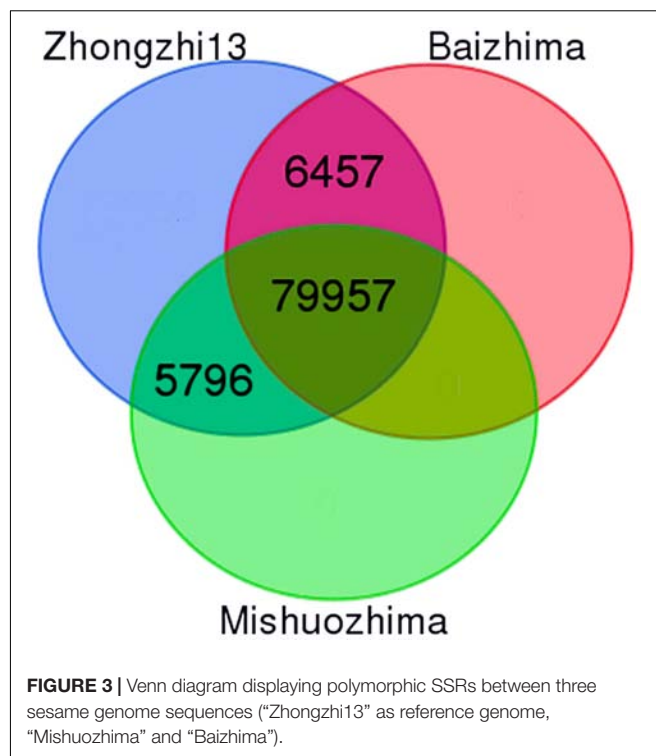


(92.5% of the whole genome SSRs) with 79% as perfect SSR types, while the remaining were in compound forms. The most prevalent motif types were A/T, accounting for 91.85% of the total mono-nucleotide repeats. For di-nucleotide motifs, the dominant motif was “AT” accounting for 50.38% of the total di-nucleotide repeats. Overall, the dominant/major motifs (A, AT, AAG/AAT, AAAT, AAAAT, and AAAAAT) were all A/T rich, whereas the absent/scarc motifs were mostly C/G rich.

From these microsatellites, 76.5% (105,880 SSRs) were successfully mapped onto the 13 pseudo-chromosomes (“chr”) of the sesame genome (Table 2 and Figure 1). Overall, SSRs are distributed throughout the “chr” with some regions exhibiting higher density than others. The chr3 displayed the highest number of SSRs (10.5% of all mapped SSRs) followed by chr6, chr8, and chr9 accounting for 9.78, 9.46, and 9.46% of the all mapped SSRs, respectively. The chr11 harbored the lowest number of SSRs (5,686; 7.74%). Based on the physical location of each SSR and the GFF files of genes or transcripts, we uncovered that 18.84% of the total mapped SSRs were located in genic regions. Next, we estimated the relationship between the “chr” length and the number of SSRs harbored on each “chr” and found a high correlation ( $r^2 = 0.94$ ) (Figure 2).

### Primer Designing and e-PCR Based Polymorphic Screening of the Developed SSRs among Three Sesame Genome Sequences

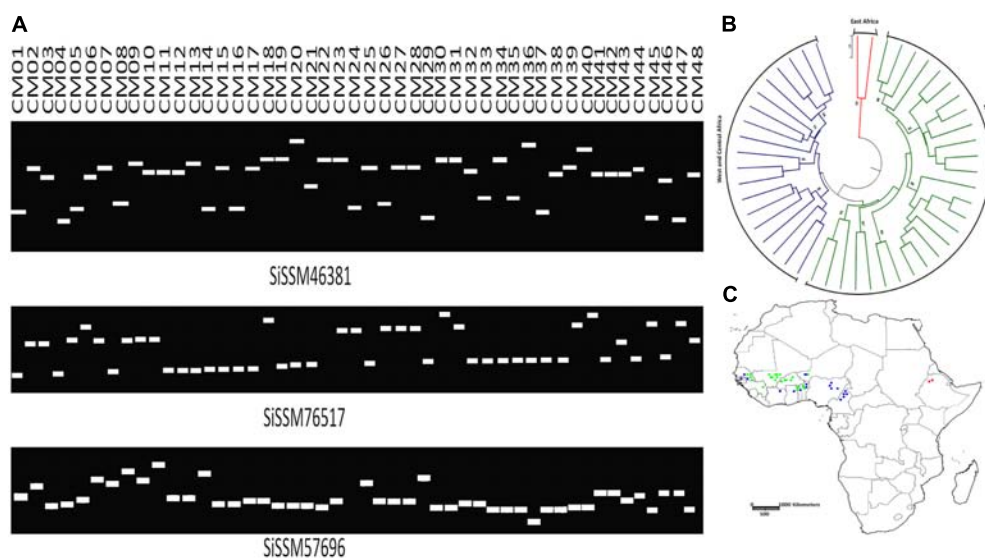
With the release of new genome sequences from two landraces (“Baizhima” and “Mishuozhima”), it is now possible to provide at the whole genome level a set of polymorphic SSRs. First, we successfully designed up to three primer pairs from flanking sequences of 104,617 SSRs (98.80% of all SSRs). Secondly, we extracted 101,930 SSRs with primers (97.4%) which were located on the 13 “chr.” Thirdly, we *in silico* amplified the three genomes mentioned above with the 101,930 SSRs. A total of 92,210 SSRs (90.5%) was conserved between the three genomes including 79.1% of total genic SSR markers. From these



SSRs, 86,414 (93.7%) were polymorphic between “Zhongzhi13” and “Baizhima,” 85,753 (93%) showed polymorphism between “Zhongzhi13” and “Mishuozhima” and finally 79,957 (86.7%) SSRs were extracted as informative markers since they were polymorphic between the three genotypes (Figure 3). It is worthy to mention that the number of SSRs exhibiting polymorphism decreased with the increase of SSR repeat-length variation.

### Amplification and Polymorphic Potential of Selected SSRs among 48 Sesame Accessions

We selected within the 79,957 informative markers, 23 SSRs from all the 13 “chr” with the aim of confirming their allelic variation between 48 sesame genotypes. Interestingly, only two markers did not amplify three accessions probably due to DNA quality issue. More importantly, all markers (100%) were polymorphic between the 48 sesame accessions. In total, 123 distinct alleles were obtained ranging from three (SiSSM105280, SiSSM11029, SiSSM35870, SiSSM61314, SiSSM59616, SiSSM78138, SiSSM91614, and SiSSM104985) to nine alleles (SiSSM46381) with an average allele number of 4.24 per locus. The mean MAF and PIC were estimated at 0.51 and 0.60, respectively (Figure 4A and Table 3). Based on the Nei’s genetic distance between the 48 accessions, we constructed a NJ tree which divided the germplasm into three main groups (Figure 4B). Some geographical clustering patterns could be observed: the first group named “East Africa” gathered together the two accessions from Ethiopia. The second cluster called “West Africa” was composed of only West African accession from Senegal, Niger, Togo, Burkina Faso, Guinea, Benin, and Mali. The



**FIGURE 4 |** Amplification and polymorphic potential of selected SSRs among 48 sesame accessions. **(A)** Gel pictures showing polymorphic potential of some selected SSRs and allelic variations among 48 sesame genotypes. **(B)** NJ tree of the accessions used based on polymorphic SSRs. Bootstrap values  $\geq 85$  are shown. **(C)** Mapping the 48 sesame accessions based on their geographical regions. The different colors assigned represent the different clusters identified by NJ tree. Green, West Africa; blue, West and Central Africa; red, East Africa.

last group named West and Central Africa, clustered together the accessions from Nigeria, Cameroon, Senegal, Ghana, Benin, Ivory Coast, Niger, and Togo (**Figure 4C**).

## SisatBase: An Online Database for SSR Functional Analysis in Sesame

In order to facilitate the exploitation of the SSRs at the whole genome level in sesame, we developed an online database with an easy-to-use interface<sup>3</sup> (**Figure 5A**). SisatBase supplied basic information for SSRs, including location on chromosomes, SSR type, SSR size and up to three primer pairs for each SSR entry, as well as the functional genes associated with the SSRs. Except that, SisatBase also provided the polymorphic SSRs among different sesame genotypes. In addition, SisatBase supplied useful search tools, including keyword, SSR type, and SSR location searches, which can help users to obtain their interested SSR information (**Figures 5B,C**). Customized BLAST and MISAweb were also embedded in SisatBase to help users to get or identify conveniently SSR with primers in their interested genomic regions or genomic sequences (**Figure 5D**).

## DISCUSSION

While the integration of molecular marker technologies have significantly improved the speed and precision of modern plant breeding, the molecular research in sesame has lagged behind other model crops mainly because sesame is a minor crop often grown by smallholders in developing countries. Hence, highly informative molecular marker systems with the advantage of easy

and low-cost detection are capital for sesame breeding research. Microsatellite markers constitute undoubtedly the best candidate and in this study, we identified 138,194 SSRs at the whole genome level, along with their primer pairs and genome location.

The number of SSRs identified and the SSR density were higher than previous reports in sesame, mainly, because the genomic sequences examined in this study are more important (Wei et al., 2014; Uncu et al., 2015; Dossa, 2016). Furthermore, by exploiting the latest version of the reference genome, we are able to provide the accurate position of SSRs in the sesame genome compared with previous reports. This would be helpful for gene fine-mapping and association analysis in sesame. Mono-nucleotide and di-nucleotide repeats accounted for 92.5% of the whole genome SSRs in sesame. Our results are in agreement with conclusions of Cardle et al. (2000) and Sonah et al. (2011), who identified mono-nucleotide and di-nucleotide repeats as the predominant repeat types in several plant genomes including *Arabidopsis thaliana*, *Brachypodium distachyon*, *Sorghum bicolor*, *Oryza sativa*, *Medicago truncatula*, and *Populus trichocarpa*. Similarly to previous reports of Wei et al. (2014), Uncu et al. (2015), and Dossa (2016), the distribution of A/T rich motif as the major motif is highly in accordance with the AT (0.68%) vs GC (0.32%) content in the sesame genome (Wang et al., 2014a). The same findings were also observed in *Brassica rapa* (Xu et al., 2010; Shi et al., 2014), *Brassica napus* (Cheng et al., 2009), *Brassica oleracea* (Li et al., 2011), cucumber (Cavagnaro et al., 2010). The high correlation of SSR number and pseudo-chromosome length suggested that this type of DNA considerably increase the length of the sesame pseudo-chromosomes.

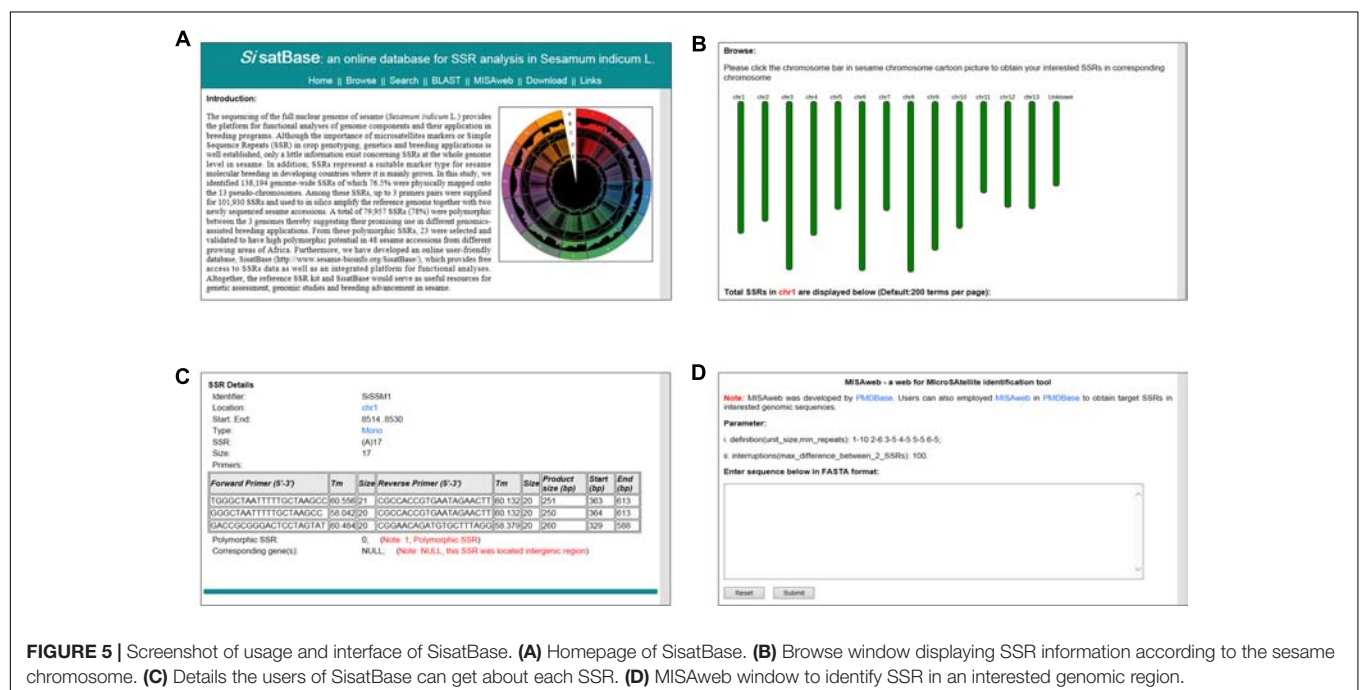
In sesame, SSRs were more concentrated in the intergenic regions compared to genic regions which is consistent with findings in *Sativa japonica* (Zhang et al., 2007), maize

<sup>3</sup><http://www.sesame-bioinfo.org/SisatBase/>



**TABLE 3 |** Polymorphism information of the 23 selected SSR markers.

Chr	SSR name	Start position (bp)	End position (bp)	Reverse primer	Forward primer	Allele number (Na)	MAF	PIC
chr1	SiSSM5580	14227382	14227544	GCTTCCACCTAGCTCGGTTAT	CCAGCAATCATGTCTGCTTAAT	4	0.61	0.6
chr1	SiSSM6522	16318566	16318595	CGTGTGCCCAATATTTGAGTT	TCAACCTCCTCCCTACACAA	4	0.74	0.6
chr2	SiSSM11029	7480687	7480700	TTGAATTTGATCTTTCCATCA	TGGACAAAGACACAATCACACA	3	0.48	0.2
chr2	SiSSM105280	5683964	5683988	GGAGATGATTGATTCTTTTGA	GAAGAACAGATCGTTGGGCT	3	0.63	0.5
chr3	SiSSM22288	15034466	15034505	GCAGTGGGAGTGAGAAGAGG	TAGTGATTCCCATCGCCCT	4	0.38	0.7
chr4	SiSSM35870	20255324	20255341	TGCATTTAAGGCTGTGCAAC	CCAGACCCAAACCAATAGA	3	0.48	0.7
chr5	SiSSM37640	3290073	3290100	TTTGGCAAACCTGCAATGAA	CATTAAACACATTACGAAACA	6	0.49	0.8
chr6	SiSSM46381	8881867	8881935	TGCACTGCATTGTCTCCTTT	TGCAAGGACAACCAAAATCA	9	0.45	0.8
chr7	SiSSM57696	13029902	13029961	GTCAAAATTGAGGGTTGCGT	TTCTGTACCAAGAATTGCG	7	0.35	0.8
chr8	SiSSM61314	4072390	4072409	TTCCAATTCTACAAGCGCAG	CCGATCAAACTAGTATGGCAA	3	0.44	0.6
chr8	SiSSM59616	391702	391727	TCATTAACCCATCATTGCGA	TGCTCACACATAACAGTTGGG	3	0.57	0.4
chr9	SiSSM76517	16006845	16006882	TCCTGAATTCAAACGCATTG	TCCTAAACCCTCTGCACCAC	8	0.59	0.7
chr9	SiSSM78138	19556070	19556099	AGCAACGATTCACGACATTG	CAACACCACCAACGCATATC	3	0.28	0.3
chr10	SiSSM84645	13921579	13921594	GATTTTGACACCTTTGCCTGA	AAAATCCTCTTTTCCGACGA	4	0.36	0.5
chr10	SiSSM86610	18019054	18019133	ACACATACGGACAGGCACAG	ATATAGCCAGTTTGGCTGCG	4	0.47	0.7
chr11	SiSSM91614	11169519	11169554	CCAGCTCTATTGTGCGTTGA	CACTGCTTTCTCTGAAAGGCT	3	0.6	0.5
chr12	SiSSM95212	6801590	6801617	AATTGGACTCCGGCTAGGAT	CGCCCTCATCCTTACAATCT	5	0.75	0.6
chr12	SiSSM95090	6354025	6354048	AGGAAGGAGGGTGTCCTAA	CCCTCTCAAATAAGCCCTC	5	0.48	0.8
chr12	SiSSM97651	12374603	12374737	CGCCTTTCTCCTCCTTATCC	CATTGCTTCTACGTCCAATTTCT	5	0.85	0.6
chr12	SiSSM97727	12601031	12601067	ACTGCACCCTCTGCATTTT	GCACGTGTGGGGTACCTTTA	5	0.36	0.6
chr13	SiSSM104985	14705386	14705400	GGCCAAACCTTTTCAGATTT	ATGCTCTGTGCTGATTGGTG	3	0.56	0.5
chr13	SiSSM100596	5393137	5393166	TCGAGTTGGAATGCAACAAA	CAAGTCGCCATCACACTCAT	5	0.55	0.7
chr13	SiSSM100938	6125899	6125920	TCCAATCAGTTAGGTCGAG	TTAAGCTTAGGGTCTGGGTT	4	0.23	0.5
Mean						4	0.51	0.60
Max						9	0.85	0.80
Min						3	0.23	0.20

**FIGURE 5 |** Screenshot of usage and interface of SsatBase. **(A)** Homepage of SsatBase. **(B)** Browse window displaying SSR information according to the sesame chromosome. **(C)** Details the users of SsatBase can get about each SSR. **(D)** MISAweb window to identify SSR in an interested genomic region.



(Xu et al., 2013), and other crops (Hancock, 1995). The landraces “Baizhima” and “Mishuozhima” exhibited similar polymorphic rates with the genome of “Zhongzhi13.” This suggested that the two landraces are much closer to each other than the elite variety “Zhongzhi13.” Our findings are in agreement with the conclusions of Wei et al. (2016) who found that the two landraces clustered together and were more closely related in the phylogenetic tree compared to “Zhongzhi13.” We further discovered that the majority of genic SSRs in the sesame genome have been found within the conserved markers between the three genotypes. This result is understandable given that SSRs within genic regions are associated with genes which constitute the genome component more conserved within species (Xiao et al., 2016). On the other hand, this implies that the conserved set of SSRs might be related to important genes which were retained during improvement from landraces to elite cultivar, as demonstrated in soybean (Zhou et al., 2015). Therefore, we infer that these genic informative microsatellites may be linked to some important biological functions and could be potential tools for sesame breeding (Lata et al., 2014; Dossa et al., 2016a,b).

In our knowledge, there are no specific molecular markers developed for other related species in the *Sesamum* genus. It has been demonstrated that SSR markers have a good transferability between species of the same genus or even in the same taxa (Fan et al., 2013; Buso et al., 2016; Huang et al., 2016; Thakur et al., 2017). In sesame, Uncu et al. (2015) uncovered a high rate of SSR marker transferability between the cultivated species *S. indicum* and the proposed wild ancestor species *S. malabaricum*. In addition, different sets of SSR markers developed in the cultivated sesame also yielded good amplicons in the wild-related species including *Sesamum radiatum*, *S. angustifolium*, *S. latifolium*, *S. angolense* (Zhang H. et al., 2012; Nyongesa et al., 2013; Wu et al., 2014). Based on these reports, we speculate that our developed informative SSR markers might be relevant for other wild-related species of the *Sesamum* genus. This will be significant for the genetic improvement of the cultivated form by exploiting the potential of the wild-related species (Dossa et al., 2017). Such transferable SSR markers between *Sesamum*-related species could be used for conducting macro-syntenic studies, genetic mapping, and molecular breeding. Therefore, in future studies, we will employ several wild-related species of the *Sesamum* genus as well as a diverse panel of the cultivated sesame to evaluate the cross-species transferability of our developed SSR markers and initiate genetic researches in the wild-related species of the *Sesamum* genus.

Although some SSR sets have been previously identified in the sesame genome, transcriptome, etc. (Spandana et al., 2012; Zhang H. et al., 2012; Wei et al., 2014; Uncu et al., 2015; Dossa, 2016), information regarding their amplification efficiency and polymorphic potential is limited. In the present study, we took advantage of the three available sequenced genomes to screen for amplification efficiency and polymorphism potential of our developed SSR markers. This led to the identification of 79,957 informative SSR markers of which 23 selected SSRs successfully discriminated 48 genotypes from Africa based on their geographical origins. This result suggested that e-PCR

is a useful strategy for a rapid screening and an effective identification of informative markers (Wang and Wang, 2016; Xiao et al., 2016). In the works of Dossa et al. (2016c), 33 polymorphic SSRs were employed to assess the genetic diversity of 96 sesame accessions from Africa and Asia which resulted in a high genetic diversity within the African germplasm. The 23 selected SSRs used in the present study to scan the diversity of 48 African accessions were all polymorphic and yielded comparable alleles number (123 vs 137) although fewer genotypes were examined here. Similarly, a high genetic diversity was also observed in the studied germplasm proving that the global 79,957 informative SSR markers could be effectively considered as the reference SSR for large-scale genotyping and molecular breeding research in sesame (Billot et al., 2012).

All SSR data were integrated into SisatBase which also supplied useful and user-friendly tools to assist users to extract more information related to SSR markers in the sesame genome. The database will be continuously updated with new versions of the sesame genome. Moreover, with the aim of extending the utility of SisatBase over other species of the *Sesamum* genus, new information about the cross-species transferable SSR markers as well as novel and specific SSRs for each species will be supplied in the future.

## CONCLUSION

In conclusion, based on the latest version of the sesame reference genome and the two newly released genome sequences, we identified 138,194 SSRs of which 79,957 are proposed as the reference SSR for future genetics/genomics and breeding studies in sesame. All microsatellite data reported in this study are integrated into a user-friendly online database (SisatBase) for a convenient exploitation and further functional analyses. These tools will undoubtedly help to speed-up sesame molecular breeding especially in the developing countries.

## AUTHOR CONTRIBUTIONS

KD and JY produced the sesame SSR data, developed the online database, and drafted the manuscript. KD performed the experiments. BL, NC, and XZ designed the project, supervised the works, and revised the draft manuscript. All authors have read and approved the final manuscript.

## FUNDING

This study was financially supported by The China Agriculture Research System (CARS-15) and The Agricultural Science and Technology Innovation Project of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2013-OCRI). The first author is grateful to the fellowship offered by the Chinese Scholarship Council (2015GXY934).

## ACKNOWLEDGMENTS

We acknowledge Mr. Li Donghua for helping in some figure configuration and Mr. Thomas Roberts for the language editing assistance.

## REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Bhat, K. V., Babrekar, P. P., and Lakhanpaul, S. (1999). Study of genetic diversity in Indian and exotic sesame (*Sesamum indicum* L.) germplasm using random amplified polymorphic DNA (RAPD) markers. *Euphytica* 110, 21–33. doi: 10.1023/A:1003724732323
- Bhawna, Chaduvula, P. K., Bonthala, V. S., Manjusha, V., Siddiq, E. A., Polumetla, A. K., et al. (2015). CmmDB: a versatile database for *Cucumis melo* microsatellite markers and other horticulture crop research. *PLoS ONE* 10:e0118630. doi: 10.1371/journal.pone.0118630
- Billot, C., Rivallan, R., Ndoye, S. M., Fonceka, D., Deu, M., Glaszmann, J. C., et al. (2012). A reference microsatellite kit to assess for genetic diversity of *Sorghum bicolor* (Poaceae). *Am. J. Bot.* 99, e245–e250. doi: 10.3732/ajb.1100548
- Buso, G. S. C., Reis, A. M. M., Amaral, Z. P. S., and Ferreira, M. E. (2016). Novel and highly informative capsicum SSR markers and their cross-species transferability. *Genet. Mol. Res.* 15, 1–13. doi: 10.4238/gmr.15038689
- Cardle, L., Ramsay, L., Milbourne, D., Macaulay, M., Marshall, D., and Waugh, R. (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* 156, 847–854.
- Cavagnaro, P. F., Senalik, D. A., Yang, L., Simon, P. W., Harkins, T. T., Kodira, C. D., et al. (2010). Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genomics* 11:569. doi: 10.1186/1471-2164-11-569
- Cheng, X., Xu, J., Xia, S., Gu, J., Yang, Y., Fu, J., et al. (2009). Development and genetic mapping of microsatellite markers from genome survey sequences in *Brassica napus*. *Theor. Appl. Genet.* 118, 1121–1131. doi: 10.1007/s00122-009-0967-8
- Doddamani, D., Katta, M. A. V. S. K., Khan, A. W., Agarwal, G., Shah, T. M., and Varshney, R. K. (2014). CicArMiSatDB: the chickpea microsatellite database. *BMC Bioinformatics* 15:212. doi: 10.1186/1471-2105-15-212
- Dossa, K. (2016). A physical map of important QTLs, functional markers and genes available for sesame breeding programs. *Physiol. Mol. Biol. Plants* 22, 613–619. doi: 10.1007/s12298-016-0385-8
- Dossa, K., Diouf, D., and Cissé, N. (2016a). Genome-wide investigation of Hsf genes in sesame reveals their segmental duplication expansion and their active role in drought stress response. *Front. Plant Sci.* 7:1522. doi: 10.3389/fpls.2016.01522
- Dossa, K., Diouf, D., Wang, L., Wei, X., Zhang, Y., Niang, M., et al. (2017). The emerging oilseed crop *Sesamum indicum* enters the “Omics” era. *Front. Plant Sci.* 8:1154. doi: 10.3389/fpls.2017.01154
- Dossa, K., Wei, X., Li, D., Zhang, Y., Wang, L., Fonceka, D., et al. (2016b). Insight into the AP2/ERF transcription factor superfamily in sesame (*Sesamum indicum*) and expression profiling of the DREB subfamily under drought stress. *BMC Plant Biol.* 16:171. doi: 10.1186/s12870-016-0859-4
- Dossa, K., Wei, X., Zhang, Y., Fonceka, D., Yang, W., Diouf, D., et al. (2016c). Analysis of genetic diversity and population structure of sesame accessions from Africa and Asia as major centers of its cultivation. *Genes* 7:14. doi: 10.3390/genes7040014
- Ercan, A. G., Taskin, M., and Turgut, K. (2004). Analysis of genetic diversity in Turkish sesame (*Sesamum indicum* L.) populations using RAPD markers. *Genet. Resour. Crop Evol.* 51, 599–607. doi: 10.1023/B:GRES.0000024651.45623.f2
- Fan, L., Zhang, M.-Y., Liu, Q.-Z., Li, L.-T., Song, Y., Wang, L.-F., et al. (2013). Transferability of newly developed pear SSR markers to other *Rosaceae* species. *Plant Mol. Biol. Rep.* 31, 1271–1282. doi: 10.1007/s11105-013-0586-z
- Hancock, J. M. (1995). The contribution of slippage-like processes to genome evolution. *J. Mol. Evol.* 41, 1038–1047.
- Huang, L., Wu, B., Zhao, J., Li, H., Chen, W., Zheng, Y., et al. (2016). Characterization and transferable utility of microsatellite markers in the wild and cultivated *Arachis* species. *PLoS ONE* 11:e0156633. doi: 10.1371/journal.pone.0156633
- Iqbal, M. A., Jaiswal, S., Angadi, U. B., Sablok, G., Arora, V., Kumar, S., et al. (2015). SBMDb: first whole genome putative microsatellite DNA marker database of sugarbeet for bioenergy and industrial applications. *Database* 2015:bav111. doi: 10.1093/database/bav111
- Iqbal, M. A., Sarika, Vasu, A. A., Verma, N., Rai, A., and Kumar, D. (2013). First whole genome based microsatellite DNA marker database of tomato for mapping and variety identification. *BMC Plant Biol.* 13:197. doi: 10.1186/1471-2229-13-197
- Jiao, Y., Jia, H. M., Li, X. W., Chai, M. L., Jia, H. J., Chen, Z., et al. (2012). Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*). *BMC Genomics* 13:201. doi: 10.1186/1471-2164-13-201
- Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., and Dhawan, A. K. (2011). Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 177, 309–334. doi: 10.1007/s10681-010-0286-9
- Kantartzi, S. K. (2013). *Microsatellites: Methods and Protocols*. New York, NY: Springer. doi: 10.1007/978-1-62703-389-3
- Kim, D. H., Zur, G., Danin-Poleg, Y., Lee, S. W., Shim, K. B., Kang, C. W., et al. (2002). Genetic relationships of sesame germplasm collection as revealed by inter-simple sequence repeats. *Plant Breed.* 121, 259–262. doi: 10.1046/j.1439-0523.2002.00700.x
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Lata, C., Mishra, A. K., Muthamilarasan, M., Bonthala, V. S., Khan, Y., and Prasad, M. (2014). Genome-wide investigation and expression profiling of AP2/ERF transcription factor superfamily in Foxtail Millet (*Setaria italica* L.). *PLoS ONE* 9:e113092. doi: 10.1371/journal.pone.0113092
- Laurentin, H. E., and Karlovsky, P. (2006). Genetic relationship and diversity in a sesame (*Sesamum indicum* L.) germplasm collection using amplified fragment length polymorphism (AFLP). *BMC Genet.* 7:10. doi: 10.1186/1471-2156-7-10
- Li, C., Miao, H., Wei, L., Zhang, T., Han, X., and Zhang, H. (2014). Association mapping of seed oil and protein content in *Sesamum indicum* L. using SSR markers. *PLoS ONE* 9:e105757. doi: 10.1371/journal.pone.0105757
- Li, H., Chen, X., Yang, Y., Xu, J., Gu, J., Fu, J., et al. (2011). Development and genetic mapping of microsatellite markers from whole genome shotgun sequences in *Brassica oleracea*. *Mol. Breed.* 28, 585–596. doi: 10.1007/s11032-010-9509-y
- Li, Y. C., Korol, A. B., Fahima, T., Beiles, A., and Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Ecol.* 11, 2453–2465.
- Li, Y. C., Korol, A. B., Fahima, T., and Nevo, E. (2004). Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21, 991–1007. doi: 10.1093/molbev/msh073
- Liu, K. J., and Muse, S. V. (2005). PowerMarker, an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282
- Morgante, M., and Olivieri, A. M. (1993). PCR-amplified microsatellites as markers in plant genetics. *Plant J.* 3, 175–182.
- Nei, M. (1972). Genetic distance between populations. *Am. Nat.* 106, 283–292.
- Nyongesa, B. O., Were, B. A., Gudu, S., Dangasuk, O. G., and Onkware, A. O. (2013). Genetic diversity in cultivated sesame (*Sesamum indicum* L.) and related wild species in East Africa. *J. Crop Sci. Biotechnol.* 16, 9–15. doi: 10.1007/s12892-012-0114-y

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.01470/full#supplementary-material>

- Parida, S. K., Verma, M., Yadav, S. K., Ambawat, S., Das, S., Garg, R., et al. (2015). Development of genome-wide informative simple sequence repeat markers for large-scale genotyping applications in chickpea and development of web resource. *Front. Plant Sci.* 6:645. doi: 10.3389/fpls.2015.00645
- Sharma, P. (2007). Mining microsatellites in eukaryotic genomes. *Trends Biotechnol.* 25, 490–498. doi: 10.1016/j.tibtech.2007.07.013
- Shi, J., Huang, S., Zhan, J., Yu, J., Wang, X., Hua, W., et al. (2014). Genome-wide microsatellite characterization and marker development in the sequenced *Brassica* crop species. *DNA Res.* 21, 53–68. doi: 10.1093/dnares/dst040
- Sonah, H., Deshmukh, R. K., Sharma, A., Singh, V. P., Gupta, D. K., Gacche, R. N., et al. (2011). Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium*. *PLoS ONE* 6:e21298. doi: 10.1371/journal.pone.0021298
- Spandana, B., Reddy, V. P., Prasanna, G. J., Anuradha, G., and Sivaramakrishnan, S. (2012). Development and characterization of microsatellite markers (SSR) in *Sesamum indicum* L. species. *Appl. Biochem. Biotechnol.* 168, 1594–1607. doi: 10.1007/s12010-012-9881-7
- Stein, L. D. (2013). Using GBrowse 2.0 to visualize and share next-generation sequence data. *Brief. Bioinform.* 14, 162–171. doi: 10.1093/bib/bbt001
- Thakur, A. K., Singh, K. H., Singh, L., Nanjundan, J., Khan, Y. J., and Singh, D. (2017). SSR marker variations in *Brassica* species provide insight into the origin and evolution of *Brassica* amphidiploids. *Hereditas* 155, 6. doi: 10.1186/s41065-017-0041-5
- Thiel, T., Michalek, W., Varshney, R. K., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106, 411–422. doi: 10.1007/s00122-002-1031-0
- Uncu, A. Ö., Gultekin, V., Allmer, J., Frary, A., and Doganlar, S. (2015). Genomic simple sequence repeat markers reveal patterns of genetic relatedness and diversity in sesame. *Plant Genome* 8, 1–12. doi: 10.3835/plantgenome2014.11.0087
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40:e115. doi: 10.1093/nar/gks596
- Wang, L., Xia, Q., Zhang, Y., Zhu, X., Zhu, X., Li, D., et al. (2016). Updated sesame genome assembly and fine mapping of plant height and seed coat color QTLs using a new high-density genetic map. *BMC Plant Biol.* 17:31. doi: 10.1186/s12864-015-2316-4
- Wang, L., Yu, J., Li, D., and Zhang, X. (2014a). Sinbase: an integrated database to study genomics, genetics and comparative genomics in *Sesamum indicum*. *Plant Cell Physiol.* 56, e2. doi: 10.1093/pcp/pcu175
- Wang, L., Yu, S., Tong, C., Zhao, Y., Liu, Y., Song, C., et al. (2014b). Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol.* 15:R39. doi: 10.1186/gb-2014-15-2-r39
- Wang, X., and Wang, L. (2016). GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. *Front. Plant Sci.* 7:1350. doi: 10.3389/fpls.2016.01350
- Wei, X., Gong, H., Yu, J., Liu, P., Wang, L., Zhang, Y., et al. (2017). Sesame FG: an integrated database for the functional genomics of sesame. *Sci. Rep.* 7:2342. doi: 10.1038/s41598-017-02586-3
- Wei, X., Liu, K., Zhang, Y., Feng, Q., Wang, L., Zhao, Y., et al. (2015). Genetic discovery for oil production and quality in sesame. *Nat. Commun.* 6:8609. doi: 10.1038/ncomms9609
- Wei, X., Wang, L., Zhang, Y., Qi, X., Wang, X., Ding, X., et al. (2014). Development of simple sequence repeat (SSR) markers of sesame (*Sesamum indicum*) from a genome survey. *Molecules* 19, 5150–5162. doi: 10.3390/molecules19045150
- Wei, X., Zhu, X., Yu, J., Wang, L., Zhang, Y., Li, D., et al. (2016). Identification of sesame genomic variations from genome comparison of landrace and variety. *Front. Plant Sci.* 7:1169. doi: 10.3389/fpls.2016.01169
- Wu, K., Yang, M., Liu, H., Tao, Y., Mei, J., and Zhao, Y. (2014). Genetic analysis and molecular characterization of Chinese sesame (*Sesamum indicum* L.) cultivars using Insertion-Deletion (InDel) and Simple Sequence Repeat (SSR) markers. *BMC Genet.* 15:35. doi: 10.1186/1471-2156-15-35
- Xiao, Y., Xia, W., Ma, J., Mason, A. S., Fan, H., Shi, P., et al. (2016). Genome-wide identification and transferability of microsatellite markers between Palmae species. *Front. Plant Sci.* 7:1578. doi: 10.3389/fpls.2016.01578
- Xu, J., Liu, L., Xu, Y., Chen, C., Rong, T., Ali, F., et al. (2013). Development and characterization of simple sequence repeat markers providing genome-wide coverage and high resolution in maize. *DNA Res.* 20, 497–509. doi: 10.1093/dnares/dst026
- Xu, J., Qian, X., Wang, X., Li, R., Cheng, X., Yang, Y., et al. (2010). Construction of an integrated genetic linkage map for the A genome of *Brassica napus* using SSR markers derived from sequenced BACs in *B. rapa*. *BMC Genomics* 11:594. doi: 10.1186/1471-2164-11-594
- Yepuri, V., Surapaneni, M., Kola, V., Vemireddy, L. R., Jyothi, B., Dineshkumar, V., et al. (2013). Assessment of genetic diversity in sesame (*Sesamum indicum* L.) genotypes, using EST-derived SSR markers. *J. Crop Sci. Biotechnol.* 16, 93–103. doi: 10.1007/s12892-012-0116-9
- Yu, J., Dossa, K., Wang, L., Zhang, Y., Wei, X., Liao, B., et al. (2016). PMDBase: a database for studying microsatellite DNA and marker development in plants. *Nucleic Acids Res.* 45, D1046–D1053. doi: 10.1093/nar/gkw906
- Zhang, H., Wei, L., Miao, H., Zhang, T., and Wang, C. (2012). Development and validation of genic-SSR markers in sesame by RNA-seq. *BMC Genomics* 13:316. doi: 10.1186/1471-2164-13-316
- Zhang, Q., Ma, B., Li, H., Chang, Y., Han, Y., Li, J., et al. (2012). Identification, characterization, and utilization of genome-wide simple sequence repeats to identify a QTL of acidity in apple. *BMC Genomics* 13:537. doi: 10.1186/1471-2164-13-537
- Zhang, Y. X., Zhang, X. R., Che, Z., Wang, L. H., Wei, W. L., and Li, D. H. (2012). Genetic diversity assessment of sesame core collection in China by phenotype and molecular markers and extraction of a mini-core collection. *BMC Genet.* 13:102. doi: 10.1186/1471-2156-13-102
- Zhang, Z., Deng, Y., Tan, J., Hu, S., Yu, J., and Xue, Q. (2007). A genome-wide microsatellite polymorphism database for the Indica and Japonica rice. *DNA Res.* 14, 37–45. doi: 10.1093/dnares/dsm005
- Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33, 408–441. doi: 10.1038/nbt.3096

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Dossa, Yu, Liao, Cisse and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The First Molecular Identification of an Olive Collection Applying Standard Simple Sequence Repeats and Novel Expressed Sequence Tag Markers

Soraya Mousavi<sup>1</sup>, Roberto Mariotti<sup>2</sup>, Luca Regni<sup>3</sup>, Luigi Nasini<sup>3</sup>, Marina Bufacchi<sup>1</sup>, Saverio Pandolfi<sup>2</sup>, Luciana Baldoni<sup>2</sup> and Primo Proietti<sup>3\*</sup>

<sup>1</sup> Consiglio Nazionale delle Ricerche – Institute for Agricultural and Forest Systems in the Mediterranean, Perugia, Italy,

<sup>2</sup> Consiglio Nazionale delle Ricerche – Institute of Biosciences and Bioresources, Perugia, Italy, <sup>3</sup> Department of Agricultural, Food and Environmental Sciences, Università degli Studi di Perugia, Perugia, Italy

## OPEN ACCESS

### Edited by:

Mariela Torres,  
Instituto Nacional de Tecnología  
Agropecuaria (INTA), Argentina

### Reviewed by:

Rosario Muleo,  
Università degli Studi della Tuscia, Italy  
Raul De La Rosa,  
Andalusian Institute of Agricultural  
and Fisheries Research and Training,  
Spain

### \*Correspondence:

Primo Proietti  
primo.proietti@unipg.it

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 21 March 2017

**Accepted:** 07 July 2017

**Published:** 19 July 2017

### Citation:

Mousavi S, Mariotti R, Regni L,  
Nasini L, Bufacchi M, Pandolfi S,  
Baldoni L and Proietti P (2017)  
The First Molecular Identification of an  
Olive Collection Applying Standard  
Simple Sequence Repeats and Novel  
Expressed Sequence Tag Markers.  
Front. Plant Sci. 8:1283.  
doi: 10.3389/fpls.2017.01283

Germplasm collections of tree crop species represent fundamental tools for conservation of diversity and key steps for its characterization and evaluation. For the olive tree, several collections were created all over the world, but only few of them have been fully characterized and molecularly identified. The olive collection of Perugia University (UNIPG), established in the years' 60, represents one of the first attempts to gather and safeguard olive diversity, keeping together cultivars from different countries. In the present study, a set of 370 olive trees previously uncharacterized was screened with 10 standard simple sequence repeats (SSRs) and nine new EST-SSR markers, to correctly and thoroughly identify all genotypes, verify their representativeness of the entire cultivated olive variation, and validate the effectiveness of new markers in comparison to standard genotyping tools. The SSR analysis revealed the presence of 59 genotypes, corresponding to 72 well known cultivars, 13 of them resulting exclusively present in this collection. The new EST-SSRs have shown values of diversity parameters quite similar to those of best standard SSRs. When compared to hundreds of Mediterranean cultivars, the UNIPG olive accessions were splitted into the three main populations (East, Center and West Mediterranean), confirming that the collection has a good representativeness of the entire olive variability. Furthermore, Bayesian analysis, performed on the 59 genotypes of the collection by the use of both sets of markers, have demonstrated their splitting into four clusters, with a well balanced membership obtained by EST respect to standard SSRs. The new OLEST (*Olea* expressed sequence tags) SSR markers resulted as effective as the best standard markers. The information obtained from this study represents a high valuable tool for *ex situ* conservation and management of olive genetic resources, useful to build a common database from worldwide olive cultivar collections, also based on recently developed markers.

**Keywords:** genetic variability, *ex situ* conservation, germplasm management, genotyping, EST-SSR, *Olea europaea* L.



## INTRODUCTION

The cultivated olive (*Olea europaea*, subsp. *europaea*, var. *europaea*, Green, 2002) is one of the most important oil crops in the world and 95% of total olive oil production derives from the Mediterranean basin (Marra et al., 2013; Trujillo et al., 2014). The olive crop counts a very rich varietal heritage, represented by more than 1,200 named cultivars, over 3,000 minor cultivars and an uncertain number of genotypes including pollinators, local ecotypes and centennial trees (El Bakkali et al., 2013; Hosseini-Mazinani et al., 2014; Mazzitelli et al., 2015; Laroussi-Mezghani et al., 2016; Mousavi et al., 2017). Since time of ancient Greece, olive cultivars have been vegetatively propagated, either by cutting or grafting, allowing the accurate reproduction of the best-performing genotypes, leading to the present varietal assortment (Breton et al., 2009; Kaniewski et al., 2012). Thus, most cultivars represent ancient pre-bred genotypes, and the limited and sporadic genetic improvement initiatives, with classical or biotechnological approaches, forced the retention of numerous traditional cultivars despite their agronomical limitations. Among these, only a few have a large area of cultivation and a clear impact on the production of oil and table olives (Ilarioni and Proietti, 2014). But the availability of a large set of well characterized and highly different cultivars is critical to increase the ability to face new agronomical challenges (De Gennaro et al., 2012; Larbi et al., 2015) and future climatic constraints (Moriondo et al., 2013; Proietti et al., 2014; Tanasijevic et al., 2014), diversifying the gene pools, preserving unique genetic traits currently available (Bracci et al., 2011; Corrado et al., 2011; Potts et al., 2012; Klepo et al., 2013) and offering different sensory profiles of extra-virgin olive oils.

Several reasons make it difficult to ensure the identification of cultivars, as the joint cultivation of native and foreign cultivars, the ambiguous plant naming, seedlings or wild plants, or the interchange of plant material over the centuries (Marra et al., 2013; Lazović et al., 2016). Furthermore, the large number of cultivars, the high degree of kinship among many of them, mainly in cases of geographic proximity, and the possible appearance of clonal variation, have raised additional identification problems (Belaj et al., 2007; Caruso et al., 2014; Ipek et al., 2015).

Olive collections represent the main tool to preserve and certify germplasm resources (Belaj et al., 2012; Caruso et al., 2014), mainly when recent trends toward establishing modern orchards exclusively based on a few highly producing and low-vigor cultivars, may potentially lead to the erosion of this germplasm. More than 100 collections of olive genetic resources have been established at international, national and regional levels for conservation and evaluation purposes (Trujillo et al., 2014). A first World Olive Germplasm Bank (WOGB) was established since the years' 70 at IFAPA (Cordoba, Spain), with about 500 accessions from 21 countries (Belaj et al., 2011; Trujillo et al., 2014). In 2003, a second WOGB was created at INRA (Marrakech, Morocco), including 560 accessions originating from 14 Mediterranean countries (Haouane et al., 2011). An international olive collection built by CNR (ISAFOM) and planted in Zagaria (Enna, Italy), includes about 400 cultivars collected worldwide (Las Casas et al., 2014). A national collection

has been built by CREA-OLI (Cosenza, Italy), consisting of approximately 500 cultivars from Italy, corresponding to 85% of total Italian olive germplasm (Muzzalupo et al., 2014). In Turkey, a national olive germplasm collection in Izmir contains 96 genotypes (Kaya et al., 2013), whereas the Greek National Olive Germplasm Collection counts on 47 olive cultivars (Xanthopoulou et al., 2014). Also new olive growing countries, such as the United States of America, have organized important olive collections (NCGR-Davis, CA, United States) (Zelasco et al., 2012), as well as Argentina, Chile, Uruguay, Australia, China, and South Africa (Trujillo et al., 2014). In addition to these important gene banks, many other minor collections were set up along the time to preserve dedicated pools of genotypes, such as cultivars with specific characteristics, wild plants, segregating progenies or core collections (Belaj et al., 2011, 2012; Díez et al., 2011; Marchese et al., 2016). Among these, the UNIPG (Perugia University, Italy) collection, established 50 years ago, represents one of the first attempts to collect and conserve *ex situ* a large number of olive cultivars. It contains genotypes of different geographical origin (although with prevalence of Italian cultivars), and holds great potential for the complete agronomic and exhaustive evaluation of cultivars, as reported by numerous previous works on agronomical, morphological or biological varietal performance (Breton et al., 2014; Portarena et al., 2015).

Simple sequence repeats were the main molecular markers used to characterize the olive germplasm collection (Haouane et al., 2011; Muzzalupo et al., 2014; Trujillo et al., 2014). In fact, SSRs represent the most popular markers for olive genotyping, due to the high polymorphism, extraordinary abundance and fast transferability (Sarri et al., 2006; Baldoni et al., 2009; Díez et al., 2011; Belaj et al., 2012; Hosseini-Mazinani et al., 2014; Mousavi et al., 2014, 2017). However, all SSR loci published so far, characterized by dinucleotide repeat motifs, have demonstrated several drawbacks due to the difficult discrimination among alleles (Baldoni et al., 2009). On the contrary, EST-SSRs derive from expressed regions of the genome, have a greater transferability among species and, since they are located within genes, their variation could find correlation with the phenotype (Duran et al., 2009). However, EST-SSRs may reveal less variations and lower polymorphic information than standard SSRs, even though sufficient for population genetic analysis and for genotyping purpose (Yang et al., 2013). For this reason, new trinucleotidic EST-SSR loci recently identified (Mariotti et al., 2016) should now be widely applied for a more clear varietal characterization.

In this work, we have provided the first molecular identification of the accessions present in the UNIPG olive varietal collection. The identification of all olive trees was performed by standard SSRs and, for the first time in olive collections, by EST-SSRs. We intended to reach numerous important goals: (1) the identification of all accessions, including those closely related or morphologically similar, (2) the evaluation of discrimination power between EST-SSRs and dinucleotide standard SSRs, and (3) establishing the level and wideness of the genetic variability inside a germplasm collection, in order to make available this important source of well-defined genotypes to all interested stakeholders and researchers.

## MATERIALS AND METHODS

### Sample Collection and Archival Records

The Olive Varietal Collection of the University of Perugia – Department of Agricultural, Food and Environmental Sciences (UNIPG) – is located in Prepo, Perugia (43°04′ 53.94″ N – 12°22′ 53.25″ E, altitude about 400 m asl), on a clay soil, with medium content of organic matter, phosphorus and potassium, temperate-Mediterranean climate, average annual temperatures of 12.8°C and annual rainfall of about 900 mm. Planting distance is 5 × 5 m and trees are grown polyconic vase-shaped. Regular agricultural practices are applied to the olive plants, without irrigation. The collection, established in 1965, has been duplicated in 1984 and enlarged by adding further local, national and international cultivars. Based on the UNIPG archive, the collection consisted of 370 olive plants, where each genotype was represented at least by three replications, randomly distributed in a single block, although, some cultivars (Carolea, Maurino, Moraiolo, Leccino, Frantoio, San Felice, Nostrale di Rigali and Manzanilla de Sevilla) were represented by at least 20 trees per cultivar, distributed in four randomized blocks, allowing for their agronomical and morpho-bio-phenological evaluation. No information was available on the original source of plant material.

### DNA Extraction and Molecular Analysis

Leaf samples were collected from each plant, for a total of 370 accessions, and plant position of each tree was recorded. For each accession, total DNA was extracted from fresh leaves following the standard manufacturer's instructions of GeneElute Plant Genomic DNA Miniprep Kit (Sigma–Aldrich).

All samples were analyzed by using nine best ranked EST-SSR markers (OLEST1-7-9-12-14-16-20-22-23) recently developed (Mariotti et al., 2016). Double step polymerase chain reactions (PCR) were performed in a volume of 25 µl containing 25 ng of DNA, 10× PCR buffer, 200 µM of each dNTP, 10 pmol of primer forward (with 18 bp tail in 5′) and reverse, and 2 U of DNA Polymerase (Q5 High Fidelity DNA Polymerase, New England Biolabs). In the second step, fluorescent tail (10 pmol) was annealed to the forward primer using a double step PCR: the first step consisting in an initial denaturation at 95°C for 5 min, followed by 35 cycles of 95°C for 30 s, 60°C for 30 s and 72°C for 25 s, the second step (for tail annealing) made up of 20 cycles, with the same conditions of the first step except for annealing temperature ( $T_m = 52^\circ\text{C}$ ), a final elongation at 72°C for 40 min closed the second step PCR.

In order to verify the identity of cultivars present in the collection, all samples were genotyped by using standard dinucleotide SSRs markers, widely applied for cultivar characterization in most olive germplasm collections (Haouane et al., 2011; Muzzalupo et al., 2014; Trujillo et al., 2014). Ten high polymorphic markers were applied, including DCA3-5-9-16-18, EMO90, GAPU71B-101-103A and UDO-043 (Sefc et al., 2000; Carriero et al., 2002; Cipriani et al., 2002), previously selected as best performing loci (Baldoni et al., 2009) and common to the other genotyping works. Forward primers carried VIC, FAM, PET, or NED labels at their 5′-end. Standard PCR amplifications

were performed in a reaction volume of 25 µl containing 25 ng of DNA, 10× PCR buffer, 200 µM of each dNTP, 10 pmol of each forward and reverse primer, and 2 U of Q5 High-Fidelity DNA Polymerase (New England Biolabs), with an initial denaturation at 95°C for 5 min, followed by 40 cycles of 95°C for 30 s, annealing temperature as suggested by authors (50–60°C) for 30 s and 72°C for 25 s, followed by a final elongation at 72°C for 40 min.

Polymerase chain reactions products were loaded on an ABI 3130 Genetic Analyzer (Applied Biosystems–Hitachi) using the internal GeneScan 500 LIZ Size Standard (Thermo Fisher Scientific). Output data were analyzed by GeneMapper 3.7 (Applied Biosystems).

In order to verify the match of the 370 olive samples with previously characterized cultivars, the data obtained for the 10 standard SSR markers were compared to those available in the database of olive SSR profiles established at CNR-IBBR of Perugia (Italy), including more than 1,000 worldwide olive cultivars, and to other available datasets (Baldoni et al., 2009; Trujillo et al., 2014), allowing to establish cultivar identity and determine all cases of identical profiles, presumably corresponding to clonal genotypes with undetermined presence of mutationsclonal replicates (Baldoni et al., 2009, 2011; Bartolini, 2009; Mousavi et al., 2017).

### Allele Frequency and Diversity Analysis

Number of alleles per locus ( $N_a$ ), number of effective alleles ( $N_e$ ), Shannon's information index ( $I$ ), observed ( $H_o$ ) and expected heterozygosity ( $H_e$ ), and fixation index ( $F$ ) were calculated at each locus for novel and standard SSRs by the use of GenAlEx 6.501 software (Peakall and Smouse, 2012). Pairwise relatedness was performed on standard and OLEST SSR markers to calculate the allelic similarity for codominant data using GenAlEx 6.501 following the  $LRM = \text{Lynch and Ritland (1999) estimator} - \text{Mean}$  multiplied by 2 to give max of 1.00. The software FreeNA (Chapuis and Estoup, 2007) was applied to detect the presence of possible null alleles ( $F_{null}$ ), to determine the genetic uniqueness of each accession and to quantify redundancy. Polymorphic information content (PIC) was calculated for each microsatellite locus using CERVUS v.3.0 software (Marshall et al., 1998). We calculated the probabilities of identity for unrelated individuals [ $P(ID)$ ] at each locus and across loci, as described by Waits et al. (2001), by using GenAlEx for both OLEST and standard SSR markers. Cumulative  $P(ID)$  was calculated by ranking the PIC values at each locus from high to low. We used the criterion of  $P(ID)$  lower than 0.001 for the estimation of the minimum number of loci required for individual identification in the study species (Waits et al., 2001).

A model-based Bayesian clustering method was applied to infer the genetic structure of 59 cultivars and to define the number of clusters in the dataset (gene pools) using the software STRUCTURE v.2.3 (Pritchard et al., 2009), for the same sample set separately for OLEST and standard SSRs. Tests were based on an admixture model with independent allele frequencies. No prior information was used to define clusters. Independent runs were done by setting the number of clusters ( $k$ ) from 1 to 10. Each run comprised a burn-in length of 100,000 followed by

100,000 MCMC (Monte Carlo Markov Chain) replicates. An *ad hoc* statistic  $\Delta K$ , based on the rate of change in the log probability of data between successive  $K$  values, as described by Evanno et al. (2005), was calculated through Structure Harvester v.0.9.93 website (Earl, 2012) and used to estimate the most likely number of clusters ( $k$ ). In order to verify the breakdown of cultivars present in the Perugia collection to the Mediterranean groups previously observed (Sarri et al., 2006), their profiles for ten standard SSRs were analyzed with those of 281 most widely cultivated cultivars of Mediterranean from the CNR-IBBR database by using the same Structure parameters. Data of 281 cultivars were already published (Baldoni et al., 2009, 2011; Mousavi et al., 2017).

## RESULTS

### Polymorphisms Detected at EST and Standard SSR Loci

The nine OLEST markers analyzed were easily scored, showed low stuttering and clear differentiation among alleles (Table 1, Supplementary Table S1 and Figure S1). Mean  $N_a$  amounted to 7.9, ranging between 5 (OLEST9) and 15 (OLEST16).  $N_e$  was 4.466 on average, while the mean  $I$  value was 1.636.  $H_e$  (0.760) was in general higher than  $H_o$  (0.718), unless for OLEST22 and 23, where  $H_o$  was significantly higher than  $H_e$ .  $F$  values were positive on average, excluding OLEST22 and 23, and a negligible or moderate amount of null alleles was observed, with

no effect on their discrimination power. PIC values were higher than 0.5 at all OLEST loci, with an average value of 0.726 and the maximum discrimination power for OLEST16 (0.848) and OLEST1 (0.804).

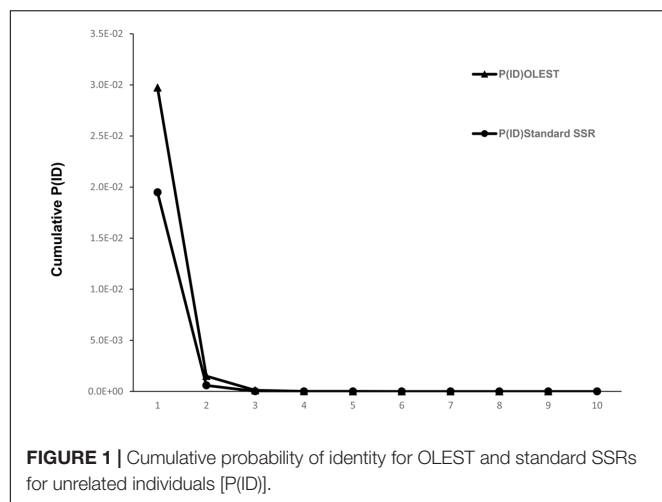
Total number of alleles for standard SSRs (Table 1) was considerably higher than for OLESTs, with 12.6 alleles per locus. Mean  $H_o$  was similar to  $H_e$  (0.808 and 0.802, respectively), and three out of 10 loci (DCA18, GAPU71B and GAPU101) with  $H_o$  higher than 0.9.  $F$  and  $F_{null}$  were slightly negative, showing  $-0.012$  and  $-0.007$ , respectively, whereas the mean value of PIC was 0.781. Cumulative probability of identity values (Figure 1) showed that a minimum of three loci was required for OLEST markers and only two for standard SSRs to reach  $P(ID) < 0.001$ . Therefore, only four and three loci were needed to distinguish all genotypes for OLEST and standard SSR markers, respectively. Nine OLEST [cumulative  $P(ID) = 2.5e^{-10}$ ] or 10 standard SSRs [cumulative  $P(ID) = 7.3e^{-14}$ ] allow for the unequivocal individual identification for this sample set with a high statistical confidence.

### Genetic Identity and Differentiation

The comparison of standard SSR profiles with the CNR-IBBR dataset and previous published data allowed for the identification of UNIPG collection's samples. Fifty nine distinct genotypes were identified, corresponding to 72 olive cultivars reported in the UNIPG archive. In fact, some samples called in the archive by different names, showed in our work identical genetic profiles (Supplementary Table S1 and Table 2). Among

**TABLE 1 |** Indices of genetic diversity at 72 cultivars for each SSR locus: number of alleles ( $N_a$ ), number of effective alleles ( $N_e$ ), Shannon's information index ( $I$ ), observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), fixation index ( $F$ ) and presence of null alleles ( $F_{null}$ ), Polymorphism Information Content ( $PIC$ ).

	$N_a$	$N_e$	$I$	$H_o$	$H_e$	$F$	$F_{null}$	$PIC$
<b>OLEST SSRs</b>								
OLEST1	7	5.750	1.835	0.806	0.826	0.025	0.013	0.804
OLEST7	8	4.056	1.581	0.681	0.753	0.097	0.043	0.715
OLEST9	5	3.660	1.421	0.694	0.727	0.044	0.026	0.681
OLEST12	7	4.056	1.537	0.542	0.753	0.281	0.174	0.713
OLEST14	11	4.852	1.921	0.681	0.794	0.143	0.083	0.775
OLEST16	15	7.210	2.221	0.819	0.861	0.049	0.021	0.848
OLEST20	6	2.742	1.270	0.583	0.635	0.082	0.057	0.591
OLEST22	6	4.213	1.546	0.819	0.763	-0.074	-0.045	0.727
OLEST23	6	3.653	1.390	0.833	0.726	-0.147	-0.068	0.679
Mean	7.9	4.466	1.636	0.718	0.760	0.055	0.034	0.726
<b>Standard SSRs</b>								
DCA3	11	6.545	2.042	0.889	0.847	-0.049	-0.027	0.830
DCA5	11	2.468	1.364	0.639	0.595	-0.074	-0.076	0.570
DCA9	16	7.551	2.214	0.861	0.868	0.007	0.002	0.853
DCA16	17	6.668	2.266	0.708	0.850	0.167	0.098	0.838
DCA18	13	6.392	2.088	0.903	0.844	-0.070	-0.034	0.826
EMO90	8	3.105	1.420	0.653	0.678	0.037	0.018	0.637
GAPU71B	7	4.324	1.607	0.944	0.769	-0.229	-0.111	0.733
GAPU101	10	6.372	2.019	0.931	0.843	-0.104	-0.049	0.825
GAPU103A	19	9.119	2.454	0.736	0.890	0.173	0.092	0.881
UDO-043	14	6.128	2.099	0.819	0.837	0.021	0.013	0.821
Mean	12.6	5.867	1.957	0.808	0.802	-0.012	-0.007	0.781



genotypes with identical profiles, the first group included the Portuguese cultivars Azeteira and Negrinha, eight cultivars resulted identical to Frantoio (Frantoio Corsini, Razzola, Casaliva, Razza, Taggiasca, Raja Sabina, and Ogliarola di Bitonto) (Group 2), Ogliarola Salentina, Mignola and Cima di Mola formed the third group, Moraiolo, Moraiolo Corsini and Corniolo the fourth, and Dritta di Moscufo and San Felice were the last case of identity. The OLEST markers showed exactly the same results, confirming all cases of identical profiles (Supplementary Table S1).

Eight different countries are represented in the collection, including Italy with 37 cultivars, Spain with nine, Greece with four, Portugal and France with three each, Morocco, Syria, and Tunisia with one each. Thirteen out of the 59 olive genotypes (Dolce d'Andria, Dritta di Loreto, Laurina, Morellona di Grecia, Negrera, Nostrale di Rigali, Olivago, Olivone, Orbetana, Pasola di Andria, Pociolo, Santagatese, Tendellone) resulted exclusive to this collection and absent in the main WOGBs. Pairwise allelic relatedness performed by GenAlEx showed 100 percent of similarity between the synonymous cultivars ( $LRM = 1.00$ ) for both set of markers. Comparing OLEST and standard SSRs for allelic similarity the highest values for non-synonymous cultivars were 0.67 and 0.57 respectively, while the minimum LRM values were  $-0.43$  for OLEST and  $-0.31$  for standard SSR markers.

## Population Genetic Structure

From the Structure analysis of data derived from 10 standard SSR loci on the 59 UNIPG cultivars run with 281 Mediterranean representative cultivars (Supplementary Figure S2), the stabilization, in terms of log-likelihood values of  $\Delta K$  values was observed at  $K = 3$  and, assigning individuals to a population for values above 70%, it was observed that 16 cultivars clustered into the Western Mediterranean group, 35 in the Central one and 12 in the Eastern population, only nine genotypes showed high levels of admixture among two or three groups.

The Structure analysis within the cultivars of the collection performed on OLEST and standard SSRs showed the most probable grouping at  $K = 4$  (Figures 2A,B). Most of the 59

**TABLE 2 |** Collection code, name of cultivars and Country of origin.

Collection code	Cultivar	Country
PER001	Arbequina	Spain
PER002	Ascolana Tenera	Italy
PER003	Azeteira <sup>G1</sup>	Portugal
PER004	Bella Di Cerignola	Italy
PER005	Canino	Italy
PER006	Carolea	Italy
PER007	Casaliva <sup>G2</sup>	Italy
PER008	Cima Di Mola <sup>G3</sup>	Italy
PER009	Cipressino	Italy
PER010	Coratina	Italy
PER011	Cordovil De Castelo Branco	Portugal
PER012	Cornicabra	Spain
PER013	Corniolo <sup>G4</sup>	Italy
PER014	Dolce D'andria	Italy
PER015	Dritta Di Loreto	Italy
PER016	Dritta Di Moscufo <sup>G5</sup>	Italy
PER017	Frantoio <sup>G2</sup>	Italy
PER018	Frantoio Corsini <sup>G2</sup>	Italy
PER019	Gentile Di Chieti	Italy
PER020	Giarrappa	Italy
PER021	Gordal Sevillana	Spain
PER022	Grappolo	Italy
PER023	Bella Di Spagna	Italy
PER024	Hojiblanca	Spain
PER025	Itrana	Italy
PER026	Kalamata	Greece
PER027	Konservolia	Greece
PER028	Koroneiki	Greece
PER029	Laurina	Italy
PER030	Leccino	Italy
PER031	Leccio Del Corno	Italy
PER032	Lechin De Sevilla	Spain
PER033	Lucio	Spain
PER034	Madural	Portugal
PER035	Manzanilla De Sevilla	Spain
PER036	Manzanilla Prieta	Spain
PER037	Mastoidis	Greece
PER038	Maurino	Italy
PER039	Mignola <sup>G3</sup>	Italy
PER040	Moraiolo <sup>G4</sup>	Italy
PER041	Moraiolo Corsini <sup>G4</sup>	Italy
PER042	Morellona Di Grecia	Italy
PER043	Moresca	Italy
PER044	Negrera	Italy
PER045	Negrinha <sup>G1</sup>	Portugal
PER046	Nocellara Del Belice	Italy
PER047	Nocellara Messinese	Italy
PER048	Nostrale Di Rigali	Italy
PER049	Ogliarola Di Bitonto <sup>G2</sup>	Italy
PER050	Ogliarola Salentina <sup>G3</sup>	Italy
PER051	Olivago	Italy
PER052	Olivone	Italy
PER053	Orbetana	Italy

(Continued)



**TABLE 2 |** Continued

Collection code	Cultivar	Country
PER054	<u>Pasola Di Andria</u>	Italy
PER055	Passalunara	Italy
PER056	Picholine De Languedoc	France
PER057	Picholine Marocaine	Morocco
PER058	Picual	Spain
PER059	<u>Pocciolo</u>	Italy
PER060	Raja Sabina <sup>G2</sup>	Italy
PER061	Razza <sup>G2</sup>	Italy
PER062	Razzola <sup>G2</sup>	Italy
PER063	Rosciola Lazio	Italy
PER064	San Felice <sup>G5</sup>	Italy
PER065	Santa Caterina	Italy
PER066	<u>Santagatese</u>	Italy
PER067	Sourani	Syria
PER068	Taggiasca <sup>G2</sup>	Italy
PER069	Tanche	France
PER070	<u>Tendellone</u>	Italy
PER071	Uovo Di Piccione	Tunisia
PER072	Verdale	France

G1, G2, G3, G4, and G5: To each group number correspond cultivars showing identical genotype. The underlined cultivars are those exclusively present in the Perugia collection and not in the World Olive Germplasm Banks.

cultivars resulted assigned to two of the four groups for standard SSRs while for OLEST the four structure population were well balanced. In fact, the proportion of membership for OLEST markers was from 0.158 (Pop2) to 0.406 (Pop1), while for standard SSRs the lowest value was 0.054 (Pop2) and for the Pop1 and Pop4 membership value were 0.423 and 0.414 respectively. Only 20 cultivars were assigned to the same population by both set of markers (**Figures 2A,B**). The expected heterozygosity individuated by Bayesian analysis within the same population was on average higher for standard than for OLEST markers (0.84 and 0.76, respectively). Furthermore, the level of population assignment for OLEST markers was lower than standard SSRs (0.75 and 0.88, respectively).

## DISCUSSION

The application of highly effective and discriminant markers may allow the correct identification of all accessions, establishing their representativeness of the species variability and justifying their conservation in *ex situ* collections. This step is crucial to avoid redundancy in germplasm repositories, reducing management costs, distributing true-to-type genotypes for propagation, ratifying reliable genetic sources for breeding programs. The management of germplasm collections, in fact, requires attention and mistakes may be introduced at many stages, from the origin of plant material, that may derive from other collections, private orchards or unreliable sources, to propagation and field planting, and each accession needs correct identification and passport data (Kato et al., 2012; Potts et al., 2012; Trujillo et al., 2014). A thorough and accurate genotype profiling represents a crucial

prerequisite to assist breeding programs, perform comparative studies and assess innovative researches.

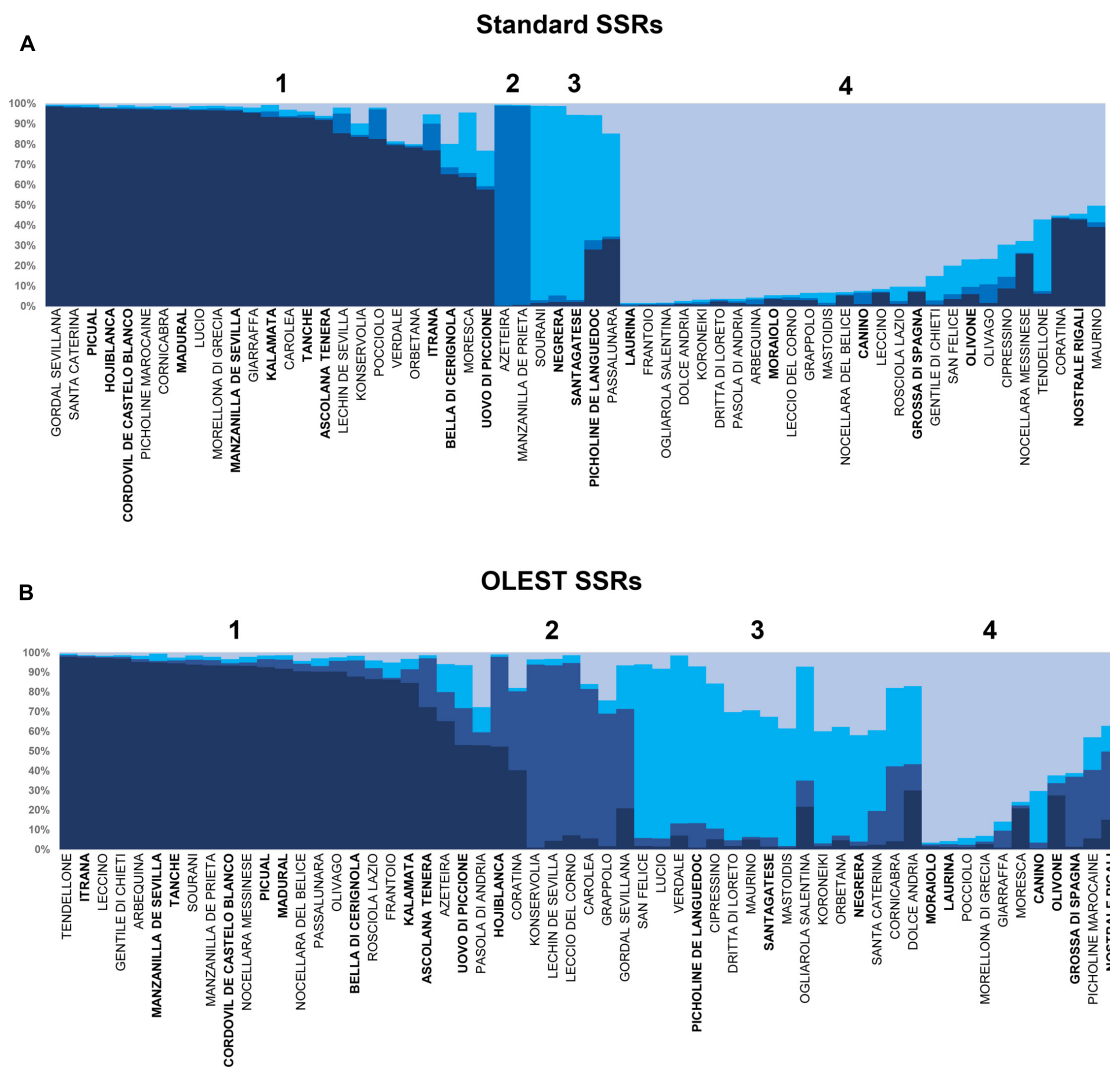
The collection of olive cultivars established at the University of Perugia represents one of the first efforts to converge into a single set deeply diverse genotypes, deriving from areas with highly different climatic and growing conditions, in order to preserve the variation of cultivated olives and evaluate their characteristics. The genetic identity of genotypes at the UNIPG olive collection was never ascertained before and we were committed to achieve a complete genotyping of all accessions.

Simple sequence repeats markers have become the preferred tool for the identification of olive cultivars, due to their high discrimination power and straightforward data reading (Haouane et al., 2011; Trujillo et al., 2014), however, the largely used dinucleotide SSRs have shown problems related to difficult discrimination between neighboring alleles and low comparability of data among different labs, severely reducing their applicability for large-scale screening (Baldoni et al., 2009) and for comparing the molecular profiles of accessions distributed in different collections (Diez et al., 2015; Torkzaban et al., 2015). For this reason, we decided to apply both, the best ranked dinucleotide SSRs and the recently developed trinucleotide EST-SSRs (OLEST) (Mariotti et al., 2016), in order to also evaluate their reliability in genotyping germplasm repositories.

To establish cultivar identity and determine all clonal replicates, 10 standard dinucleotide SSR markers were preliminarily applied and allele profiles were compared with previously published data (Baldoni et al., 2009, 2011; Hosseini-Mazinani et al., 2014; Trujillo et al., 2014; Mousavi et al., 2017), or included in the CNR-IBBR database. Results derived from these analyses highlighted the presence of 59 distinct genotypes, including five groups of cultivars sharing identical SSR profiles (Bartolini, 2009; Trujillo et al., 2014), but coming from different areas of cultivation and carrying different names.

The same results were obtained when the analysis was independently performed with the new OLEST SSRs: 59 genotypes were distinguished and identified, and the same groups with identical profiles were displayed. Also the values of diversity parameters resulted quite similar to those of best ranked dinucleotide SSRs, particularly for the discrimination power and observed heterozygosity values, with a negligible presence of null alleles. The pairwise relatedness analysis demonstrated the same single-profile groups and highlighted that OLEST markers were more efficient to discriminate among the most polymorphic genotypes, showing the minimum values of allelic similarity.

The occurrence of cases of identical genotype under different cultivar names represents a primary source of problems for identification and a major challenge to the management of germplasm collections (Belaj et al., 2007; Abdessemed et al., 2015). In the olive case (Bradai et al., 2016), as for many other long living trees (Vezzulli et al., 2012; Urrestarazu et al., 2012; Fresnedo-Ramírez et al., 2013; Jiao et al., 2013; Frank and Chitwood, 2016), it can not be theoretically excluded that plant genotypes clonally propagated and living for thousands of years, may accumulate somatic mutations, over the time or as a result of environmental shocks. But these mutations could not be easily



**FIGURE 2 |** Genetic structure of the 59 cultivars identified at the UNIPG Olive Collection based on data derived from standard (A) and new OLEST (B) SSR markers. Each vertical bar represents single accessions and colors distinguish the four detected groups. Olive samples with more than one color indicate admixture in their genetic composition. Cultivars assigned to the same group by both kind of markers are reported in bold below every population (from Pop 1 to Pop 4).

revealed by the use of a restricted set of SSR markers and, for this reason, we decided to leave the original names of cultivars, even if they showed the same SSR profile, making them available for future in-depth genomic analyses that would highlight eventual polymorphisms otherwise undetectable (Wu et al., 2014).

By using only three OLEST markers it was possible to discriminate 96.6% of all genotypes. Moreover, OLEST SSRs resulted more easily scorable than dinucleotide SSRs, and didn't show stuttering problems due to the higher distance among similar alleles and lower slippage during replication. Using the three OLEST markers with the highest PIC values (OLEST1, OLEST14 and OLEST16), 57 out of 59 genotype were discriminated, whereas applying the three most discriminant standard SSRs (DCA09, DCA16 and GAPU103A), all 59 genotypes were completely recognized. In fact, the individual identification estimator [P(ID)] indicates two different accessions

may have the same genotype at one specific locus in a population by chance rather than through inheritance, we found that both set of markers were able to clearly distinguish all 59 olive genotypes in the Perugia olive collection.

The Bayesian structure analysis of genotypes present in the Perugia assortment with the wide set of other important cultivars of Mediterranean basin, has shown that the collection well represents the groups in which the cultivated Mediterranean olives were previously splitted (Haouane et al., 2011; Diez et al., 2015), with a higher membership to the Central Mediterranean group, likely due to the prevalence of Italian cultivars. Furthermore, this repository owns 13 cultivars not present in the main international olive germplasm banks (Haouane et al., 2011; Trujillo et al., 2014), strengthening its relevant function for conservation, evaluation and protection of specific genotypes potentially endangered.

When the same analysis was exclusively performed on the UNIPG genotypes, 34% of cultivars resulted assigned to the same population by both sets of markers. The Bayesian results clearly highlighted the differences between OLEST and standard SSRs in the cultivar's assignment into the structure populations. This dissimilarity was evidenced by the values of expected heterozygosity, the overall proportion of membership and admixture level. Therefore, the results of the present study suggest that, for phylogenetic studies, by using different set of markers could achieve unbalanced assignments. The different ability of both kinds of markers to group cultivars into different clusters could be explained by the nature of OLEST markers as mutations residing in the sequence of transcribed genes, and their alleles could display a higher frequency at regional level, where cultivars were selected based on common characteristics (Biton et al., 2015; Mariotti et al., 2016). Considering that olive domestication process has implied a selection of cultivars for certain agronomic characters, resulting in a loss of genetic variation due to genetic bottlenecks and, in some cases, episodes of founder effect (Cao et al., 2014; Hosseini-Mazinani et al., 2014; Mousavi et al., 2017), EST-SSRs could be related to agronomical traits more than neutral standard SSRs. The very long history of olive growing with several trading events, introduction of alien cultural practices and changes of dietary habits, may have blurred the fingerprints of independent domestication events and led to complex relationships among cultivars (Sarri et al., 2006; Soleri et al., 2010; Díez et al., 2011; Koehmstedt et al., 2011).

The Perugia collection represents the first study case of a real olive germplasm repository validated by standard SSRs and characterized by EST-SSRs. The work has allowed to confirm the OLEST markers as effective genotyping tools, as good as best standard markers for cultivar identification, allowing to avoid

the application of other unreliable dinucleotide SSRs. The use of the OLEST markers on a wide set of olive cultivars will help establishing a common fingerprint database without miscalling and binning, exploitable for several molecular investigations, representing a valuable resource for comparative genomics, evolutionary analyses and population studies.

## AUTHOR CONTRIBUTIONS

SM, LB, LR, PP, MB, RM, LN, and SP contributed substantially to the conception and design of the work; SM, LR, LN, RM, and SP contributed to plant material collection; SM, LB, and RM, performed all molecular work and genotype scoring; SM, LB, LR, RM, LN, and SP interpretation of data; SM, LB, LR, PP, MB, RM, LN, and SP drafted the text; SM, LB, LR, PP, MB, RM, LN, and SP approved the version to be published; SM, LB, LR, PP, MB, RM, LN, and SP agreed to be accountable for all aspects of the work.

## ACKNOWLEDGMENTS

The study has been partially performed within the Project “BeFOre – Bioresources for Oliviculture,” 2015–2019, H2020-MSCA-RISE- Marie Skłodowska-Curie Research and Innovation Staff Exchange, Grant Agreement N. 645595.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.01283/full#supplementary-material>

## REFERENCES

- Abdessemed, S., Muzzalupo, I., and Benbouza, H. (2015). Assessment of genetic diversity among Algerian olive (*Olea europaea* L.) cultivars using SSR marker. *Sci. Hortic.* 192, 10–20. doi: 10.1016/j.scienta.2015.05.015
- Baldoni, L., Cultrera, N., Mariotti, R., Pandolfi, S., Blanco, A., Montemurro, C. (eds). et al. (2011). *Tematica I - Manuale per l'identificazione e il Riordino del Patrimonio Varietale di Olivo*. Bari: Università degli Studi di Bari.
- Baldoni, L., Cultrera, N. G., Mariotti, R., Ricciolini, C., Arcioni, S., Vendramin, G. G., et al. (2009). A consensus list of microsatellite markers for olive genotyping. *Mol. Breed.* 24, 213–231. doi: 10.1007/s11032-009-9285-8
- Bartolini, G. (2009). “Banca dati mondiale: germoplasma dell'olivo (www.oleadb.eu) 1,” in *Banca Dati del Germoplasma di Olivo (Olea europaea L.): Cultivar, Sinonimi, Aree di Coltivazione, Descrittori, Collezioni*, Vol. 5, ed. Accademia dei Georgofili (Florence: Atti della Accademia dei Georgofili), 699–710. doi: 10.7349/OLEA\_databases
- Belaj, A., del Carmen Dominguez-García, M., Atienza, S. G., Urdíroz, N. M., De la Rosa, R., Satovic, Z., et al. (2012). Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DARs, SSRs, SNPs) and agronomic traits. *Tree Genet. Genomes* 8, 365–378. doi: 10.1007/s11295-011-0447-6
- Belaj, A., León, L., Satovic, Z., and De la Rosa, R. (2011). Variability of wild olives (*Olea europaea* subsp. *europaea* var. *sylvestris*) analyzed by agro-morphological traits and SSR markers. *Sci. Hortic.* 129, 561–569. doi: 10.1016/j.scienta.2011.04.025
- Belaj, A., Muñoz-Díez, C., Baldoni, L., Porceddu, A., Barranco, D., and Satovic, Z. (2007). Genetic diversity and population structure of wild olives from the north-western Mediterranean assessed by SSR markers. *Ann. Bot.* 100, 449–458. doi: 10.1093/aob/mcm132
- Biton, I., Doron-Faigenboim, A., Jamwal, M., Mani, Y., Eshed, R., Rosen, A., et al. (2015). Development of a large set of SNP markers for assessing phylogenetic relationships between the olive cultivars composing the Israeli olive germplasm collection. *Mol. Breed.* 35, 107. doi: 10.1007/s11032-015-0304-7
- Bracci, T., Busconi, M., Fogher, C., and Sebastiani, L. (2011). Molecular studies in olive (*Olea europaea* L.): overview on DNA markers applications and recent advances in genome analysis. *Plant Cell Rep.* 30, 449–462. doi: 10.1007/s00299-010-0991-9
- Bradai, F., Pliego-Alfaro, F., and Sánchez-Romero, C. (2016). Somaclonal variation in olive (*Olea europaea* L.) plants regenerated via somatic embryogenesis: influence of genotype and culture age on phenotypic stability. *Sci. Hortic.* 213, 208–215. doi: 10.1016/j.scienta.2016.10.031
- Breton, C., Farinelli, D., Shafiq, S., Heslop-Harrison, J. S., Sedgley, M., and Bervillé, A. (2014). The self-incompatibility mating system of the olive (*Olea europaea* L.) functions with dominance between S-alleles. *Tree Genet. Genomes* 10, 1055–1067. doi: 10.1007/s11295-014-0742-0
- Breton, C., Terral, J. F., Pinatel, C., Médail, F., Bonhomme, F., and Bervillé, A. (2009). The origins of the domestication of the olive tree. *C. R. Biol.* 332, 1059–1064. doi: 10.1016/j.crv.2009.08.001
- Cao, K., Zheng, Z., Wang, L., Liu, X., Zhu, G., Fang, W., et al. (2014). Comparative population genomics reveals the domestication history of the peach, *Prunus*

- persica*, and human influences on perennial fruit crops. *Genome Biol.* 15, 415. doi: 10.1186/s13059-014-0415-1
- Carriero, F., Fontanazza, G., Cellini, F., and Giorio, G. (2002). Identification of simple sequence repeats (SSRs) in olive (*Olea europaea* L.). *Theor. Appl. Genet.* 104, 301–307. doi: 10.1007/s001220100691
- Caruso, T., Marra, F. P., and Costa, F. (2014). Genetic diversity and clonal variation within the main Sicilian olive cultivars based on morphological traits and microsatellite markers. *Sci. Hortic.* 180, 130–138. doi: 10.1016/j.scienta.2014.10.019
- Chapuis, M. P., and Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* 24, 621–631. doi: 10.1093/molbev/msl191
- Cipriani, G., Marrazzo, M. T., Marconi, R., Cimato, A., and Testolin, R. (2002). Microsatellite markers isolated in olive (*Olea europaea* L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. *Theor. Appl. Genet.* 104, 223–228. doi: 10.1007/s001220100685
- Corrado, G., Imperato, A., La Mura, M., Perri, E., and Rao, R. (2011). Genetic diversity among olive varieties of Southern Italy and the traceability of olive oil using SSR markers. *J. Hortic. Sci. Biotechnol.* 86, 461. doi: 10.1080/14620316.2011.11512789
- De Gennaro, B., Notarnicola, B., Roselli, L., and Tassielli, G. (2012). Innovative olive-growing models: an environmental and economic assessment. *J. Clean. Prod.* 28, 70–80. doi: 10.1016/j.jclepro.2011.11.004
- Diez, C. M., Trujillo, I., Barrio, E., Belaj, A., Barranco, D., and Rallo, L. (2011). Centennial olive trees as a reservoir of genetic diversity. *Ann. Bot.* 108, 797–807. doi: 10.1093/aob/mcr194
- Diez, C. M., Trujillo, I., Martínez-Urdiroz, N., Barranco, D., Rallo, L., Marfil, P., et al. (2015). Olive domestication and diversification in the Mediterranean Basin. *New Phytol.* 206, 436–447. doi: 10.1111/nph.13181
- Duran, C., Appleby, N., Edwards, D., and Batley, J. (2009). Molecular genetic markers: discovery, applications, data storage and visualisation. *Curr. Bioinform.* 4, 16–27. doi: 10.2174/157489309787158198
- Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- El Bakkali, A., Haouane, H., Moukhli, A., Costes, E., Van Damme, P., and Khadari, B. (2013). Construction of core collections suitable for association mapping to optimize use of Mediterranean olive (*Olea europaea* L.) genetic resources. *PLoS ONE* 8:e61265. doi: 10.1371/journal.pone.0061265
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Frank, M. H., and Chitwood, D. H. (2016). Plant chimeras: the good, the bad, and the 'Bizzaria'. *Dev. Biol.* 419, 41–53. doi: 10.1016/j.ydbio.2016.07.003
- Fresnedo-Ramírez, J., Martínez-García, P. J., Parfitt, D. E., Crisosto, C. H., and Gradziel, T. M. (2013). Heterogeneity in the entire genome for three genotypes of peach [*Prunus persica* (L.) Batsch] as distinguished from sequence analysis of genomic variants. *BMC Genomics* 14:750. doi: 10.1186/1471-2164-14-750
- Green, P. S. (2002). A revision of *Olea* L. (Oleaceae). *Kew Bull.* 57, 91–140. doi: 10.2307/4110824
- Haouane, H., El Bakkali, A., Moukhli, A., Tollon, C., Santoni, S., Oukabli, A., et al. (2011). Genetic structure and core collection of the World Olive Germplasm Bank of Marrakech: towards the optimised management and use of Mediterranean olive genetic resources. *Genetica* 139, 1083–1094. doi: 10.1007/s10709-011-9608-7
- Hosseini-Mazinani, M., Mariotti, R., Torkzaban, B., Sheikh-Hassani, M., Ataei, S., Cultrera, N. G., et al. (2014). High genetic diversity detected in olives beyond the boundaries of the Mediterranean Sea. *PLoS ONE* 9:e93146. doi: 10.1371/journal.pone.0093146
- Ilarioni, L., and Proietti, P. (2014). "Olive tree cultivars," in *The Extra-Virgin Olive Oil Handbook*, Vol. 5, ed. C. Peri (Hoboken, NJ: John Wiley & Sons, Ltd), 59–67.
- Ipek, M., Seker, M., Ipek, A., and Gul, M. K. (2015). Identification of molecular markers associated with fruit traits in olive and assessment of olive core collection with AFLP markers and fruit traits. *Genet. Mol. Res.* 14, 2762–2774. doi: 10.4238/2015.March.31.6
- Jiao, W. B., Huang, D., Xing, F., Hu, Y., Deng, X. X., Xu, Q., et al. (2013). Genome-wide characterization and expression analysis of genetic variants in sweet orange. *Plant J.* 75, 954–964. doi: 10.1111/tpj.12254
- Kaniewski, D., Van Campo, E., Boiy, T., Terral, J. F., Khadari, B., and Besnard, G. (2012). Primary domestication and early uses of the emblematic olive tree: palaeobotanical, historical and molecular evidence from the Middle East. *Biol. Rev.* 87, 885–899. doi: 10.1111/j.1469-185X.2012.00229.x
- Kato, S., Matsumoto, A., Yoshimura, K., Katsuki, T., Iwamoto, K., Tsuda, Y., et al. (2012). Clone identification in Japanese flowering cherry (*Prunus* subgenus *Cerasus*) cultivars using nuclear SSR markers. *Breed. Sci.* 62:248. doi: 10.1270/jsbbs.62.248
- Kaya, H. B., Cetin, O., Kaya, H., Sahin, M., Sefer, F., Kahraman, A., et al. (2013). SNP discovery by Illumina-based transcriptome sequencing of the olive and the genetic characterization of Turkish olive genotypes revealed by AFLP, SSR and SNP markers. *PLoS ONE* 8:e73674. doi: 10.1371/journal.pone.0073674
- Klepeš, T., De la Rosa, R., Satovic, Z., León, L., and Belaj, A. (2013). Utility of wild germplasm in olive breeding. *Sci. Hortic.* 152, 92–101. doi: 10.1016/j.scienta.2012.12.010
- Koehmstedt, A. M., Aradhya, M. K., Soler, D., Smith, J. L., and Polito, V. S. (2011). Molecular characterization of genetic diversity, structure, and differentiation in the olive (*Olea europaea* L.) germplasm collection of the United States Department of Agriculture. *Genet. Res. Crop Evol.* 58, 519–531. doi: 10.1007/s10722-010-9595-z
- Larbi, A., Vázquez, S., El-Jendoubi, H., Msallem, M., Abadía, J., Abadía, A., et al. (2015). Canopy light heterogeneity drives leaf anatomical, eco-physiological, and photosynthetic changes in olive trees grown in a high-density plantation. *Photosynth. Res.* 123, 141–155. doi: 10.1007/s11120-014-0052-2
- Laroussi-Mezghani, S., Le Dréau, Y., Molinet, J., Hammami, M., Grati-Kamoun, N., and Artaud, J. (2016). Biodiversity of Tunisian virgin olive oils: varietal origin classification according to their minor compounds. *Eur. Food Res. Technol.* 242, 1087–1099. doi: 10.1007/s00217-015-2613-9
- Las Casas, G., Scollo, F., Distefano, G., Continella, A., Gentile, A., and La Malfa, S. (2014). Molecular characterization of olive (*Olea europaea* L.) Sicilian cultivars using SSR markers. *Biochem. Syst. Ecol.* 57, 15–19. doi: 10.1016/j.bse.2014.07.010
- Lazović, B., Adakalić, M., Pucci, C., Perović, T., Bandelj, D., and Belaj, A. (2016). Characterizing ancient and local olive germplasm from Montenegro. *Sci. Hortic.* 209, 117–123. doi: 10.1016/j.scienta.2016.06.022
- Lynch, M., and Ritland, K. (1999). Estimation of pairwise relatedness with molecular markers. *Genetics* 152, 1753–1766.
- Marchese, A., Marra, F. P., Caruso, T., Mhelembe, K., Costa, F., Fretto, S., et al. (2016). The first high-density sequence characterized SNP-based linkage map of olive (*Olea europaea* L. subsp. *europaea*) developed using genotyping by sequencing. *Aust. J. Crop Sci.* 10, 857–863. doi: 10.21475/ajcs.2016.10.06.p7520
- Mariotti, R., Cultrera, N. G. M., Mousavi, S., Baglivo, F., Rossi, M., Albertini, E., et al. (2016). Development, evaluation, and validation of new EST-SSR markers in olive (*Olea europaea* L.). *Tree Genet. Genomes* 12, 120. doi: 10.1007/s11295-016-1077-9
- Marra, F. P., Caruso, T., Costa, F., Di Vaio, C., Mafrica, R., and Marchese, A. (2013). Genetic relationships, structure and parentage simulation among the olive tree (*Olea europaea* L. subsp. *europaea*) cultivated in Southern Italy revealed by SSR markers. *Tree Genet. Genomes* 9, 961–973. doi: 10.1007/s11295-013-0609-9
- Marshall, T. C., Slate, J., Kruuk, L. E. B., and Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7, 639–655. doi: 10.1046/j.1365-294x.1998.00374.x
- Mazzitelli, O., Calleja, A., Sardella, D., Farrugia, C., and Zammit-Mangion, M. (2015). Analysis of the molecular diversity of *Olea europaea* in the Mediterranean Island of Malta. *Genet. Res. Crop Evol.* 62, 1021–1027. doi: 10.1007/s10722-014-0205-3
- Moriondo, M., Trombi, G., Ferrise, R., Brandani, G., Dibari, C., Ammann, C. M., et al. (2013). Olive trees as bioindicators of climate evolution in the Mediterranean Basin. *Glob. Ecol. Biogeogr.* 22, 818–833. doi: 10.1111/geb.12061
- Mousavi, S., Mariotti, R., Bagnoli, F., Costantini, L., Cultrera, N. G. M., Arzani, K., et al. (2017). The eastern part of the Fertile Crescent concealed an unexpected route of olive (*Olea europaea* L.) differentiation. *Ann. Bot.* 119, 1305–1318. doi: 10.1093/aob/mcx027
- Mousavi, S., Mazinani, M. H., Arzani, K., Ydollahi, A., Pandolfi, S., Baldoni, L., et al. (2014). Molecular and morphological characterization of Golestan (Iran) olive ecotypes provides evidence for the presence of promising genotypes. *Genet. Res. Crop Evol.* 61, 775–785. doi: 10.1007/s10722-013-0071-4



- Muzzalupo, I., Vendramin, G. G., and Chiappetta, A. (2014). Genetic biodiversity of Italian olives (*Olea europaea*) germplasm analyzed by SSR markers. *Sci. World J.* 2014:296590. doi: 10.1155/2014/296590
- Peakall, R., and Smouse, P. E. (2012). GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460
- Portarena, S., Farinelli, D., Lauteri, M., Famiani, F., Esti, M., and Brugnoli, E. (2015). Stable isotope and fatty acid compositions of monovarietal olive oils: implications of ripening stage and climate effects as determinants in traceability studies. *Food Control* 57, 129–135. doi: 10.1016/j.foodcont.2015.03.052
- Potts, S. M., Han, Y., Khan, M. A., Kushad, M. M., Rayburn, A. L., and Korban, S. S. (2012). Genetic diversity and characterization of a core collection of *Malus* germplasm using simple sequence repeats (SSRs). *Plant Mol. Biol. Rep.* 30, 827–837. doi: 10.1007/s11105-011-0399-x
- Pritchard, J., Wen, X., and Falush, D. (2009). *Documentation for Structure Software: Version 2.3*. Available at: <http://pritch.bsd.uchicago.edu/structure.html>
- Proietti, S., Sdringola, P., Desideri, U., Zepparelli, F., Brunori, A., Ilarioni, L., et al. (2014). Carbon footprint of an olive tree grove. *Appl. Energy* 127, 115–124. doi: 10.1016/j.apenergy.2014.04.019
- Sarri, V., Baldoni, L., Porceddu, A., Cultrera, N. G. M., Contento, A., Frediani, M., et al. (2006). Microsatellite markers are powerful tools for discriminating among olive cultivars and assigning them to geographically defined populations. *Genome* 49, 1606–1615. doi: 10.1139/g06-126
- Sefc, K. M., Lopes, M. S., Mendonça, D., Santos, M. R. D., Machado, L. M., and Machado, A. D. C. (2000). Identification of microsatellite loci in olive (*Olea europaea*) and their characterization in Italian and Iberian olive trees. *Mol. Ecol.* 9, 1171–1173. doi: 10.1046/j.1365-294x.2000.00954.x
- Soleri, D., Koehmstedt, A., Aradhya, M. K., Polito, V., and Pinney, K. (2010). Comparing the historic olive trees (*Olea europaea* L.) of Santa Cruz Island with contemporaneous trees in the Santa Barbara, CA area: a case study of diversity and structure in an introduced agricultural species conserved in situ. *Genet. Resour. Crop Evol.* 57, 973–984. doi: 10.1007/s10722-010-9537-9
- Tanasijevic, L., Todorovic, M., Pereira, L. S., Pizzigalli, C., and Lionello, P. (2014). Impacts of climate change on olive crop evapotranspiration and irrigation requirements in the Mediterranean region. *Agric. Water Manage.* 144, 54–68. doi: 10.1016/j.agwat.2014.05.019
- Torkzaban, B., Kayvanjoo, A. H., Ardalan, A., Mousavi, S., Mariotti, R., Baldoni, L., et al. (2015). Machine learning based classification of microsatellite variation: an effective approach for phylogeographic characterization of olive populations. *PLoS ONE* 10:e0143465. doi: 10.1371/journal.pone.0143465
- Trujillo, I., Ojeda, M. A., Urdiroz, N. M., Potter, D., Barranco, D., Rallo, L., et al. (2014). Identification of the Worldwide Olive Germplasm Bank of Córdoba (Spain) using SSR and morphological markers. *Tree Genet. Genomes* 10, 141–155. doi: 10.1007/s11295-013-0671-3
- Urrestarazu, J., Miranda, C., Santesteban, L. G., and Royo, J. B. (2012). Genetic diversity and structure of local apple cultivars from Northeastern Spain assessed by microsatellite markers. *Tree Genet. Genomes* 8, 1163–1180. doi: 10.1007/s11295-012-0502-y
- Vezzulli, S., Leonardelli, L., Malossini, U., Stefanini, M., Velasco, R., and Moser, C. (2012). Pinot blanc and Pinot gris arose as independent somatic mutations of Pinot noir. *J. Exp. Bot.* 63, 6359–6369. doi: 10.1093/jxb/ers290
- Waits, L. P., Luikart, G., and Taberlet, P. (2001). Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Mol. Ecol.* 10, 249–256. doi: 10.1046/j.1365-294X.2001.01185.x
- Wu, G. A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., et al. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat. Biotechnol.* 32, 656–662. doi: 10.1038/nbt.2906
- Xanthopoulou, A., Ganopoulos, I., Koubouris, G., Tsafaris, A., Sergendani, C., Kalivas, A., et al. (2014). Microsatellite high-resolution melting (SSR-HRM) analysis for genotyping and molecular characterization of an *Olea europaea* germplasm collection. *Plant Genet. Resour.* 12, 273–277. doi: 10.1017/S147926211400001X
- Yang, J., Dai, P., Zhou, T., Huang, Z., Feng, L., Su, H., et al. (2013). Genetic diversity and structure of wintersweet (*Chimonanthus praecox*) revealed by EST-SSR markers. *Sci. Hortic.* 150, 1–10. doi: 10.1016/j.scienta.2012.11.004
- Zelasco, S., Salimonti, A., Baldoni, L., Mariotti, R., Preece, J. E., Aradhya, M., et al. (2012). Efficiency of SSR markers for exploring olive germplasm diversity through a genetic comparison between the USDA-NCGR and the CRA-OLI olive collections. *Acta Hortic.* 1057, 585–592. doi: 10.17660/ActaHortic.2014.1057.75

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Mousavi, Mariotti, Regni, Nasini, Bufacchi, Pandolfi, Baldoni and Proietti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Emerging Oilseed Crop *Sesamum indicum* Enters the “Omics” Era

Komivi Dossa<sup>1,2,3\*</sup>, Diaga Diouf<sup>2\*</sup>, Linhai Wang<sup>3</sup>, Xin Wei<sup>3</sup>, Yanxin Zhang<sup>3</sup>, Mareme Niang<sup>1</sup>, Daniel Fonceka<sup>1,4</sup>, Jingyin Yu<sup>3</sup>, Marie A. Mmadi<sup>1,2,3</sup>, Louis W. Yehouessi<sup>1</sup>, Boshou Liao<sup>3</sup>, Xiurong Zhang<sup>3\*</sup> and Ndiaga Cisse<sup>1\*</sup>

<sup>1</sup> Centre d'Etudes Régional Pour l'Amélioration de l'Adaptation à la Sécheresse, Thiès, Sénégal, <sup>2</sup> Laboratoire Campus de Biotechnologies Végétales, Département de Biologie Végétale, Faculté des Sciences et Techniques, Université Cheikh Anta Diop, Dakar, Sénégal, <sup>3</sup> Key Laboratory of Biology and Genetic Improvement of Oil Crops, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Ministry of Agriculture, Wuhan, China, <sup>4</sup> Centre de Coopération Internationale en Recherche Agronomique Pour le Développement, UMR AGAP, Montpellier, France

## OPEN ACCESS

### Edited by:

Mariela Torres,  
Instituto Nacional de Tecnología  
Agropecuaria (INTA), Argentina

### Reviewed by:

Sarita K. Pandey,  
International Crops Research Institute  
for the Semi-Arid Tropics, India  
Ioannis Ganopoulos,  
Institute of Plant Breeding  
and Genetic Resources-ELGO  
DEMETER, Greece

### \*Correspondence:

Diaga Diouf  
diaga.diouf@ucad.edu.sn  
Xiurong Zhang  
zhangxr@oilcrops.cn  
Ndiaga Cisse  
ncisse@refer.sn  
Komivi Dossa  
dossakomivi@gmail.com

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 19 April 2017

**Accepted:** 15 June 2017

**Published:** 30 June 2017

### Citation:

Dossa K, Diouf D, Wang L, Wei X,  
Zhang Y, Niang M, Fonceka D, Yu J,  
Mmadi MA, Yehouessi LW, Liao B,  
Zhang X and Cisse N (2017)  
The Emerging Oilseed Crop  
*Sesamum indicum* Enters the “Omics”  
Era. *Front. Plant Sci.* 8:1154.  
doi: 10.3389/fpls.2017.01154

Sesame (*Sesamum indicum* L.) is one of the oldest oilseed crops widely grown in Africa and Asia for its high-quality nutritional seeds. It is well adapted to harsh environments and constitutes an alternative cash crop for smallholders in developing countries. Despite its economic and nutritional importance, sesame is considered as an orphan crop because it has received very little attention from science. As a consequence, it lags behind the other major oil crops as far as genetic improvement is concerned. In recent years, the scenario has considerably changed with the decoding of the sesame nuclear genome leading to the development of various genomic resources including molecular markers, comprehensive genetic maps, high-quality transcriptome assemblies, web-based functional databases and diverse draft genome sequences. The availability of these tools in association with the discovery of candidate genes and quantitative trait loci for key agronomic traits including high oil content and quality, waterlogging and drought tolerance, disease resistance, cytoplasmic male sterility, high yield, pave the way to the development of some new strategies for sesame genetic improvement. As a result, sesame has graduated from an “orphan crop” to a “genomic resource-rich crop.” With the limited research teams working on sesame worldwide, more synergic efforts are needed to integrate these resources in sesame breeding for productivity upsurge, ensuring food security and improved livelihood in developing countries. This review retraces the evolution of sesame research by highlighting the recent advances in the “Omics” area and also critically discusses the future prospects for a further genetic improvement and a better expansion of this crop.

**Keywords:** *Sesamum indicum*, Omic resources, molecular breeding, large-scale re-sequencing, improvement

## INTRODUCTION

Since the beginning of agriculture, humans have been selecting and cultivating crops that would serve their taste, energy, and health requirements (Nagaraj, 2009). Oilseeds are crops in which energy is stored mainly in the form of oil and are a very important component of semi-tropical and tropical agriculture, providing easily available and highly nutritious human and animal food

(Weiss, 2000). Among the important oilseed crops widely grown in the world such as rapeseed, peanut, soybean, sunflower, sesame (*Sesamum indicum* L.) provides one of the highest and richest edible oils (Pathak et al., 2014). Sesame is a diploid species ( $2n = 2x = 26$ ), an annual plant principally grown for its seeds. The seed contains 50–60% oil which has an excellent stability due to the presence of natural antioxidants such as sesamol, sesamin, and sesamol (Anilakumar et al., 2010). The chemical composition of sesame oil characterized by a low level of saturated fatty acids (SFAs) (less than 15%) and the presence of antioxidants has been reported to have health promoting effects such as lowering cholesterol levels and hypertension in humans (Noguchi et al., 2001; Sankar et al., 2005), neuroprotective effects against hypoxia or brain damage (Cheng et al., 2006) and reducing the incidence of certain cancers (Hibasami et al., 2000; Miyahara et al., 2001). With increasing knowledge on the dietary and health benefits of sesame, the market demand for its seed and oil has enlisted a continuous steep increase. Likewise, sesame by virtue of low irrigation requirement, adaption to different types of soil and weather conditions, not being labor intensive and being instead a highly remunerative crop, is ideally suited to replace low-yield crops, especially in the current scenario of global warming affecting crop productivity in more and more traditional agricultural areas. As a result, sesame production is rapidly increasing over the years and is becoming an alternative important cash crop for smallholders, thus helping to alleviate rural poverty (Figure 1). In 2014, more than 6 million tons of sesame seeds have been produced under nearly 11 million ha classifying sesame at the ninth rank among the major oil crops (Food and Agriculture Organization Statistical Databases [FAOSTAT], 2015).

Despite its importance, sesame is considered as an orphan crop because it has received very little support from science, industry and policy makers. As a consequence, it lags behind the other major oilseed crops as concerns genetic improvement (Dossa, 2016). Cultivated sesame still has some wild characters including seed shattering, indeterminate growth habit and asynchronous capsule ripening leading to a very weak seed yield (300–400 Kg/ha) (Islam et al., 2016). Furthermore, sesame is often grown in harsh environments and exposed to various biotic and abiotic stresses that heavily impair its productivity (Witcombe et al., 2007). Hence, it has become crucial to enhance sesame germplasms for higher productivity and seed quality to efficiently cope with the growing demand of its oil.

Limited progress has been made in these directions through conventional breeding methods due to a lack of genomic tools and resources for deep insights into the underlying molecular background of the important agronomic traits. In addition, few scientific groups are engaged in sesame research worldwide resulting in a slow pace of sesame improvement strategies. However, in recent years, significant breakthroughs in the “Omics” area have taken sesame research into another higher stage. This has then thus propelled us to review the notable achievements made so far in this field as well as the future perspectives to speed-up sesame improvement.

## EVOLUTIONARY HISTORY OF SESAME RESEARCH

Sesame is a very ancient crop thought to be one of the oldest oil crops known by humankind (Bedigian and Harlan, 1986; Ashri, 1998). Its research history followed three major periods viz. the “germplasm collection and genebank constitution” era, the “classical breeding and genetics” era, and currently the “Omics” era (Figure 2). During the first era (before year 2000), genetic materials of cultivated sesame as well as wild related species were collected from many growing areas, morphologically characterized and different seedbanks have been set up in several countries (Hiltebrandt, 1932; Kinman and Martin, 1954; Bedigian and Harlan, 1986; Bisht et al., 1998). Meanwhile during that period, questions related to the origin and domestication process of the cultivated sesame were the source of long debate and investigations (Hiltebrandt, 1932; Nayar and Mehra, 1970).

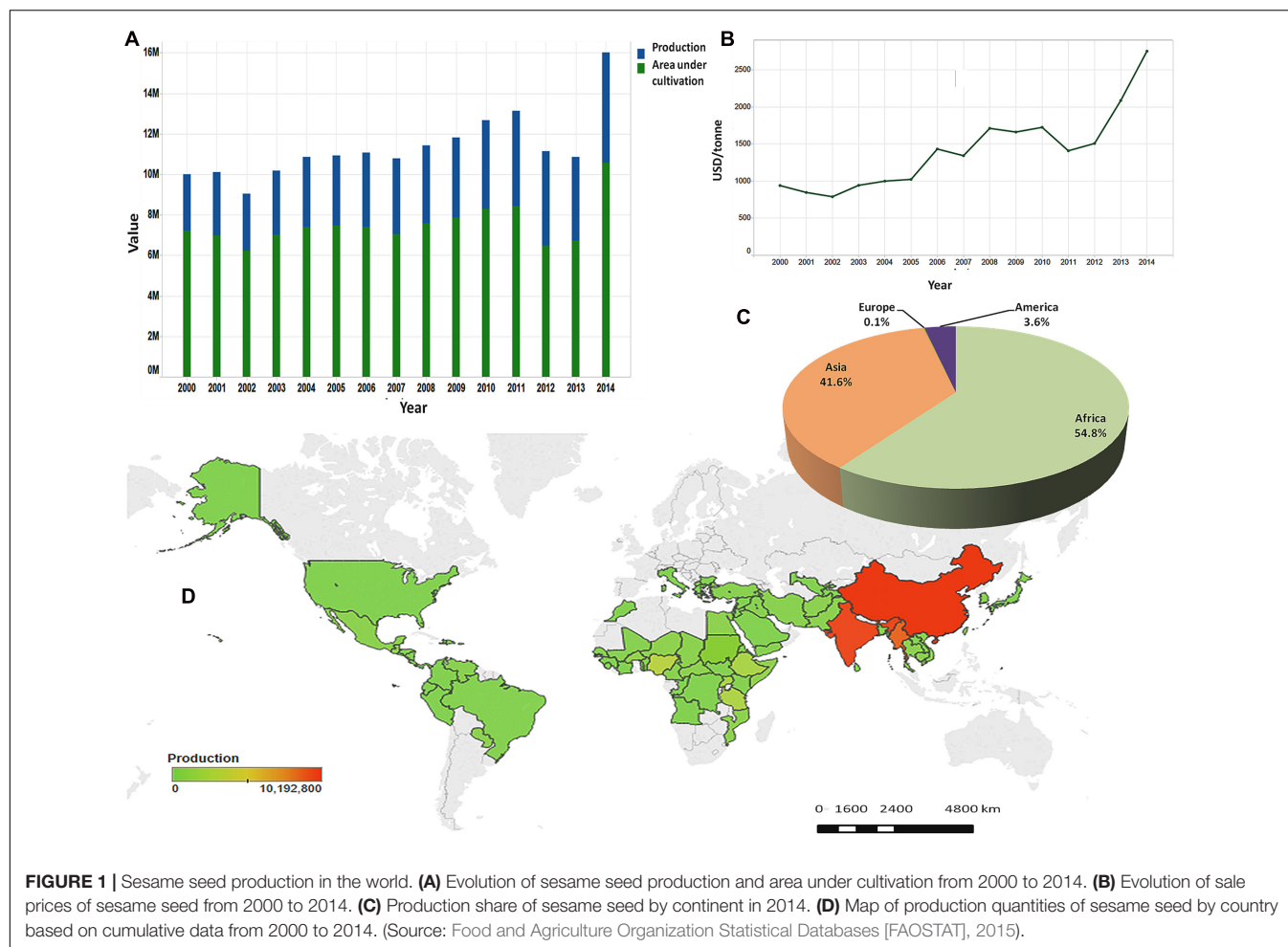
The second era in sesame research (2000–2013) was characterized first, by the employment of classical breeding methods including induced mutation and screening of genotype for desirable characters (Wongyai et al., 2001; Uzun et al., 2003; Boureima et al., 2012). Afterward, sesame research has witnessed a rapid development of genetic tools particularly molecular markers and their application in genetic diversity studies and marker assisted breeding (Dixit et al., 2005; Wang et al., 2012b; Wei X. et al., 2014). In addition, during that period of time, several studies have been undertaken on sesame oil properties that shed more light on the nutritional, pharmaceutical and engineering applications of this untapped crop (Miyahara et al., 2001; Cheng et al., 2006; Saydut et al., 2008; Anilakumar et al., 2010).

Finally, since 2013, the sesame research enters the “Omics” era. With the completion of the nuclear and chloroplast genome sequencing as well as the release of various transcriptomic data, enormous genomic resources have been generated and being applied for sesame improvement.

## GENETIC RESOURCES

The *Sesamum* genus belongs to the Eudicotyledon clade, Lamiales order, Pedaliaceae family (Mabberley, 1997) and *S. indicum* is the well-known and widely grown species within this genus (Ashri, 1998). Kobayashi et al. (1990) proposed 36 species belonging to this genus including 22 species found exclusively in the African continent, five in Asia, seven commonly found in Africa and Asia and one species each in Brazil and a Greek island. Later on, based on works of Bedigian, the list of *Sesamum* species has been revisited to 23 species (IPGRI and NBPGR, 2004) (Table 1). Beside *S. indicum*, the species *S. radiatum* is also cultivated in some African countries as leafy vegetables.

Because most of wild species of the *Sesamum* genus exist only in Africa, sesame has been thought to be originated from this continent (Hiltebrandt, 1932). However, according to evidence in studies Bedigian’ (2003, 2004), it is assumed that the crop has been domesticated from its wild relative species *S. malabaricum* native to south Asia and spread west to Mesopotamia before 2000



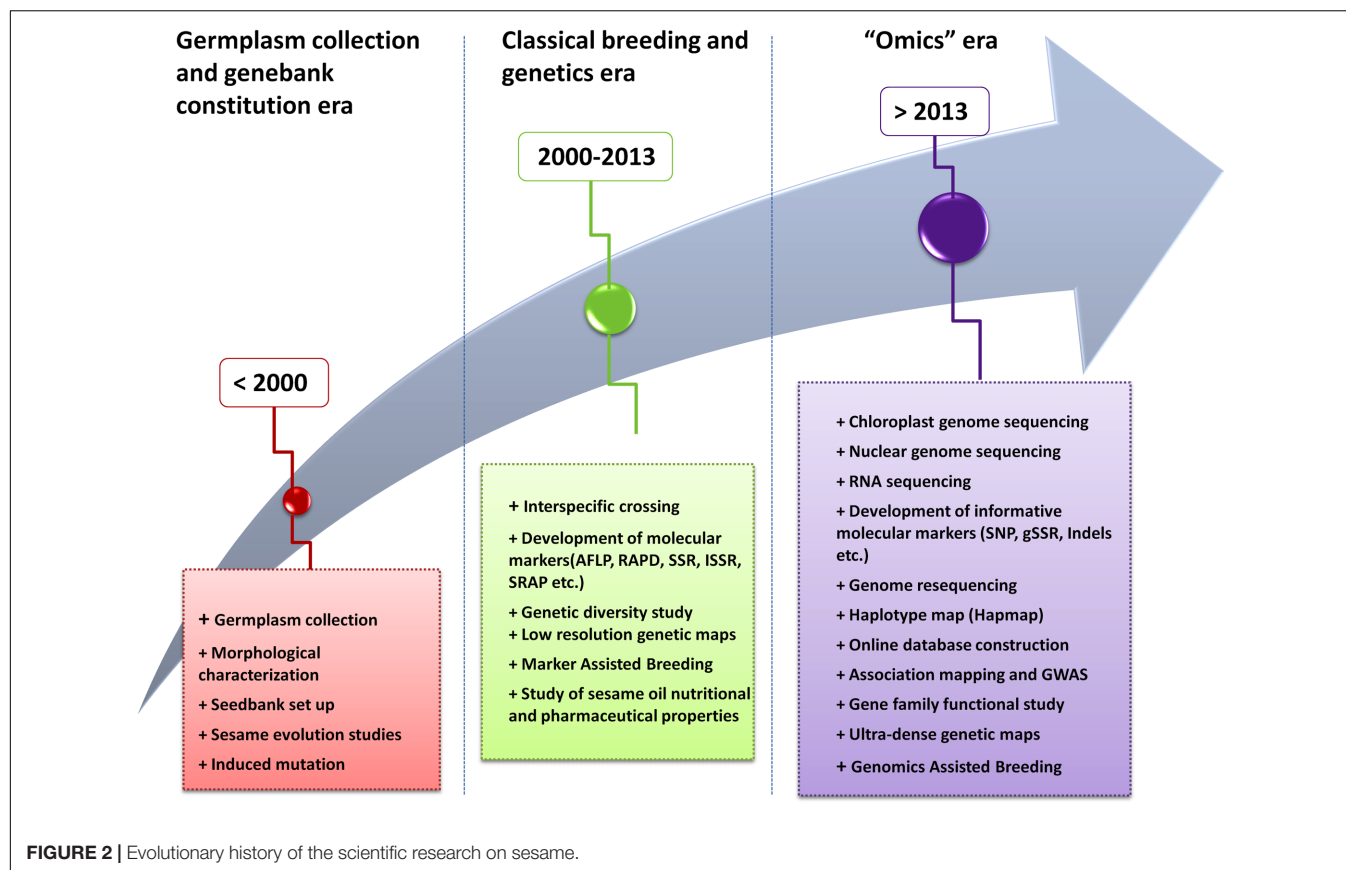
B.C. (Fuller, 2003). Sesame harbors a huge diversity probably because of an adaptation to the various environments where its presence has been recorded coupled with long-term natural and artificial selections (Bedigian and Harlan, 1986; Wei et al., 2015). In total, five major centers of diversity have been proposed for sesame including India, China, Central Asia, the Middle-East and Ethiopia (Zeven and Zhukovsky, 1975). Thanks to the meaningful efforts of the scientific community in sesame germplasm collection, characterization and conservation, huge genetic materials of cultivated sesame along with wild related species are currently preserved in several genebanks around the world mainly in Asia (Zhang Y. et al., 2012) (**Table 2**). The principal genebanks of sesame held in India (NBPGR National Gene Bank), in South Korea (National Agrobiodiversity Center, Rural Development Administration; Park et al., 2015), in China (Oil Crops Research Institute, Chinese Academy of Agricultural Sciences; Wei et al., 2015) and in United States (USDA, ARS, PGRU), have preserved about 25,000 genetic materials (**Table 2**). Moreover, several small-scale genebanks exist in some African countries including Nigeria, Ethiopia, Sudan, etc. Since these genebanks harbor important quantity of genetic resources, it is important to establish core collections (CC) which is a favored approach for the efficient exploration and utilization

of novel variations in genetic resources (Hodgkin et al., 1995). In this vein, researches on sesame CC establishment have been conducted resulting in 362 accessions for Indian germplasm (Bisht et al., 1998), 453 accessions for Chinese germplasm (Zhang et al., 2000) and 278 accessions for Korean germplasm (Park et al., 2015). These are the reservoirs of genetic resources for the present and future sesame improvement programs. Unfortunately, utilization of these wealthy genetic resources for sesame improvement is very limited and most of diversity existing in the germplasm remains unexplored (Dossa et al., 2016a). Furthermore, it becomes apparent that sesame genetic resources from Asia have been well characterized and preserved in contrast to African germplasm which also harbors a valuable diversity (Dossa et al., 2016a). Therefore, further exertions are needed to gather locally available sesame accessions and wild related species from Africa and constitute an extensive genebank for their efficient conservation and exploitation.

## "OMICS" RESOURCES

The genetic and molecular biology study of sesame began very late with only one genetic map published and no report on



**TABLE 1 |** Revised list of *Sesamum* species and their chromosomes number ( $2n$ ).

$2n = 26$	$2n = 32$	$2n = 64$	$2n = \text{indeterminate}$
<i>S. alatum</i> Thonn.	<i>S. capense</i> Burm.f. ssp. lepidotum Schinz	<i>S. radiatum</i> Schum. & Thonn.	<i>S. abbreviatum</i> Merxm.
<i>S. capense</i> Burm.f.	<i>S. angolense</i> Welw.		<i>S. calycinum</i> Welw. ssp. calycinum
<i>S. indicum</i> L.	<i>S. angustifolium</i> Engl.		<i>S. calycinum</i> Welw. ssp. baumii (Stapf) Seidenst. ex. Ihlenf
<i>S. malabaricum</i> Burm.	<i>S. laciniatum</i> Wild.		<i>S. calycinum</i> Welw.ssp. pseudoangolense Seidenst ex.Ihlent
	<i>S. latifolium</i> Gillet		<i>S. marlothii</i> Engl.
	<i>S. prostratum</i> Retz.		<i>S. parviflorum</i> Grabow-Seidenst
			<i>S. pedalioides</i> Heirn
			<i>S. rigidum</i> Peyr. ssp. rigidum
			<i>S. rigidum</i> ssp. merenksyanum Ihlenf. & Seidenst
			<i>S. schinzianum</i> Aschers. ex. Schinz
			<i>S. triphyllum</i> Welw. ex. Aschers
			<i>S. triphyllum</i> Welw. ex. Aschers. var. <i>grandiflorum</i> (Schinz) Merxm

Source: IPGRI and NBPGR, 2004.

**TABLE 2 |** List of worldwide major genebanks available for sesame species.

Country	Institut	Accession numbers	Website
India	NBPGR National Gene Bank	~10,000	www.nbpgr.ernet.in
South Korea	National Agrobiodiversity Center, Rural Development Administration	~7,698	http://www.rda.go.kr/foreign/ten/
China	Oil Crops Research Institute	~7000	http://www.sesame-bioinfo.org/phenotype/index.html
United States	USDA-ARS- PGRU	~1,226	www.ars.usda.gov

quantitative trait loci (QTL) mapping before 2013. However, over the last few years, some significant progress has been made in the development of large-scale genomic resources including informative molecular markers, ultra-dense genetic maps, transcriptome assemblies, multi-omics online platforms etc. In addition, the release of the draft genome of sesame (Wang et al., 2014a) triggered functional analyses of candidate genes related to key agronomic traits. With these invaluable efforts, sesame holds some important genomic resources and platforms for its improvement which for the time being are inexistent in some important oilseed crops such as groundnut. Similarly like pigeonpea (Pazhamala et al., 2015), chickpea, millets (Varshney et al., 2009, 2010), sesame has graduated from an “orphan crop” to a “resource-rich crop”.

## Molecular Markers

Molecular marker technologies have significantly speeded up modern plant breeding in enhancing the genetic gain and reducing the breeding cycles in many crop species. Different types of molecular marker systems have been developed and applied to sesame genotyping and breeding efforts. The first class of molecular markers including Random Amplified Polymorphic DNA (RAPD; Bhat et al., 1999) and Amplified Fragment Length Polymorphism (AFLP; Laurentin and Karlovsky, 2006) were designed and employed mainly for genetic diversity studies. The second class of markers involved basically Simple Sequence Repeat (SSR) types such as Inter-Simple Sequence Repeats (ISSR; Kim et al., 2002), Expressed Sequence Tags-SSR (EST-SSR; Wei et al., 2008, 2011; Badri et al., 2014; Sehr et al., 2016), cDNA-SSR (Spandana et al., 2012; Wang et al., 2012b; Zhang H. et al., 2012; Surapaneni et al., 2014; Wu et al., 2014b), Genome sequence-SSR (gSSR; Dixit et al., 2005; Wei X. et al., 2014; Uncu et al., 2015; Dossa, 2016; Yu et al., 2016), Chloroplast SSR (cpSSR, Sehr et al., 2016). By compiling all developed SSR marker resources, there are in total more than 7,000 validated and 100,000 non-validated SSR markers available for sesame research. Interestingly, a new study is underway to set up an online database gathering all SSR information and providing an integrated platform for functional analyses in sesame. Many of these markers were used for genetic and association mapping, molecular breeding and genetic diversity studies in sesame (Wei et al., 2013; Li et al., 2014; Liu et al., 2015; Uncu et al., 2015; Dossa et al., 2016a). Finally, in recent years with the next generation sequencing (NGS) technology, the third class of molecular markers came into existence. SNPs are more useful as genetic markers than many conventional markers because they are the most abundant and stable form of genetic variation in most genomes. Therefore, the available high-throughput methods for SNP discovery and genotyping have been employed in sesame research including Restriction site-Associated DNA sequencing (RAD-seq; Wu et al., 2014a; Wang et al., 2016a), Specific Length Amplified Fragment Sequencing (SLAF-seq; Zhang et al., 2013), RNA-Seq (Wei L. et al., 2014), Whole-Genome Sequencing (WGS; Wang et al., 2014b; Wei et al., 2015; Zhang et al., 2016), Genotyping by sequencing (GBS; Uncu et al., 2016). Another important marker system referred as insertion/deletions (Indels) has also been reported in sesame (Wei L. et al., 2014; Wu et al., 2014a).

As a whole, molecular marker technologies in sesame are witnessing considerable progress and it is obvious that sesame is no longer lagging far behind major crops in this field.

## Genome Sequence Resources

A high-quality reference genome sequence provides access to the relatively complete gene catalog for a species, the regulatory elements that control their function and a framework for understanding genomic variation. As such, it is a prerequisite resource for fully understanding the role of genes in development, driving genomic-based approaches to systems biology and efficiently exploiting the natural and induced genetic diversity of an organism (Feuillet et al., 2011). Researchers from Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, BGI and other institutes have successfully cracked the nuclear genome of sesame, generating 54.5 Gb of high-quality data from the elite cultivar “Zhongzhi No.13” using the Illumina HiSeq2000 platform. This has been the major breakthrough in the sesame research for decades (Wang et al., 2014a). The high-quality draft genome encompassing 27,148 genes distributed on 16 Linkage Groups (LG) with 274 Mb of size, has become the reference genome for biology study in sesame<sup>1</sup>. This genome with a contig N50 of 52.2 kb and a scaffold N50 of 2.1 Mb has been recently upgraded to reach 13 pseudochromosomes, 94.3% of the estimated genome size and 97.2% of the predicted gene models in sesame (Wang et al., 2016a). In parallel, another genome sequencing project was initiated under the auspices of the Sesame Genome Working Group (SGWG). By 2013, they assembled from the variety “Yuwhi 11” a genome size of 293.7 Mb out of the 354 Mb estimated in sesame and predicted the function of 23,713 genes<sup>2</sup>. More recently, two new genome sequences from sesame landraces (“Baizhima” and “Mishuozhima”) have also been released, increasing the genome sequence resources available for this crop (Wei et al., 2016).

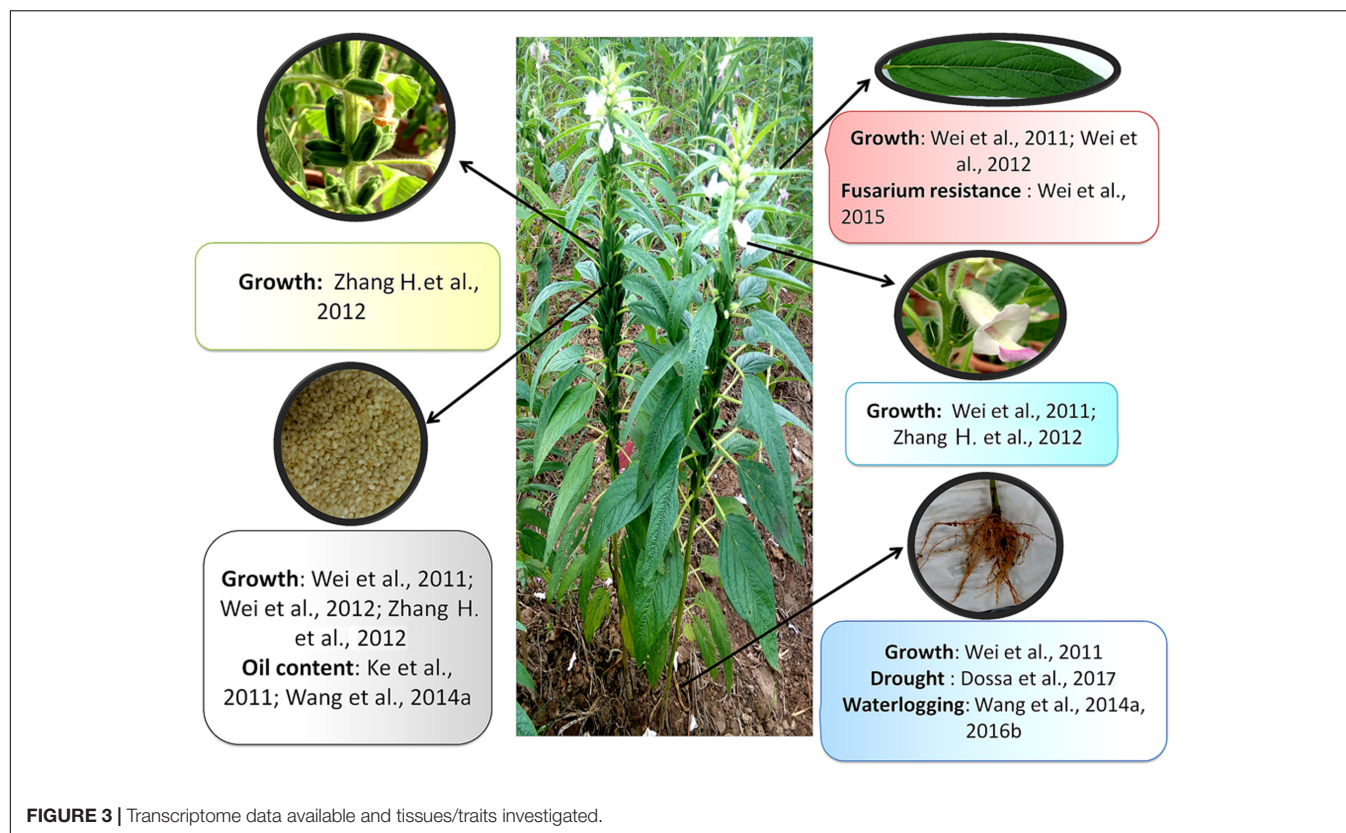
In addition, works carried out by a team of National Bureau of Plant Genetic Resources from India resulted in the genome sequencing of the Indian variety<sup>3</sup> “Swetha”. Of note, nearly 1000 sesame accessions and mapping population have been re-sequenced providing tremendous and inestimable genome-wide information (Zhang et al., 2013a, 2016; Wang et al., 2014b, 2016a; Uncu et al., 2015, 2016; Wei et al., 2015). Nowadays, gene family study, gene fine-mapping, gene cloning and molecular breeding, genome wide association studies (GWAS), genome variation and evolution studies are feasible (Wei et al., 2015, 2016; Dossa et al., 2016b,d; Yu et al., 2017). Novel breeding approaches such as genomic selection (GS) could be implemented in sesame and accelerate the crop improvement.

Beyond the nuclear genome sequences of sesame, the chloroplast genome has also been decrypted first, in a black-seeded cultivar “Ansanggae” (Yi and Kim, 2012) and subsequently, in a white-seeded cultivar “Yuzhi11” (Zhang et al., 2013b). These studies indicated that *Sesamum* (Pedaliaceae family) is a sister genus to the *Olea* and *Jasminum* (Oleaceae

<sup>1</sup><https://www.ncbi.nlm.nih.gov/biosample/SAMN02981519>

<sup>2</sup><https://www.ncbi.nlm.nih.gov/biosample/SAMN04574066>

<sup>3</sup><https://www.ncbi.nlm.nih.gov/biosample/SAMN02357081>



family) clade and represents the core lineage of the Lamiales families.

## Transcriptome Assembly

Transcriptome or EST sequencing is the first step to access the gene contents of a species and has emerged to be an efficient way to generate functional genomic data for non-model organisms. Like genome sequence resources, sesame holds several transcriptome data generated from various organs of the plant (**Figure 3**). The first transcriptome profiling began with works of Suh et al. (2003) who obtained 3,328 ESTs from a cDNA library of 5–25 days old immature sesame seeds. This study shed light on the metabolic pathways involved in lignan biosynthesis in sesame including sesamin and sesamol. Wei et al. (2011) sequenced five tissues using for the first time the high-throughput Illumina paired-end sequencing technology. Likewise, works of Wei et al. (2012), Zhang H. et al. (2012), Wang et al. (2014a) yielded various transcriptome resources related to sesame growth and developmental stages using various sequencing technologies including Illumina HiSeq 2000 and GAI. These studies increased our understanding on the genomic background underpinning sesame growth and development.

On the other hand, given that sesame productivity is seriously hampered by different biotic and abiotic stresses, studies have been designed to unravel the molecular basis of stress tolerance and find out some potential genes to impart stress tolerance in sesame genotypes. The transcriptional response

of root tissues to waterlogging stress was investigated in sesame root (Wang et al., 2012c). More recently based on the Illumina 2500 platform, a time-course transcriptome profiling in two sesame genotypes displaying contrasting tolerance levels provided substantial gene expression data for sesame responses to waterlogging stress (Wang et al., 2016b). Another important abiotic stress impairing sesame productivity is drought stress which is still poorly characterized at the molecular level (Dossa et al., 2016b). For this purpose, Dossa et al. (2017) explored gene expression changes in two sesame genotypes (tolerant and sensitive) based on Illumina HiSeq 4000 sequencing platform, under progressive drought and after re-watering so as to find out some potential genes associated with drought tolerance. Currently, efforts are underway to decipher sesame molecular responses to salt stress, another important abiotic stress.

Concerning the biotic stress very little research has been undertaken at the molecular level, urging more investigations for the development of disease resistant crop. RNA-seq study has been performed on resistant and susceptible sesame cultivars inoculated with *Fusarium oxysporum* f. sp. sesami to clarify the molecular mechanism of sesame resistance to Fusarium Wilt which is one of the main worldwide diseases in sesame, resulting in 15–30% losses of yield (Li et al., 2012; Wei et al., 2016).

Finally, transcriptome profiling approach has also been deployed to explore other important traits. For example, Liu et al. (2016) compared two near-isogenic lines [W1098A with dominant genic male sterility (DGMS) characteristic and its



fertile counterpart, W1098B] to identify differentially expressed genes related to male sterility.

To sum up, huge transcriptome data based on state-of-the-art sequencing technologies from various tissue samples and experimental conditions are now available in sesame and have increased our knowledge on sesame biology. These resources could assist in upgrading the current reference genome in the near future. In this way, a functional transcriptome database should be built to facilitate the exploitation of these incommensurable resources.

## Development of Genetic Maps and Breeding Populations

A genetic linkage map is a prerequisite to better understand the inheritance of traits at the genome-wide level (Verma et al., 2015). It helps to identify molecular markers associated with relevant traits that can be used in breeding programs. The first genetic map in sesame was constructed in 2009 using a combination of 220 EST-SSR, AFLP and Random Selective Amplification of Microsatellite Polymorphic Loci (RSAMPL) markers (Wei et al., 2009). It was built from a  $F_2$  population COI1134  $\times$  RXBS and encompassed 30 LGs covering a genetic length of 936.72 cM with an average marker interval of 4.93 cM (Table 3). At that period, the number of informative molecular markers was limited hence, the development of linkage maps with a good resolution was challenging. Subsequently, this map has been improved to obtain 14 LGs spanning a genome distance of 1,216 cM. In total, 653 markers were successfully mapped with a marker density of 1.86 cM per marker interval. Thanks to the good resolution reached in this linkage map, QTL identification for seed coat color has been conducted for the first time in sesame (Zhang et al., 2013c). During the same period, another linkage map was released based on Specific-Locus Amplified Fragment sequencing (SLAF-seq) technology (Zhang et al., 2013). This map based on an  $F_2$  population Shandong Jiaxiang Sesame  $\times$  Zhongzhi No.13, comprised 1,233 SLAF markers that are distributed on 15 LGs, and was 1,474.87 cM in length with average marker spacing of 1.20 cM. The density of this map was higher than the previous ones and it was the first linkage map to integrate SNP markers in sesame. Surprisingly, all these three published linkage maps lacked common markers. Furthermore, an important knot is that they were constructed from temporary population ( $F_2$ ) that renders repeated phenotyping unfeasible, hence were not ideal for quantitative traits mapping. In this regard, another high-throughput sequencing technology named RAD-seq, has been adopted by Wu et al. (2014a) to map in a high-density linkage map nearly 1,230 markers and identified several QTLs linked to grain yield-related traits based on a recombinant inbred line (RIL) population from Zhongzhi14  $\times$  Miaoqianzhima.

Zhang et al. (2014) also developed a RIL population from Zhongzhi No.13  $\times$  Yiyangbai to analyzed QTLs related to waterlogging tolerance. Furthermore, linkage analysis strategy has been applied in an inter-specific population Ezhi1 (*S. indicum* L.,  $2n = 26$ )  $\times$  Yezhi2 (*S. mulayanum* Nair,  $2n = 26$ ) and a RIL from 95 ms-5A  $\times$  95 ms-5B to study genic male sterility (GMS) traits (Zhao et al., 2013; Liu et al., 2015).

Although the resolution in sesame genetic mapping steeply increased over the years, none of the published maps included the same number of LGs as the number of chromosomes known in sesame. It was just recently that the significant works of Wang et al. (2016a), Zhang et al. (2016), and Mei et al. (2017) have permitted to reach this milestone. The first map was based on 430 RILs from Zhongzhi No.13  $\times$  ZZM2748 (semi-dwarf) and included 1,522 bins anchored on 13 LGs spanning 1090.99 cM genome length with a mean interval distance of 0.72 cM between adjacent bins. This bin map was used to identify several QTLs for sesame plant height and seed coat color (Wang et al., 2016a). The second one was constructed on the basis of a  $F_2$  population from Yuzhi11 (indeterminate growth)  $\times$  Yuzhi DS899 (determinate growth). This SNP map was comprised of 3,041 bins including 30,193 SNPs in 13 LGs with an average marker density of 0.10 cM. At present, this is densest linkage map in sesame and was efficiently applied for map-based gene identification (Zhang et al., 2016). The third one was built on 150 BC<sub>1</sub> from Yuzhi4  $\times$  Bengal Small-seed and encompassed 9,378 SLAF markers anchored onto 13 LGs spanning a total genetic distance of 1,974.23 cM with an average genetic distance of 0.22 cM.

Worth noting, a recent linkage map has been constructed from a RIL population Acc. No. 95–223  $\times$  Acc. No. 92–3091 and yielded 13 LGs encompassing 914 cM with 432 markers including 420 SNPs and 12 SSRs (Uncu et al., 2016). Besides, a haplotype map (Hapmap) has been constructed from 705 worldwide accessions providing 5,407,981 SNPs information with an average LD decay rate of 88 kb in the whole sesame genome. This genotyped worldwide panel has been selected as a “training population” which constitutes a tremendous resource for GWAS, Genomic Selection (GS) and evolution studies in sesame (Wei et al., 2015).

Definitively, from the first low-resolution genetic map constructed in 2009 to the ultra-dense and high-resolution recent ones, there is evidence that the genetic research in sesame has impressively improved and has reached a new dimension. Nonetheless, in order to use these map resources to their maximum advantage, it would be ideal to construct a consensus map that will provide a framework of unprecedented marker density and genome coverage for fine QTL analysis, association mapping, thus facilitating the application of molecular breeding strategies in diverse sesame germplasms. In this regard, an effective collaboration between the different sesame working teams will help achieve this goal.

## Online Functional Database Resources

Owing to the multitude genomic sequence resources being released for sesame research, several integrative online databases have been created to gather sesame information and providing user-friendly platforms for researchers to easily study the molecular function of genome components, for comparative genomics and breeding applications (Table 4). The first web-based functional platform on sesame was set by the Sesame Genome Project but it does not supply detailed sesame genomic or genetic data at the time. Hence, after the release of the



**TABLE 3 |** Genetic mapping and association mapping studies in sesame.

Authors	Traits	Marker systems	Marker number	Population	Major results
Wei et al., 2009	–	EST-SSR, AFLP and RSAMPL	220	96 F <sub>2</sub>	First genetic map
Zhang et al., 2013b	Seed coat color	SSR, AFLP and RSAMPL	653	BC <sub>1</sub> , BC <sub>2</sub> , F <sub>2</sub>	4 QTLs
Zhang et al., 2013c	–	SLAFseq	1,223	107F <sub>2</sub>	First dense genetic map
Wei et al., 2013	Seed quality traits	SRAP, SSR and AFLP	79	216 Chinese collection	10 markers associated with oil, oleic acid, linoleic acid and protein
Wu et al., 2014a	Grain yield traits	SNP	1,230	224 RIL	30 QTLs
Li et al., 2014	Oil and protein content	SSR	112	369 worldwide accessions	19 SSR linked to oil content and 24 SSR associated with protein content
Wei et al., 2015	56 agronomic traits	SNP	1,800,000	705 worldwide accessions	549 associated loci and 46 causative genes linked to key agronomic traits
Wang et al., 2016a	Plant height and seed coat color	SNP	1,522 bins	430 RIL	41 QTLs linked to plant height and 9 QTLs linked to First genetic map with 13 LGs
Wei et al., 2016	Seed coat color	SSR	400	550 F <sub>6</sub>	6 QTLs linked to seed coat color and Identification of <i>PPO</i> gene related to black seed coat color
Zhang et al., 2016	Determinate growth habit	SNP	30,193	120 F <sub>2</sub>	First ultra-dense genetic map with 13 LGs Identification of the determinacy gene <i>SIDt</i>
Uncu et al., 2016	–	SNP and SSR	432	93 RIL	Genetic map with 13 LGs
Mei et al., 2017	Number of flowers per axil and Branching habit	SLAF	9,378	150 BC <sub>1</sub>	High-density genetic map with 13 LGs, identification of <i>SIFA</i> and <i>SIBH</i> linked to Number of flowers per axil and Branching habit, respectively

**TABLE 4 |** List of available online databases for functional genomics in sesame.

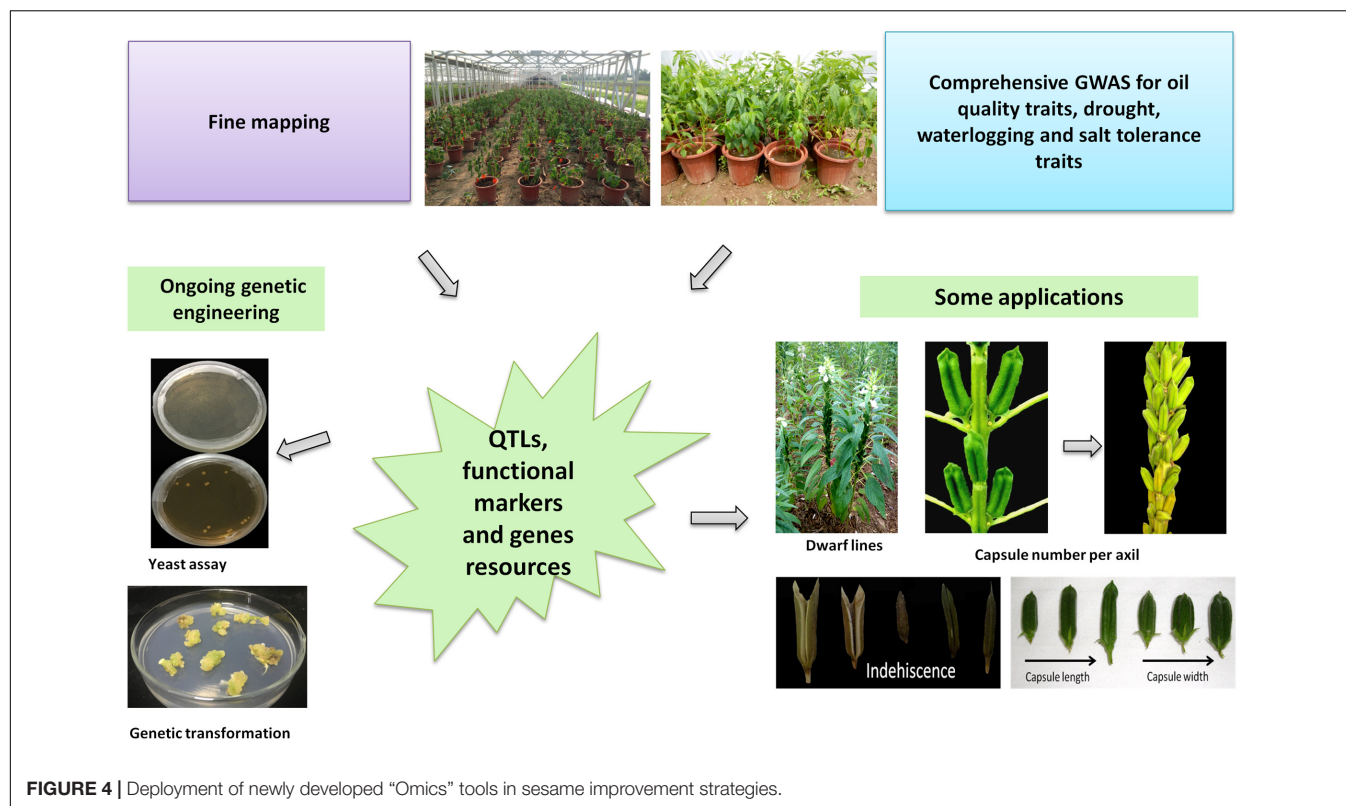
Database name	Website	Utility	Reference
Sinbase	<a href="http://www.ocri-genomics.org/Sinbase/index.html">http://www.ocri-genomics.org/Sinbase/index.html</a>	Genomics/Comparative genomics/Genetics/Phenotypes etc.	Wang et al., 2014c
SesameHapMap	<a href="http://202.127.18.228/SesameHapMap/">http://202.127.18.228/SesameHapMap/</a>	Genome wide SNP	Wei et al., 2015
SesameFG	<a href="http://www.ncgr.ac.cn/SesameFG">http://www.ncgr.ac.cn/SesameFG</a>	Genomics/Evolution/breeding/comparative genomics/Molecular markers/Phenotypes/Transcriptomics	Wei et al., 2017
SisatBase	<a href="http://www.sesame-bioinfo.org/SisatBase/">http://www.sesame-bioinfo.org/SisatBase/</a>	Genome wide SSR	–
The Sesame Genome Project	<a href="http://www.sesamegenome.org">http://www.sesamegenome.org</a>	Genomics	Zhang et al., 2013
Sesame Germplasm Resource Information Database	<a href="http://www.sesame-bioinfo.org/phenotype/index.html">http://www.sesame-bioinfo.org/phenotype/index.html</a>	Plant phenotype	–
NCBI*	<a href="http://www.ncbi.nlm.nih.gov/genome/?term=sesame">http://www.ncbi.nlm.nih.gov/genome/?term=sesame</a>	Versatile	–
ocsESTdb*	<a href="http://www.ocri-genomics.org/ocsESTdb/index.html">http://www.ocri-genomics.org/ocsESTdb/index.html</a>	Seed expression sequence tags/comparative genomics	Ke et al., 2015
PTGBase*	<a href="http://www.ocri-genomics.org/PTGBase/index.html">http://www.ocri-genomics.org/PTGBase/index.html</a>	Tandem duplication/evolution	Yu et al., 2015
PMDBase*	<a href="http://www.sesame-bioinfo.org/PMDBase">http://www.sesame-bioinfo.org/PMDBase</a>	SSR information	Yu et al., 2016

\*These database involved several species including sesame.

reference genome, a versatile online database referred to as Sinbase<sup>4</sup> was developed and provides digestible information related to sesame genomics and genetics (Wang et al., 2014c). This database included genomic component annotations to allow users to study sesame more thoroughly; genetic linkage groups for gene cloning; QTL for genetic linkage analyses; colinear blocks and orthologous genes to perform comparative and evolutionary analyses in sesame and other related species. Additionally, extensive phenotyping data were

supplied describing variations in sesame plant (growth habits, branching styles, number of flowers and capsules per axil, number of carpels in a capsule, flower colors, capsule length etc.). Overall, 10 comprehensive online databases are currently available to study sesame biology including five platforms focusing solely in sesame on aspects related to genome functional components, gene expression, SSR, SNP, Indels, Transposons, QTL and functional genes, gene family, comparative genomics, genetic maps, haplotype map, phenotypes, etc. (Zhang et al., 2013a; Wang et al., 2014c; Ke et al., 2015; Yu et al., 2015, 2016; Wei et al., 2016, 2017).

<sup>4</sup><http://ocri-genomics.org/Sinbase>



## APPLICATION OF NEWLY DEVELOPED “OMICS” TOOLS IN SESAME RESEARCH

With mounting development of versatile “Omics” tools in sesame, their application and deployment for the crop improvement strategies have also yielded conspicuous results. Nowadays, these tools enable high efficiency and resolution in genetic diversity study, gene-trait association analysis using bi-parental or natural diverse populations, gene family study, RNA-seq based candidate gene identification. Various traits of the plant have been tagged by researchers including oil content and quality traits, yield components, tolerance to drought and waterlogging stresses, disease resistance, good plant architecture etc. A glimpse of the diverse applications of “Omics” tools in sesame research is presented in **Figure 4**.

### Germplasm Characterization

The availability of molecular markers has paved the way for several genetic diversity studies in sesame. Knowledge on the genetic diversity and population structure of germplasm collections is an important foundation for crop improvement and a key component of effective conservation and breeding strategies (Thomson et al., 2007). Pathak et al. (2014) have previously reviewed in detail the numerous genetic diversity studies conducted on sesame germplasms. It emerges that sesame harbors a valuable diversity and incongruence between geographical proximity and genetic distance has been reported (Bhat et al., 1999; Laurentin and Karlovsky, 2006; Abdellatef

et al., 2008; Pham et al., 2009; Zhang Y. et al., 2012; Alemu et al., 2013; Abate et al., 2015). However, few informative and polymorphic markers were used in these studies and might not accurately distinguish the samples. Since the genome sequence of sesame becomes available, thousands of highly informative markers including gSSRs, SNPs and Indels covering the entire genome, have been developed and applied for genetic diversity studies (Wang et al., 2014b; Wei X. et al., 2014; Wu et al., 2014b; Wei et al., 2015; Uncu et al., 2015; Dossa et al., 2016a). These studies in contrast, have proved certain patterns of association between genetic similarity and geographical proximity in sesame. Even though the exchange of sesame materials between diverse locations especially through seed trading increases gene flow, in most of cases, some locally unique gene pools still exist. For example, Dossa et al. (2016a) investigated for the first time the unknown sesame accessions from West Africa and found that they were unique and distinct from all the rest. These observations suggested that newly developed molecular markers would bring more precision and efficiency in both genetic studies and breeding programs. Furthermore, combination of molecular and morphological characterizations has been employed to assess the diversity in sesame germplasm. A total of 137 Turkish germplasm has been characterized using both morphological and molecular data which led to the selection of a core collection useful for sesame preservation and breeding (Frary et al., 2014). Ugandan sesame landraces were also investigated and results showed incongruence between morphological data and molecular data (Sehr et al., 2016). Similarly, Pandey et al. (2015) analyzed a worldwide germplasm dominated by Indian

accessions. They detected a high genetic diversity within the germplasm but there was an insignificant correlation between phenotypic and molecular marker information which highlighted the importance to associate both genetic and phenotypic diversity to efficiently inform on the extent of the variation present in sesame germplasm.

Concerning wild related species, very few genetic studies have been conducted for their characterization. Uncu et al. (2015) uncovered a high rate of marker transferability between *S. indicum* and *S. malabaricum*, supporting the designation of the two taxa as cultivated and wild forms of the same species. In earlier works, Nyongesa et al. (2013) found that, there is a high genetic diversity within wild sesame species. The clustering pattern of wild and the cultivated forms indicated that there is no cross-pollination between them during domestication. In addition, it was proved that the genetic diversity of sesame had been eroded due to selection after domestication (Wu et al., 2014b; Wei et al., 2015; Pathak et al., 2015; Mondal et al., 2016). Therefore, future sesame cultivation would benefit from the incorporation of alleles from sesame's wild relatives. Wild species of sesame possess genes for resistance to major biotic and abiotic stresses and adaptability to different environments (Joshi, 1961). Unfortunately, in contrast to several crops such as groundnut, cotton, sunflower, rice, maize, wheat, tomato, soybean, etc., which are profiting from their wild related species for the improvement of cultivars (Zamir, 2001), the introgression of useful genes from wild species into cultivars via conventional breeding has not been so far successful in sesame mainly due to post-fertilization barriers (Tiwari et al., 2011).

## Functional Genomics Research for Key Agronomic Traits

### Oil and Quality Traits

Sesame is primarily grown for its oil-bearing seed. Beside the high oil content, sesame seeds contain almost 18% proteins and among the fatty acid compositions, oleic acid (39.6%) and linoleic acid (46%) are the two main components with the ideal ratio of almost 1:1 (Anilakumar et al., 2010). Wherefore, numerous studies have early attempted to decipher the genetic basis of the oil yield and quality which are some key agronomic traits in sesame breeding. Even so, until 2013, the molecular mechanisms of the high oil content and quality in sesame seeds were still unclear (Jin et al., 2001; Chun et al., 2003; Suh et al., 2003; Ke et al., 2011). An association mapping of oil content, protein content, oleic acid concentration, and linoleic acid concentration based on multi-environment trials was conducted using 79 SSR, SRAP, and AFLP markers in 216 Chinese sesame accessions (Wei et al., 2013). Only one associated marker (*M15E10-3*) was identified for oil content in two environments suggesting inadequate molecular markers and/or germplasm resources. In this regard, Li et al. (2014) analyzed 369 sesame accessions with larger phenotypic variation under 5 environments using 112 informative SSRs. A total of 19 and 24 SSRs were detected for oil content and protein content, respectively. From these, 19 markers were shared by both traits suggesting that oil and protein contents are controlled mostly by similar and major genes. By combining genome information

(Zhang et al., 2013a) to the association mapping results, 36 candidate genes related to lipid pathway including fatty acid elongation gene and a gene encoding Stearoyl-ACP Desaturase were identified. Later, Chen et al. (2014) investigated the sequence divergence in the coding region of the Fatty acid desaturase (FAD) gene between wild and cultivated sesame. They found some nucleotide polymorphisms located in enzyme active site between the wild and cultivated forms which may contribute to the higher fatty acid composition in the cultivated sesame. The specific primers linked to these functional SNPs would be for a great importance in molecular breeding toward high fatty acid content in sesame varieties.

The release of the reference genome sequence has provided an unparalleled opportunity to further excavate the molecular basis of high oil content and quality traits in sesame (Wang et al., 2014a). The sesame genome was found to harbor low copy of lipid-related genes (708) compared to related species such as soybean (1,298). This finding was unexpected since there is an obvious difference in oil contents between sesame (~55%) and soybean (~20%). More interestingly, by combining comparative genomic and transcriptomic analyses, authors have discovered that some lipid gene families, especially the transfer protein type 1 (LTP1) genes beneficial for high oil accumulation have been expanded and retained during domestication, while lipid degradation-related families were found reduced in sesame compared to soybean and this may underlay sesame high oil content. Additionally, important genes in the triacylglycerol biosynthesis pathway were found highly implicated in the oil accumulation during early stages of sesame seed development. Finally, two potential key genes *SiDIR* (*SIN\_1015471*) and *SiPSS* (*SIN\_1025734*) were detected for sesamin production in sesame (Wang et al., 2014a).

Genome wide association studies takes full advantage of ancient recombination events to identify the genetic loci underlying traits at a relatively high resolution (Huang and Han, 2014). In a comprehensive GWAS for oil and quality traits in 705 sesame accessions under 4 environments, 13 significant associations were unraveled for oilseed compounds including oil, protein, sesamin, sesamolin, SFA, Unsaturated Fatty Acid (USFA) and the ratio SFA/USFA. Several causative genes were uncovered for oil content [*SIN\_1003248*, *SIN\_1013005*, *SIN\_1019167*, *SIN\_1009923*, *SiPPO* (*SIN\_1016759*) and *SiNST1* (*SIN\_1005755*)], for fatty acid composition [*SiKASI* (*SIN\_1001803*), *SiKASII* (*SIN\_1024652*), *SiACNA* (*SIN\_1005440*), *SiDGAT2* (*SIN\_1019256*), *SiFATA* (*SIN\_1024296*), *SiFATB* (*SIN\_1022133*), *SiSAD* (*SIN\_1008977*), *SiFAD2* (*SIN\_1009785*)], for sesamin and sesamolin content [*SiNST1* (*SIN\_1005755*)] and protein content [*SiPPO* (*SIN\_1016759*)].

Further studies of Zhang et al. (2013), Wang et al. (2016a) resulted in 4 QTLs (*QTL1-1*, *QTL11-1*, *QTL11-2* and *QTL13-1*) and 9 QTLs (*qSCa-8.2*, *qSCb-4.1*, *qSCb-8.1*, *qSCb-11.1*, *qSCL-4.1*, *qSCL-8.1*, *qSCL-11.1*, *qSCa-4.1* and *qSCa-8.1*) detected for seed coat color, respectively. Additionally, the gene *SiPPO* (*SIN\_1016759*) has been recently detected through fine mapping, as the candidate gene that controlling seed coat color in sesame (Wei et al., 2016). Seed coat color is an important agronomic

trait in sesame, as it has been shown that white sesame seeds typically have higher oil, sesamin or sesamolin content (Wang et al., 2012a), whereas black sesame seeds usually have higher ash and carbohydrate content and lower protein, oil, and moisture ratios (Kanu, 2011). Though, these QTLs harbor dozen of genes, further screening will help to pinpoint the candidate genes.

Compiling all these meaningful results regarding oil and quality traits, researchers actually have substantial genomic information at their disposal for breeding and releasing higher nutritional cultivars to meet the various demands of oil markets.

### Waterlogging and Drought Tolerance

Sesame is highly susceptible to waterlogging stress. The crop experiences a reduction in growth and yield after 2–3 days of waterlogging, which frequently occurs when they are grown on soils that are poorly drained (Ucan et al., 2007). Wang et al. (2012c) found 13,307 differentially expressed genes (DEGs) in sesame under waterlogging stress. In a more comprehensive study, a total of 1,379 genes were found as the core gene that functions in response to waterlogging. News worthily, they reported 66 genes that may be candidate for improving sesame tolerance to waterlogging (Wang et al., 2016b). Meanwhile, 6 QTLs (*qEZ09ZCL13*, *qWH09CHL15*, *qEZ10ZCL07*, *qWH10ZCL09*, *qEZ10CHL07*, and *qWH10CHL09*) linked to waterlogging traits were identified and a SSR marker (ZM428) closely linked to *qWH10CHL09* was further reported as effective marker for marker-assisted selection (MAS) toward waterlogging tolerance (Zhang et al., 2014). Currently, studies are being implemented to unveil genomic variants associated with waterlogging tolerance in sesame.

Concerning drought stress in sesame, few molecular researches have been conducted so far in this field. Using a comparative genomic approach, Dossa et al. (2016b) identified in the whole sesame genome a set of 75 candidate genes for drought tolerance enriched in transcription factors (TFs). Hence, they afterward dissected two important TF families (AP2/ERF and HSF) and proposed some candidate TFs for drought tolerance improvement in sesame (Dossa et al., 2016c,d). Evolutionary analyses of these families showed that sesame has retained most of its drought-related genes similarly as uncovered for its oil-related genes. This may thus explain the relative high drought tolerance observed in this crop. A recent RNA-seq analysis demonstrated that 722 genes act as the core gene set involved in drought responses and 61 candidate genes conferring higher drought tolerance were discovered (Dossa et al., 2017). In another very recent report, an Osmotin-like gene (*SindOLP*) has been uncovered to enhance tolerance to drought, salinity, oxidative stresses, and the charcoal rot pathogen in transgenic sesame (Chowdhury et al., 2017). Finally, a study is underway to decipher SNP variants significantly linked to various drought tolerance traits through an inclusive GWAS.

### Productivity Enhancement

Although sesame has been domesticated since long time, the yield in most of the growing areas is still very low, thus, hampering its adoption and expansion in the world. Grain yield of sesame per plant is considered to be composed of three

components, i.e., the number of capsules per plant, the number of grains per capsule and the grain weight. Some other factors, including plant height, length of capsules, number of capsules per axil and axis height of the first capsule were found to be strongly associated with grain yield of sesame (Biabani and Pakniyat, 2008). Concerning the plant height trait, some QTLs have been reported including *Qph-6* and *Qph-12* (Wu et al., 2014a); 41 QTLs were further identified and the major QTL *qPH-3.3*, was predicted to be responsible for the semi-dwarf sesame plant phenotype (Wang et al., 2016a). However, it contains 102 candidate genes and thus needs further excavation to pinpoint the causative gene. A semi-dwarf gene [*SiGA20ox1* (*SIN\_1002659*)] has been recently detected through fine-mapping strategy (Wei et al., 2016). Moreover, two important candidate genes for plant height *SiDFL1* (*SIN\_1014512*) and *SiILR1* (*SIN\_1018135*) were found in works of Wei et al. (2015). These findings will undoubtedly assist in efforts to create mechanized cultivation varieties with super high yield.

Quantitative trait loci were also identified for the capsule related trait including capsule number per plant (*Qcn-11*), First capsule height (*Qfch-4*, *Qfch-11*, and *Qfch-12*), capsule axis length (*Qcal-5* and *Qcal-9*), capsule length (*Qcl-3*, *Qcl-4*, *Qcl-7*, *Qcl-8*, and *Qcl-12*) (Wu et al., 2014a). Similarly, the gene *SiACS* (*SIN\_1006338*) coding for the number of capsule per axil was discovered and may be an important asset for yield improvement in sesame (Wei et al., 2015). Since the number of capsules per axil is related to the number of flowers per axil, Mei et al. (2017) successfully mapped a gene *SiFA* (mono-flower vs. triple-flower) onto the LG11 flanked by the markers *Marker58311* and *Marker36337*.

In regard to the grain yield, few QTLs are actually available including *Qgn-1*, *Qgn-6*, and *Qgn-12* for grain number per capsule and *Qtgw-11* for thousand grain weight.

Flowering time is also an important trait for adaptation of crops to different agro-climatic conditions that significantly affects the yield. Two candidate genes at flowering-time loci *SiDOG1* (*SIN\_1022538*) and *SiIAA14* (*SIN\_1021838*) have been discovered (Wei et al., 2015). Sesame's indeterminate growth habit is one of the reasons of its low yielding capacity compared to other oilseed crops (Yol and Uzun, 2012). Recently, the gene *SiDt* (*DS899s00170.023*) was detected as a target gene for conferring the determinate trait in sesame cultivar (Zhang et al., 2016). Also, the branching habit which is an important trait in sesame as it plays a cardinal role in grain yield, cultivation practices and mechanized harvest, has been investigated. A gene *SiBH* controlling the branching habit (uniculm vs. branched type) was mapped onto the LG5 flanked by the markers *Marker129539* and *Marker31462*. This QTL region will be useful for developing unbranched sesame varieties fit for mechanized harvest (Mei et al., 2017).

In another realm, AFLP markers *P01MC08*, *P06MG04* and *P12EA14* were found to be linked to the recessive GMS gene *SiMs1* (Zhao et al., 2013) and 13 SSRs including *SBM298* and *GB50* were associated to the dominant GMS gene *Ms* (Liu et al., 2015). These markers will be valuable for marker-aided breeding of GMS hybrids and harnessing heterosis as one of the promising approaches for yield improvement in sesame (Murty, 1975).



## CURRENT HOT-TOPICS IN SESAME RESEARCH AND FUTURE DIRECTIONS

Recent years have witnessed a continuously increasing number of functional genes discovered for key agronomic traits in sesame thanks to the availability of versatile “Omics” tools. In this regard, the next logical step in the sesame research is the functional validation of these gene resources through genetic engineering approaches. Genetic transformation would be an ideal opportunity to quickly transfer the functional genes into sesame elite cultivars. Actually, several successful attempts of sesame genetic transformation through *Agrobacterium* have led to up 42.66% of transformation efficiency (Yadav et al., 2010; Al-Shafeay et al., 2011; Chowdhury et al., 2014). However, improvements of the sesame genetic transformation protocol for reaching higher efficiency are imperative. Nonetheless, studies are in the offing to transfer candidate genes for oil quality traits as well as abiotic stress tolerance into elite cultivars and unveil their molecular mechanisms. In this vein, the first study of the functional analysis in transgenic sesame came into being very recently (Chowdhury et al., 2017). This report presages a bright future for the genetic engineering era of sesame.

Alternatively, the available “Omics” tools actually spur us to enquire scientific questions that remain unexplored or weakly investigated in sesame and could considerably aid in efforts toward its enhancement. These issues constitute the present hot-topics in the sesame research and involve various domains of the crop:

- (a) **Origin and domestication of sesame:** Further investigations into the evolution of sesame have been hampered by the absence of detailed molecular data across multiple sesame accessions and wild related species. There is a need for more characterization of wild germplasm including African species, inter-specific crosses and molecular studies to efficiently harness potentialities of sesame wild species. Therefore, the ongoing sequencing projects of wild sesame species and landraces from different origins would help to definitely state the domestication process of sesame.
- (b) **Study to solve constraints related to the mechanized harvesting in sesame cultivation:** Plant architecture, dehiscence and indeterminate growth habit are some key factors affecting mechanization of sesame harvest. Many QTLs and genes related to these traits are now available and could be applied through molecular breeding to improve sesame ability for mechanized harvesting.
- (c) **Instigation of molecular studies related to root system architecture (RSA), dehiscence character, resistance to disease, tolerance to salt and heat as well as other key agronomic traits for yield improvement:** Sesame root system is thought to play a foremost role in its adaptation to different environments and weather conditions. However, till date, no study has been undertaken to dissect the RSA of sesame and look for some candidate genes linked to this major trait. Furthermore, gene identification for the indehiscence trait in sesame should be a priority in

future researches. Also, the available genomic tools need to be more effectively applied for other no less important agronomic traits that could lead to the enhancement of sesame productivity especially the resistance to biotic and abiotic stresses.

- (d) **Genetic dissection and exploitation of sesame seed's bioactive components:** Sesame seeds contain several bioactive components including tocopherols, sesamin, and sesamol which have tremendous potential for the valorization and value addition of the crop. Hence further studies to investigate the genetic basis of these traits would help to develop nutritionally superior cultivars.
- (e) **More “Omics” technologies to support genomic research:** Although several transcriptome data have been released in sesame, future investigations should combine additional “Omics” technologies including proteomics and metabolomics as they are capital for a full understanding of a biological system. Moreover, the available “draft” reference genome sequence of sesame has to be updated to reach a “full sequence” since it may affect the ability to accurately link sequence variations to phenotypes, allele mining or other biological issues.
- (f) **Application of advanced biotechnological technologies:** Functional genomics and biotechnological methodologies, such as genetic transformation, yeast assay system, new mutation approach such as Target Induced Local Lesions in Genomes (TILLING) and genome-editing technologies using CRISPR/Cas9 system, are much needed to validate the effects of the available functional genes and their functional variants on the key agronomic traits.

## CONCLUSION

Sesame has become an emerging crop in the world and its entrance into the “Omics” era has raised it at the “genomic resource-rich crop” level. Invaluable efforts during recent years have engendered several genetic/genomic tools and resources that provide an impetus to research and nurture sesame production for the benefit of smallholder farmers in developing countries. The major traits in sesame including oil and quality, yield related traits, abiotic stress resistance have been thoroughly explored and our understanding of the molecular basis underlying these traits has deeply increased. More importantly, several functional genes, QTLs and molecular markers linked to these traits are now available and could be employed in sesame breeding programs. The current scope in the sesame research concerns the exploitation of these available genomic information for the effective sesame improvement through molecular- or genomics-assisted breeding. However, the road to raise sesame as one of the major oilseed crops in the world is still long and will need more synergetic efforts, more applications of MAS and inter-disciplinary researches. Other related research fields for the valorization and expansion of sesame should also follow the same trend as the molecular field. Finally and perhaps one of the most important recommendations is to enhance partnerships between national and international sesame teams, so that major issues

of sesame production could be addressed through international projects and effective breeding strategies could be implemented.

## AUTHOR CONTRIBUTIONS

KD, BL, NC, XZ, and DD conceived and designed the paper. KD, LW, XW, YZ, MN, DF, JY, MM, LY, and DD collected and analyzed the literature. KD, LY, MM, and DD drafted the paper. DF, JY, LW, and DD prepared the figures. DF, YZ, LW, XW, MN,

DD, XZ, and NC revised the manuscript. All authors have read and approved the final version of the manuscript.

## ACKNOWLEDGMENTS

We apologize to colleagues whose works and original articles are not cited in this review owing to space limitations. We thank Mr. Sobowale Soremi for his assistance in language editing of the manuscript.

## REFERENCES

- Abate, M., Mekbib, F., Ayana, A., and Nigussie, M. (2015). Assessment of genetic diversity in Ethiopian sesame (*Sesamum indicum* L.) germplasm using ISSR markers. *Br. Biotechnol. J.* 8, 1–13. doi: 10.9734/BBJ/2015/18481
- Abdellatif, E., Sirelkhatem, R., Mohamed, A. M. M., Radwan, K. H., and Khalafalla, M. M. (2008). Study of genetic diversity in Sudanese sesame (*Sesamum indicum* L.) germplasm using random amplified polymorphic DNA (RAPD) markers. *Afr. J. Biotechnol.* 7, 4423–4427.
- Alemu, A., Petros, Y., and Tesfaye, K. (2013). Genetic distance of sesame (*Sesamum indicum* L.) cultivars and varieties from northwestern Ethiopia using inter simple sequence repeat markers. *East Afr. J. Sci.* 7, 31–40. doi: 10.1007/s12298-016-0385-8
- Al-Shafeay, A. F., Ibrahim, A. S., Nesiem, M. R., and Tawfik, M. S. (2011). Establishment of regeneration and transformation system in Egyptian sesame (*Sesamum indicum* L.) cv Sohag1. *GM Crops* 2, 182–192. doi: 10.4161/gmcr.2.3.18378
- Anilakumar, K. R., Pal, A., Khanum, F., and Bawas, A. S. (2010). Nutritional, medicinal and industrial uses of sesame (*Sesamum indicum* L.) seeds. *Agric. Conspec. Sci.* 75, 159–168.
- Ashri, A. (1998). Sesame breeding. *Plant Breed. Rev.* 16, 179–228.
- Badri, J., Yepuri, V., Ghanta, A., Siva, S., and Siddiq, E. A. (2014). Development of microsatellite markers in sesame (*Sesamum indicum* L.). *Turk. J. Agric. For.* 38, 603–614. doi: 10.3906/tar-1312-104
- Bedigian, D. (2003). Evolution of sesame revisited: domestication, diversity and prospects. *Genet. Resour. Crop Evol.* 50, 779–787. doi: 10.1023/A:1025029903549
- Bedigian, D. (2004). History and lore of sesame in Southwest Asia. *Econ. Bot.* 58, 329–353. doi: 10.1663/0013-0001
- Bedigian, D., and Harlan, J. R. (1986). Evidence for cultivation of sesame in the ancient world. *Econ. Bot.* 40, 137–154. doi: 10.1007/BF02859136
- Bhat, K. V., Babrekar, P. P., and Lakhanpaul, S. (1999). Study of genetic diversity in Indian and exotic sesame (*Sesamum indicum* L.) germplasm using random amplified polymorphic DNA (RAPD) markers. *Euphytica* 110, 21–33. doi: 10.1023/A:1003724732323
- Biabani, A. R., and Pakniyat, H. (2008). Evaluation of seed yield-related characters in sesame (*Sesamum indicum* L.) using factor and path analysis. *Pak. J. Biol. Sci.* 11, 1157–1160. doi: 10.3923/pjbs.2008.1157.1160
- Bisht, I. S., Mahajan, R. K., Loknathan, T. R., and Agarwal, R. C. (1998). Diversity in Indian sesame collection and stratification of germplasm accessions in different diversity groups. *Genet. Resour. Crop Evol.* 45, 325–335. doi: 10.1023/A:1008652420477
- Boureima, S., Oukarroum, A., Diouf, M., Cissé, N., and Van Damme, P. (2012). Screening for drought tolerance in mutant germplasm of sesame (*Sesamum indicum*) probing by chlorophyll a fluorescence. *Environ. Exp. Bot.* 81, 37–43. doi: 10.1016/j.envexpbot.2012.02.015
- Chen, Z., Tonniss, B., Morris, B., Wang, R. B., Zhang, A. L., Pinnow, D., et al. (2014). Variation in seed fatty acid composition and sequence divergence in the FAD2 gene coding region between wild and cultivated sesame. *J. Agri. Food Chem.* 62, 11706–11710. doi: 10.1021/jf503648b
- Cheng, F. C., Jinn, T. R., Hou, R. C. W., and Tzen, J. T. C. (2006). Neuroprotective effects of sesamin and sesamolin on gerbil brain in cerebral ischemia. *Int. J. Biomed. Sci.* 2, 284–288.
- Chowdhury, S., Basu, A., and Kundu, S. (2014). A new high-frequency *Agrobacterium*-mediated transformation technique for *Sesamum indicum* L. using de-embryonated cotyledon as explant. *Protoplasma* 251, 1175–1190. doi: 10.1007/s00709-014-0625-0
- Chowdhury, S., Basu, A., and Kundu, S. (2017). Overexpression of a new osmotin-like protein gene (SindOLP) confers tolerance against biotic and abiotic stresses in sesame. *Front. Plant Sci.* 8:410. doi: 10.3389/fpls.2017.00410
- Chun, J. A., Jin, U. H., Lee, J. W., Yi, Y. B., Hyung, N. I., Kang, M. H., et al. (2003). Isolation and characterization of a myo-inositol 1-phosphate synthase cDNA from developing sesame (*Sesamum indicum* L.) seeds: functional and differential expression, and salt-induced transcription during germination. *Planta* 216, 874–880. doi: 10.1007/s00425-002-0940-0
- Dixit, A., Jin, M. H., Chung, J. W., Yu, J. W., Chung, H. K., Ma, K. H., et al. (2005). Development of polymorphic microsatellite markers in sesame (*Sesamum indicum* L.). *Mol. Ecol. Notes* 5, 736–738. doi: 10.1111/j.1471-8286.2005.01048.x
- Dossa, K. (2016). A physical map of important QTLs, functional markers and genes available for sesame breeding programs. *Physiol. Mol. Biol. Plants* 22, 613–619. doi: 10.1007/s12298-016-0385-8
- Dossa, K., Li, D., Wang, L., Zheng, X., Yu, J., Wei, X., et al. (2017). Dynamic transcriptome landscape of sesame (*Sesamum indicum* L.) under progressive drought and after rewatering. *Genom. Data* 11, 122–124. doi: 10.1016/j.gdata.2017.01.003
- Dossa, K., Diouf, D., and Cissé, N. (2016d). Genome-wide investigation of Hsf genes in sesame reveals their segmental duplication expansion and their active role in drought stress response. *Front. Plant Sci.* 7:1522. doi: 10.3389/fpls.2016.01522
- Dossa, K., Niang, M., Assogbadjo, A. E., Cisse, N., and Diouf, D. (2016c). Whole genome homology-based identification of candidate genes for drought resistance in (*Sesamum indicum* L.). *Afr. J. Biotechnol.* 15, 1464–1475. doi: 10.5897/AJB2016.15420
- Dossa, K., Wei, X., Li, D., Zhang, Y., Wang, L., Fonckea, D., et al. (2016b). Insight into the AP2/ERF transcription factor superfamily in sesame (*Sesamum indicum*) and expression profiling of the DREB subfamily under drought stress. *BMC Plant Biol.* 16:171. doi: 10.1186/s12870-016-0859-4
- Dossa, K., Wei, X., Zhang, Y., Fonckea, D., Yang, W., Diouf, D., et al. (2016a). Analysis of genetic diversity and population structure of sesame accessions from Africa and Asia as major centers of its cultivation. *Genes* 7:14. doi: 10.3390/genes7040014
- Feuillet, C., Leach, E. J., Rogers, J., Schnable, P. S., and Eversole, K. (2011). Crop genome sequencing: lessons and rationales. *Trends Plant Sci.* 16, 77–88. doi: 10.1016/j.tplants.2010.10.005
- Food and Agriculture Organization Statistical Databases (FAOSTAT) (2015). *FAOSTAT Provides Free Access to Food and Agriculture Data for Over 245 Countries and Territories and Covers All FAO Regional Groupings*. Available at: <http://faostat.fao.org/> [accessed December 19, 2016].
- Frary, A., Tekin, P., Celik, I., Furat, S., Uzun, B., and Doganlar, S. (2014). Morphological and molecular diversity in Turkish sesame germplasm and selection of a core set for inclusion in the national collection. *Crop Sci.* 54, 1–10. doi: 10.2135/cropsci2012.12.0710
- Fuller, D. Q. (2003). Further evidence on the prehistory of sesame. *Asian Agrihist.* 7, 127–137.
- Hibasami, H., Fujikawa, T., Takeda, H., Nishibe, S., Satoh, T., Fujisawa, T., et al. (2000). Induction of apoptosis by *Acanthopanax senticosus* HARMS and its

- component, sesamin in human stomach cancer KATO III cells. *Oncol. Rep.* 7, 1213–1216.
- Hiltebrandt, V. M. (1932). Sesame (*Sesamum indicum* L.). *Bull. Appl. Bot. Gen. Plant Breed.* 9, 1–114.
- Hodgkin, T., Brown, A. H. D., Hintum, T. J. L. V., and Morales, E. A. V. (1995). *Core collections of Plant Genetic Resources*. London: A co-publication with the International Plant Genetic Resources Institute (IPGRI) and Sayce publishing.
- Huang, X., and Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* 65, 531–551. doi: 10.1146/annurev-arplant-050213-035715
- IPGRI and NBPGR (2004). *International Crop Network Series 5*. Rome: International Plant Genetic Resources Institute, and *Report of An International Workshop on Okra Genetic Resources*. New Delhi: National Bureau of Indian Plant Genetic Resources, 61.
- Islam, F., Gill, R. A., Ali, B., Farooq, M. A., Xu, L., Najeeb, U., et al. (2016). "Sesame," in *Breeding Oilseed Crop for Sustainable Production: Opportunities and Constraints*, ed. S. K. Gupta (Cambridge, MA: Academic Press), 135–147.
- Jin, U., Lee, J., Chung, Y., Lee, J., Yi, Y., Kim, Y., et al. (2001). Characterization and temporal expression of a  $\alpha$ -6 fatty acid desaturase cDNA from sesame (*Sesamum indicum* L.) seeds. *Plant Sci.* 161, 935–941.
- Joshi, A. B. (1961). *Sesamum*. Hyderabad: Indian Central Oilseed Committee.
- Kanu, P. J. (2011). Biochemical analysis of black and white sesame seeds from China. *Am. J. Biochem. Mol. Biol.* 1, 145–157. doi: 10.3923/ajbmb.2011.145.157
- Ke, T., Dong, C., Han, M., Zhao, Y., Chen, H., Liu, H., et al. (2011). Analysis of expression sequence tags from a full-length-enriched cDNA library of developing sesame seeds (*Sesamum indicum*). *BMC Plant Biol.* 11:180. doi: 10.1186/1471-2229-11-180
- Ke, T., Yu, J., Dong, C., Mao, H., Hua, W., and Liu, S. (2015). ocsESTdb: a database of oil crop seed EST sequences for comparative analysis and investigation of a global metabolic network and oil accumulation metabolism. *BMC Plant Biol.* 15:19. doi: 10.1186/s12870-014-0399-8
- Kim, D. H., Zur, G., Danin-Poleg, Y., Lee, S. W., Shim, K. B., Kang, C. W., et al. (2002). Genetic relationships of sesame germplasm collection as revealed by inter-simple sequence repeats. *Plant Breed.* 121, 259–262. doi: 10.1046/j.1439-0523.2002.00700.x
- Kinman, M. L., and Martin, J. A. (1954). Present status of sesame breeding in the United States. *Agron. J.* 46, 24–27.
- Kobayashi, T., Kinoshita, M., Hattori, S., Ogawa, T., Tsuboi, Y., Ishida, M., et al. (1990). Development of the sesame metallic fuel performance code. *Nucl. Technol.* 89, 183–193.
- Laurentin, H. E., and Karlovsky, P. (2006). Genetic relationship and diversity in a sesame (*Sesamum indicum* L.) germplasm collection using amplified fragment length polymorphism (AFLP). *BMC Genet.* 7:10. doi: 10.1186/1471-2156-7-10
- Li, C., Miao, H., Wei, L., Zhang, T., Han, X., and Zhang, H. (2014). Association mapping of seed oil and protein in *Sesamum indicum* L. using SSR Markers. *PLoS ONE* 9:e105757. doi: 10.1371/journal.pone.0105757
- Li, D., Wang, L., Zhang, Y., Lv, H., Qi, X., Wei, W., et al. (2012). Pathogenic variation and molecular characterization of *Fusarium* species isolated from wilted sesame in China. *Afr. J. Microbiol. Res.* 6, 149–154. doi: 10.5897/AJMR11.1081
- Liu, H. M., Tan, H., Yu, L., Li, F., Zhou, M., Yang, T., et al. (2016). Comparative transcriptome profiling of the fertile and sterile flower buds of a dominant genic male sterile line in sesame (*Sesamum indicum* L.). *BMC Plant Biol.* 16:250. doi: 10.1186/s12870-016-0934-x
- Liu, H. Y., Zhou, X. A., Wu, K., Yang, M. M., and Zhao, Y. Z. (2015). Inheritance and molecular mapping of a novel dominant genic male-sterile gene in *Sesamum indicum* L. *Mol. Breed.* 35:9. doi: 10.1007/s11032-015-0189-5
- Mabberley, D. J. (1997). *The Plant-Book. A Portable Dictionary of the Higher Plants*, 2nd Edn. Cambridge: Cambridge University Press.
- Mei, H., Liu, Y., Du, Z., Wu, K., Cui, C., Jiang, X., et al. (2017). High-density genetic map construction and gene mapping of basal branching habit and flowers per leaf axil in sesame. *Front. Plant Sci.* 8:636. doi: 10.3389/fpls.2017.00636
- Miyahara, Y., Hibasami, H., Katsuzaki, H., Imai, K., and Komiya, T. (2001). Sesamol from sesame seed inhibits proliferation by inducing apoptosis in human lymphoid leukemia Molt 4B cells. *Int. J. Mol. Med.* 7, 369–371. doi: 10.3892/ijmm.7.4.369
- Mondal, N., Bhat, K. V., Srivastava, P. S., and Sen, S. K. (2016). Effects of domestication bottleneck and selection on fatty acid desaturases in Indian sesame germplasm. *Plant Genet. Resour.* 14, 81–90. doi: 10.1017/S1479262115000106
- Murty, D. S. (1975). Heterosis, combining ability and reciprocal effects for agronomic and chemical characters in *Sesamum*. *Theor. Appl. Genet.* 45, 294–299. doi: 10.1007/BF00276682
- Nagaraj, G. (2009). *Oilseeds: Properties, Processing, Products and Procedures*, 2nd Edn. New Delhi: New India Publication.
- Nayar, N. M., and Mehra, K. L. (1970). Sesame. Its uses, botany, World. 2nd edn. Oxford University Press, Oxford. Cytogenetics and origin. *Econ. Bot.* 24, 20–31. doi: 10.1007/BF02860629
- Noguchi, T., Ikeda, K., Sasaki, Y., Yamamoto, J., and Yamori, Y. (2001). Effects of vitamin E and sesamin on hypertension and cerebral thrombogenesis in stroke-prone spontaneously hypertensive rats. *Hypertens. Res.* 24, 735–742.
- Nyongesa, B. O., Were, B. A., Gudu, S., Dangasuk, O. G., and Onkware, A. O. (2013). Genetic diversity in cultivated sesame (*Sesamum indicum* L.) and related wild species in East Africa. *J. Crop Sci. Biotechnol.* 16, 9–15.
- Pandey, S. K., Das, A., Rai, P., and Dasgupta, T. (2015). Morphological and genetic diversity assessment of sesame (*Sesamum indicum* L.) accessions differing in origin. *Physiol. Mol. Biol. Plants* 21, 519–529. doi: 10.1007/s12298-015-0322-2
- Park, J., Suresh, S., Raveendar, S., Baek, H., Kim, C., Lee, S., et al. (2015). Development and evaluation of core collection using qualitative and quantitative trait descriptor in sesame (*Sesamum indicum* L.) germplasm. *Korean J. Crop Sci.* 60, 75–84.
- Pathak, N., Bhaduri, A., Bhat, K. V., and Rai, A. K. (2015). Tracking sesamin synthase gene expression through seed maturity in wild and cultivated sesame species – a domestication footprint. *Plant Biol.* 17, 1039–1046. doi: 10.1111/plb.12327
- Pathak, N., Rai, A. K., Kumari, R., and Bhat, K. V. (2014). Value addition in sesame: a perspective on bioactive components for enhancing utility and profitability. *Pharmacogn. Rev.* 8, 147–155. doi: 10.4103/0973-7847.134249
- Pazhamala, L., Saxena, R. K., Singh, V. K., Sameerkumar, C. V., Kumar, V., Sinha, P., et al. (2015). Genomics-assisted breeding for boosting crop improvement in pigeonpea (*Cajanus cajan*). *Front. Plant Sci.* 6:50. doi: 10.3389/fpls.2015.00050
- Pham, D. T., Bui, M. T., Werlemark, G., Bui, C. T., Merker, A., and Carlsson, A. S. (2009). A study of genetic diversity of sesame (*Sesamum indicum* L.) in vietnam and cambodia estimated by RAPD markers. *Genet. Resour. Crop. Evol.* 56, 679–690. doi: 10.1007/s10722-008-9393-z
- Sankar, D., Sambandam, G., Ramakrishna, R. M., and Pugalendi, K. V. (2005). Modulation of blood pressure, lipid profiles and redox status in hypertensive patients taking different edible oils. *Clin. Chem. Acta* 355, 97–104.
- Saydut, A., Zahir-Duz, M., Kaya, C., Kafadar, A. B., and Hamamci, C. (2008). Transesterified sesame (*Sesamum indicum* L.) seed oil as a biodiesel fuel. *Bioresour. Technol.* 99, 6656–6660. doi: 10.1016/j.biortech.2007.11.063
- Sehr, E. M., Okello-Anyanga, W., Hasel-Hohl, K., Burg, A., Gaubitzer, S., Rubaihayo, P. R., et al. (2016). Assessment of genetic diversity amongst Ugandan sesame (*Sesamum indicum* L.) landraces based on agromorphological traits and genetic markers. *J. Crop Sci. Biotechnol.* 19, 117–129. doi: 10.1007/s12892-015-0105-x
- Spandana, B., Reddy, V. P., Prasanna, G. J., Anuradha, G., and Sivaramakrishnan, S. (2012). Development and characterization of microsatellite markers (SSR) in *Sesamum indicum* L. species. *Appl. Biochem. Biotechnol.* 168, 1594–1607. doi: 10.1007/s12010-012-9881-7
- Suh, M. C., Kim, M. J., Hur, C. G., Bae, J. M., Park, Y. I., Chung, C. H., et al. (2003). Comparative analysis of expressed sequence tags from *Sesamum indicum* and *Arabidopsis thaliana* developing seeds. *Plant Mol. Biol.* 52, 1107–1123. doi: 10.1023/B:PLAN.0000004304.22770.e9
- Surapaneni, M., Yepuri, V., Vemireddy, L. R., Ghanta, A., and Siddiq, E. A. (2014). Development and characterization of microsatellite markers in Indian sesame (*Sesamum indicum* L.). *Mol. Breed.* 34, 1185–1200. doi: 10.1007/s11032-014-0109-0
- Thomson, M. J., Septiningsih, E. M., Swardjo, F., Santoso, T. J., Silitonga, T. S., and McCouch, S. R. (2007). Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using microsatellite markers. *Theor. Appl. Genet.* 114, 559–568. doi: 10.1007/s00122-006-0457-1
- Tiwari, S., Kumar, S., and Gontia, I. (2011). Biotechnological approaches for sesame (*Sesamum indicum* L.) and Niger (*Guizotia abyssinica* L.f. Cass.). *As. Pac. J. Mol. Biol. Biotechnol.* 19, 2–9.



- Ucan, K., Killi, F., Gencoglan, C., and Merdun, H. (2007). Effect of irrigation frequency and amount on water use efficiency and yield of sesame (*Sesamum indicum* L.) under field conditions. *Field Crops Res.* 101, 249–258.
- Uncu, A. O., Frary, A., Karlovsky, P., and Doganlar, S. (2016). High-throughput single nucleotide polymorphism (SNP) identification and mapping in the sesame (*Sesamum indicum* L.) genome with genotyping by sequencing (GBS) analysis. *Mol. Breed.* 36:173. doi: 10.1007/s11032-016-0604-6
- Uncu, A. O., Gultekin, V., Allmer, J., Frary, A., and Doganlar, S. (2015). Genomic simple sequence repeat markers reveal patterns of genetic relatedness and diversity in sesame. *Plant Genome* 8, 1–12. doi: 10.3835/plantgenome2014.11.0087
- Uzun, B., Lee, D., Donini, P., and Cagirgan, M. I. (2003). Identification of a molecular marker linked to the closed capsule mutant trait in sesame using AFLP. *Plant Breed.* 122, 95–97. doi: 10.1046/j.1439-0523.2003.00787.x
- Varshney, R. K., Glaszmann, J.-C., Leung, H., and Ribaut, J. M. (2010). More genomic resources for less-studied crops. *Trends Biotechnol.* 28, 452–460. doi: 10.1016/j.tibtech.2010.06.007
- Varshney, R. K., Nayak, S. N., May, G. D., and Jackson, S. A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 27, 522–530. doi: 10.1016/j.tibtech.2009.05.006
- Verma, P., Goyal, R., Chahota, R. K., Sharma, T. R., Abidin, M. Z., and Bhatia, S. (2015). Construction of a genetic linkage map and identification of QTLs for seed weight and seed size traits in Lentil (*Lens culinaris* Medik.). *PLoS ONE* 10:e0139666. doi: 10.1371/journal.pone.0139666
- Wang, L., Han, X., Zhang, Y., Li, D., Wei, X., Ding, X., et al. (2014c). Deep resequencing reveals allelic variation in *Sesamum indicum*. *BMC Plant Biol.* 14:225. doi: 10.1186/s12870-014-0225-3
- Wang, L., Li, D., Zhang, Y., Gao, Y., Yu, J., Wei, X., et al. (2016b). Tolerant and susceptible sesame genotypes reveal waterlogging stress response patterns. *PLoS ONE* 11:e0149912. doi: 10.1371/journal.pone.0149912
- Wang, L., Xia, Q., Zhang, Y., Zhu, X., Zhu, X., Li, D., et al. (2016a). Updated sesame genome assembly and fine mapping of plant height and seed coat color QTLs using a new high-density genetic map. *BMC Genome* 17:31. doi: 10.1186/s12864-015-2316-4
- Wang, L., Yu, J., Li, D., and Zhang, X. (2014b). Sinbase: an integrated database to study genomics, genetics and comparative genomics in *Sesamum indicum*. *Plant Cell Physiol.* 56:e2. doi: 10.1093/pcp/pcu175
- Wang, L., Yu, S., Tong, C., Zhao, Y., Liu, Y., Song, C., et al. (2014a). Genome sequencing of the high oil crop sesame. *Genome Biol.* 15:R39. doi: 10.1186/gb-2014-15-2-r39
- Wang, L., Zhang, Y., Li, P., Wang, X., Zhang, W., Wei, W., et al. (2012a). HPLC analysis of seed sesamin and sesamol variation in a sesame germplasm collection in China. *J. Am. Oil Chem. Soc.* 8, 1011–1020. doi: 10.1007/s11746-011-2005-7
- Wang, L., Zhang, Y., Qi, X., Gao, Y., and Zhang, X. (2012b). Development and characterization of 59 polymorphic cDNA-SSR markers for the edible oil crop *Sesamum indicum* (Pedaliaceae). *Am. J. Bot.* 99, 394–398. doi: 10.3732/ajb.1200081
- Wang, L., Zhang, Y., Qi, X., Li, D., Wei, W., and Zhang, X. (2012c). Global gene expression responses to waterlogging in roots of sesame (*Sesamum indicum* L.). *Acta Physiol. Plant.* 34, 2241–2249. doi: 10.1007/s11738-012-1024-9
- Wei, L., Miao, H., Li, C., Duan, Y., Niu, J., Zhang, T., et al. (2014). Development of SNP and InDel markers via de novo transcriptome assembly in *Sesamum indicum* L. *Mol. Breed.* 34, 2205–2217. doi: 10.1007/s11032-014-0174-4
- Wei, L., Miao, H., and Zhang, H. (2012). Transcriptomic analysis of sesame development. *Sci. Agric. Sin.* 45, 1246–1256. doi: 10.3864/j.issn.0578-1752.2012.07.002
- Wei, L., Zhang, H., Zheng, Y., Guo, W., and Zhang, T. (2008). Development and utilization of EST-derived microsatellites in sesame (*Sesamum indicum* L.). *Acta Agron. Sin.* 34, 2077–2084.
- Wei, L., Zhang, H., Zheng, Y., Miao, H., Zhang, T., and Guo, W. (2009). A genetic linkage map construction for sesame (*Sesamum indicum* L.). *Genes Genomics* 31, 199–208. doi: 10.1007/BF03191152
- Wei, W., Qi, X., Wang, L., Zhang, Y., Hua, W., Li, D., et al. (2011). Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired end sequencing and development of EST SSR markers. *BMC Genomics* 12:451. doi: 10.1186/1471-2164-12-451
- Wei, W., Zhang, Y., Lv, H., Li, D., Wang, L., and Zhang, X. (2013). Association analysis for quality traits in a diverse panel of Chinese sesame (*Sesamum indicum* L.) germplasm. *J. Integr. Plant Biol.* 55, 745–758. doi: 10.1111/jipb.1204
- Wei, X., Gong, H., Yu, J., Liu, P., Wang, L., Zhang, Y., et al. (2017). Sesame FG: an integrated database for the functional genomics of sesame. *Sci. Rep.* 7:2342. doi: 10.1038/s41598-017-02586-3
- Wei, X., Liu, K., Zhang, Y., Feng, Q., Wang, L., Zhao, Y., et al. (2015). Genetic discovery for oil production and quality in sesame. *Nat. Commun.* 6, 8609. doi: 10.1038/ncomms9609
- Wei, X., Wang, L., Zhang, Y., Qi, X., Wang, X., Ding, X., et al. (2014). Development of simple sequence repeat (SSR) markers of sesame (*Sesamum indicum*) from a genome survey. *Molecules* 19, 5150–5162. doi: 10.3390/molecules19045150
- Wei, X., Zhu, X., Yu, J., Wang, L., Zhang, Y., Li, D., et al. (2016). Identification of sesame genomic variations from genome comparison of landrace and variety. *Front. Plant Sci.* 7:1169. doi: 10.3389/fpls.2016.01169
- Weiss, E. A. (2000). *Sesame. Oilseed Crops*, 2nd Edn. London: Blackwell Science.
- Witcombe, J. R., Hollington, P. A., Howarth, C. J., Reader, S., and Steele, K. A. (2007). Breeding for abiotic stresses for sustainable agriculture. *Philos. Trans. R. Soc. B* 363, 703–716. doi: 10.1098/rstb.2007.2179
- Wongyai, W., Saengkaewsook, W., and Veerawudh, J. (2001). "Sesame mutation induction: improvement of non-shattering capsule by using gamma rays and EMS," in *Sesame Improvement by Induced Mutations*, IAEA-TECDOC-1195, ed. L. van Zanten (Vienna: IAEA), 71–78.
- Wu, K., Liu, H., Yang, M., Tao, Y., Ma, H., Wu, W., et al. (2014a). High-density genetic map construction and QTLs analysis of grain yield-related traits in sesame (*Sesamum indicum* L.) based on RAD-Seq technology. *BMC Plant Biol.* 14:274. doi: 10.1186/s12870-014-0274-7
- Wu, K., Yang, M., Liu, H., Tao, Y., Mei, J., and Zhao, Y. (2014b). Genetic analysis and molecular characterization of Chinese sesame (*Sesamum indicum* L.) cultivars using Insertion-Deletion (InDel) and Simple Sequence Repeat (SSR) markers. *BMC Genet.* 15:35. doi: 10.1186/1471-2156-15-35
- Yadav, M., Sainger, D. C. M., and Jaiwal, P. K. (2010). *Agrobacterium tumefaciens*-mediated genetic transformation of sesame (*Sesamum indicum* L.). *Plant Cell Tissue Organ Cult* 103, 377–386. doi: 10.1007/s11240-010-9791-8
- Yi, D. K., and Kim, K. J. (2012). Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PLoS ONE* 7:e35872. doi: 10.1371/journal.pone.0035872
- Yol, E., and Uzun, B. (2012). Geographical patterns of sesame (*Sesamum indicum* L.) accessions grown under Mediterranean environmental conditions, and establishment of a core collection. *Crop Sci.* 52, 2206–2214. doi: 10.2135/cropsci2011.07.0355
- Yu, J., Dossa, K., Wang, L., Zhang, Y., Wei, X., Liao, B., et al. (2016). PMDBase: a database for studying microsatellite DNA and marker development in plants. *Nucleic Acids Res.* 45, D1046–D1053. doi: 10.1093/nar/gkw906
- Yu, J., Ke, T., Tehrim, S., Sun, F., Liao, B., and Hua, W. (2015). PTGBase: an integrated database to study tandem duplicated genes in plants. *Database (Oxford)* 2015:bav017. doi: 10.1093/database/bav017
- Yu, J., Wang, L., Guo, H., Liao, B., King, G., and Zhang, X. (2017). Genome evolutionary dynamics followed by diversifying selection explains the complexity of the *Sesamum indicum* genome. *BMC Genomics* 18:257. doi: 10.1186/s12864-017-3599-4
- Zamir, D. (2001). Improving plant breeding with exotic genetic libraries. *Nat. Rev. Genet.* 2, 983–989. doi: 10.1038/35103590
- Zeven, A., and Zhukovsky, P. (1975). *Dictionary of Cultivated Plants and Their Centres of Diversity*. Wageningen: PUDOC.
- Zhang, H., Miao, H., Li, C., Wei, L., Duan, Y., Ma, Q., et al. (2016). Ultra-dense SNP genetic map construction and identification of SiDt gene controlling the determinate growth habit in *Sesamum indicum* L. *Sci. Rep.* 6:31556. doi: 10.1038/srep31556
- Zhang, H., Li, C., Miao, H., and Xiong, S. (2013b). Insights from the complete chloroplast genome into the evolution of *Sesamum indicum* L. *PLoS ONE* 8:e80508. doi: 10.1371/journal.pone.0080508
- Zhang, H., Miao, H., Wang, L., Qu, L., Liu, H., Wang, Q., et al. (2013a). Genome sequencing of the important oilseed crop *Sesamum indicum* L. *Genome Biol.* 14:401. doi: 10.1186/gb-2013-14-1-401



- Zhang, H., Miao, H., Wei, L., Li, C., Zhao, R., and Wang, C. (2013c). Genetic analysis and QTL mapping of seed coat color in sesame (*Sesamum indicum* L.). *PLoS ONE* 8:e63898. doi: 10.1371/journal.pone.0063898
- Zhang, H., Wei, L., Miao, H., Zhang, T., and Wang, C. (2012). Development and validation of genic-SSR markers in sesame by RNA-seq. *BMC Genomics* 13:316. doi: 10.1186/1471-2164-13-316
- Zhang, X., Zhao, Y., Cheng, Y., Feng, X., Guo, Q., Zhou, M., et al. (2000). Establishment of sesame germplasm core collection in China. *Genet. Resour. Crop Evol.* 47, 273–279. doi: 10.1023/A:1008767307675
- Zhang, Y., Wang, L., Li, D., Gao, Y., Lu, H., and Zhang, X. (2014). Mapping of sesame waterlogging tolerance QTL and identification of excellent waterlogging tolerant germplasm. *Sci. Agric. Sin.* 47, 422–430.
- Zhang, Y., Wang, L., Xin, H., Li, D., Ma, C., Ding, X., et al. (2013). Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. *BMC Plant Biol.* 13:141. doi: 10.1186/1471-2229-13-141
- Zhang, Y., Zhang, X., Che, Z., Wang, L., Wei, W., and Li, D. (2012). Genetic diversity assessment of sesame core collection in China by phenotype and molecular markers and extraction of a mini-core collection. *BMC Genet.* 13:102. doi: 10.1186/1471-2156-13-102
- Zhao, Y., Yang, M., Wu, K., Liu, H., Wu, J., and Liu, K. (2013). Characterization and genetic mapping of a novel recessive genic male sterile gene in sesame (*Sesamum indicum* L.). *Mol. Breed.* 32, 901–908. doi: 10.1007/s11032-013-9919-8
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Dossa, Diouf, Wang, Wei, Zhang, Niang, Fonceka, Yu, Mmadi, Yehouessi, Liao, Zhang and Cisse. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Recent Genetic Gains in Nitrogen Use Efficiency in Oilseed Rape

Andreas Stahl<sup>1\*</sup>, Mara Pfeifer<sup>1</sup>, Matthias Frisch<sup>2</sup>, Benjamin Wittkop<sup>1</sup> and Rod J. Snowdon<sup>1</sup>

<sup>1</sup> Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University Giessen, Giessen, Germany, <sup>2</sup> Department of Biometry and Population Genetics, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University Giessen, Giessen, Germany

## OPEN ACCESS

### Edited by:

Dragana Miladinović,  
Institute of Field and Vegetable Crops,  
Serbia

### Reviewed by:

Liezhao Liu,  
Southwest University, China  
Shengwu Hu,  
Northwest A&F University, China

### \*Correspondence:

Andreas Stahl  
andreas.stahl@agr.uni-giessen.de

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 07 April 2017

**Accepted:** 22 May 2017

**Published:** 07 June 2017

### Citation:

Stahl A, Pfeifer M, Frisch M, Wittkop B  
and Snowdon RJ (2017) Recent  
Genetic Gains in Nitrogen Use  
Efficiency in Oilseed Rape.  
Front. Plant Sci. 8:963.  
doi: 10.3389/fpls.2017.00963

Nitrogen is essential for plant growth, and N fertilization allows farmers to obtain high yields and produce sufficient agricultural commodities. On the other hand, nitrogen losses potentially cause adverse effects to ecosystems and to human health. Increasing nitrogen use efficiency (NUE) is vital to solve the conflict between productivity, to secure the demand of a growing world population, and the protection of the environment. To ensure this, genetic improvement is considered to be a paramount aspect toward ecofriendly crop production. Winter oilseed rape (*Brassica napus* L.) is the second most important oilseed crop in the world and is cultivated in many regions across the temperate zones. To our knowledge, this study reports the most comprehensive field-based data generated to date for an empirical evaluation of genetic improvement in winter oilseed rape varieties under two divergent nitrogen fertilization levels (NFLs). A collection of 30 elite varieties registered between 1989 and 2014, including hybrids and open pollinated varieties, was tested in a 2-year experiment in 10 environments across Germany for changes in seed yield and seed quality traits. Furthermore, NUE was calculated. We observed a highly significant genetics-driven increase in seed yield *per-se* and, thus, increased NUE at both NFLs. On average, seed yield from modern open-pollinated varieties and modern hybrids was higher than from old open-pollinated varieties and old hybrids. The annual yield progress across all tested varieties was  $\sim 35 \text{ kg ha}^{-1} \text{ year}^{-1}$  at low nitrogen and  $45 \text{ kg ha}^{-1} \text{ year}^{-1}$  under high nitrogen fertilization. Furthermore, in modern varieties an increased oil concentration and decreased protein concentration was observed. Despite the significant effects of nitrogen fertilization, a surprisingly low average seed yield gap of  $180 \text{ kg N ha}^{-1}$  was noted between high and low nitrogen fertilization. Due to contrary effects of N fertilization on seed yield *per-se* and seed oil concentration an oil yield of  $2.04 \text{ t ha}^{-1}$  was measured at both N levels. Collectively, the data reveal that genetic improvement through modern breeding techniques in conjunction with reduced N fertilizer inputs has a tremendous potential to increase NUE of oilseed rape.

**Keywords:** yield, breeding progress, *Brassica napus*, hybrid varieties, nitrogen, fertilization, oil, sustainable intensification

**Abbreviations:** GHG, Greenhouse gas; HN, High nitrogen; LN, Low nitrogen; N, Nitrogen;  $N_{\min}$ , Soil mineral nitrogen; NFL, Nitrogen fertilization level; NUE, Nitrogen use efficiency (as Seed yield over N supplied); OilConc, Seed oil concentration (at 91% dry matter content); OilY, Oil yield; OP, Open pollinated; ProteinConc, Seed protein concentration (at 91% dry matter content); ProteinConc DFF, ProteinConc in defatted fraction (at 91% dry matter content); ProteinY, Protein yield; SumProOilConc, Sum of protein and oil concentration (at 91% dry matter content); SumProOilY, Sum of protein and oil yield; SDH, Semi dwarf hybrid; SY, Seed yield (at 91% dry matter content).

## INTRODUCTION

An increase in global crop demand of up to 110% is expected by 2050 compared to 2005, requiring a tremendous increase in production (Tilman et al., 2011). Competitive crop production crucially depends on the adequate application of nitrogen (N). It is well-known that the application of N fertilizer is a substantial driver of yield increases in the last century. At the same time, it has also been found that 50–70% of the applied N is not recovered in the harvested plant organs (seeds) and can cause severe damages to the surrounding ecosystems (Sylvester-Bradley and Kindred, 2009; Liu et al., 2010; Galloway et al., 2013). Estimations have revealed that the current global conversion of atmospheric N<sub>2</sub> into reactive N and its application on fields have already transgressed the boundaries of sustainable development (Rockström et al., 2009; Steffen et al., 2015). Moreover, the lost N is not only an ecological problem but also uneconomical for farmers (Rothstein, 2007). Therefore, in the decades ahead, agricultural crop production faces the unprecedented challenge of enhancing crop yields to match the increasing demands, while simultaneously reducing environmental damages caused by unused nitrogen. To master this dichotomy, a dramatic increase in nitrogen use efficiency (NUE) is inevitable. Besides a more precise fertilizer application (Henke et al., 2008; Müller et al., 2012), the use of genetic potential by breeding and cultivation of most suitable and efficient varieties is considered to play an important role in the sustainable intensification of agriculture (Hirel et al., 2007; Kant et al., 2011; Hawkesford, 2014).

Oilseed rape (*Brassica napus* L.) is the principle European oil crop and the second most important oilseed crop in the world after soybean, with a wide dissemination across countries in the moderate climatic zone (Fischer et al., 2014). It is mainly cultivated because of its high quality vegetable oil used for human nutrition purposes (Kumar et al., 2016) and as a renewable source for fuels and technical oils. Not at least, the residual meal is winning wide use as a protein feed for animals and is also discussed for human nutrition recently (Fleddermann et al., 2013). Despite several positive agronomic effects, including its integral role as a break crop in cereal crop rotations, its ability to improve soil fertility, and its strong ability to capture nutrients during vegetative growth stages, oilseed rape cultivation is often associated with a relatively high N balance surplus (Sieling and Kage, 2006; Weiser et al., 2017). Consequently, a more competitive and eco-friendly production of oil and protein would benefit from increased NUE.

In the face of this challenge the question arises whether or not breeding is directing NUE improvement. A critical assessment of the genetic progress for SY under divergent N inputs is of great value in order to analyze the extent of the breeding effect on NUE improvement in the past and in what way a course correction in breeding programs is required to reach future sustainability goals. Breeding is a long-term procedure and environmental conditions as well as management practices have also changed over time. Therefore, the isolation of the genetic contribution by comparing the direct performance between varieties from different period of registration requires multilocation field trials with simultaneous

cultivation of the varieties under a very similar N fertilization management. Owing to the tedious trial setup, experimental data of those comparisons are seldom presented.

In the case of oilseed rape Kessel et al. (2012) evaluated 36 genotypes in the vegetation period 1997–1998. This study included 3 hybrids, 6 resynthesized lines, 8 old cultivars, and 19 modern cultivars. It was found that modern cultivars and hybrids outperform older varieties and resynthesized lines. However, the varieties designated as modern in that study no longer have any relevance for present farming practice. Moreover, since hybrids varieties have gained enormous importance in the last two decades, today comprising the vast majority of available varieties, this type of variety must be considered for yield and NUE monitoring with an up-to-date European benchmark. In this regard, Rathke and Diepenbrock (2006) investigated the energy balance of winter oilseed rape depending on the N fertilizer inputs and suggested that SY levels should be assessed through further studies to document progress in plant breeding including the advent of hybrid varieties. Although further field studies on NUE in oilseed rape have since been conducted, most focussed on genetic mapping approaches, either in experimental populations (Bouchet et al., 2014; Nyikako et al., 2014; Miersch et al., 2016) or in case of association studies (Bouchet et al., 2016) do not address the breeding progress.

The present study aims to (i) assess the SY change through breeding during the last decades, (ii) examine the response to nitrogen of older and modern varieties in multilocation field trials, (iii) investigate the change and interrelationship of seed quality traits affected by environment, N fertilization, and breeding activities, and (iv) provide evidence-based suggestions for further orientations in oilseed rape breeding and cultivation.

## MATERIALS AND METHODS

### Plant Material

Diverse varieties ( $n = 30$ ) from different breeding companies and a period of registration between 1989 and 2014 were investigated (Table 1). All of them were selected varieties that were adapted to the northern European growing conditions and have passed the official varieties test by official varieties agency. In order to keep the number of genotype on a manageable size, major varieties from the different periods of registration were selected as proposed in a similar study on wheat by Lopes et al. (2012). According to the year of registration, varieties were grouped as older and modern varieties.

### Field Experiments

Since environmental factors heavily affect the phenotype, 10 experiments were conducted in total at six different locations in Germany during two subsequent growing seasons (Table 2). Environmental specifications are given in Supplementary Tables 1, 2. Experiments were conducted in a split plot design for N treatments. Within each nitrogen fertilization level (NFL, as main plot), the genotypes were arranged in three replicates with eight sub blocks each, according to an alpha lattice designed with R package *agricolae* (De Mendiburu and Simon, 2015). The sowing rate was adjusted for the previous tested germination rate to

**TABLE 1** | Groups of investigated varieties according to year of registration and type of variety.

Varieties	Year of registration
<b>NEW HYBRID VARIETIES</b>	
Thure (SDH)	2014
Marathon	2013
Mercedes	2013
Avatar	2011
DK Exstorm	2011
Inspiration	2011
Genie	2011
Mascara	2011
Troy (SDH)	2011
Artoga	2010
Sherpa	2010
Compass	2009
NK Linus	2009
Visby	2007
<b>OLD HYBRID VARIETIES</b>	
Exocet	2005
Taurus	2004
Baldur	2002
Elektra	2002
Ryder	2000
Artus	1997
<b>NEW OP VARIETIES</b>	
Patron	2012
Trinity	2012
Adriana	2007
Lorenz	2005
Oase	2004
<b>OLD OP VARIETIES</b>	
Pacific	2003
Californium	2002
Aviso	2000
Express	1993
Lirajet	1989

SDH, Semi dwarf hybrid.

target a final plant density of 50 plants per square meter. At all locations, a plot in plot system was used, and only the middle part of each plot was harvested in order to avoid side effects from the neighboring plots. Organic fertilizers were applied neither during the experiment nor during cultivation of the preceding crops. Moreover, no intercrop was cultivated before the experiment. In the first year of experiment, 120 kg N ha<sup>-1</sup> was applied in both low N (LN) and high N treatment (HN) at the beginning of the spring vegetation. Another 100 kg N ha<sup>-1</sup> was applied only in HN during bolting. In the second year, 65 kg N ha<sup>-1</sup> was applied in LN and 120 kg N ha<sup>-1</sup> in HN at the first application. The second application during bolting comprised of 55 kg N ha<sup>-1</sup> in LN and 100 kg N ha<sup>-1</sup> in HN. In both years, the first application of fertilizer was reduced for soil mineral N (N<sub>min</sub>) present in the soil at the end of winter (Supplementary Table 1). The intention of the high N application was to simulate intensive crop production

conditions as they are common in Northern Europe. The low N treatment was to investigate the response of the genotypes during reduced fertilizer application. All other nutrients were applied at the same level across both NFL at the particular locations. Furthermore, full control of weeds, pest (insects), and diseases (fungi) and straw stiffness were conducted at each side on a local appropriate level. Thus, only N was varied and all other factors were kept on an intensive level in both NFL levels.

## Data Collection

Seed yield was determined by threshing from standing mature canopy in the second half of July or first days of August (Table 2). The water content of seeds was determined immediately after seed harvest, and SY was corrected to a standard water content of 9%. From each individual plot, an aliquot of at least 100 g were used to determine seed oil concentration (OilConc) and seed protein concentration (ProteinConc) in duplicates on the same machine (Unity SpectraStar 2500, Brookfield, USA) via near-infrared reflectance spectroscopy (Tkachuk, 1981; Reinhardt, 1992; Tillmann and Paul, 1998; Tillmann et al., 2000). Subsequently, OilConc and ProteinConc were also corrected to 9% water content and multiplied with SY in order to determine seed oil yield (OilY) and seed protein yield (ProteinY). Since one old hybrid variety showed a strongly reduced germination rate at all locations in the second experimental year, it was excluded from further data analysis.

In environments ASE15, ASE16, NIE16, and MOS16, lodging was evaluated before harvest on each individual plot on a scale from 1 (no lodging at all) to 9 (plot is completely lodging).

## Data Analysis

Adjusted means for each variety at each NFL across all locations were estimated using a linear model with varieties, NFL, and their interaction as fixed factors; and interactions of year, location, NFL, replicate, and block considered as random factors (Equation 1). *P*-values for significance of fixed effects were derived from an analysis of variance (ANOVA) using the models described in Equation (1).

$$P_{ijklmn} = \mu + \mathbf{g}_i + \mathbf{n}_j + \mathbf{gn}_{ij} + B_{jklmn} + R_{jkmn} + W_{jkm} + E_{km} + e_{ijklmn} \quad (1)$$

with  $P_{ijklmn}$  as the observed phenotype of the *i*th variety, the *j*th nitrogen fertilization, the *k*th year, the *m*th location, the *n*th replicate, and the *l*th block.  $\mu$  is the general mean of the experiment,  $\mathbf{g}_i$  is the *i*th fixed effect of variety,  $\mathbf{n}_j$  is the *j*th fixed effect of NFL, and  $\mathbf{gn}_{ij}$  is the fixed effect of variety by NFL interaction.  $B_{jklmn}$  is the random effect of the *l*th block within the *n*th replicate, within the *j*th main plot, at the *m*th location, and at the *k*th year.  $R_{jkmn}$  is the random effect of the *n*th replicate, within the *j*th main plot, at the *m*th location, and at the *k*th year.  $W_{jkm}$  is the random effect of the *j*th main plot, at the *m*th location, and at the *k*th year.  $E_{km}$  is the random effect of the environment at the *m*th location in the *k*th year.  $e_{ijklmn}$  is the error term. Fixed effects are written in bold lowercase letters.

To estimate the adjusted means of each variety at individual year (Equation 2) and location (Equation 3), the linear model



**TABLE 2** | Overview of single experiments and side specific conditions.

Experiment	Location	GPS Position	Date of sowing	Date of harvest	Preceding crop/Pre-preceding crop
ASE15	Asendorf	52.763480, 8.996599	26.08.2014	01.08.2015	Winterbarley/Winterwheat
ASE16	Asendorf	52.719760, 8.962428	26.08.2015	20/21.07.2016	Winterbarley/Winterwheat
BOV16	Bovenau	54.360205, 9.806736	26.08.2015	24.07.2016	Winter Barley/Winter Barley
MOS15	Moosburg	48.500523, 11.936888	25.08.2014	16.07.2015	Winterwheat/Potatos
MOS16	Moosburg	48.500473, 11.941212	28.08.2015	20.07.2016	Winterbarley/Winterwheat
NIE16	Nienstädt	52.260731, 9.091446	04.09.2015	14.07.2016	Winterbarley/Winterwheat
RHH15	Rauischholzhausen	50.779556, 8.889699	03.09.2014	21/22.07.2015	Winterwheat/Maize
RHH16	Rauischholzhausen	50.778334, 8.868113	26.08.2015	21.07.2016	Winterwheat/Maize
ROS15	Rosenthal	52.306046, 10.164338	25.08.2014	21.07.2015	Winterwheat/Winterwheat
ROS16	Rosenthal	52.298665, 10.117582	25.08.2015	19.07.2016	Winterwheat/Winterwheat

was modified and only the interactions of the location, main plot, replicate, and block were considered to be random factors. Fixed effects are written in bold lowercase letters.

$$P_{ijlmn} = \mu + \mathbf{g}_i + \mathbf{n}_j + \mathbf{gn}_{ij} + B_{jlmn} + R_{jmn} + W_{jm} + E_m + e_{ijlmn} \quad (2)$$

$$P_{ijln} = \mu + \mathbf{g}_i + \mathbf{n}_j + \mathbf{gn}_{ij} + B_{jln} + R_{jn} + W_j + e_{ijln} \quad (3)$$

The model shown in Equation (4) was used to estimate the variance for broad sense heritability estimation. In contrast to Equation (1),  $\mathbf{g}_i$  is the  $i$ th random effect of variety,  $\mathbf{n}_j$  is the  $j$ th fixed effect of NFL, and  $\mathbf{gn}_{ij}$  is the random effect of variety by NFL interaction.  $GE_{ikm}$  is the random effect of the  $G \times E$  interaction. Fixed effect is written in bold. Equation (5) was used to estimate the broad sense heritability.

$$P_{ijklmn} = \mu + G_i + \mathbf{n}_j + GN_{ij} + GE_{ikm} + B_{jklmn} + R_{jkmn} + W_{jkm} + E_{km} + e_{ijklmn} \quad (4)$$

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{G \times N}^2}{j} + \frac{\sigma_{G \times E}^2}{p} + \frac{\sigma_e^2}{j \times p \times l \times n}} \quad (5)$$

with  $j$  the number of NFL,  $p$  the number of the investigated environments,  $l$  the number of blocks, and  $n$  the number of replicates.

All analyses were conducted by statistical software R (R Core Team, 2013) by using the packages lmerTest (Kuznetsova et al., 2016), lsmeans (Lenth, 2016), and lme4 (Bates et al., 2015).

Adjusted data across all environments were used to determine the NUE, which is expressed as SY over N fertilizer application. In addition, the amount of N fertilization required to produce 1 ton of rapeseed oil was determined. Therefore, 1 ton was divided by the particular OilY of individual varieties and multiplied with the NFL. Production losses and inefficiency in post-harvest proceedings were neglected in this calculation.

Pairwise Pearson correlation coefficients ( $r$ ) were estimated between individual traits and year of registration, as well as between the trait value observed at HN and LN.

Packages ggplot2 were used to design diagrams (Wickham, 2009).

## RESULTS

### Seed Yield

Across all 10 environments, SY varied between 3.79 and 5.00 t ha<sup>-1</sup> for LN and between 3.91 and 5.13 t ha<sup>-1</sup> for HN (Table 3). At some location the variation was even bigger (Supplementary Table 4). The average SY in experimental year 2014–2015 was about 220 kg and 480 kg ha<sup>-1</sup> higher in LN and HN respectively, compared to the experimental year 2015–2016. While the genetic effect on yield was highly significant in each of the environments, except RHH15 ( $p = 0.0181$ ), NFL was only significant in some environments (Supplementary Table 4). Over all tested environments, difference in N fertilization resulted in a yield difference of 180 kg ha<sup>-1</sup>, which was only significant on the 10% error level, according to ANOVA.  $G \times N$  interaction was significant only at several individual environments (MOS and ASE in both years, Supplementary Table 4). Generally, the correlation between the environments at HN is high, ranking between  $r = 0.17$  and  $r = 0.81$ , and was significant in most cases (Supplementary Figure 1B). At LN, it was observed that the environment RHH15 did not show significant correlation to most other environment, while all the other environments were positive and significantly correlated to each other. The highest correlation between environments was  $r = 0.77$  (Supplementary Figure 1A).

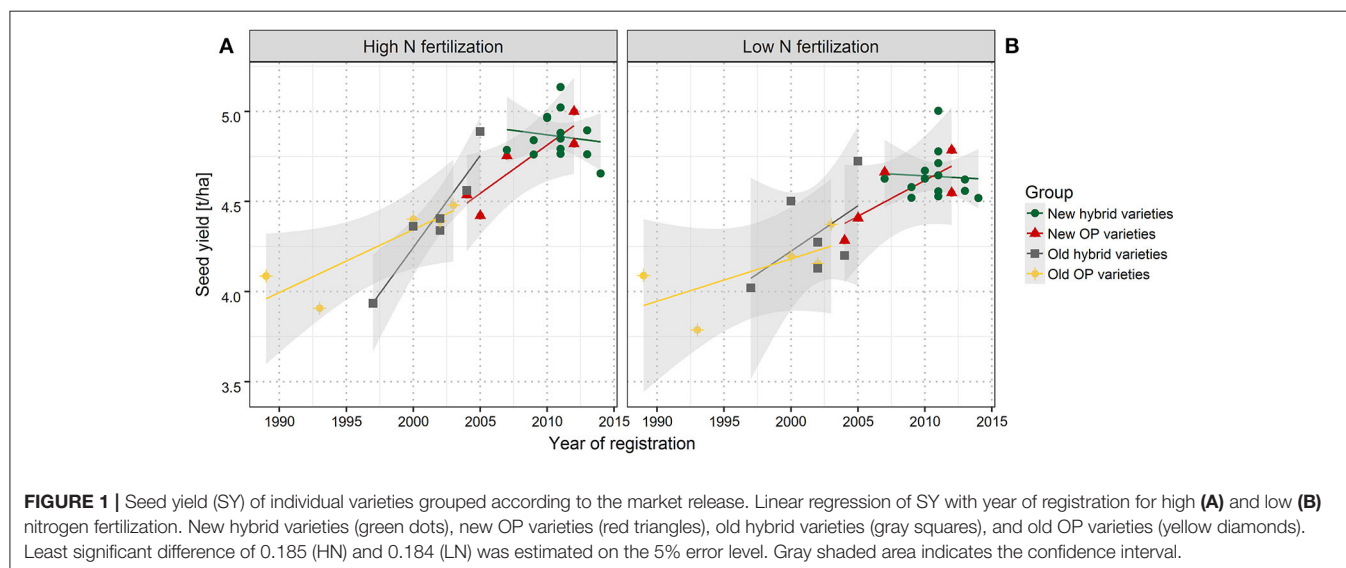
Across the entire set of tested varieties, correlation between SY and the year of registration was  $r = 0.81$  for LN and  $r = 0.88$  for HN, reflecting a strong breeding progress over the period of observation (Figures 1A,B, 2). This is further underlined by a very high value for broad sense heritability in the investigated variety set ( $h^2 = 0.92$ , Table 4). The annual yield progress was ~35 kg ha<sup>-1</sup> year<sup>-1</sup> at LN and 45 kg ha<sup>-1</sup> year<sup>-1</sup> at HN.

If the investigated set is split into groups of hybrid and OP varieties and in older and modern varieties, it becomes obvious that according to the respective arithmetic means, modern hybrid varieties outperform old hybrid varieties as modern OP varieties outperform old OP varieties in all the 10 environments and in both the NFL. There is just one exception at HN in ROS16, where the arithmetic mean of older OP varieties is marginally higher than in modern OP varieties. As indicated in Figure 1, at HN, Pearson coefficient of correlation between SY and year of

**TABLE 3 |** Descriptive statistics for investigated traits

		Low nitrogen fertilization				High nitrogen fertilization			
		Min	Max	Mean	CoV	Min	Max	Mean	CoV
Seed Yield (t/ha)	Mean 2015	3.94	4.94	4.60	0.05	4.00	5.39	4.94	0.07
	Mean 2016	3.70	5.06	4.38	0.07	3.67	5.06	4.46	0.07
	Mean	3.79	5.00	4.47	0.06	3.91	5.13	4.65	0.07
NUE (kg/kg)	Mean 2015	32.86	41.13	38.32	0.05	18.19	24.49	22.46	0.07
	Mean 2016	30.81	42.19	36.53	0.07	16.66	23.01	20.26	0.07
	Mean	31.56	41.70	37.25	0.06	17.76	23.34	21.12	0.07
Protein Conc (%)	Mean 2015	13.90	15.98	14.97	0.04	15.97	18.09	16.89	0.04
	Mean 2016	14.90	17.09	15.73	0.04	16.61	18.49	17.43	0.03
	Mean	14.64	16.52	15.43	0.03	16.35	18.28	17.22	0.03
Oil Conc (%)	Mean 2015	43.96	47.97	45.80	0.02	41.48	45.84	43.96	0.03
	Mean 2016	43.47	47.57	45.55	0.02	41.60	45.67	44.00	0.02
	Mean	43.72	47.56	45.64	0.02	41.53	45.64	43.96	0.02
Protein Yield (t/ha)	Mean 2015	0.62	0.74	0.69	0.05	0.69	0.93	0.84	0.06
	Mean 2016	0.62	0.78	0.69	0.06	0.67	0.89	0.78	0.06
	Mean	0.62	0.76	0.69	0.05	0.70	0.89	0.80	0.06
Oil Yield (t/ha)	Mean 2015	1.80	2.29	2.11	0.06	1.69	2.40	2.17	0.08
	Mean 2016	1.68	2.33	1.99	0.08	1.59	2.24	1.96	0.08
	Mean	1.73	2.30	2.04	0.07	1.63	2.27	2.04	0.08
Protein Conc DFF (%)	Mean 2015	26.28	29.50	27.55	0.02	28.78	32.29	30.09	0.03
	Mean 2016	28.01	30.60	28.86	0.02	30.19	32.76	31.12	0.02
	Mean	27.27	30.09	28.29	0.02	29.67	32.49	30.67	0.02

Adjusted values of highest (Max) and lowest (Min) performing variety is depicted along with the arithmetic mean of the evaluated variety set. CoV, Coefficient of variation.



registration was  $r = 0.89$  for new OP varieties,  $r = 0.87$  for old OP varieties, and  $r = 0.94$  for old hybrids (all significant on the 5% error level). In contrast, the correlations at LN of  $r = 0.76$

for new OP varieties,  $r = 0.67$  for old OP varieties, and  $r = 0.55$  for old hybrids are not significant and lower, indicating a weaker association than at HN. Within the group of modern hybrids, no

**TABLE 4 |** Analysis of variance and broad sense heritability for investigated traits.

Trait	Environment	Variety	NFL	Variety x NFL	$h^2$
Seed Yield (t/ha)	Mean 2015	***	.		0.92
	Mean 2016	***			
	Mean	***	*	.	
ProteinConc (%)	Mean 2015	***	**		0.96
	Mean 2016	***	***	.	
	Mean	***	***	.	
OilConc (%)	Mean 2015	***	**		0.97
	Mean 2016	***	***	*	
	Mean	***	***		
ProteinY (t/ha)	Mean 2015	***	*	*	0.86
	Mean 2016	***	***	.	
	Mean	***	***	**	
OilY (t/ha)	Mean 2015	***			0.94
	Mean 2016	***			
	Mean	***			
ProteinConcDFF (%)	Mean 2015	***	**	.	0.96
	Mean 2016	***	***	.	
	Mean	***	***	.	

NFL, Nitrogen fertilization level. Level of significance is indicated by . for  $p < 0.1$ , \* $p < 0.01$ , \*\* $p < 0.005$  and \*\*\* $p < 0.001$ .

significant correlation between SY and year of registration could be determined either at HN or at LN.

Within the group of old hybrids, it was noticeable that two varieties showed a higher yield of  $0.09 \text{ t ha}^{-1}$  and  $0.14 \text{ t ha}^{-1}$  under LN than under HN (Figure 3). However, it has to be mentioned that this is only a relative performance. On an absolute level, these varieties are only average or second lowest yielding, respectively. Interestingly, the lower-yielding variety appears to have an extreme above-average lodging score (Supplementary Figure 2). Calculations of NUE, expressed as SY per unit of applied nitrogen fertilizer reveal (i) a much higher NUE at LN than at HN and (ii) a gradient of NUE from modern hybrids over modern OP varieties and old hybrids to old OP varieties (Figure 4).

## Oil

The ANOVA indicated a significant effect of NFL on OilConc (Table 4). In all environments, OilConc was expectedly higher in LN than in HN treatment. On average, it was 45.64 and 43.96% at LN and HN, respectively, across the entire study, and was almost constant between the investigated years (Table 3). OilConc was found to be very heritable ( $h^2 = 0.97$ ). A highly significant variety effect was observed in all environments (Supplementary Table 4). With the exception of RHH15 at LN, all individual environments were very high and significantly correlated with each other, ranging between  $r = 0.64$  and  $r = 0.93$  (Supplementary Figure 3). Furthermore, in RHH15, and BOV16, significant  $G \times N$

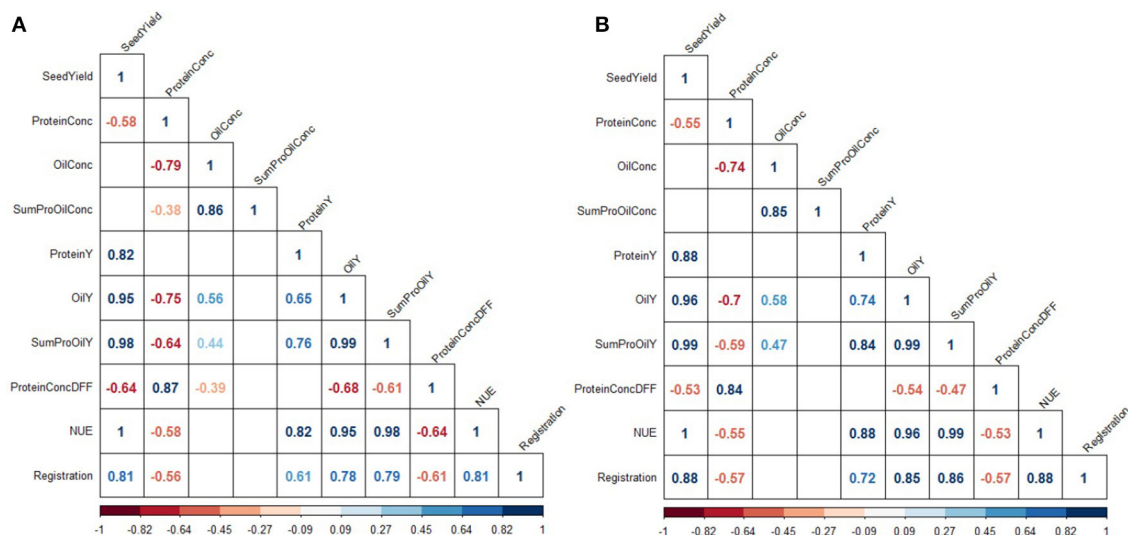
interactions were detected. The correlation between the year of registration and OilConc was only  $r = 0.28$  at LN and  $r = 0.30$  at HN and not significant. Figure 5 indicates that modern varieties, especially the OP varieties, clearly outperform the older groups at both NFL (but not at LN in RHH15).

Furthermore, the results show that LN fertilization was sufficient to produce the same amount of oil than produced through HN. Since, OilY is the product of SY and OilConc, and both traits showed a contrary response to N fertilization, OilY is the same at both NFL ( $2.04 \text{ t ha}^{-1}$  at LN and HN). The highest OilY difference was observed with  $210 \text{ kg ha}^{-1}$  at ROS15 (Supplementary Table 3). At both NFL, results have proven that modern varieties require less N fertilization to achieve the same OilY compared to varieties registered in olden times (Supplementary Figure 4). As correlations in Figure 2 illustrated, the OilY is much more determined by SY (LN  $r = 0.95$ ; HN  $r = 0.96$ ) than by oil concentration (LN  $r = 0.56$ ; HN  $r = 0.58$ ). Moreover, correlations of OilY between environments at LN and HN are in most cases above a correlation of  $r = 0.5$  and significant, except for RHH15 at LN (Supplementary Figures 5A,B).

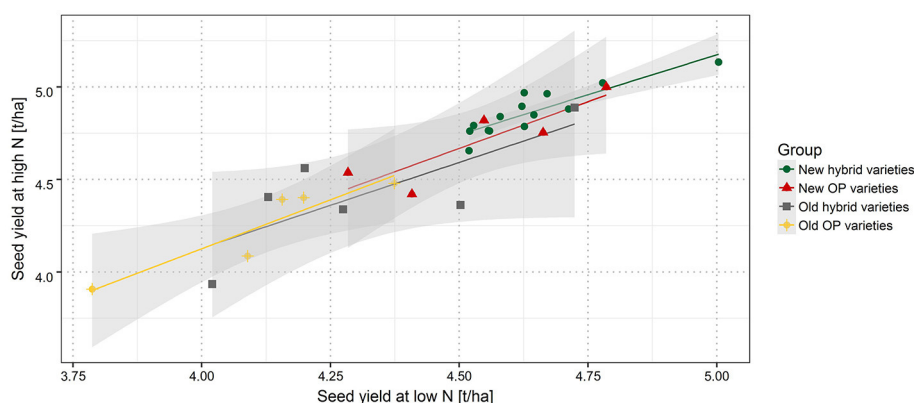
## Protein

ProteinConc was significantly affected by variety and NFL across the entire study and at each of the 2 years (Table 4). At LN, ProteinConc was 15.43% at LN, and 90% of the level achieved at HN (17.22%). Also, in each single experiment, the variety and NFL significantly affected ProteinConc, and only in RHH15 and BOV16,  $G \times N$  interactions were observable (as for OilConc). In summary, Pearson coefficient of correlation between SY and ProteinY of  $r = 0.82$  for LN and  $r = 0.88$  for HN revealed that latter was through the overwhelming part determined by SY. No significant relationship was found between ProteinConc and ProteinY (Figure 2), indicating that SY was the relevant parameter for N extraction and removal from the field. For experimental year 2015–2016, ProteinConc at LN was with 15.73% higher compared to the first year's trials (14.97%). The ProteinY at LN was with  $0.69 \text{ t ha}^{-1}$ , just 86% of the ProteinY at HN ( $0.80 \text{ t ha}^{-1}$ ). ProteinConc and ProteinY were the traits with the strongest alteration due to divergent NFL, stronger than oil-related traits (Table 3). As indicated in Figure 2, across the entire set in both NFL, a negative relationship between the year of registration (LN  $r = -0.56$ ; HN  $r = -0.57$ ) was observed. In addition, ProteinConc had an exceptional strong negative relationship with OilConc at LN ( $r = -0.79$ ) and HN ( $r = -0.74$ ). Except for RHH15 at LN protein concentration show a medium to high correlation between environments (Supplementary Figure 6). In contrast, for ProteinY also low correlations between random environments were observed (Supplementary Figure 7).

If the diversity set is separated into the four variety groups, it become obvious that the oldest varieties, in most cases the old OP varieties, have the highest protein concentration. On the contrary, modern varieties, predominantly modern OP varieties, have the lowest protein concentration (Figure 6). Old OP varieties (LN  $r = -0.61$ ; HN  $r = -0.47$ ) and modern OP varieties (LN  $r = -0.50$ ; HN  $r = +0.64$ ) have weaker correlation



**FIGURE 2 |** Inter-trait phenotypic correlations at low (A) and high (B) nitrogen fertilization. Colors indicate the strength of correlations. Only correlations significant at a confidence level of 95% are depicted.



**FIGURE 3 |** Correlation between SY at high and low nitrogen fertilization. New hybrid varieties (green dots), new OP varieties (red triangles), old hybrid varieties (gray squares), and old OP varieties (yellow diamonds). Gray shaded area indicates the confidence interval.

between OilConc and ProteinConc, while old hybrids (LN  $r = -0.81$ ; HN  $r = -0.88$ ) and modern hybrids (LN  $r = -0.81$ ; HN  $r = -0.76$ ) have the strongest negative correlation.

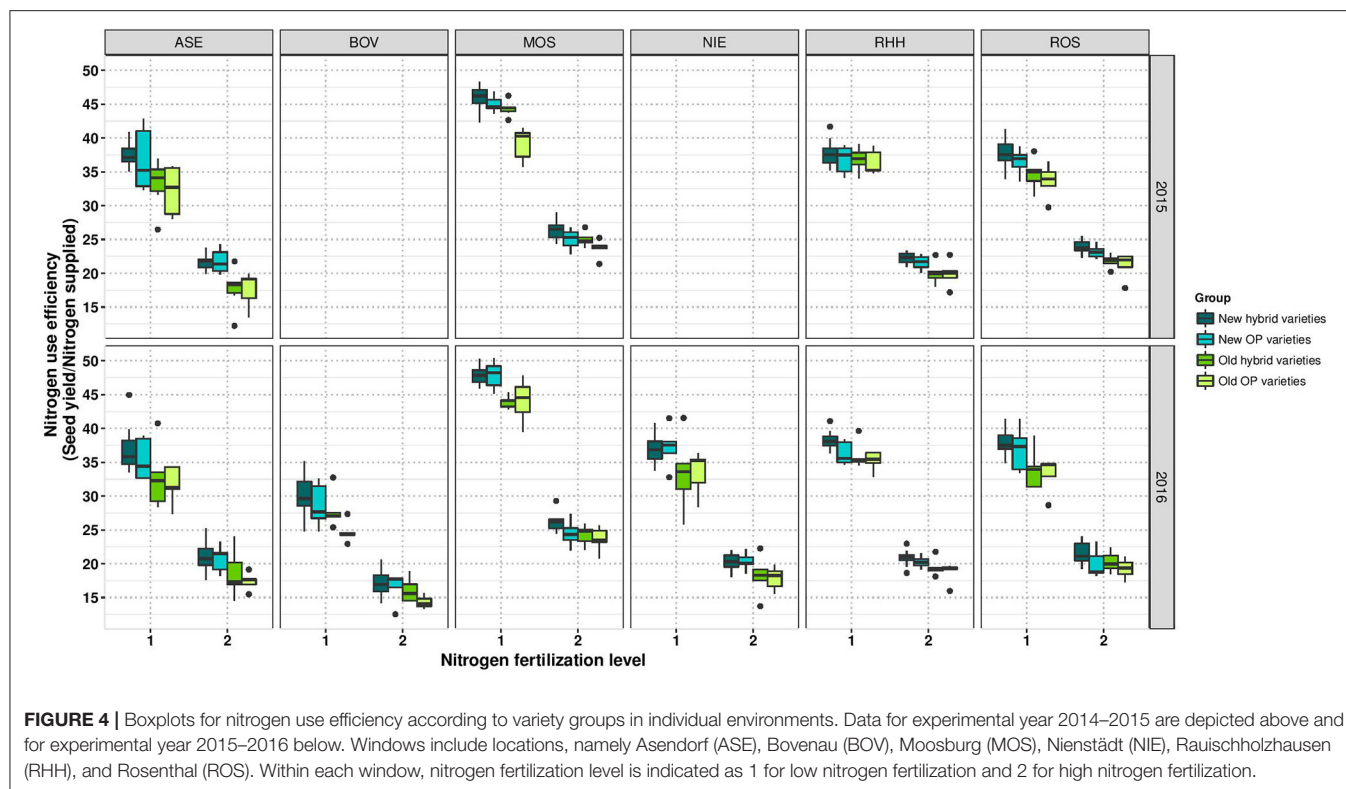
## DISCUSSION

### Measurement of Breeding Progress in a Complex Interacting Growth System

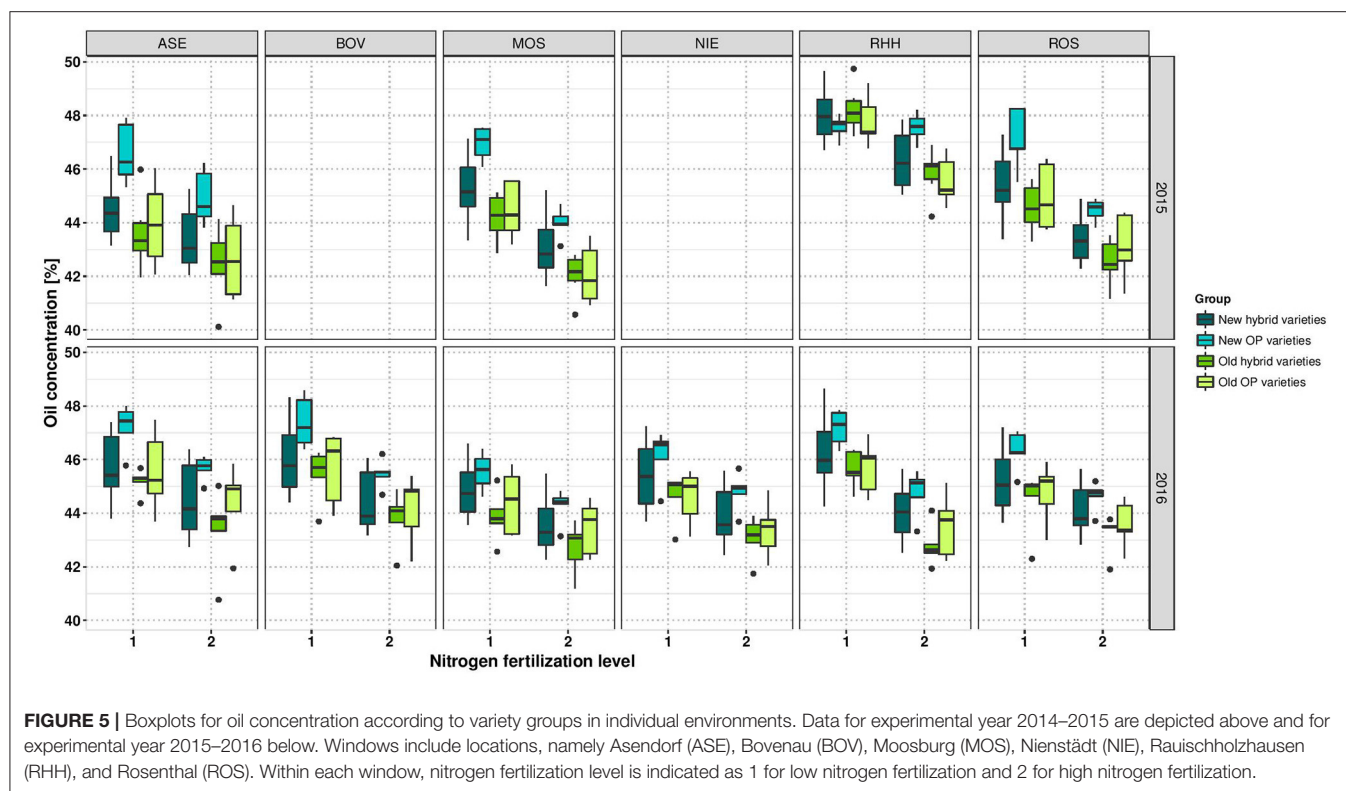
NUE is the final outcome of a complex cropping system with Genotype  $\times$  Environment  $\times$  Management ( $G \times E \times M$ ) interactions (Dresbøll and Thorup-Kristensen, 2014; Thorup-Kristensen and Kirkegaard, 2016). While environmental conditions such as rainfall, temperature, radiation, and others cannot be influenced, the management of crop rotation and decision on soil tillage, as well as dosage and timing of sowing,

fertilization, and responsive use of plant protection agents can be modified by farmers. In addition, since the effect of the preceding crop on the following crop has to be taken into account, the realized NUE has to be measured within the cropping system level as the relevant benchmark (Dresbøll and Thorup-Kristensen, 2014). Within this system, the use of most appropriate genotypes is an important question. In this regard, the aim of our study was to understand the relevance of the genotype of oilseed rape and how breeding influences the progress of its NUE. Precise extraction of genetic contribution for alteration of NUE is extremely challenging due to a lack of comparable environmental conditions. This is not only because SY is determined over a lengthy maturation period, but also because numerous environmental factors and management decisions, including N fertilizer applications, influence various components of NUE throughout the growing season. Therefore,

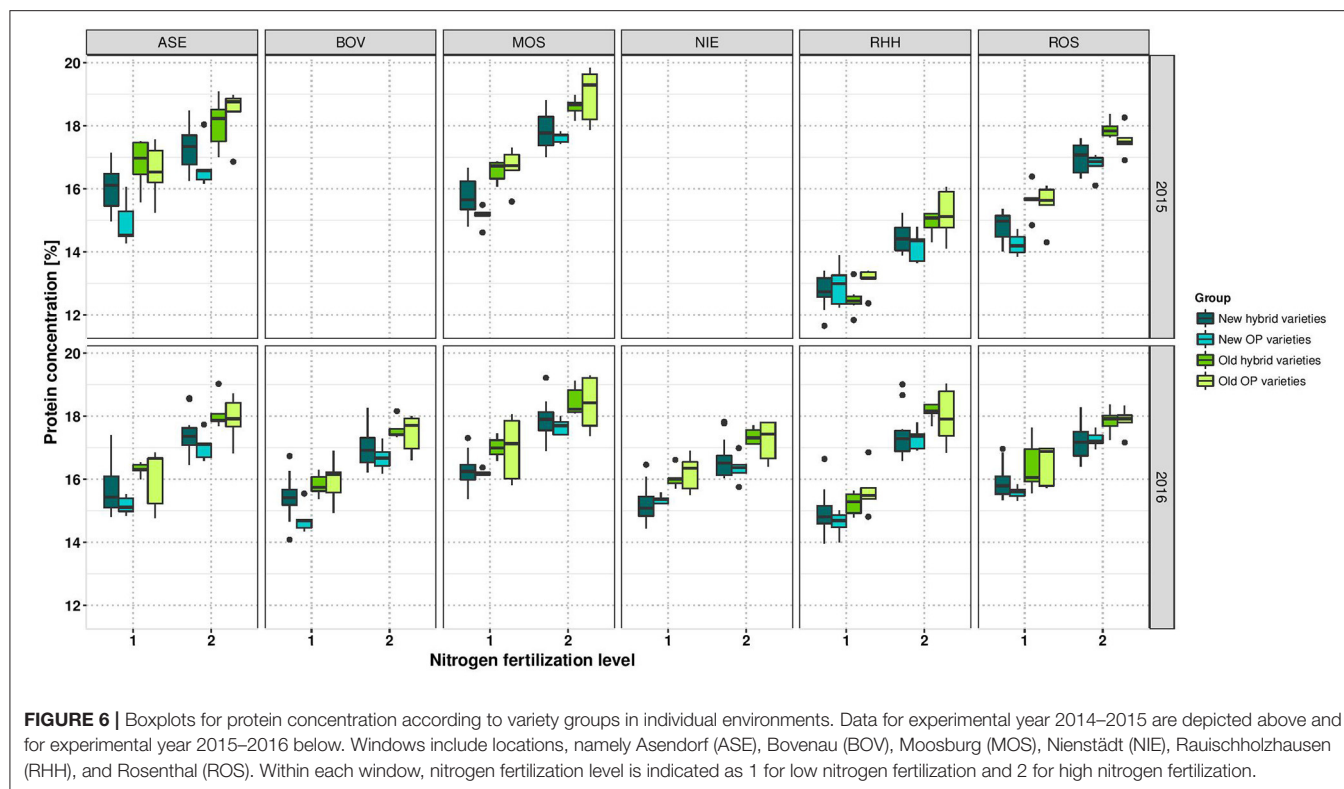




**FIGURE 4 |** Boxplots for nitrogen use efficiency according to variety groups in individual environments. Data for experimental year 2014–2015 are depicted above and for experimental year 2015–2016 below. Windows include locations, namely Asendorf (ASE), Bovenau (BOV), Moosburg (MOS), Nienstädt (NIE), Rauschholzhausen (RHH), and Rosenthal (ROS). Within each window, nitrogen fertilization level is indicated as 1 for low nitrogen fertilization and 2 for high nitrogen fertilization.



**FIGURE 5 |** Boxplots for oil concentration according to variety groups in individual environments. Data for experimental year 2014–2015 are depicted above and for experimental year 2015–2016 below. Windows include locations, namely Asendorf (ASE), Bovenau (BOV), Moosburg (MOS), Nienstädt (NIE), Rauschholzhausen (RHH), and Rosenthal (ROS). Within each window, nitrogen fertilization level is indicated as 1 for low nitrogen fertilization and 2 for high nitrogen fertilization.



investigation of breeding progress requires testing of varieties from different periods of registration under exactly the same conditions, however conditions always differ between individual experiments or replicates. In this study, the experiments were conducted after wheat or barley and crop management at all locations was oriented on a commonly used, relatively intensive Northern European oilseed rape production system.

The results show that the locations chosen for this study were subject to contrasting environmental influences. For example, in BOV16, the yields were extremely low due to a combination of pest and disease pressure at a historic low level (C. Algermissen, personal communication). On the contrary, MOS15 and MOS16 were characterized by favorable growth conditions, resulting in significant above-average yield levels.

Despite the divergent environmental influences and complex interactions of NUE, due to its highly quantitative inheritance, the high heritability determined in our study ( $h^2 = 0.92$  for SY and  $h^2 = 0.94$  for OilY) indicates that the enhancement of NUE can be achieved through genetic improvement. Thus, over a long period of almost 25 years, breeding was revealed to be a successful strategy, even when gains from year to year are rather small. This finding is in line with modeling studies by Dresbøll and Thorup-Kristensen (2014).

Several studies have described a genotype by nitrogen (G  $\times$  N) interaction for several major crops (Foulkes et al., 2009; Gaju et al., 2014). In this study the overall picture (Figure 3) suggests that there is a very strong correlation between HN and LN. Also at most locations the correlation between HN and LN is higher than correlations between locations. This suggests

that soil, weather and climate conditions often had a stronger effect than variety specific reaction to NFL. For example, the best performing variety at HN in MOS16 performs only on average in ASE16 and vice versa (Supplementary Figure 1). This finding is in line with previous studies on oilseed rape in France, where most quantitative trait loci were constitutive under LN and HN fertilization for yield parameters but not between environments (Bouchet et al., 2014, 2016).

## Yield Increase Drives NUE

In this work, a collection of winter type oilseed rape varieties with major market importance in Germany and neighboring Northern-European countries was investigated. In contrast to a previous study on NUE in oilseed rape (Stahl et al., 2016), where genetic diversity for NUE was investigated in a broad collection of winter type oilseed rape associations, this study was conducted only on elite varieties that were registered between 1989 and 2014. Thus, the results are not biased through tipping points in breeding history of oilseed rape, as the introduction of zero erucic acid and low glucosinolate content varieties (Downey et al., 1969). In contrast to an earlier study (Kessel et al., 2012) investigating the breeding progress before 1999, the present study focuses on a very recent time window. The fact that varieties such as Express and Lirajet, labeled as modern varieties in Kessel et al. (2012) but are by far the oldest varieties in our study, is indicative of the sliding window in breeding history. Finally, since hybrid varieties have become important and meanwhile dominate oilseed rape production in Northern Europe, a reassessment of breeding progress was overdue. This issue is especially relevant

due to several indications for a heterosis effect on yield, which is often pronounced particularly under limited N conditions (Gehring et al., 2007; Koeslin-Findeklee et al., 2014; Wang et al., 2016).

The ultimate yardstick for mitigating N losses and to create a most N sustainable oilseed rape production is the amount of nitrogen, which is added to the cultivation system required to harvest one unit seed or oil (and protein). The same relationship is expressed by the reciprocal relationship, which is the widely used definition of NUE, as the SY over supplied N (Moll et al., 1982; Good et al., 2004). The above presented results provide evidence for a tremendously successful increase in SY and OilY, and are in the overall picture with many previously described data for oilseed rape (Kessel et al., 2012; Koeslin-Findeklee et al., 2014), wheat (Austin et al., 1980; Fischer and Edmeades, 2010; Cormier et al., 2013; Laidig et al., 2017a), maize (Tollenaar, 1989), rye (Laidig et al., 2017b); rice (Zhu et al., 2016), and triticale (Losert et al., 2017). Therefore, the sometimes appearing hypothesis that cultivation of older varieties could be beneficial for better nutrient use, since they might have comparative advantages in adaptation to low input systems, can unambiguously be rejected.

Since modern high-yielding varieties achieve much higher yields (SY, OilY, and ProteinY), while N inputs and management were the same as those for the old low-yielding varieties, it can be concluded that N losses are significantly lower in modern varieties. In conclusion, the amount of N fertilizer required to produce 1 ton of oilseed rape has dramatically declined within the last 25 years (Supplementary Figure 4). Therefore, our results support the statement of Burney et al. (2010) that increasing the yield is a highly effective instrument to reduce the negative effects on the environment. This is not only because of a higher yield to fertilizer ratio but also due to a lower cultivation area that is required to produce the same amount of commodities. Thus, increasing yields reduces indirect land-use change due to oilseed rape production (Don et al., 2012).

In our study, a strong increase in SY and NUE was observed for both HN and LN, with a high correlation between both the treatments. Bouchet et al. (2014) also found only small  $G \times N$  interactions when NFL differs by 80–90 kg N ha<sup>-1</sup>. Nevertheless, the stronger correlation between the year of registration and SY indicates that the progress was slightly more pronounced at HN. Similarly, the majority of studies (e.g., Brancourt-Hulmel et al., 2005; Brisson et al., 2010) pointed out that progress in SY was higher at HN than at LN. Over the last few decades, farmers became used to rather high amounts of N fertilization, while breeders tried to treat their selection environments in a manner similar to common farming practice in order to select the genotypes for the target environment. This suggests that the selection for HN conditions was probably a direct selection, while selection for LN was rather indirect, and thus not quite so effective as selection under HN conditions. The results of our study should encourage breeders to select their varieties directly at reduced NFL, in order to speed up the breeding progress for low N inputs.

The breeding progress described here is still an ongoing process. Although, **Figure 1** shows a complete absence or even

a negative slope for modern hybrids, unpublished data from varieties released to the market after 2014 helps to conclude that breeding for high yielding and more efficient varieties has not come to an end and is rather still an ongoing process.

## Seed Quality in Light of Efficient Nitrogen Use

The analysis of the breeding success depends not only on yield quantities but also on the changes in the specific seed quality composition. The strong positive correlation between the year of registration and OilY can be explained by the fact that oil was an economically relevant component, with farmers in Germany paid a premium for a high-oil crop, hence breeders have selected genotypes high in oil (Abbadi and Leckband, 2011). OilY increased in an over-proportional manner along with a simultaneous increase in SY. Although, the new OP varieties investigated in this study are, on average, not higher yielding than new hybrids, we observed that they outperformed all other varieties in terms of OilConc under both HN and LN (**Figure 5**).

On the contrary, the improvement of OilConc correlated with a decrease in ProteinConc (**Figure 2**). The negative correlation of  $-0.79$  for LN and  $-0.74$  for HN is consistent with studies reported earlier (Bouchet et al., 2014, 2016; Nyikako et al., 2014). Since, oil and protein synthesis are supposed to rely on the same carbon sources, and protein synthesis is preferentially enhanced with increasing N availability, OilConc declines with increased N fertilization (Rathke et al., 2006; Zhao et al., 2006). However, ANOVA (**Table 4**) illustrates that the effect of NFL is more pronounced on ProteinConc than on OilConc.

From the perspective of resource efficiency, a high ProteinY is desired, since it determines the proportion of N that is captured in the harvested plant organs (as the sink), removed from the field, and thus, protected from losses. Although a strong reduction of ProteinConc is evident in groups of modern varieties, the proportion of N removed from the field compared to proportion invested has improved significantly by overcompensation of a SY-driven enhancement of ProteinY. This finding is in agreement with previously published results from Koeslin-Findeklee et al. (2014).

While OilY and ProteinY are positively correlated to each other (**Figure 2**), since both are predominantly explained by the common factor SY, the improvement of ProteinY through ProteinConc is hampered by the trade-off. Nevertheless, although the notion that the strong negative correlation between OilConc and ProteinConc makes it impossible to increase OilY to N fertilization ratio through enhanced ProteinConc, it does not necessarily mean that breeders are unable to increase the sink capacity due to selection of quality traits. A selection of higher ProteinConcDFF can be a promising strategy to increase the amount of N stored and harvested in seeds without neglecting the achievements in high OilConc (Potter et al., 2016). In this case, the correlation is weakly negatively correlated to OilConc ( $r = -0.39$  of ProteinConcDFF vs.  $r = -0.79$  of ProteinConc at LN, **Figure 2**). However, this strategy would require that the protein content in the meal is included as a couple product in ecological footprint calculations and furthermore receives the

necessary economic attention to justify breeders' attempts to select genotypes superior in this trait.

## Scope for Reduced Fertilizer Inputs without Drastic Yield Penalties

Choosing NFL that are suitable to phenotype responses to N is a non-trivial question for farming practices, selection decisions by breeders, and for experimental setup in crop research. Han et al. (2015) reviewed that phenotypic data collected under a severe N stress are not comparable to mild N stress. In some studies (Kessel et al., 2012; Miersch et al., 2016), N stress was maximized by zero N treatment in order to observe genotypes' response to provoked severe stress. On the other hand, zero N is not a realistic scenario for future oilseed rape production. Even if one considers a reduction in the maximum allowance of N application under upcoming stricter environmental policies, a certain stock application of N is always inevitable to achieve crop yields that allow rentable crop cultivation. Furthermore, simply for the reason, that soil exploitation has to be avoided N fertilization will always be essential to re-deliver the removed N in sustainable farming practice. For this reason, the application of 120 kg N ha<sup>-1</sup> instead of zero N is much closer to reality. The low NFL represents a below-average application rate that is clearly below the maximum NFL that will legally be applicable according to future fertilizer ordinance. The high NFL of 220 kg ha<sup>-1</sup> was oriented on today's common farming practice (Rathke et al., 2006).

The average SY difference between both NFL of 180 kg ha<sup>-1</sup> found in our study was surprisingly low. Even the highest SY difference of 690 kg ha<sup>-1</sup> is comparably small considering the delta of 100 kg N fertilizer between both the treatments. Thus, in our study, the additional fertilizer application did not result in higher yields but rather contributed to a higher N balance surplus. For the first experimental year, one might speculate that observation is explainable due to equal fertilization dosage at the first application date in conjunction with a potentially low availability of N after the second application due to limitations in rainfall in central Germany. Although, the interaction between water limitation and nitrogen uptake (Albert et al., 2012; Sadras and Lawson, 2013) is a reasonable explanation, this logic does not hold true for MOS15, where high rainfalls were observed during spring, and not for the second experimental year, where fertilizer application was reduced at both application dates. Therefore, we have to conclude in all environments of this study that N dosage of much <220 kg N ha<sup>-1</sup> is sufficient to produce yield levels that are usually achieved in agricultural farming practice. Since, there is a linear relationship between N fertilizer inputs and energy inputs, the reduction of NFL not only provides an advantageous effect on the mitigation of greenhouse gas emissions but also provides a drastic profit for energy balance of oilseed rape production.

Even if our results cannot be generalized and might not be applicable to all future growth scenarios, our findings that the yield difference between both NFL was very low, in 10 independent experiments conducted across Germany in 2 years, suggest a remarkable potential for reduction in N fertilization

without dramatic yield penalties. However, this is depending on the weather conditions, which are not known in advance. The fact that farmers have to make their fertilizing decisions in absence of knowledge about further growth conditions makes it difficult to precisely adjust NFL to the real demand (Henke et al., 2007). For further field-based research experiments with contrasting NFL, we suggest to lower the N fertilization in both treatments and agree with Miersch et al. (2016), who suggested to use a delta of at least 100 kg N ha<sup>-1</sup> between NFL.

## CONCLUSION

The study was designed to evaluate the effect of plant breeding on NUE in the past decades and the adjustments in selection required to address future sustainability goals. The experimental data provide strong evidence that direct selection for SY and seed oil concentration leads not only to an enormous (oil) yield gain in highly fertilized environments, but also to a selection of genotypes with superior performance in low N-input cultivation systems. Thus, genetic improvement increases NUE in oilseed rape and reduces the reliance on fertilizer inputs, as already suggested for other crops (Hawkesford, 2014). From this perspective, we concur with Burney et al. (2010) that yield improvement should play a predominant role in strategies toward GHG emission mitigation and enhancement of sustainability of crop production.

The surprisingly low yield gap between high and low nitrogen fertilization provides promising hints toward further N fertilizer-saving potential. However, transfer of these results into knowledge-based farming practices remains challenging. For a more precise and directed selection of even more efficient varieties, a better understanding of G × E × M interaction and physiological determinants of NUE are essential tasks for future research.

## AUTHOR CONTRIBUTIONS

RS, BW, and AS conceived the research; AS and MP performed the experiments, AS and MF performed data analysis; AS and RS wrote the manuscript.

## FUNDING

The work was funded by Federal Ministry for Food and Agriculture grant 22020013 (Federal Agency for Renewable Resources). Additional support was provided by the German Society for the Promotion of Plant Innovation (GFPI, Bonn).

## ACKNOWLEDGMENTS

The authors thank Guido Koglin, Wolfgang Sauermann, Christoph Algermissen (Landwirtschaftskammer Schleswig-Holstein, Bovenau, Germany), Dorothee Varelmann, Jutta Ahlemeyer (Deutsche Saatveredelung AG, Asendorf, Germany), Rüdiger Beißner, Vadym Avramenko, Simone Sendke, Pia Roppel (Monsanto Deutschland GmbH, Nienstädt, Germany),



Stefan Abel, Maximilian Leps (Limagrain, Rosenthal/Peine, Germany), Lothar Behle-Schalk, Karl Heinz Balzer, Mechthild Schwarte (Justus Liebig-University Giessen, Rauischholzhausen, Germany), Stephan Priglmeier and Franz-Xaver Zellner (Saaten-Union, Moosburg, Germany) for excellent technical conductance of field experiments, and Petra Degen, Sabine Frei and Stjepan Vukasovic (Justus Liebig-University Giessen, Giessen, Germany) for valuable help with seed quality analysis. Furthermore we would like to thank Dieter Stelling (Deutsche Saatveredelung

AG, Lippstadt, Germany), Martin Frauen (NPZ, Hohenlieth, Germany) and Amine Abbadi (NPZ Innovation, Hohenlieth, Germany) for valuable comments on the study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00963/full#supplementary-material>

## REFERENCES

- Abbadi, A., and Leckband, G. (2011). Rapeseed breeding for oil content, quality, and sustainability. *Eur. J. Lipid Sci. Technol.* 113, 1198–1206. doi: 10.1002/ejlt.201100063
- Albert, B., Le Cahérec, F., Niogret, M.-F., Faes, P., Avicé, J.-C., Lepout, L., et al. (2012). Nitrogen availability impacts oilseed rape (*Brassica napus* L.) plant water status and proline production efficiency under water-limited conditions. *Planta* 236, 659–676. doi: 10.1007/s00425-012-1636-8
- Austin, R. B., Bingham, J., Blackwell, R. D., Evans, L. T., Ford, M. A., Morgan, C. L., et al. (1980). Genetic improvements in winter wheat yields since 1900 and associated physiological changes. *J. Agric. Sci.* 94, 675. doi: 10.1017/S0021859600028665
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bouchet, A.-S., Laperche, A., Bissuel-Belaygue, C., Baron, C., Morice, J., Rousseau-Gueutin, M., et al. (2016). Genetic basis of nitrogen use efficiency and yield stability across environments in winter rapeseed. *BMC Genet.* 17:131. doi: 10.1186/s12863-016-0432-z
- Bouchet, A.-S., Nesi, N., Bissuel, C., Bregeon, M., Lariépe, A., Navier, H., et al. (2014). Genetic control of yield and yield components in winter oilseed rape (*Brassica napus* L.) grown under nitrogen limitation. *Euphytica* 199, 183–205. doi: 10.1007/s10681-014-1130-4
- Brancourt-Hulmel, M., Heumez, E., Pluchard, P., Beghin, D., Depatureaux, C., Giraud, A., et al. (2005). Indirect versus direct selection of winter wheat for low-input or high-input levels. *Crop Sci.* 45, 1427–1431. doi: 10.2135/cropsci2003.0343
- Brisson, N., Gate, P., Gouache, D., Charmet, G., Oury, F.-X., and Huard, F. (2010). Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *Field Crops Res.* 119, 201–212. doi: 10.1016/j.fcr.2010.07.012
- Burney, J. A., Davis, S. J., and Lobell, D. B. (2010). Greenhouse gas mitigation by agricultural intensification. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12052–12057. doi: 10.1073/pnas.0914216107
- Cormier, F., Faure, S., Dubreuil, P., Heumez, E., Beauchene, K., Lafarge, S., et al. (2013). A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 126, 3035–3048. doi: 10.1007/s00122-013-2191-9
- De Mendiburu, F., and Simon, R. (2015). Agricolae - ten years of an open source statistical tool for experiments in breeding, agriculture and biology. *PeerJ PrePrints* 3:e1404v1. doi: 10.7287/peerj.preprints.1404v1
- Don, A., Osborne, B., Hastings, A., Skiba, U., Carter, M. S., Drewer, J., et al. (2012). Land-use change to bioenergy production in Europe: implications for the greenhouse gas balance and soil carbon. *Glob. Change Biol. Bioenergy* 4, 372–391. doi: 10.1111/j.1757-1707.2011.01116.x
- Downey, R. K., Craig, B. M., and Youngs, C. G. (1969). Breeding rapeseed for oil and meal quality. *J. Am. Oil Chem. Soc.* 46, 121–123. doi: 10.1007/BF02635712
- Dresbøll, D. B., and Thorup-Kristensen, K. (2014). Will breeding for nitrogen use efficient crops lead to nitrogen use efficient cropping systems? A simulation study of G×E×M interactions. *Euphytica* 199, 97–117. doi: 10.1007/s10681-014-1199-9
- Fischer, R. A., Byerlee, D., and Edmeades, G. O. (2014). *Crop Yields and Global Food Security: Will Yield Increase Continue to Feed the World?* ACIAR Monograph No. 158. Australian Centre for International Agricultural Research, Canberra, ACT.
- Fischer, R. A., and Edmeades, G. O. (2010). Breeding and cereal yield progress. *Crop Sci.* 50, S-85–S-98. doi: 10.2135/cropsci2009.10.0564
- Fleddermann, M., Fechner, A., Rossler, A., Bahr, M., Pastor, A., Liebert, F., et al. (2013). Nutritional evaluation of rapeseed protein compared to soy protein for quality, plasma amino acids, and nitrogen balance—a randomized cross-over intervention study in humans. *Clin. Nutr.* 32, 519–526. doi: 10.1016/j.clnu.2012.11.005
- Foulkes, M. J., Hawkesford, M. J., Barraclough, P. B., Holdsworth, M. J., Kerr, S., Kightley, S., et al. (2009). Identifying traits to improve the nitrogen economy of wheat: recent advances and future prospects. *Field Crops Res.* 114, 329–342. doi: 10.1016/j.fcr.2009.09.005
- Gaju, O., Allard, V., Martre, P., Le Gouis, J., Moreau, D., Bogard, M., et al. (2014). Nitrogen partitioning and remobilization in relation to leaf senescence, grain yield and grain nitrogen concentration in wheat cultivars. *Field Crops Res.* 155, 213–223. doi: 10.1016/j.fcr.2013.09.003
- Galloway, J. N., Leach, A. M., Bleeker, A., and Erisman, J. W. (2013). A chronology of human understanding of the nitrogen cycle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368:20130120. doi: 10.1098/rstb.2013.0120
- Gehring, A., Snowdon, R., Spiller, T., Basunanda, P., and Friedt, W. (2007). New oilseed rape (*Brassica napus*) hybrids with high levels of heterosis for seed yield under nutrient-poor conditions. *Breed. Sci.* 57, 315–320. doi: 10.1270/jsbbs.57.315
- Good, A. G., Shrawat, A. K., and Muench, D. G. (2004). Can less yield more? Is reducing nutrient input into the environment compatible with maintaining crop production? *Trends Plant Sci.* 9, 597–605. doi: 10.1016/j.tplants.2004.10.008
- Han, M., Okamoto, M., Beatty, P. H., Rothstein, S. J., and Good, A. G. (2015). The genetics of nitrogen use efficiency in crop plants. *Annu. Rev. Genet.* 49, 269–289. doi: 10.1146/annurev-genet-112414-055037
- Hawkesford, M. J. (2014). Reducing the reliance on nitrogen fertilizer for wheat production. *J. Cereal Sci.* 59, 276–283. doi: 10.1016/j.jcs.2013.12.001
- Henke, J., Böttcher, U., Neukam, D., Sieling, K., and Kage, H. (2008). Evaluation of different agronomic strategies to reduce nitrate leaching after winter oilseed rape (*Brassica napus* L.) using a simulation model. *Nutr. Cycl. Agroecosyst.* 82, 299–314. doi: 10.1007/s10705-008-9192-0
- Henke, J., Breustedt, G., Sieling, K., and Kage, H. (2007). Impact of uncertainty on the optimum nitrogen fertilization rate and agronomic, ecological and economic factors in an oilseed rape based crop rotation. *J. Agric. Sci.* 145, 455. doi: 10.1017/S0021859607007204
- Hirel, B., Le Gouis, J., Ney, B., and Gallais, A. (2007). The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *J. Exp. Bot.* 58, 2369–2387. doi: 10.1093/jxb/erm097
- Kant, S., Bi, Y.-M., and Rothstein, S. J. (2011). Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency. *J. Exp. Bot.* 62, 1499–1509. doi: 10.1093/jxb/erq297
- Kessel, B., Schierholt, A., and Becker, H. C. (2012). Nitrogen use efficiency in a genetically diverse set of winter oilseed rape (*Brassica napus* L.). *Crop Sci.* 52, 2546–2554. doi: 10.2135/cropsci2012.02.0134
- Koeslin-Findeklee, F., Meyer, A., Girke, A., Beckmann, K., and Horst, W. J. (2014). The superior nitrogen efficiency of winter oilseed rape (*Brassica napus* L.)

- hybrids is not related to delayed nitrogen starvation-induced leaf senescence. *Plant Soil* 384, 347–362. doi: 10.1007/s11104-014-2212-8
- Kumar, A., Sharma, A., and Upadhyaya, K. C. (2016). Vegetable oil: nutritional and industrial perspective. *Curr. Genomics* 17, 230–240. doi: 10.2174/1389202917666160202220107
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2016). *lmerTest: Tests in Linear Mixed Effects Models*. R Package version 2.0–30.
- Laidig, F., Piepho, H. P., Rentel, D., Drobek, T., Meyer, U., and Huesken, A. (2017a). Breeding progress, environmental variation and correlation of winter wheat yield and quality traits in German official variety trials and on-farm during 1983–2014. *Theor. Appl. Genet.* 130, 223–245. doi: 10.1007/s00122-016-2810-3
- Laidig, F., Piepho, H. P., Rentel, D., Drobek, T., Meyer, U., and Huesken, A. (2017b). Breeding progress, variation, and correlation of grain and quality traits in winter rye hybrid and population varieties and national on-farm progress in Germany over 26 years. *Theor. Appl. Genet.* 130, 981–998. doi: 10.1007/s00122-017-2865-9
- Lenth, V. R. (2016). Least-squares means: the R package lsmeans. *J. Stat. Softw.* 69, 1–33. doi: 10.18637/jss.v069.i01
- Liu, J., You, L., Amini, M., Obersteiner, M., Herrero, M., Zehnder, A. J. B., et al. (2010). A high-resolution assessment on global nitrogen flows in cropland. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8035–8040. doi: 10.1073/pnas.0913658107
- Lopes, M. S., Reynolds, M. P., Manes, Y., Singh, R. P., Crossa, J., and Braun, H. J. (2012). Genetic yield gains and changes in associated traits of CIMMYT spring bread wheat in a “historic” set representing 30 years of breeding. *Crop Sci.* 52, 1123–1131. doi: 10.2135/cropsci2011.09.0467
- Losert, D., Maurer, H. P., Marulanda, J. J., and Würschum, T. (2017). Phenotypic and genotypic analyses of diversity and breeding progress in European triticale ( $\times$  Triticosecale Wittmack). *Plant Breed.* 136, 18–27. doi: 10.1111/pbr.12433
- Miersch, S., Gertz, A., Breuer, F., Schierholt, A., and Becker, H. C. (2016). Influence of the Semi-dwarf growth type on seed yield and agronomic parameters at low and high nitrogen fertilization in winter oilseed rape. *Crop Sci.* 56, 1573–1585. doi: 10.2135/cropsci2015.09.0554
- Moll, R. H., Kamprath, E. J., and Jackson, W. A. (1982). Analysis and interpretation of factors which contribute to efficiency of nitrogen utilization. *Agron. J.* 74, 562. doi: 10.2134/agronj1982.00021962007400030037x
- Müller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., and Foley, J. A. (2012). Closing yield gaps through nutrient and water management. *Nature* 490, 254–257. doi: 10.1038/nature11420
- Nyikako, J., Schierholt, A., Kessel, B., and Becker, H. C. (2014). Genetic variation in nitrogen uptake and utilization efficiency in a segregating DH population of winter oilseed rape. *Euphytica* 199, 3–11. doi: 10.1007/s10681-014-1201-6
- Potter, T., Burton, W., Edwards, J., Wratten, N., Mailer, R., Salisbury, P., et al. (2016). Assessing progress in breeding to improve grain yield, quality and blackleg (*Leptosphaeria maculans*) resistance in selected Australian canola cultivars (1978–2012). *Crop Pasture Sci.* 67, 308–316. doi: 10.1071/CP15290
- Rathke, G., Behrens, T., and Diepenbrock, W. (2006). Integrated nitrogen management strategies to improve seed yield, oil content and nitrogen efficiency of winter oilseed rape (*Brassica napus* L.): a review. *Agric. Ecosyst. Environ.* 117, 80–108. doi: 10.1016/j.agee.2006.04.006
- Rathke, G.-W., and Diepenbrock, W. (2006). Energy balance of winter oilseed rape (*Brassica napus* L.) cropping as related to nitrogen supply and preceding crop. *Eur. J. Agron.* 24, 35–44. doi: 10.1016/j.eja.2005.04.003
- R Core Team (2013). *A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reinhardt, T.-C. (1992). *Entwicklung und Anwendung von Nah-Infrarot-Spektroskopischen Methoden für die Bestimmung von Öl-, Protein-, Glucosinolat-, Feuchte- und Fettsäure- Gehalten in intakter Rapssaat*. Ph.D. dissertation. University of Göttingen.
- Rockström, J., Steffen, W., Noone, K., Persson, A., Chapin, F. S. III., Lambin, E. F., et al. (2009). A safe operating space for humanity. *Nature* 461, 472–475. doi: 10.1038/461472a
- Rothstein, S. J. (2007). Returning to our roots: making plant biology research relevant to future challenges in agriculture. *Plant Cell* 19, 2695–2699. doi: 10.1105/tpc.107.053074
- Sadras, V. O., and Lawson, C. (2013). Nitrogen and water-use efficiency of Australian wheat varieties released between 1958 and 2007. *Eur. J. Agron.* 46, 34–41. doi: 10.1016/j.eja.2012.11.008
- Sieling, K., and Kage, H. (2006). N balance as an indicator of N leaching in an oilseed rape – winter wheat – winter barley rotation. *Agric. Ecosyst. Environ.* 115, 261–269. doi: 10.1016/j.agee.2006.01.011
- Stahl, A., Friedt, W., Wittkop, B., and Snowdon, R. J. (2016). Complementary diversity for nitrogen uptake and utilisation efficiency reveals broad potential for increased sustainability of oilseed rape production. *Plant Soil* 400, 245–262. doi: 10.1007/s11104-015-2726-8
- Steffen, W., Richardson, K., Rockstrom, J., Cornell, S. E., Fetzer, I., Bennett, E. M., et al. (2015). Sustainability. Planetary boundaries: guiding human development on a changing planet. *Science* 347:1259855. doi: 10.1126/science.1259855
- Sylvester-Bradley, R., and Kindred, D. R. (2009). Analysing nitrogen responses of cereals to prioritize routes to the improvement of nitrogen use efficiency. *J. Exp. Bot.* 60, 1939–1951. doi: 10.1093/jxb/erp116
- Thorup-Kristensen, K., and Kirkegaard, J. (2016). Root system-based limits to agricultural productivity and efficiency: the farming systems context. *Ann. Bot.* 118, 573–592. doi: 10.1093/aob/mcw122
- Tillmann, P., and Paul, C. (1998). The repeatability file—a tool for reducing the sensitivity of near infrared spectroscopy calibrations to moisture variation. *J. Near Infrared Spectrosc.* 6:61. doi: 10.1255/jnirs.122
- Tillmann, P., Reinhardt, T.-C., and Paul, C. (2000). Networking of near infrared spectroscopy instruments for rapeseed analysis: a comparison of different procedures. *J. Near Infrared Spectrosc.* 8:101. doi: 10.1255/jnirs.269
- Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20260–20264. doi: 10.1073/pnas.1116437108
- Tkachuk, R. (1981). Oil and protein analysis of whole rapeseed kernels by near infrared reflectance spectroscopy. *J. Am. Oil Chem. Soc.* 58, 819–822. doi: 10.1007/BF02665588
- Tollenaar, M. (1989). Genetic improvement in grain yield of commercial maize hybrids grown in Ontario from 1959 to 1988. *Crop Sci.* 29, 1365–1371. doi: 10.2135/cropsci1989.0011183X002900060007x
- Wang, L., Mühling, K.-H., and Schulte auf'm Erley, S. (2016). Nitrogen efficiency and leaf nitrogen remobilisation of oilseed rape lines and hybrids. *Ann. Appl. Biol.* 169, 125–133. doi: 10.1111/aab.12286
- Weiser, C., Fuß, R., Kage, H., and Flessa, H. (2017). Do farmers in Germany exploit the potential yield and nitrogen benefits from preceding oilseed rape in winter wheat cultivation? *Arch. Agron. Soil Sci.* 1–13. doi: 10.1080/03650340.2017.1326031
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.
- Zhao, J., Becker, H. C., Zhang, D., Zhang, Y., and Ecke, W. (2006). Conditional QTL mapping of oil content in rapeseed with respect to protein content and traits related to plant development and grain yield. *Theor. Appl. Genet.* 113, 33–38. doi: 10.1007/s00122-006-0267-5
- Zhu, G., Peng, S., Huang, J., Cui, K., Nie, L., and Wang, F. (2016). Genetic improvements in rice yield and concomitant increases in radiation- and nitrogen-use efficiency in middle reaches of Yangtze River. *Sci. Rep.* 6:21049. doi: 10.1038/srep21049

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Stahl, Pfeifer, Frisch, Wittkop and Snowdon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Use of Blue-Green Fluorescence and Thermal Imaging in the Early Detection of Sunflower Infection by the Root Parasitic Weed *Orobanche cumana* Wallr.

Carmen M. Ortiz-Bustos<sup>1</sup>, María L. Pérez-Bueno<sup>2</sup>, Matilde Barón<sup>2</sup> and Leire Molinero-Ruiz<sup>1\*</sup>

<sup>1</sup> Department of Crop Protection, Institute for Sustainable Agriculture, CSIC, Cordoba, Spain, <sup>2</sup> Estación Experimental del Zaidín, CSIC, Granada, Spain

## OPEN ACCESS

### Edited by:

Juan Moral,  
Kearney Agricultural Center  
(UC-Davis), United States

### Reviewed by:

Khawar Jabran,  
Düzce University, Turkey  
Grama Nanjappa Dhanapal,  
University of Agricultural Sciences,  
Bangalore, India

### \*Correspondence:

Leire Molinero-Ruiz  
leire.molinero@csic.es

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 06 March 2017

**Accepted:** 04 May 2017

**Published:** 18 May 2017

### Citation:

Ortiz-Bustos CM, Pérez-Bueno ML,  
Barón M and Molinero-Ruiz L (2017)  
Use of Blue-Green Fluorescence and  
Thermal Imaging in the Early Detection  
of Sunflower Infection by the Root  
Parasitic Weed *Orobanche cumana*  
Wallr. *Front. Plant Sci.* 8:833.  
doi: 10.3389/fpls.2017.00833

Although the impact of *Orobanche cumana* Wallr. on sunflower (*Helianthus annuus* L.) becomes evident with emergence of broomrape shoots aboveground, infection occurs early after sowing, the host physiology being altered during underground parasite stages. Genetic resistance is the most effective control method and one of the main goals of sunflower breeding programmes. Blue-green fluorescence (BGF) and thermal imaging allow non-destructive monitoring of plant diseases, since they are sensitive to physiological disorders in plants. We analyzed the BGF emission by leaves of healthy sunflower plantlets, and we implemented BGF and thermal imaging in the detection of the infection by *O. cumana* during underground parasite development. Increases in BGF emission were observed in leaf pairs of healthy sunflowers during their development. Lower BGF was consistently detected in parasitized plants throughout leaf expansion and low pigment concentration was detected at final time, supporting the interpretation of a decrease in secondary metabolites upon infection. Parasite-induced stomatal closure and transpiration reduction were suggested by warmer leaves of inoculated sunflowers throughout the experiment. BGF imaging and thermography could be implemented for fast screening of sunflower breeding material. Both techniques are valuable approaches to assess the processes by which *O. cumana* alters physiology (secondary metabolism and photosynthesis) of sunflower.

**Keywords:** broomrape, carotenoids, early diagnosis, *Helianthus annuus* L., multicolor fluorescence, thermal detection

## INTRODUCTION

Sunflower (*Helianthus annuus* L.) oil is a major commodity in world trade mainly in Europe, where 60% of the total world production is obtained every year (FAOSTAT, 2016). The main biotic constraint for sunflower oil production in all the countries where sunflowers are grown - with the only exception of the Americas- is broomrape, caused by the achlorophyllous parasitic plant *Orobanche cumana* Wallr. Currently, *O. cumana* is present in 15–20% of the area cropped

to sunflower in the world, causing close to 100% yield losses under high infestations (Fernández-Martínez et al., 2015). After parasite seed germination in the soil early after sowing, penetration of its intrusive cells into the host root tissues triggers their division, leading to the formation of a subterranean shoot that grows outside the root of sunflower. Then, parasite shoots emerge from the soil and form a flowering spike that rapidly produces a large amount of tiny seeds. *Orobanche cumana* parasitizes sunflower all through the crop growing season, and flowering of both, host and parasite, are coincident in time (Molinero-Ruiz et al., 2015).

Vascular connections established between broomrapes (*Orobanche* spp. and *Phelipanche* spp.) and their hosts force a complete disruption of the vascular system of the latter during the initial weeks of parasitism (Westwood, 2013). As a result, strong sink strength diverts water and nutrients necessary for the growth of the host (Stewart and Press, 1990). From the emergence of *O. cumana* shoots aboveground onwards, a reduced growth of sunflower is evident (Alcántara et al., 2006) and the decrease of the yield is due to absent or smaller sized capitula, low number and small seeds, and even plant death (Molinero-Ruiz et al., 2015). Thus, most of the metabolic imbalance and water flow impairment might be produced irreversibly in sunflower during underground parasite stages. Since it is physiologically closely linked to sunflower, *O. cumana* is extremely difficult to control in its parasitic stage. Some herbicide treatments may be effective, but depend on the development of herbicide resistance in the crop (Molinero-Ruiz et al., 2015). With regard to the free-standing stage of the parasite, thousands of tiny seeds are produced each season, their longevity contributing to the build-up of populations in soils. Since soil fumigation and solarisation are not economically feasible, use of inducers of suicidal broomrape germination requires an appropriate application technology and agronomic practices and biological control agents have limited efficacy, the effective control of *O. cumana* relies on genetic resistance (Molinero-Ruiz et al., 2015; Fernández-Aparicio et al., 2016).

Imaging techniques, especially blue-green fluorescence and thermal imaging among others, are powerful tools for use in plant stress detection (Lichtenthaler and Miché, 1997; Chaerle and Van Der Straeten, 2000), including determination of plant diseases (Nilsson, 1995). They are based on the assessment of optical properties of plants within different regions of the electromagnetic spectrum and are able to utilize information beyond the visible range (Chaerle and Van Der Straeten, 2000). Under UV-excitation, plants can emit a wide fluorescence spectrum ranging from about 400 to 800 nm. This spectrum is the sum of two distinct types of fluorescence: blue-green fluorescence (BGF) characterized by a peak around 440 nm (F440) and a shoulder near 520 nm (F520), and fluorescence in the red and far-red regions with two characteristic peaks at about 680 nm (F680) and 740 nm (F740), respectively. The intensity of the BGF after UV light excitation is constant on a short time scale (minutes) and it has been proven to be very sensitive to single stress factors in plants (Buschmann and Lichtenthaler, 1998; Buschmann et al., 2000). The BGF from intact leaves is emitted by cinnamic acids, mainly ferulic acid (Morales et al., 1996),

covalently bound to the cell walls of the epidermis, and by other phenolic compounds (Buschmann et al., 2000). In addition, BGF can emanate also from other secondary metabolites as reviewed by Cerovic et al. (1999). Moreover, quantitative variation of plant secondary metabolites associated with BGF emission can be the consequence, among others, of leaf aging (Morales et al., 2005), water stress (Kautz et al., 2014) or pathogen infection (Chaerle et al., 2007; Granum et al., 2015; Pérez-Bueno et al., 2015). Additionally, F680 and F740 are emitted by chlorophyll *a* (Chl *a*) and they are dependent on the pigment content and other factors such as optics of the leaf (Buschmann and Lichtenthaler, 1998; Gitelson et al., 1998).

On the other hand, radiation emitted by plants in the thermal infrared range from 8 to 12–14  $\mu\text{m}$  can be detected by thermographic and infrared cameras. Infrared thermography is an indicator of transpiration and stomatal conductance (Eaton and Belden, 1929; Jackson et al., 1977; Nilsson, 1995). It has been correlated, among others, with plant water status (Zarco-Tejada et al., 2012; Raza et al., 2014; Grant et al., 2016; Mangus et al., 2016) and canopy microclimate (Lenthe et al., 2007; Leuzinger and Korner, 2007), as well as with early infections by airborne (Lindenthal et al., 2005; Baranowski et al., 2015; Pérez-Bueno et al., 2015) and soilborne (Wang et al., 2012; Calderón et al., 2013; Granum et al., 2015) plant pathogens.

Both fluorescence and thermal imaging of whole leaves and whole plants have become essential techniques from the perspective of non-destructive monitoring of plant diseases from far (remote sensing) and from near distance, since sensors are sensitive to physiological disorders in plants associated with pathogen attack and with disease resulting from that attack (Buschmann et al., 2000; Chaerle and Van Der Straeten, 2000; Mahlein, 2016). Techniques for near distance monitoring of plant diseases can have a direct application in plant phenotyping of breeding programmes as those of sunflower. In this regard, fluorescence imaging in the red and far-red region has been recently applied in the early diagnosis of the infection of sunflower by *O. cumana* (Ortiz-Bustos et al., 2016a).

The objectives of this study were to: (a) analyse the distribution of BGF emitted by sunflower leaves during early growth stages of healthy sunflower plants in order to identify the adequate areas and times for BGF imaging, (b) evaluate the use of BGF imaging as an indicator of the infection of sunflower by *O. cumana* during initial underground development of the parasite and compare the BGF information with the pigments concentration in leaves, and (c) analyse the effect of underground infection by *O. cumana* in early stages of sunflower growth using thermal imaging of the leaves.

## MATERIALS AND METHODS

### Growth of healthy Sunflower

Four sunflower seeds of the inbred line NR5 were surface sterilized by immersion in 20% household bleach (50 g of active chlorine per liter) for 5 min, then thoroughly rinsed in deionized water and incubated at 25°C in the dark at saturation humidity until radicles were 2–5 mm long. Thereafter, individual sunflower seedlings were transplanted into pots with 250 g of a soil mixture



SSP (sand:silt:peat moss 2:1:1, V). Pots were kept in a glasshouse at 12–22°C without additional lighting (15/9 h light/dark regime) for 5 weeks. Plants were watered as needed and, when they were 2 weeks old and until the end of the experiment 3 weeks later, they were fertilized once a week with 15 ml/pot of a nutrient solution with N:P:K (7:5:6).

## Inoculation with *O. cumana*, Growth Conditions and Disease Development

Sunflower plants were inoculated with population LPA13 of *O. cumana* (race F) (García-Carneros et al., 2014) following the methodology by Molinero-Ruiz et al. (2014). Eight sunflower seedlings (replications) of the inbred line NR5, which is the differential for the race F (Molinero-Ruiz et al., 2015), were individually transferred into pots with 250 g of SSP uniformly infested with 10 mg of parasite seeds. Eight non inoculated plants were established as controls. Plants were grown in the glasshouse under the same conditions described above for 5 weeks.

Sunflower plants were watched for development of wilting symptoms caused by the infection by *O. cumana*, but no visible differences were detected in inoculated plants as compared to the controls. At the end of the experiment, each plant was removed from the soil and its root system was washed and air dried at room temperature. Then, roots were weighed and the number and weight of *O. cumana* tubercles attached to the roots were subsequently recorded using a stereoscope and a precision balance, respectively.

## Blue-Green Fluorescence Imaging

Blue-green fluorescence images were acquired sequentially with an Open FluorCamFC 800-O using UV (355 nm) excitation light. Images for F440 and F520 as well as images corresponding to the F440/F520 ratio were analyzed with Fluorcam7 software (Photon Systems Instruments, Brno, Czech Republic) according to Pérez-Bueno et al. (2015). Nine images were captured during 18 s in order to obtain each averaged fluorescence image. Image size was 640 × 480 pixels with a resolution of 96 pixels per inch. Measurements were taken on attached and unshaded leaves and always at the same time of day. Two replicates (one from each blade) were considered in each leaf pair (LP). Numerical data from a representative area in each replicate were analyzed.

The time span of BGF in healthy sunflower plants was determined by capturing images of the first four LPs each, every 3–4 days from 2 to 5-weeks-old. For each measurement date, LPs were independently analyzed, and between 6 and 8 replicates were considered for each of them.

Concerning the effect of *O. cumana* on the BGF emission of leaves of infected sunflower, measurements of F440 and F520, as well as those of F440/F520, F440/F680, and F440/F740, were acquired in all leaves of control and inoculated plants twice a week since they were 1 cm long until their complete expansion. Measurements were made during the V1–V4 vegetative stages of the plants (Schneider and Miller, 1981) and, at each point in time, leaves of the same developmental stage were compared.

## Spectrophotometrical Determination of Pigments Concentration in Leaves

Total chlorophyll [Chl (*a*+*b*)] and carotenoids (xanthophylls and carotenes) [Car (*x*+*c*)] contents of parasitized sunflower were determined by spectrophotometry, and compared to those of the controls. Extractions were carried out on the second, third and fourth LP of the plants 5 weeks after inoculation (wai). From each LP blade, a 1 cm disk was punched, weighted and placed in vials containing liquid nitrogen. Disks were ground and photosynthetic pigments were extracted by adding 4 ml of 80% acetone (v/v). Absorbance at 470, 647, and 663 nm was measured from the resulting extracts with a Shimadzu UV-1800 spectrophotometer (Shimadzu Corporation, Tokyo, Japan). Concentrations of Chl *a*, Chl *b* and Car (*x*+*c*) were calculated according to Lichtenthaler and Buschmann (2001) and expressed as µg/g fresh leaf weight:

$$Chl_a(\mu g/ml) = 12.25A_{663.2} - 2.79A_{646.8}$$

$$Chl_b(\mu g/ml) = 21.50A_{646.8} - 5.10A_{663.2}$$

$$Car_{(x+c)}(\mu g/ml) = (1000A_{470} - 1.82C_a - 85.02C_b)/198$$

The Chl (*a*+*b*)/Car (*x*+*c*) ratio was calculated according to Konanz et al. (2014).

## Thermal Imaging

Infrared images of leaves of inoculated and control sunflowers were taken using a FLIR A305sc camera (FLIR Systems, Wilsonville, Oregon, USA) that operates in the 7.5–13.5 µm wavelength range and has a thermal sensitivity <0.05°C at +30°C and accuracy of ±2°C. The thermal camera was vertically positioned at approximately 0.3 m from the canopy and produced images of 320 × 240 pixels, with a viewing field of 45°. Digital video data were analyzed by the Research & Development software by FLIR.

Images were taken twice a week, from 2 and until 5 wai. At each time point, all fully developed leaves in control and in inoculated plants were measured. Measurements were performed on attached and unshaded leaves at the same time of day. Two hours before each measurement time, inoculated and control plants were moved from the glasshouse to a growth chamber in order to avoid glasshouse thermal fluctuations.

Digital color (RGB) images (2,048 × 1,536 pixels) were obtained simultaneous to the thermal images. One representative area in the midsection of each leaf was selected for the analysis after comparison of thermal and RGB images.

## Data Processing and Statistics

Data of the spatial distribution and time span of BGF on healthy sunflower were analyzed as means and their corresponding standard errors. When *O. cumana* was inoculated to sunflower, BGF emission, pigments concentration, leaf temperature and broomrape incidence [BI, transformed according to sqrt (BI + 0.5)] were statistically analyzed. The experiment was set up as a completely randomized design and was conducted twice. Similarity between repetitions was tested by analysis of variance (ANOVA) with repetitions as blocks. Since no significant differences were found, the data were pooled. When,

after ANOVA, differences of any of the considered variables between inoculated and control plants were significant, mean values were compared by Fisher's protected Least Significant Difference tests ( $P = 0.05$ ). STATISTIX 8.0 software (Analytical software, Tallahassee, FL, USA) was used for all the analyses.

## RESULTS

### Blue-Green Fluorescence Emission in Healthy Sunflower Plants

A clear increase in the intensity of F440 and F440/F520 was observed in the first LP during its expansion and development. A similar trend was also observed in the case of the second LP in both parameters from the third week onwards. Values of F440 and F440/F520 of the two upper LPs evolved similarly to those of the lower LPs along the time, although smaller increases were observed in the third and fourth LP throughout the measurement period. By contrast, and with the exception of a slight increase in the F520 signal in the first LP, that of the rest of leaves remained fairly constant throughout the experiment, and even a decrease was detected in the second LP in the last week (Figure 1).

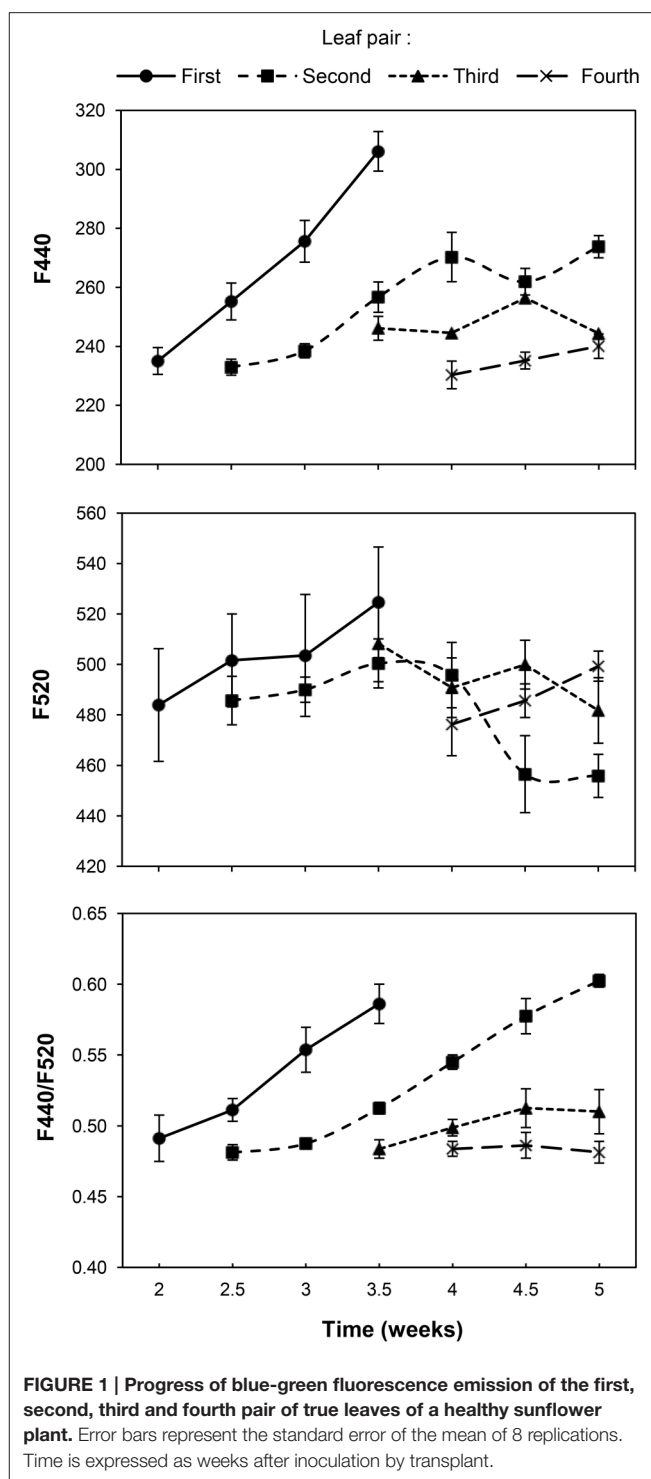
### Effect of *O. cumana* on BGF Emission of Sunflower

The effect of the infection of sunflower by *O. cumana* on the BGF emission of leaves was examined by comparison with that of leaves from the control plants at five moments during the 3 week period, and is presented in Figure 2.

A significantly lower fluorescence at 440 nm was consistently detected in parasitized plants throughout the experiment. Initially, leaves showed lower values of F440 than those at the end of the experiment, in both inoculated and control sunflowers (Figure 2A). Similarly, the F520 values in young leaves were significantly decreased by the parasite attack. As with F440, leaves of inoculated plants had lower F520, although F520 of older leaves did not significantly differ from that of the controls (Figure 2B).

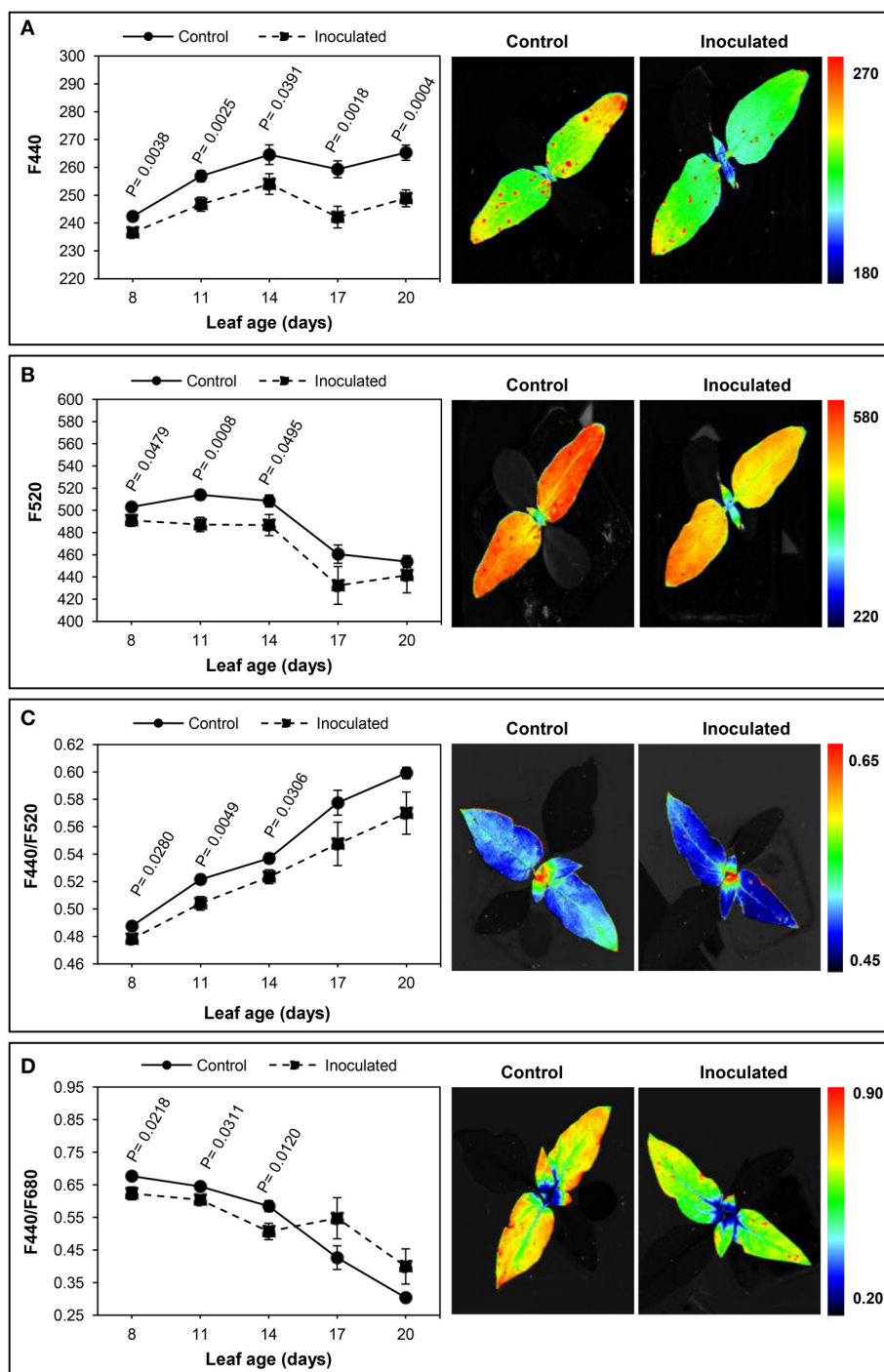
Concerning the F440/F520 ratio, significant decreases were obtained in up to 14-day-old leaves of infected plants as compared to leaves from healthy plants. A progressive increase in F440/F520 ratio was evident in leaves of inoculated and control sunflowers throughout the experiment, although no significant differences were found between treatments in the two last measurements (Figure 2C).

On the other hand, the effect of the inoculation by *O. cumana* was also observed in the F440/F680 emission, being this fluorescence variable significantly low in 8 to 14-day-old leaves of parasitized sunflowers. Higher F440/F680 ratios were obtained in older leaves of inoculated plants, although differences with those in leaves of the control plants were not significant (Figure 2D). Similar results were obtained in the case of the F440/F740 ratio, young leaves of inoculated plants presenting significantly lower values than those of the control plants. When leaves were older than 14 days, the F440/F740 ratio was higher in the inoculated sunflowers, but no significant differences were found (data not shown).



### Carotenoids and Total Chlorophyll Contents in Sunflower Leaves upon Infection by *O. cumana*

The effect of the inoculation by *O. cumana* on chlorophyll and carotenoids contents at the end of the experiment is presented in Table 1. As previously observed by our research group



**FIGURE 2 | Time-course of multicolour fluorescence emission of *Orobancha cumana*-inoculated sunflower plants and non-inoculated (control) plants throughout leaf development.** Mean measurements of F440 (A), F520 (B), F440/F520 (C), and F440/F680 (D) of leaves and representative fluorescence images of 8 day-old leaves. Vertical bars represent the standard error of means of 14–48 replications. Analyses of variance at  $P < 0.05$  between inoculated and control plants were conducted and values of significance are shown.

(Ortiz-Bustos et al., 2016a), significant differences of chlorophyll content occurred in sunflower plants upon inoculation with *O. cumana*. Conversely, on the second and third LPs, no

significant differences in either the carotenoids content or the chlorophyll/carotenoids ratio were detected between inoculated and control plants. Significantly lower values of both variables

**TABLE 1 | Measurements of pigments concentration (chlorophyll content, carotenoids content and chlorophyll/carotenoids ratio) for the second, third, and fourth leaf pair of sunflower plants inoculated with *Orobancha cumana* (I) and control plants (C) at 5 weeks after inoculation.**

Pigments Concentration <sup>a</sup>	Treatment	Leaf pair (LP)		
		2nd	3rd	4th
Chl (a+b)/Fw (μg/g fresh weight)	C	1359.50 ± 56.58 <sup>b</sup>	1443.90 ± 51.90	1442.00 ± 101.04
	I	1548.60 ± 63.53	1515.90 ± 94.95	1028.40 ± 67.51
	<i>P</i> <sup>c</sup>	0.0462	0.5184	0.0052
Car (x+c)/Fw (μg/g fresh weight)	C	239.96 ± 8.24	244.74 ± 15.84	274.30 ± 19.40
	I	263.57 ± 16.01	257.71 ± 22.38	203.20 ± 12.13
	<i>P</i>	0.2141	0.6434	0.0091
Chl (a+b)/Car (x+c)	C	5.92 ± 0.11	6.00 ± 0.17	5.80 ± 0.11
	I	5.93 ± 0.16	5.75 ± 0.22	5.07 ± 0.17

<sup>a</sup>Chlorophyll and carotenoids (xanthophylls and carotenes) content is expressed as content of chlorophyll a and b [Chl (a+b)], and [Car (x+c)] respectively, and chlorophyll/carotenoids ratio is expressed as Chl (a+b)/Car (x+c).

<sup>b</sup>Mean ± standard error (SE), *n* = 7–8.

<sup>c</sup>Level of significance of differences in the variables between control (C) and inoculated (I) plants obtained after analyses of variance according to a completely randomized statistical design.

were obtained in the fourth LP of inoculated plants (203.20 μg/g and 5.07 for total carotenoids and for the chlorophyll/carotenoids ratio, respectively) compared with those in the same LP of healthy plants (274.30 μg/g and 5.80 for total carotenoids and for the chlorophyll/carotenoids ratio, respectively) (Table 1).

## Effect of *O. cumana* on the Temperature of Sunflower Leaf

Figure 3 shows the progress over time of the average leaf temperature of the inoculated and control plants.

Leaves of inoculated sunflowers were warmer than those of the controls at all the measurement dates. Significant temperature increases were observed 2.5 wai and persisted until the end of the experiment (5 wai) with the exception of measurements at 3 and 4 wai. Differences of leaf temperature ranged from 0.4 to 0.8°C at 5 and 2.5 wai respectively (Figure 3).

Leaf temperature showed an irregular time course irrespective whether the plants were inoculated or not, although a similar trend was observed in both situations. Thus, leaf temperature remained fairly constant during the two first weeks until 4 wai, when a clear increase was observed in the inoculated and also in the control leaves of sunflower (0.8 and 0.9°C respectively).

## Assessment of the Infection of *O. cumana* in Sunflower

Parasitized sunflowers reached the same growth stage than the controls. No symptoms (i.e., wilting) due to *O. cumana* were observed in aboveground parts of the plants throughout the experiment. Nonetheless, the infection was confirmed by tubercle formation in roots of inoculated sunflowers at the end of the experiment.

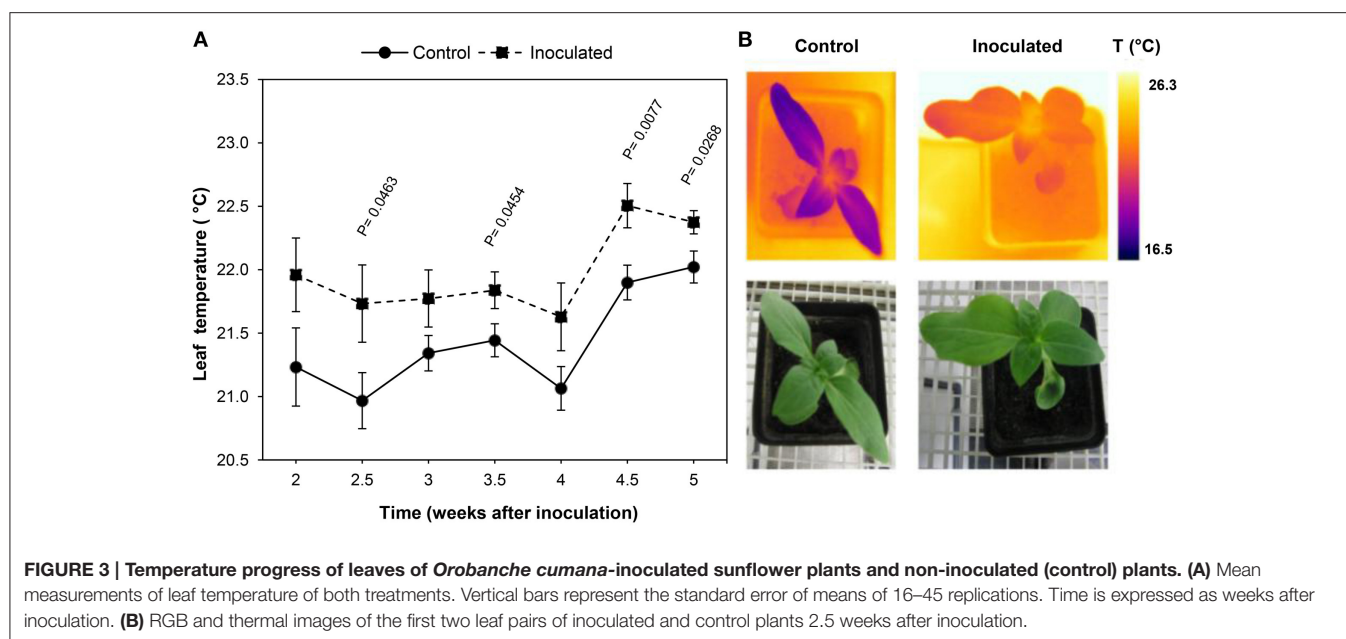
Although no significant differences in root weight were observed between inoculated and control plants, up to 0.29 g and 17 tubercles per plant were recorded in inoculated sunflowers. Tubercles were absent in the controls (Table 2).

## DISCUSSION

Our results showed that the BGF emission of individual LPs of healthy sunflower increases in time. Besides, the BGF emission depends on leaf development: the same pattern of the blue-green signal emitted by a LP (i.e., the first LP) until its complete expansion was presented, 1 week later, by the following LP (i.e., the second LP). An increase of BGF of leaves throughout their elongation was observed in wheat by Meyer et al. (2003), who proposed BGF as a signature of leaf aging. Indeed, aging of leaf implies transition from source to sink tissue by diversion of carbon from primary to secondary metabolism. On the other hand, different BGF emissions in artichoke leaves with age were also observed by Morales et al. (2005), although old leaves emitted lower BGF than young ones. Interestingly, although in most plant species ferulic acid is the major blue-green fluorophore (Morales et al., 1996), in the case of sunflower there is no ferulic acid bound to cell walls. Small amounts of caffeic acid and trichome secretions (e.g., sesquiterpenes lactones and coumarins) have been described as blue-green fluorophores in sunflower (Spring and Schilling, 1989; Olson and Roseland, 1991; Morales et al., 1996; Tourvieille de Labrouhe et al., 1997; Lichtenthaler and Schweiger, 1998; Rowe et al., 2012). These compounds could be responsible for BGF emission in our experiments. In fact, the coumarins content in sunflowers grown under optimal non-limiting conditions tends to be very low, thus their presence can be considered as a good marker of a stress event (Prats-Perez et al., 2000).

Apart from monitoring plant physiological status in response to growth, BGF has been successfully used to detect nutrient deficiencies (McMurtrey et al., 1996; Cadet and Samson, 2011), water and temperature stress (Lang et al., 1996; Hura et al., 2007), pathogen attack (Granum et al., 2015) and a simultaneous combination of stress events (Bürling et al., 2011). Preliminary studies by our group proved the potential of BGF to discriminate *O. cumana* infection in sunflower plants at early stages (3 wai)





**TABLE 2 | Effect of *Orobancha cumana* infection on roots of sunflower and on the development of parasite tubercles 5 weeks after inoculation.**

Treatment	Root fresh weight (g)	Tubercle fresh weight (g/plant)	No. tubercles per sunflower plant
C	1.511 ± 0.265 <sup>a</sup>	0 ± 0.000	0 ± 0.000
I	1.163 ± 0.176	0.2932 ± 0.115	17.38 ± 4.132
<i>p</i> <sup>b</sup>	0.2926		

<sup>a</sup>Mean ± standard error,  $n = 8$ .

<sup>b</sup>Level of significance of differences of root fresh weight between control (C) and inoculated (I) plants obtained after analysis of variance according to a completely randomized statistical design.

(Pérez-Bueno et al., 2014). The present findings allow us to distinguish infected from non-infected sunflowers 1 week earlier by means of decreases in F440 and F520 emissions, as well as by lower values of the F440/F520, F440/F680, and F440/F740 ratios. Spectral features of BGF and their intensity ratios have been investigated not only for their diagnostic value, but also for understanding the physiological changes that take place during stress development (Buschmann and Lichtenthaler, 1998). Our BGF imaging results suggest that *O. cumana* alters the secondary metabolism of sunflower, e.g., accumulation of caffeic acid and coumarins. Furthermore, the spectrophotometric pigment quantification provided evidence of decreased contents of carotenoids in leaves of infected plants as compared to those of the controls in agreement with the results by Shen et al. (2013) on *Mikania micrantha* infected by *Cuscuta campestris*. Many secondary metabolites are not only major contributors to specific odors, tastes, and colors of plants, but they also play a key role in defense against herbivores and pathogens (Berger et al., 2007; Wink, 2010; Ouzounis et al., 2015). Coumarins are excreted by roots of sunflowers with resistance to the parasite (Serghini et al., 2001), as well as other toxic (Zélicourt et al., 2007) and phenolic

(Echevarría-Zomeño et al., 2006) compounds, all of them having a defensive role against *O. cumana*.

Since the F440/F520 ratio is affected by chlorophyll content (Stober and Lichtenthaler, 1992; Morales et al., 1994), decreases upon infection by *O. cumana* suggest that a lower content of chlorophyll is present in leaves. Recent findings by our research group also suggested that, when sunflowers are infected by *O. cumana*, the chlorophyll content in young leaves is decreased (Ortiz-Bustos et al., 2016a). Reductions in chlorophyll content are also induced by *O. foetida* attack on chickpea (*Cicer arietinum* L.) (Nefzi et al., 2016) and by *C. australis* infection on *M. micrantha* (Le et al., 2015). Our results also support the value of the fluorescence ratios F440/F680 and F440/F740 as very early stress indicators (Lichtenthaler and Miehe, 1997; Buschmann and Lichtenthaler, 1998; Buschmann et al., 2000). Low values of these ratios during early stages of the infection of sunflower may not only be due to decreased F440 and F520 emission in leaves, but also to higher F680 and F740 signals (Ortiz-Bustos et al., 2016a). Beyond that, our findings show that BGF and its intensity ratios with chlorophyll bands could be used, in addition to directly detecting *O. cumana* in sunflower, as an indirect approach to the alteration of plant photosynthesis and to the impairment of the secondary metabolism as a result of parasite infection.

The applicability of thermal infrared imaging in determining plant temperature as an early response to biotic or abiotic stresses is widely documented and a topic of hectic research activity (Nilsson, 1995; Chaerle and Van Der Straeten, 2001; Raza et al., 2014; Baranowski et al., 2015; Grant et al., 2016; Mahlein, 2016; Mangus et al., 2016). In this work we observed increases in leaf temperature of parasite infected sunflower from 2 to 5 wai. Recently, increases in canopy temperature allowed the detection of late wilt disease caused by the soil borne fungus *Harpophora maydis* up to 17 days earlier than symptoms development in maize (Ortiz-Bustos et al., 2016b). Also, significant increases in

crown temperature have allowed the differentiation of olive trees infected by *Verticillium dahliae* in the field (Calderón et al., 2013). Leaf thermal increases are closely related to reduced transpiration rates due to either activation of stomatal closure or inhibition of stomatal opening. Previous studies revealed that *Orobancha ramosa* and *C. campestris* parasitism reduced stomatal conductance and transpiration rate and, consequently, slowed down host photosynthesis and host growth (Mauromicale et al., 2008; Chen et al., 2011). The effects of parasite-induced stomatal closure and transpiration reduction on the decreased development of sunflower (Alcántara et al., 2006) should be investigated in the future. To the best of our knowledge, this work constitutes the first approach to the diagnosis of parasite infection in crops by means of thermal imaging and could be further implemented in field conditions.

Although the damage done by broomrape species to crops is directly attributed to parasitic sink activity (Barker et al., 1996; Manschadi et al., 1996; Hibberd et al., 1998; Draie et al., 2011; Péron et al., 2012), our results have evidenced that, in the *O. cumana*—sunflower interaction, the damage might extend beyond assimilate diversion. Many parasitic angiosperms display a pathogenic nature promoting disease in the crop mainly through negative effects on the photosynthesis, physiology and hormonal balance of the host (Stewart and Press, 1990; Watling and Press, 2001; Mauromicale et al., 2008). The present work provides valuable and essential clues toward the understanding of the processes by which *O. cumana* seems not only to cause changes in sunflower secondary metabolism but also to alter its photosynthetic capacity and unbalance its carbohydrate metabolism. Nevertheless, additional research will be required to clarify how both physiological processes are affected.

## CONCLUSION

The outstanding significance of our BGF imaging and thermography results, from a diagnosis point of view, is

that the establishment of a soil borne pathogen (*O. cumana*) in a below-ground organ (root) of the plant can be detected prior to the development of visual symptoms in far distant and above-ground organs (leaves). Diagnostic fluorescence and thermal signals are related to host physiology alterations upon infection, and continuous measurements for long periods of time are possible. Therefore, these techniques enable not only the detection of stress onset by *O. cumana*, but also the monitoring of its development in sunflower over time, providing an additional tool for basic research about holoparasite-host plant interactions. Finally, and as a useful outcome of this work, a fast phenotyping of sunflower lines could be achieved by means of the implementation of BGF imaging and thermography in breeding programmes for resistance to *O. cumana*.

## AUTHOR CONTRIBUTIONS

LM, MP, and MB conceived and designed the experiments. CO, LM, and MP conducted experiments. CO and LM analyzed data and interpreted the results. CO mounted images. LM and MB contributed materials, equipment and analysis tools. CO and LM wrote the manuscript and all the authors reviewed it and approved the final version.

## ACKNOWLEDGMENTS

Financial support for this research was provided by “Consejería de Economía, Innovación y Ciencia” of the Andalusian Government (P12-AGR1281 to LM and P12-AGR370 to MB) and the European Social Fund, and RECUPERA 2020 (grant number 20134R060 to MB) and FEDER Funds. We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).

## REFERENCES

- Alcántara, E., Morales-García, M., and Díaz-Sánchez, J. (2006). Effects of broomrape parasitism on sunflower plants: growth, development, and mineral nutrition. *J. Plant Nutr.* 29, 1199–1206. doi: 10.1080/01904160600767351
- Baranowski, P., Jedryczka, M., Mazurek, W., Babula-Skowronska, D., Siedliska, A., and Kaczmarek, J. (2015). Hyperspectral and thermal imaging of oilseed rape (*Brassica napus*) response to fungal species of the genus *Alternaria*. *PLoS ONE* 10:e0122913. doi: 10.1371/journal.pone.0122913
- Barker, E. R., Press, M. C., Scholes, J. D., and Quick, W. P. (1996). Interactions between the parasitic angiosperm *Orobancha aegyptiaca* and its tomato host: growth and biomass allocation. *New Phytol.* 133, 637–642. doi: 10.1111/j.1469-8137.1996.tb01932.x
- Berger, S., Sinha, A. K., and Roitsch, T. (2007). Plant physiology meets phytopathology: plant primary metabolism and plant-pathogen interactions. *J. Exp. Bot.* 58, 4019–4026. doi: 10.1093/jxb/erm298
- Bürling, K., Hunsche, M., and Noga, G. (2011). Use of blue-green and chlorophyll fluorescence measurements for differentiation between nitrogen deficiency and pathogen infection in winter wheat. *J. Plant Physiol.* 168, 1641–1648. doi: 10.1016/j.jplph.2011.03.016
- Buschmann, C., Langsdorf, G., and Lichtenthaler, H. K. (2000). Imaging of the blue, green and red fluorescence emission of plants: an overview. *Photosynthetica* 38, 483–491. doi: 10.1023/A:1012440903014
- Buschmann, C., and Lichtenthaler, H. K. (1998). Principles and characteristics of multicolour fluorescence imaging of plants. *J. Plant Physiol.* 152, 297–314. doi: 10.1016/S0176-1617(98)80144-2
- Cadet, É., and Samson, G. (2011). Detection and discrimination of nutrient deficiencies in sunflower by blue-green and chlorophyll-*a* fluorescence imaging. *J. Plant Nutr.* 34, 2114–2126. doi: 10.1080/01904167.2011.618572
- Calderón, R., Navas-Cortés, J. A., Lucena, C., and Zarco-Tejada, P. J. (2013). High-resolution airborne hyperspectral and thermal imagery for early detection of *Verticillium* wilt of olive using fluorescence, temperature and narrow-band spectral indices. *Remote Sens. Environ.* 139, 231–245. doi: 10.1016/j.rse.2013.07.031
- Cerovic, Z. G., Samson, G., Morales, F., Tremblay, N., and Moya, I. (1999). Ultraviolet-induced fluorescence for plant monitoring: present state and prospects. *Agronomie* 19, 543–578. doi: 10.1051/agro:19990701
- Chaerle, L., Lenk, S., Hagenbeek, D., Buschmann, C., and Van Der Straeten, D. (2007). Multicolor fluorescence imaging for early detection of the hypersensitive reaction to tobacco mosaic virus. *J. Plant Physiol.* 164, 253–262. doi: 10.1016/j.jplph.2006.01.011

- Chaerle, L., and Van Der Straeten, D. (2000). Imaging techniques and the early detection of plant stress. *Trends Plant Sci.* 5, 495–501. doi: 10.1016/S1360-1385(00)01781-7
- Chaerle, L., and Van Der Straeten, D. (2001). Seeing is believing: imaging techniques to monitor plant health. *Biochim. Biophys. Acta Gene Struct. Expression* 1519, 153–166. doi: 10.1016/S0167-4781(01)00238-X
- Chen, H., Shen, H., Ye, W., Cao, H., and Wang, Z. (2011). Involvement of ABA in reduced photosynthesis and stomatal conductance in *Cuscuta campestris*–*Mikania micrantha* association. *Biol. Plant.* 55, 545–548. doi: 10.1007/s10535-011-0122-7
- Draie, R., Péron, T., Pouvreau, J. B., Véronési, C., Jégou, S., Delavault, P., et al. (2011). Invertases involved in the development of the parasitic plant *Phelipanche ramosa*: characterization of the dominant soluble acid isoform, PrSA11. *Mol. Plant Pathol.* 12, 638–652. doi: 10.1111/J.1364-3703.2010.00702.X
- Eaton, F. M., and Belden, G. O. (1929). Leaf temperatures of cotton and their relation to transpiration, varietal differences, and yields. *U.S.D.A Techn. Bull.* 91, 1–39.
- Echevarría-Zomeño, S., Pérez-de-Luque, A., Jorrín, J., and Maldonado, A. M. (2006). Pre-haustorial resistance to broomrape (*Orobancha cumana*) in sunflower (*Helianthus annuus*): cytochemical studies. *J. Exp. Bot.* 57, 4189–4200. doi: 10.1093/jxb/erl195
- FAOSTAT (2016). Available online at: <http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567&anchor>
- Fernández-Aparicio, M., Reboud, X., and Gibot-Leclerc, S. (2016). Broomrape weeds: underground mechanisms of parasitism and associated strategies for their control: a review. *Front. Plant Sci.* 7:135. doi: 10.3389/fpls.2016.00135
- Fernández-Martínez, J. M., Pérez-Vich, B., and Velasco, L. (2015). “Sunflower broomrape (*Orobancha cumana* Wallr.),” in *Sunflower Oilseed. Chemistry, Production, Processing and Utilization*, eds E. Martínez-Force, N. T. Dunford, and J. J. Salas (Champaign, IL: AOCS Press), 129–156.
- García-Carneros, A. B., Dedic, B., Miladinovic, D., and Molinero-Ruiz, L. (2014). “Pathogenic comparison of highly virulent *O. cumana* affecting sunflower in Moldova, the South of Russian Federation, Serbia and Spain,” in *Proceedings of the Third International Symposium of Orobancha spp. of Sunflower*, ed International Sunflower Association (Paris: International Sunflower Association), 173–177.
- Gitelson, A. A., Buschmann, C., and Lichtenthaler, H. K. (1998). Leaf chlorophyll fluorescence corrected for re-absorption by means of absorption and reflectance measurements. *J. Plant Physiol.* 152, 283–296. doi: 10.1016/S0176-1617(98)80143-0
- Grant, O. M., Ochagavia, H., Baluja, J., Diago, M., and Tardáguila, J. (2016). Thermal imaging to detect spatial and temporal variation in the water status of grapevine (*Vitis vinifera* L.). *J. Horticult. Sci. Biotechnol.* 91, 43–54. doi: 10.1080/14620316.2015.1110991
- Granum, E., Pérez-Bueno, M. L., Calderón, C. E., Ramos, C., de Vicente, A., Cazorla, F. M., et al. (2015). Metabolic responses of avocado plants to stress induced by *Rosellinia necatrix* analysed by fluorescence and thermal imaging. *Eur. J. Plant Pathol.* 142, 625–632. doi: 10.1007/s10658-015-0640-9
- Hibberd, J. M., Quick, W. P., Press, M. C., and Scholes, J. D. (1998). Can source–sink relations explain responses of tobacco to infection by the root holoparasitic angiosperm *Orobancha cernua*? *Plant Cell Environ.* 21, 333–340. doi: 10.1046/j.1365-3040.1998.00272.x
- Hura, T., Hura, K., Grzesiak, M., and Rzepka, A. (2007). Effect of long-term drought stress on leaf gas exchange and fluorescence parameters in C3 and C4 plants. *Acta Physiol. Plant.* 29, 103–113. doi: 10.1007/s11738-006-0013-2
- Jackson, R. D., Reginato, R. J., and Idso, S. B. (1977). Wheat canopy temperature: a practical tool for evaluating water requirements. *Water Res. Res.* 13, 651–656. doi: 10.1029/WR013i003p00651
- Kautz, B., Noga, G., and Hunsche, M. (2014). Controlled long-term water deficiency and its impact on the fluorescence emission of tomato leaves during stress and re-watering. *Eur. J. Horticult. Sci.* 79, 60–69.
- Konanz, S., Kocsányi, L., and Buschmann, C. (2014). Advanced multi-color fluorescence imaging system for detection of biotic and abiotic stresses in leaves. *Agriculture* 4, 79–95. doi: 10.3390/agriculture4020079
- Lang, M., Lichtenthaler, H. K., Sowinska, M., Heisel, F., and Miehe, J. A. (1996). Fluorescence imaging of water and temperature stress in plant leaves. *J. Plant Physiol.* 148, 613–621. doi: 10.1016/S0176-1617(96)80082-4
- Le, Q. V., Tennakoon, K. U., Metali, F., Lim, L. B., and Bolin, J. F. (2015). Impact of *Cuscuta australis* infection on the photosynthesis of the invasive host, *Mikania micrantha*, under drought condition. *Weed Biol. Manag.* 15, 138–146. doi: 10.1111/wbm.12077
- Lenthe, J. H., Oerke, E. C., and Dehne, H. W. (2007). Digital infrared thermography for monitoring canopy health of wheat. *Precis. Agric.* 8, 15–26. doi: 10.1007/s11119-006-9025-6
- Leuzinger, S., and Korner, C. (2007). Tree species diversity affects canopy leaf temperatures in a mature temperate forest. *Agric. For. Meteorol.* 146, 29–37. doi: 10.1016/j.agrformet.2007.05.007
- Lichtenthaler, H. K., and Buschmann, C. (2001). Chlorophylls and carotenoids: measurement and characterization by UV-VIS spectroscopy. *Curr. Protoc. Food Anal. Chem.* (Suppl. 1), F4.3.1–F4.3.8. doi: 10.1002/0471142913.faf0403s01
- Lichtenthaler, H. K., and Miehe, J. A. (1997). Fluorescence imaging as a diagnostic tool for plant stress. *Trends Plant Sci.* 2, 316–320. doi: 10.1016/S1360-1385(97)89954-2
- Lichtenthaler, H. K., and Schweiger, J. (1998). Cell wall bound ferulic acid, the major substance of the blue-green fluorescence emission of plants. *J. Plant Physiol.* 152, 272–282. doi: 10.1016/S0176-1617(98)80142-9
- Lindenthal, M., Steiner, U., Dehne, H. W., and Oerke, E. C. (2005). Effect of downy mildew development on transpiration of cucumber leaves visualized by digital infrared thermography. *Phytopathology* 95, 233–240. doi: 10.1094/PHYTO-95-0233
- Mahlein, A. K. (2016). Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis.* 100, 241–251. doi: 10.1016/S1360-1385(97)89954-2
- Mangus, D. L., Sharda, A., and Zhang, N. (2016). Development and evaluation of thermal infrared imaging system for high spatial and temporal resolution crop water stress monitoring of corn within a greenhouse. *Comput. Electron. Agric.* 121, 149–159. doi: 10.1016/j.compag.2015.12.007
- Manschadi, A. M., Kroschel, J., and Sauerborn, J. (1996). Dry matter production and partitioning in the host-parasite association *Vicia faba*–*Orobancha crenata*. *Angew. Bot.* 70, 224–229.
- Mauromicale, G., Monaco, A. L., and Longo, A. M. (2008). Effect of branched broomrape (*Orobancha ramosa*) infection on the growth and photosynthesis of tomato. *Weed Sci.* 56, 574–581. doi: 10.1614/WS-07-147.1
- McMurtrey, J. E., Chappelle, E. W., Kim, M. S., Corp, L. A., and Daughtry, C. S. (1996). Blue-green fluorescence and visible-infrared reflectance of corn (*Zea mays* L.) grain for *in situ* field detection of nitrogen supply. *J. Plant Physiol.* 148, 509–514. doi: 10.1016/S0176-1617(96)80069-1
- Meyer, S., Cartelat, A., Moya, I., and Cerovic, Z. G. (2003). UV-induced blue-green and far-red fluorescence along wheat leaves: a potential signature of leaf ageing. *J. Exp. Bot.* 54, 757–769. doi: 10.1093/jxb/erg063
- Molinero-Ruiz, L., Delavault, P., Pérez-Vich, B., Pacureanu-Joita, M., Bulos, M., Altieri, E., et al. (2015). History of the race structure of *Orobancha cumana* and the breeding of sunflower for resistance to this parasitic weed: a review. *Span. J. Agric. Res.* 13:e10Re01. doi: 10.5424/sjar/2015134-8080
- Molinero-Ruiz, L., García-Carneros, A. B., Collado-Romero, M., Raranciuc, S., Domínguez, J., and Melero-Vara, J. M. (2014). Pathogenic and molecular diversity in highly virulent populations of the parasitic weed *Orobancha cumana* (sunflower broomrape) from Europe. *Weed Res.* 54, 87–96. doi: 10.1111/wre.12056
- Morales, F., Cartelat, A., Alvarez-Fernández, A., Moya, I., and Cerovic, Z. G. (2005). Time-resolved spectral studies of blue-green fluorescence of artichoke (*Cynara cardunculus* L. var. Scolymus) leaves: identification of chlorogenic acid as one of the major fluorophores and age-mediated changes. *J. Agric. Food Chem.* 53, 9668–9678. doi: 10.1021/jf051842q
- Morales, F., Cerovic, Z. G., and Moya, I. (1994). Characterization of blue-green fluorescence in the mesophyll of sugar beet (*Beta vulgaris* L.) leaves affected by iron deficiency. *Plant Physiol.* 106, 127–133. doi: 10.1104/pp.106.1.127
- Morales, F., Cerovic, Z. G., and Moya, I. (1996). Time resolved blue-green fluorescence of sugar beet (*Beta vulgaris* L.) leaves: spectroscopic evidence for the presence of ferulic acid as the main fluorophore in the epidermis. *BBA Bioenerget.* 1273, 251–262. doi: 10.1016/0005-2728(95)00153-0
- Nefzi, F., Trabelsi, I., Amri, M., Triki, E., Kharrat, M., and Abbes, Z. (2016). Response of some chickpea (*Cicer arietinum* L.) genotypes to *Orobancha foetida* Poir. parasitism. *Chil. J. Agric. Res.* 76, 170–178. doi: 10.4067/S0718-58392016000200006

- Nilsson, H. E. (1995). Remote sensing and image analysis in plant pathology. *Annu. Rev. Phytopathol.* 15, 489–527. doi: 10.1146/annurev.py.33.090195.002421
- Olson, M. M., and Roseland, C. R. (1991). Induction of the coumarins scopoletin and ayapin in sunflower by insect-feeding stress and effects of coumarins on the feeding of sunflower beetle (Coleoptera: Chrysomelidae). *Environ. Entomol.* 20, 1166–1172. doi: 10.1093/ee/20.4.1166
- Ortiz-Bustos, C. M., Pérez-Bueno, M., Barón, M., and Molinero-Ruiz, L. (2016a). Fluorescence imaging in the red and far-red region during growth of sunflower plantlets. Diagnosis of the early infection by the parasite *Orobanche cumana*. *Front. Plant Sci.* 7:884. doi: 10.3389/fpls.2016.00884
- Ortiz-Bustos, C. M., Testi, L., García-Carneros, A. B., and Molinero-Ruiz, L. (2016b). Geographic distribution and aggressiveness of *Harpophora maydis* in the Iberian Peninsula, and thermal detection of maize late wilt. *Eur. J. Plant Pathol.* 144, 383–397. doi: 10.1007/s10658-015-0775-8
- Ouzounis, T., Rosenqvist, E., and Ottosen, C. O. (2015). Spectral effects of artificial light on plant physiology and secondary metabolism: a review. *Hortscience* 8, 1128–1135.
- Pérez-Bueno, M. L., Barón, M., García-Carneros, A. B., and Molinero-Ruiz, L. (2014). Diagnosis of the infection of sunflower by *Orobanche cumana* using multicolour fluorescence imaging. *Helia* 37, 161–174. doi: 10.1515/helia-2014-0015
- Pérez-Bueno, M. L., Pineda, M., Díaz-Casado, E., and Barón, M. (2015). Spatial and temporal dynamics of primary and secondary metabolism in *Phaseolus vulgaris* challenged by *Pseudomonas syringae*. *Physiol. Plant.* 153, 161–174. doi: 10.1111/ppl.12237
- Péron, T., Véronési, C., Mortreau, E., Pouvreau, J. B., Thoiron, S., Leduc, N., et al. (2012). Role of the sucrose synthase encoding PrSus1 gene in the development of the parasitic plant *Phelipanche ramosa* L. (Pomel). *Mol. Plant Microbe Interact.* 25, 402–411. doi: 10.1094/MPMI-10-11-0260
- Prats-Perez, E., Bazzalo, M. E., León, A., and Jorrín-Novo, J. V. (2000). Agronomic aspects of the sunflower 7-hydroxylated simple coumarins. *Helia* 33, 105–112.
- Raza, S. E. A., Smith, H. K., Clarkson, G. J. J., Taylor, G., Thompson, A. J., Clarkson, J., et al. (2014). Automatic detection of regions in spinach canopies responding to soil moisture deficit using combined visible and thermal imagery. *PLoS ONE* 9:e97612. doi: 10.1371/journal.pone.0097612
- Rowe, H. C., Ro, D. K., and Rieseberg, L. (2012). Response of sunflower (*Helianthus annuus* L.) leaf surface defenses to exogenous methyl jasmonate. *PLoS ONE* 7:e37191. doi: 10.1371/journal.pone.0037191
- Schneider, A. A., and Miller, J. F. (1981). Description of sunflower growth stages. *Crop Sci.* 21, 901–903. doi: 10.2135/cropsci1981.0011183X002100060024x
- Serghini, K., Pérez-de-Luque, A., Castejón-Muñoz, M., García-Torres, L., and Jorrín, J. V. (2001). Sunflower (*Helianthus annuus* L.) response to broomrape (*Orobanche cernua* Loefl.) parasitism: induced synthesis and excretion of 7-hydroxylated simple coumarins. *J. Exp. Bot.* 52, 2227–2234. doi: 10.1093/jexbot/52.364.2227
- Shen, H., Xu, S. J., Hong, L., Wang, Z. M., and Ye, W. H. (2013). Growth but not photosynthesis response of a host plant to infection by a holoparasitic plant depends on nitrogen supply. *PLoS ONE* 8:e75555. doi: 10.1371/journal.pone.0075555
- Spring, O., and Schilling, E. E. (1989). Chemosystematic investigation of the annual species of *Helianthus* (Asteraceae). *Biochem. Syst. Ecol.* 17, 519–528. doi: 10.1016/0305-1978(89)90094-X
- Stewart, G. R., and Press, M. C. (1990). The physiology and biochemistry of parasitic angiosperms. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 41, 127–151. doi: 10.1146/annurev.pp.41.060190.001015
- Stober, F., and Lichtenthaler, H. K. (1992). Changes of the laser-induced blue, green and red fluorescence signatures during greening of etiolated leaves of wheat. *J. Plant Physiol.* 140, 673–680. doi: 10.1016/S0176-1617(11)81022-9
- Tourvieille de Labrouhe, D., Mondolot-Cosson, L., Walser, P., Andary, C., and Serieys, H. (1997). Relation entre teneurs en derives cafeoylquiniques des feuilles et la resistance de *Helianthus* spp. a *Sclerotinia sclerotiorum*. *Helia* 20, 39–50.
- Wang, M., Ling, N., Dong, X., Zhu, Y., Shen, Q., and Guo, S. (2012). Thermographic visualization of leaf response in cucumber plants infected with the soil-borne pathogen *Fusarium oxysporum* f. sp. *cucumerinum*. *Plant Physiol. Biochem.* 61, 153–161. doi: 10.1016/j.plaphy.2012.09.015
- Watling, J. R., and Press, M. C. (2001). Impacts of infection by parasitic angiosperms on host photosynthesis. *Plant Biol.* 3, 244–250. doi: 10.1055/s-2001-15195
- Westwood, J. H. (2013). “The physiology of the established parasite-host association” in *Parasitic Orobanchaceae*, eds D. M. Joel, J. Gressel, and L. J. Musselman (Berlin:Springer), 87–114.
- Wink, M. (2010). *Annual Plant Reviews Volume 39: Functions and Biotechnology of Plant Secondary Metabolites*, 2th Edn., ed. M. Wink (Oxford: Wiley-Blackwell).
- Zarco-Tejada, P. J., González-Dugo, V., and Berni, J. A. (2012). Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera. *Remote Sens. Environ.* 117, 322–337. doi: 10.1016/j.rse.2011.10.007
- Zélicourt, A., Letousey, P., Thoiron, S., Campion, C., Simoneau, P., Elmorjani, K., et al. (2007). Ha-DEF1, a sunflower defensin, induces cell death in *Orobanche* parasitic plants. *Planta* 22, 591–600. doi: 10.1007/s00425-007-0507-1

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Ortiz-Bustos, Pérez-Bueno, Barón and Molinero-Ruiz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Genome-Wide Identification of the MIKC-Type MADS-Box Gene Family in *Gossypium hirsutum* L. Unravels Their Roles in Flowering

Zhongying Ren<sup>1,2†</sup>, Daoqian Yu<sup>1,2†</sup>, Zhaoen Yang<sup>1,2</sup>, Changfeng Li<sup>2,3</sup>, Ghulam Qanmber<sup>2</sup>, Yi Li<sup>2</sup>, Jie Li<sup>2</sup>, Zhao Liu<sup>2</sup>, Lili Lu<sup>2</sup>, Lingling Wang<sup>2</sup>, Hua Zhang<sup>1</sup>, Qianjia Chen<sup>1</sup>, Fuguang Li<sup>1,2\*</sup> and Zuoren Yang<sup>1,2\*</sup>

<sup>1</sup> Xinjiang Research Base, State Key Laboratory of Cotton Biology, Xinjiang Agriculture University, Urumqi, China, <sup>2</sup> Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, China, <sup>3</sup> Cotton Research Institute, Anhui Academy of Agricultural Sciences, Hefei, China

## OPEN ACCESS

### Edited by:

Leire Molinero-Ruiz,  
Instituto de Agricultura Sostenible  
(CSIC), Spain

### Reviewed by:

Antonio José Monforte,  
Instituto de Biología Molecular y  
Celular de Plantas (CSIC), Spain  
Ankica Kondic-Spika,  
Institute of Field and Vegetable Crops,  
Serbia

### \*Correspondence:

Zuoren Yang  
yangzuoren4012@163.com  
Fuguang Li  
aylifug@163.com

<sup>†</sup> These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 25 December 2016

**Accepted:** 06 March 2017

**Published:** 22 March 2017

### Citation:

Ren Z, Yu D, Yang Z, Li C,  
Qanmber G, Li Y, Li J, Liu Z, Lu L,  
Wang L, Zhang H, Chen Q, Li F and  
Yang Z (2017) Genome-Wide  
Identification of the MIKC-Type  
MADS-Box Gene Family in *Gossypium*  
*hirsutum* L. Unravels Their Roles in  
Flowering. *Front. Plant Sci.* 8:384.  
doi: 10.3389/fpls.2017.00384

Cotton is one of the major world oil crops. Cottonseed oil meets the increasing demand of fried food, ruminant feed, and renewable bio-fuels. MADS intervening keratin-like and C-terminal (MIKC)-type MADS-box genes encode transcription factors that have crucial roles in various plant developmental processes. Nevertheless, this gene family has not been characterized, nor its functions investigated, in cotton. Here, we performed a comprehensive analysis of MIKC-type MADS genes in the tetraploid *Gossypium hirsutum* L., which is the most widely cultivated cotton species. In total, 110 *GhMIKC* genes were identified and phylogenetically classified into 13 subfamilies. The Flowering locus C (*FLC*) subfamily was absent in the *Gossypium hirsutum* L. genome but is found in *Arabidopsis* and *Vitis vinifera* L. Among the genes, 108 were distributed across the 13A and 12 of the D genome's chromosomes, while two were located in scaffolds. *GhMIKC*s within subfamilies displayed similar exon/intron characteristics and conserved motif compositions. According to RNA-sequencing, most MIKC genes exhibited high flowering-associated expression profiles. A quantitative real-time PCR analysis revealed that some crucial MIKC genes determined the identities of the five flower organs. Furthermore, the overexpression of *GhAGL17.9* in *Arabidopsis* caused an early flowering phenotype. Meanwhile, the expression levels of the flowering-related genes *CONSTANS* (*CO*), *LEAFY* (*LFY*) and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOC1*) were significantly increased in these lines. These results provide useful information for future studies of *GhMIKC*s' regulation of cotton flowering.

**Keywords:** *Gossypium hirsutum* L., *GhMIKC*s, phylogeny, structure, expression patterns, flower

## INTRODUCTION

Transcription factors play an indispensable role in growth and development, and MADS transcription factor family members have been detected in the genomes of plants, animals, and fungi (Becker et al., 2000; Becker and Theissen, 2003; Messenguy and Dubois, 2003). In monophyletic evolution, they are divided into two classes: type I and type II (Alvarez-Buylla et al., 2000). The type I MADS-box genes are serum response factor-like genes in animals and fungi, while

they are M-type genes in plants. They are characterized by a highly conserved MADS domain of the 58–60 amino acids, located in the N-terminal region of the proteins, which are involved in DNA binding and dimerization. Functional investigations have been mainly restricted to Arabidopsis (Parenicová et al., 2003). The type II family plays a significant role in regulating flowering during plant development (Mondragon-Palomino, 2013). Type II genes are closely related to the myocyte enhancer factor-2-like genes of animals and yeast. However, MADS intervening keratin-like and C-terminal (MIKC)-type MADS-box genes are found only in plants.

The MIKC-type plant genes contain three additional domains other than the MADS (M): Intervening (I), Keratin (K), and the C-terminal (C) domains (Theissen et al., 1996; Kaufmann et al., 2005). The I domain forms DNA-binding dimers, which is less conserved (Riechmann et al., 1996). The K domain, which consists of ~70 amino acids, is mainly responsible for dimerization by a coiled-coil structure (Ma et al., 1991; Fan et al., 1997). The C domain exhibits transactivation and mediates protein–protein interactions (Kramer and Irish, 1999; Honma and Goto, 2001). Based on structure divergence at the I domain, the MIKC-type genes are classified into two subgroups, MIKC<sup>C</sup> and MIKC\*. Earlier investigations found 39 and 37 MIKC<sup>C</sup> genes in Arabidopsis and *O. sativa*, respectively (Parenicová et al., 2003; Arora et al., 2007). The MIKC<sup>C</sup> type plays a crucial role in the flowering time, floral organ identity determination and fruit ripening in plant growth and development (Theissen, 2001; Becker and Theissen, 2003; Theissen and Melzer, 2007; Li et al., 2016).

The genetic floral organ model is derived from the analysis of homeotic floral mutants. The ABC model was named after three classes of genes (A, B, and C) (Coen and Meyerowitz, 1991), and has developed into the more exact ABCDE model. MIKC<sup>C</sup> family genes have combined and determined the identities of the floral organs: sepals (A + E), petals (A + B + E), stamens (B + C + E), carpels (C + E), and ovules (D + E) (Bowman et al., 1991; Coen and Meyerowitz, 1991; Ma and Depamphilis, 2000; Zahn et al., 2006; Silva et al., 2015). In Arabidopsis, the functional genes were divided into five classes: Class A: *APETALA1* (*AP1*); Class B: *PISTILATA* (*PI*) and *AP3*; Class C: *AGAMOUS* (*AG*) (Acri-Nunes-Miranda and Mondragon-Palomino, 2014); Class D: *SEEDSTICK/AGAMOUS-LIKE11* (*STK/AGL11*); and Class E: *SEPALLATA* (*SEP1*, *SEP2*, *SEP3*, and *SEP4*) (Ferrándiz et al., 2000; Pinyopich et al., 2003). Other MIKC<sup>C</sup> genes that regulated flowering time and flower initiation have been identified as follows: *Suppressor of Overexpression Of Constans1* (*SOC1*) (Lee et al., 2000; Hepworth et al., 2002); *Flowering Locus c* (*FLC*) (Michaels and Amasino, 1999; Searle et al., 2006; Reeves et al., 2007); *AGAMOUSLIKE GENE 24* (*AGL24*) (Michaels et al., 2003; Liu et al., 2008) and *Short Vegetative Phase* (*SVP*) (Hartmann et al., 2000; Michaels et al., 2003; Lee et al., 2007). Others are involved in fruit ripening, such as *SHATTERPROOF 1–2* and *FUL* (Ferrándiz et al., 2000; Liljegren et al., 2000), in seed pigmentation and endothelium development, such as *TRANSPARENT TESTA16* (Nesi et al., 2002), and in root development such as *AGL12* and *AGL17* (Rounsley et al., 1995; Tapia-López et al., 2008). Studies of the evolutionary history of

MIKC genes have explored the internal mechanisms behind their functional diversification in plant growth and development.

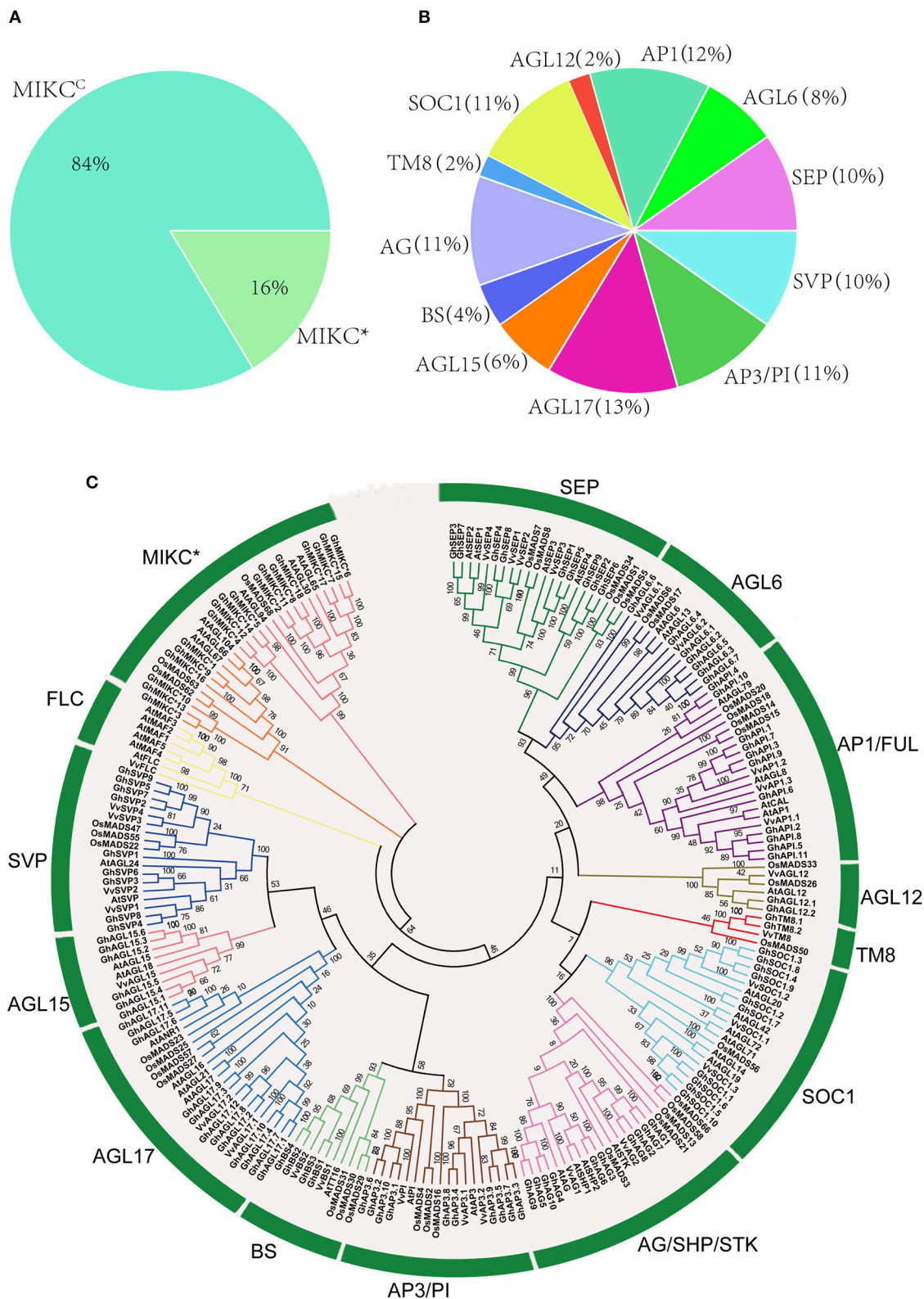
Cotton is not only the most important source of natural fiber for textile industry (Pang et al., 2010), but also a major contributor in world oilseed economy. The extracted cottonseed oil has long been considered to be a good vegetable oil (Michaelk et al., 2010; Sawan, 2014; Zhang et al., 2014). Simultaneously, as an alternative and sustainable oil source, cottonseed oil has been developed into biodiesel and used as substitutes for petroleum (Carlsson, 2009; Alhassan et al., 2014). As the top five oil crops in the world (Wang et al., 2016), cottonseed oil occupies about 21% of the cottonseed production (Malik and Ahsan, 2016; Wang et al., 2016; Yang and Zheng, 2016). The formation of cotton seed originates from ovule which is an important part of floral organs. *G. hirsutum*'s MIKC functions are highly significant in plant developmental processes. Especially, a number of genes could involve in the development of flower morphology (Honma and Goto, 2001; Messenguy and Dubois, 2003). For example, *GhMADS3*, a homolog of Arabidopsis *AG* and putative C function gene, overexpression can improve sepal-to-carpel and petal-to-stamen transformations in transgenic tobacco (Guo et al., 2007). *GhMADS13*, a high homolog of Arabidopsis *AGL6*, overexpression significantly promotes flower buds in cotton (Wu et al., 2009), and *GhMADS14* is enhanced gradually during the early stages of fiber elongation (Zhou et al., 2014). In previous study, 53 members of the *G. hirsutum* MIKC<sup>C</sup> gene family were identified based on the *G. raimondii* genome (Jiang et al., 2014). However, owing to the lack of *G. hirsutum* genome sequences, a comprehensive analysis of MIKC-type MADS genes in *G. hirsutum* has not yet been reported.

Recently, the *G. hirsutum* genome was sequenced. To systematically analyze the MIKC-Type MADS family genes in *G. hirsutum*, 92 MIKC<sup>C</sup>, and 18 MIKC\* members of the MIKC family were identified from the whole *G. hirsutum* genome. Phylogeny, structures, locations and expression patterns were comprehensively analyzed. *AGL17* is the biggest subgroup, and the involvement of *GhAGL17* subfamily gene in regulating flowering was confirmed by ectopic expression in Arabidopsis. Our findings provide a foundation for the genetic improvement of cotton flowering.

## MATERIALS AND METHODS

### Identification of MIKC Genes in *Gossypium hirsutum* L

To identify members of the MIKC gene family in *G. hirsutum*, Arabidopsis MIKC sequences were obtained from the TAIR database (<http://www.arabidopsis.org>) and used as queries for a BLASTP algorithm-based against the *G. hirsutum* genome database ([https://www.cottongen.org/species/Gossypium\\_hirsutum/nbi-AD1\\_genome\\_v1.1](https://www.cottongen.org/species/Gossypium_hirsutum/nbi-AD1_genome_v1.1)) (Zhang et al., 2015). The MIKC protein domain was analyzed using the Hidden Markov Model (HMM) from the Pfam database (<http://pfam.xfam.org/>). The SRF-TF and K-box domains were confirmed by Pfam accessions (PF00319 and PF01486, respectively). All of the candidate proteins were manually



**FIGURE 1 | (A,B).** Classification and percentage of *GhMIKC* genes. **(C).** Neighbor-joining phylogenetic tree using MIKC proteins of *Gossypium hirsutum* L., *Arabidopsis*, *Oryza sativa* L. and *Vitis vinifera* L. Full-length protein sequences were aligned using the MEGA 6.06 program with 1,000 bootstrap replicates. The numbers of the MIKC proteins are listed in Supplementary S1.



checked using the above described methods to remove the redundant sequences.

## Phylogenetic Tree Construction

To construct a MIKC-protein phylogenetic tree using MEGA 6.06, MIKC proteins from four plant species, *Arabidopsis*, *O. sativa*, *V. vinifera*, and *G. hirsutum*, were employed. The neighbor-joining method with amino acid p-distance was applied to construct the tree (Tamura et al., 2011), and the reliability was obtained by bootstrapping with 1,000 replicates.

## Exon/Intron Structure, Motif and Chromosomal Location Analyses

The exon/intron structures of MIKC genes were retrieved by the alignment of predicted coding sequences with corresponding genomic sequences using the gene structure display server (GSDS) program (<http://gsds.cbi.pku.edu.cn/>).

The online program MEME (<http://meme-suite.org/>) was employed to determine the conserved motifs in GhMIKCs with the following optimum parameters: a motif width of 8–200 amino acids and a maximum of 13 motifs. The identified motifs were annotated using the program InterProScan (Quevillon et al., 2005).

The chromosomal distributions of MIKC genes were obtained based on genome annotation data. The MapInspect software was applied to draw images of their physical locations in *G. hirsutum*.

## Gene Expression Analysis

The expression of MIKC family genes were measured using RNA-sequencing method. The raw RNA-sequencing data of *G. hirsutum* TM-1 seven different tissues (root, stem, leaf, flower, ovule, seed, and fiber) was downloaded from the NCBI Gene Expression repository under the accession number PRJNA248163 (Table S4) (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA248163/>). The relative data were normalized to calculate the expression levels. Hierarchical clustering was performed using Genesis 1.7.7 (Sturn et al., 2002).

## RNA Isolation and the qRT-PCR Analysis

*Gossypium hirsutum* L. (cv CCRI24) was cultivated in the field in Zhengzhou, China. Five different tissue parts of flower: sepal, petal, stamen, carpel, and ovule were sampled, respectively at full bloom stage. *Arabidopsis* (Columbia-0) was used as wild type; the leaves of wild type and transgenic lines grown for 25 days were harvested. All samples were frozen immediately in liquid nitrogen and kept at  $-80^{\circ}\text{C}$  for total RNA extraction. Total RNA was extracted from each sample using the TRIzol reagent (TIANGEN, Beijing, China) and treated with RNase-free DNase I. Gel electrophoresis and a Nanodrop2000 nucleic acid analyzer were employed to detect the quality of RNA. The first cDNA strand was synthesized from 1  $\mu\text{g}$  total RNA using the Transcriptor First Strand cDNA Synthesis Kit version DRR047A (TaKaRa, Dalian, China). The cDNA was diluted five times for the next experiments.

The gene-specific primers used for qRT-PCR were listed in Supplementary Table S2 and S3. The *G. hirsutum* *His3* gene and *Arabidopsis* *Actin2* gene were used as an internal control

respectively. The qRT-PCR was performed using SYBR Green (Roche) on a LightCycler480 system (Roche). Each reaction was conducted in a 96-well plate with a volume of 20  $\mu\text{l}$ . The PCR cycling parameters were as follows:  $95^{\circ}\text{C}$  for 5 min, 40 cycles of  $95^{\circ}\text{C}$  for 10 s,  $60^{\circ}\text{C}$  for 10 s, and  $72^{\circ}\text{C}$  for 10 s, followed by an increase from 60 to  $95^{\circ}\text{C}$ . The relative expression levels were analyzed using the LightCycler<sup>®</sup> 480 gene scanning software. Three biological replicates were measured and each biological replicate was run three times.

## Isolation of GhAGL17.9 and Transformation of Arabidopsis

We amplified *GhAGL17.9* using cDNA templates from the mix of CCRI24 root, stem, leaf and flower. The amplified product was cloned into vector *pCambia2301* (CAMBIA) containing the *CAULIFLOWER MOSAIC VIRUS* (CaMV) 35S constitutive promoter, and then, the constructed vector was introduced into *Agrobacterium tumefaciens* GV3101 (Clough and Bent, 1998). Floral dip method was used for *Agrobacterium*-mediated transformation of *Arabidopsis*. Positive transgenic lines were selected on MS medium containing kanamycin. To grow the transgenic lines, seedlings were sown in plastic pots filled with a nutrient soil and vermiculite mix. Then, they were grown in a culturing room at  $22^{\circ}\text{C}$  under a 16-h light/8-h dark cycle for 1 month.

## RESULTS

### Identification of MIKC Genes in *Gossypium hirsutum* L

HMMER and BLASTP algorithm-based searches were used to identify MIKC protein HMM profiles based on the highly conserved MADS and K-box domains. To identify the maximum number of MIKC genes in *G. hirsutum*, HMMs of SRF-TE, and K-box domains (PF00319 and PF01486, respectively) were extracted from Pfam database to use as queries against protein sequences from the *G. hirsutum* genome ([https://www.cottongen.org/species/Gossypium\\_hirsutum/nbi-AD1\\_genome\\_v1.1](https://www.cottongen.org/species/Gossypium_hirsutum/nbi-AD1_genome_v1.1)). A total of 145 putative MIKC proteins were identified. To verify the results, we conducted a multiple sequence alignment and removed 35 redundant sequences. Finally, 110 MIKC protein sequences were identified by confirming their conserved domains using the Pfam web server (Figure S3). From the sequences, 92 MIKC<sup>C</sup> genes and 18 MIKC<sup>\*</sup> genes were identified. Thus, 84% of the MIKC genes were MIKC<sup>C</sup> in *G. hirsutum* (Figure 1A). The identified MIKC genes were listed with their corresponding locus tag (Table 1). We named the MIKC<sup>C</sup> genes on the basis of their assignment to the 13 previously classified *Arabidopsis*, *O. sativa* and *P. tremula* subfamilies (Parenicová et al., 2003; Leseberg et al., 2006; Arora et al., 2007). Subgroup AGL17 had the greatest (13%) number of *GhMIKC<sup>C</sup>* genes; however, subgroups TM8 and AGL12 had the lowest (2%) number of *GhMIKC<sup>C</sup>* genes (Figure 1B). The *GhMIKCs*' encoding amino acids were relatively conserved, and MIKC<sup>C</sup> proteins were highly conserved, ranging from 200 to 300 amino acids in most cases. MIKC<sup>\*</sup> proteins generally possessed more than 300 amino acids. The



**TABLE 1 | MIKC genes identified in *Gossypium hirsutum* L.**

Gene name	Locus ID	Arabidopsis ortholog/locus	ORF length	Introns	Chro	Chromosome location	
GhSVP1	Gh_A01G1089	AT2G22540	142	3	A01	41156433	41165546
GhAP3.1	Gh_A01G1608	AT5G20240	252	7	A01	93811588	93817184
GhAP3.2	Gh_A02G0736	AT5G20240	212	6	A02	13128453	13131418
GhAP3.3	Gh_A02G1617	AT3G54340	181	5	A02	82656247	82659133
GhAGL15.1	Gh_A02G1782	AT5G13790	151	3	A02	36467	38800
GhAPI.1	Gh_A03G0634	AT5G60910	231	7	A03	17989506	17995387
GhSEP1	Gh_A03G1085	AT1G24260	243	7	A03	78292489	78296031
GhSVP2	Gh_A03G1551	AT2G22540	211	6	A03	96584815	96588356
GhAGL17.1	Gh_A03G1563	AT2G14210	235	6	A03	96795941	96813515
GhSOC1.1	Gh_A03G2004	AT4G22950	209	6	A03	123532	129632
GhBS1	Gh_A04G0934	AT5G23260	237	5	A04	58360668	58362528
GhAPI.2	Gh_A04G1264	AT1G69120	208	5	A04	62683429	62686813
GhSEP2	Gh_A04G1265	AT2G03710	240	7	A04	62706600	62709888
GhAG1	Gh_A05G2136	AT4G09960	223	7	A05	24318090	24321470
GhAP3.4	Gh_A05G2191	AT3G54340	225	6	A05	25250408	25253113
GhAG2	Gh_A05G2334	AT4G09960	224	7	A05	28279032	28282673
GhAG3	Gh_A05G3267	AT2G42830	234	6	A05	85617722	85626860
GhAGL17.2	Gh_A06G0244	AT3G57230	287	9	A06	3003556	2987706
GhSVP3	Gh_A06G1875	AT2G22540	222	7	A06	12525	16115
GhAPI.3	Gh_A07G0605	AT5G60910	241	7	A07	8448050	8466601
GhAPI.4	Gh_A07G0722	AT1G26310	237	7	A07	11122961	11127898
GhAGL6.1	Gh_A07G1339	AT2G45650	279	8	A07	33028973	33034276
GhSEP3	Gh_A07G1615	AT3G02310	244	7	A07	63573195	63577955
GhAGL6.2	Gh_A08G1148	AT2G45650	246	7	A08	80735208	80752525
GhAGL15.2	Gh_A08G1275	AT5G13790	251	6	A08	84295318	84298231
GhSEP4	Gh_A09G2157	AT3G02310	247	7	A09	74603850	74608301
GhAG4	Gh_A10G2220	AT4G18960	267	7	A10	5463	9549
GhAG5	Gh_A10G2221	AT4G18960	246	6	A10	15954	26053
GhSOC1.2	Gh_A11G0077	AT5G62165	198	6	A11	733666	740007
GhTM8.1	Gh_A11G0343	AT2G45650	236	6	A11	3157934	3160687
GhAGL17.3	Gh_A11G0462	AT4G37940	194	6	A11	4463913	4467818
GhAGL6.3	Gh_A11G0754	AT2G45650	243	7	A11	7460283	7463442
GhSOC1.3	Gh_A11G0755	AT2G45660	219	6	A11	7469708	7466793
GhAGL17.4	Gh_A12G0150	AT4G37940	235	6	A12	2190001	2216459
GhAP3.5	Gh_A12G0570	AT3G54340	224	6	A12	14074225	14075827
GhSVP4	Gh_A12G0775	AT2G22540	220	6	A12	43064710	43067081
GhAGL15.3	Gh_A12G0910	AT5G13790	254	7	A12	59195226	59199007
GhSOC1.4	Gh_A12G0936	AT2G45660	221	6	A12	59611866	59616247
GhSOC1.5	Gh_A12G2048	AT4G22950	240	7	A12	83409840	83379376
GhAGL17.5	Gh_A13G0423	AT3G57230	241	6	A13	5975824	6005673
GhSVP5	Gh_A13G0442	AT2G22540	210	6	A13	6384500	6389623
GhBS2	Gh_A13G0524	AT5G23260	234	5	A13	12156955	12159494
GhAPI.5	Gh_A13G0751	AT1G69120	245	7	A13	28430363	28435475
GhAGL12.1	Gh_A13G0981	AT1G71692	197	6	A13	54345900	54358603
GhAP3.6	Gh_D02G0779	AT5G20240	267	4	D02	12433523	12436663
GhAPI.6	Gh_D02G1311	AT1G69120	220	6	D02	43248417	43251158
GhSEP5	Gh_D02G1502	AT1G24260	243	7	D02	51916080	51919792
GhAGL17.7	Gh_D02G2012	AT3G57230	235	6	D02	64384346	64401650
GhAP3.7	Gh_D03G0105	AT3G54340	224	6	D03	776473	779332

(Continued)

TABLE 1 | Continued

Gene name	Locus ID	Arabidopsis ortholog/locus	ORF length	Introns	Chro	Chromosome location	
GhAGL15.4	Gh_D03G0626	AT5G13790	151	3	D03	16902671	16900320
GhAPI.7	Gh_D03G0922	AT5G60910	229	7	D03	31494711	31500162
GhSOC1.6	Gh_D03G1493	AT4G22950	209	6	D03	44044063	44050384
GhAG6	Gh_D04G0341	AT2G42830	234	6	D04	5195072	5203980
GhBS3	Gh_D04G1451	AT5G23260	237	5	D04	46216303	46218157
GhAPI.8	Gh_D04G1891	AT1G69120	208	5	D04	51250784	51254164
GhSEP6	Gh_D04G1892	AT2G03710	239	7	D04	51278196	51281401
GhAG7	Gh_D05G2375	AT4G09960	249	6	D05	23662023	23665217
GhAP3.8	Gh_D05G2452	AT3G54340	225	6	D05	24620795	24622467
GhAG8	Gh_D05G2596	AT4G09960	224	7	D05	26719974	26723647
GhAGL17.8	Gh_D06G0245	AT3G57230	278	7	D06	2585610	2602790
GhSVP6	Gh_D06G0267	AT2G22540	222	7	D06	2967937	2971772
GhAPI.9	Gh_D07G0671	AT5G60910	249	7	D07	7915954	7933806
GhAPI.10	Gh_D07G0780	AT5G60910	237	7	D07	9802387	9807245
GhAGL6.4	Gh_D07G1448	AT2G45650	205	6	D07	24514728	24517915
GhSEP7	Gh_D07G1814	AT3G02310	258	7	D07	43509836	43514333
GhAGL6.5	Gh_D08G1430	AT2G45650	246	7	D08	47171916	47179378
GhAGL6.6	Gh_D09G0390	AT2G45650	241	7	D09	14291576	14276476
GhSEP8	Gh_D09G2362	AT3G02310	246	7	D09	50587466	50591473
GhAG9	Gh_D10G0308	AT4G18960	270	6	D10	2675977	2685321
GhAG10	Gh_D10G0309	AT4G18960	267	6	D10	2691566	2695530
GhTM8.2	Gh_D11G0400	AT2G42830	209	7	D11	3355660	3359530
GhAGL17.9	Gh_D11G0534	AT3G57230	217	6	D11	4716974	4720878
GhSOC1.7	Gh_D11G0082	AT5G62165	198	6	D11	761555	756004
GhAGL6.7	Gh_D11G0882	AT2G45650	243	7	D11	7626649	7629794
GhSOC1.8	Gh_D11G0883	AT2G45660	226	6	D11	7646051	7640905
GhAGL15.5	Gh_D11G3150	AT3G57390	253	7	D11	64012119	64007167
GhSVP7	Gh_D12G0156	AT2G22540	217	6	D12	1988003	1993542
GhAGL17.10	Gh_D12G0163	AT4G37940	235	6	D12	2078670	2093157
GhAP3.9	Gh_D12G0585	AT3G54340	371	7	D12	10852076	10878409
GhSVP8	Gh_D12G0778	AT2G22540	220	6	D12	21444908	21447279
GhAGL15.6	Gh_D12G1000	AT5G13790	257	7	D12	35612730	35616501
GhAGL17.11	Gh_D13G0472	AT3G57230	239	6	D13	5578532	5572141
GhSVP9	Gh_D13G0489	AT2G22540	215	6	D13	5951231	5956667
GhBS4	Gh_D13G0605	AT4G18960	223	11	D13	8417153	8434999
GhSEP9	Gh_D13G0877	AT2G03710	244	7	D13	16474813	16478208
GhAPI.11	Gh_D13G0878	AT1G69120	248	7	D13	16637933	16642915
GhAGL12.2	Gh_D13G1226	AT1G71692	197	6	D13	37196257	37208865
GhAGL17.12	Gh_Sca004768G07	AT3G57230	304	7		77172	100229
GhAP3.10	Gh_Sca007246G01	AT5G20240	252	7		1411	8231
GhMIKC*1	Gh_A02G0780	AT1G22130	328	9	A02	15645054	15648072
GhMIKC*2	Gh_A03G0884	AT2G03060	353	9	A03	56790026	56793057
GhMIKC*3	Gh_A05G1797	AT1G22130	308	9	A05	18899832	18897846
GhMIKC*4	Gh_A05G2981	AT1G69540	192	7	A05	73507812	73509545
GhMIKC*5	Gh_A06G0748	AT1G18750	188	6	A06	25502870	25501592
GhMIKC*6	Gh_A07G0593	AT1G18750	380	7	A07	8207711	8211325
GhMIKC*7	Gh_A12G1618	AT1G18750	377	10	A12	77294489	77290589
GhMIKC*8	Gh_A13G0671	AT1G69540	353	8	A13	20300605	20303927
GhMIKC*9	Gh_D02G0829	AT1G77980	336	10	D02	14113941	14116904
GhMIKC*10	Gh_D02G0895	AT1G22130	319	8	D02	17340192	17341980

(Continued)

TABLE 1 | Continued

Gene name	Locus ID	Arabidopsis ortholog/locus	ORF length	Introns	Chro	Chromosome location	
GhMIKC <sup>+</sup> 11	Gh_D02G1268	AT2G03060	357	9	D02	41714661	41717693
GhMIKC <sup>+</sup> 12	Gh_D04G0771	AT1G69540	192	7	D04	15890346	15900308
GhMIKC <sup>+</sup> 13	Gh_D05G1992	AT1G22130	310	9	D05	18333572	18331617
GhMIKC <sup>+</sup> 14	Gh_D06G0878	AT1G18750	188	6	D06	16346541	16347795
GhMIKC <sup>+</sup> 15	Gh_D07G0660	AT1G18750	358	9	D07	7708262	7712331
GhMIKC <sup>+</sup> 16	Gh_D11G2216	AT1G77950	329	9	D11	37092959	37090156
GhMIKC <sup>+</sup> 17	Gh_D12G1758	AT1G18750	352	11	D12	49950909	49947011
GhMIKC <sup>+</sup> 18	Gh_D13G0785	AT1G69540	399	8	D13	13393230	13397325

chromosomal locations of the 108 *GhMIKCs* were distributed in different subgroups of the A and D genomes, while *GhAGL17.12* and *GhAP3.10* were located on scaffolds.

## Phylogenetic Analysis of the MIKC Gene Family

To examine the phylogenetic relationships among *G. hirsutum* MIKC proteins and to categorize them within the established subfamilies from other plants, we performed a multiple alignment analysis using the neighbor-joining method of 110 full-length MIKC proteins from *G. hirsutum*, 44 MIKC proteins from *V. vinifera*, 46 MIKC proteins from Arabidopsis, and 41 MIKC proteins from *O. sativa* (Table S1). The MIKC<sup>C</sup> proteins were divided into 13 subfamilies (SVP, BS, AGL17, AGL15, AP3-PI, AGL12, SOC1, AG/SHP/STK, AP1/FUL, AGL6, SEP, TM8, and FLC; **Figure 1C**). The AGL17 subgroup was the largest, and the FLC subgroup was absent in the *G. hirsutum* genome. Additionally, no TM8 family members were found in Arabidopsis. TM8 constituted the smallest clade, having only four members, including two GhMIKCs, GhTM8.1, and GhTM8.2. The MIKC<sup>\*</sup> proteins were divided into two subfamilies.

## Gene Structure and Protein Motif Analysis

A phylogenetic analysis revealed that our tree corresponded to those reported recently in *V. vinifera* and *C. sativus* (Díaz-Riquelme et al., 2009; Hu and Liu, 2012). The structures of the MIKC genes also helped to determine phylogenetic relationships (**Figure 2**). Most members had significant sequence identities in the same subfamily and similar exon-intron structures, indicating close evolutionary relationships. The most important differences were in the exon-intron lengths (**Figure 2B**). In general, most members contained eight exons in the SEP, AGL6, and AP1 gene families (except *GhAGL6.1*, *GhAGL6.4*, *GhAP1.2*, *GhAP1.6*, and *GhAP1.8*). The SVP (other than *SVP1*) and AGL12 subgroups had seven exons, whereas *GhAGL15.1* and *GhAGL15.4* of the AGL15 subgroup had four exons, which was consistent with *GhSVP1* of the SVP subgroup. The AGL17 genes displayed relatively longer lengths compared with other subgroup genes. Additionally, *GhBS4* had 11 introns and the first exon was meaningfully shorter, while in *GhSOC1.5*, the second of seven introns was longer than the others. The MIKC<sup>\*</sup> had much shorter gene lengths and more introns than the MIKC<sup>C</sup>. *GhMIKC<sup>+</sup>12*

had the fourth longest intron, which distinguished it from other members of the MIKC<sup>\*</sup> family.

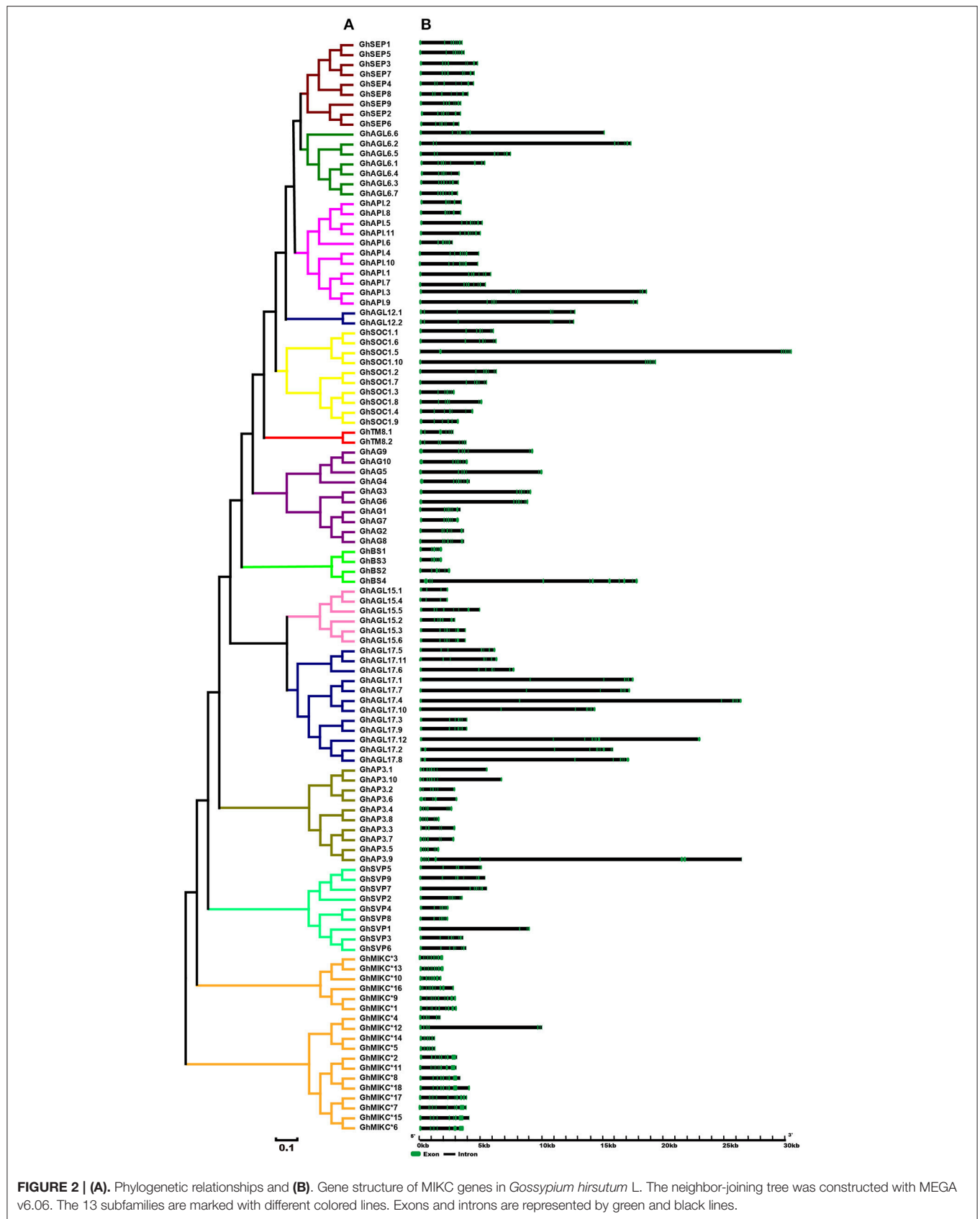
We used MEME to analyze MIKC proteins, and 13 conserved motifs were identified (**Figure 3B**). Most of the closely related MIKC proteins had similar motif type distributions in the same subfamily (**Figure 3A**). The most striking divergence among the subgroups was in the composition of the C-terminal domains. Motif 1 contained the MADS domain in all of the MIKC families, except GhSEP8. The highly conserved sequence logs were showed in Figure S2. The differences between I regions and K-box domains were distinctly shown in the MIKC<sup>C</sup> and MIKC<sup>\*</sup> proteins (**Figure 3B**). The K-box domain contained three motifs, 2, 4, and 9, in GhMIKC<sup>C</sup>. However, motif 4, 5, 8, and 9 were present in the GhMIKC<sup>\*</sup> K-box domain, depending on the lengths. The I region in the MIKC<sup>C</sup> subfamily contained Motifs 3 and 6, while members of the MIKC<sup>\*</sup> contained motifs 6 and 11, which resulted in a longer I region.

## Chromosomes Distributions of GhMIKC Genes

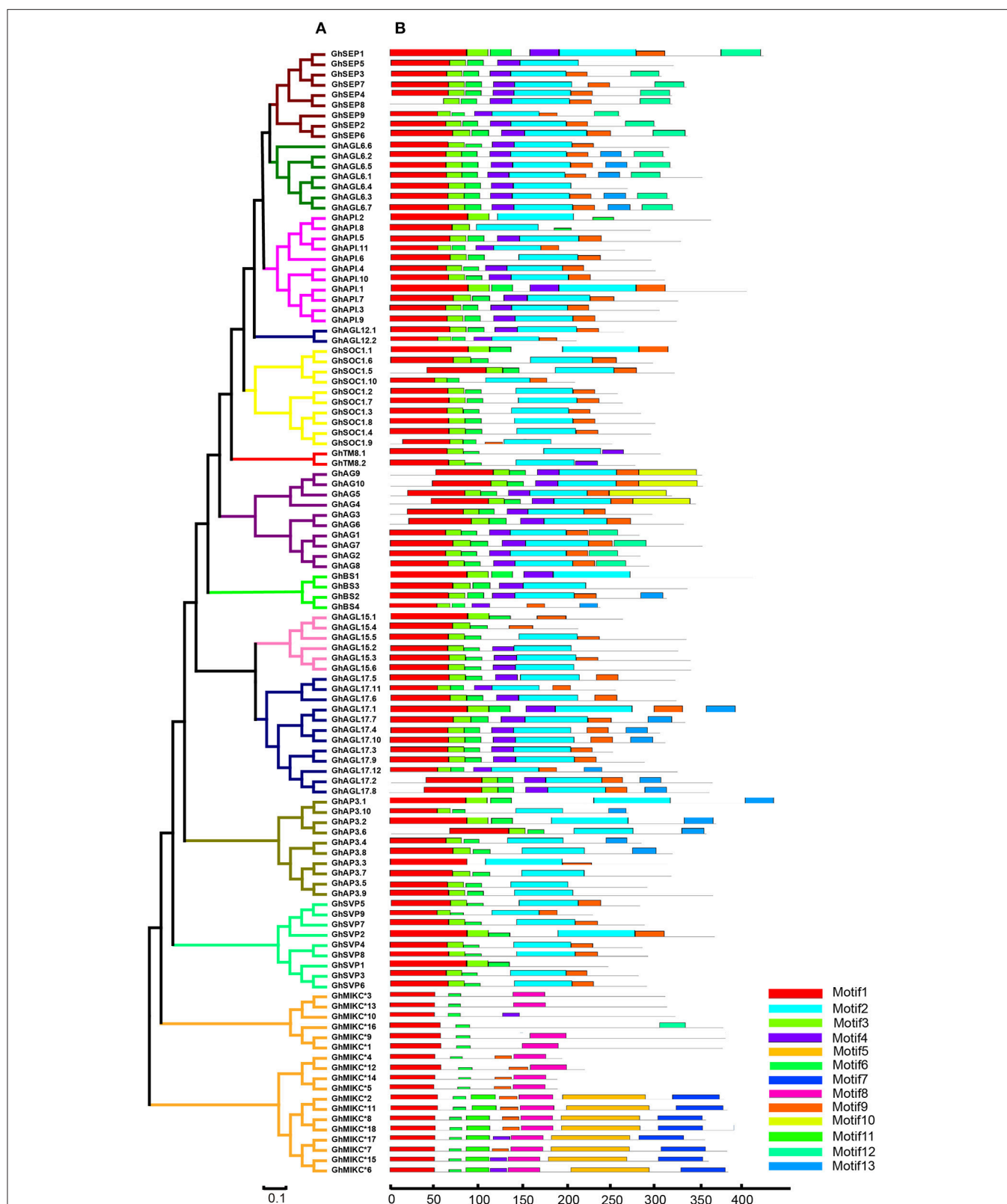
Among the 25 *G. hirsutum* chromosomes, MIKC genes were physically located on all of the 13 A chromosomes and on 12 of the 13 D chromosomes (**Figure 4**). Among the 110 MIKC genes, two genes, *GhAP3.10* and *GhAGL17.12*, could not be distributed on the *G. hirsutum* chromosomes, but were located on unmapped scaffolds (7,246 and 4,768, respectively). The greatest numbers of genes were located on Dt-chr12 (eight genes), followed by Dt-chr2, At-chr12, At-chr13, Dt-chr11, and Dt-chr13 (seven genes on each). In contrast, two genes were located on chromosomes At-chr1, At-chr8, At-chr10, Dt-chr9, and Dt-chr10. Only one gene was mapped on At-chr9 and Dt-chr8, and no genes were located on Dt-chr1.

## Expression Pattern Analyses of MIKC Genes

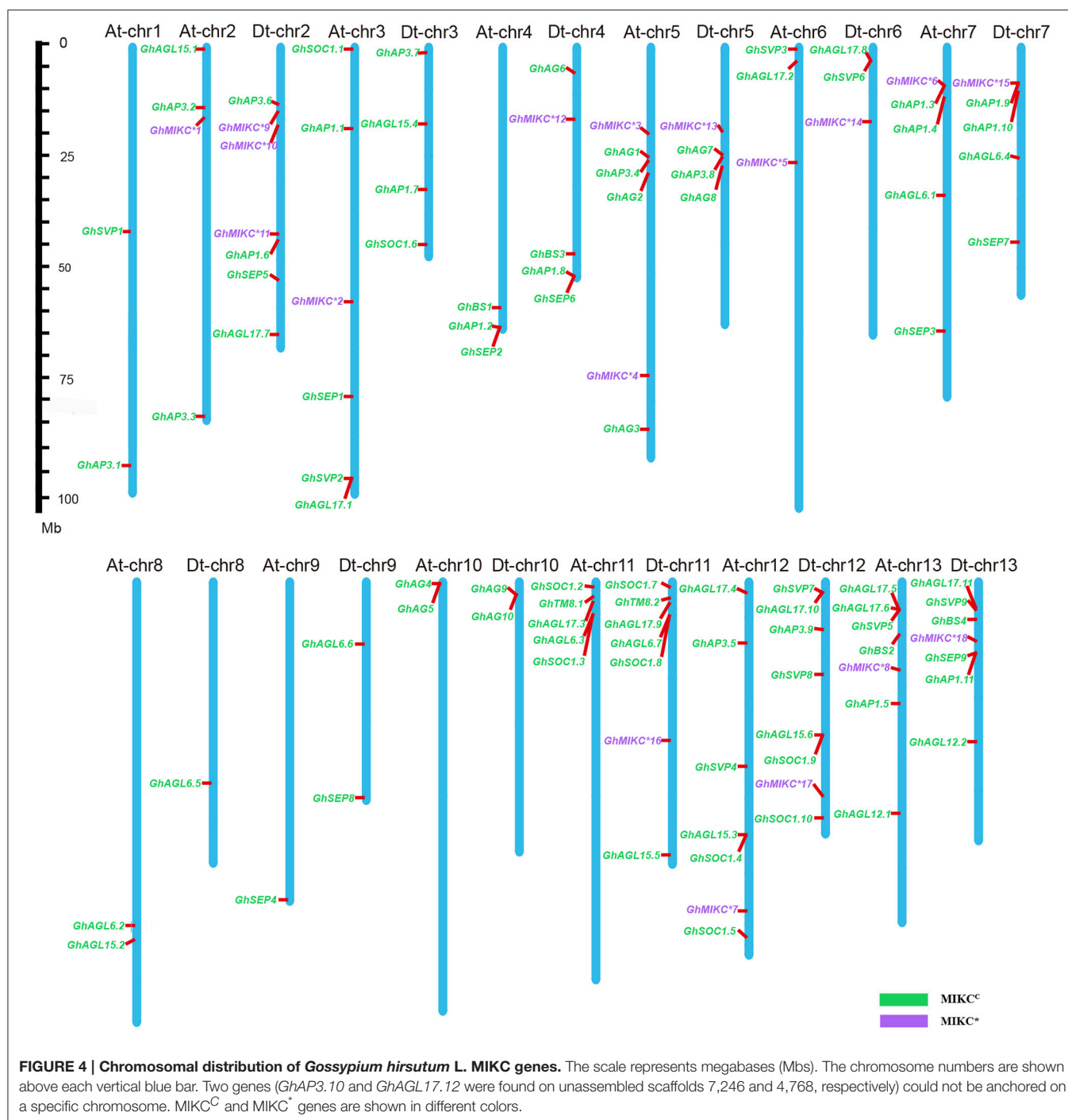
To explore the expression patterns of the MIKC family genes in *G. hirsutum*-specific developmental processes, the 110 genes' expression profiles were detected in seven different tissues (root, stem, leaf, flower, ovule, seed, and fiber) by transcriptome sequencing (**Figure 5**). A heat map showed that different genes shared similar expression patterns within subfamilies. For example, the SEP, AG, AP1, AP3/PI, TM8, and AGL6 subgroups were preferentially expressed in flowers. Similarly, the





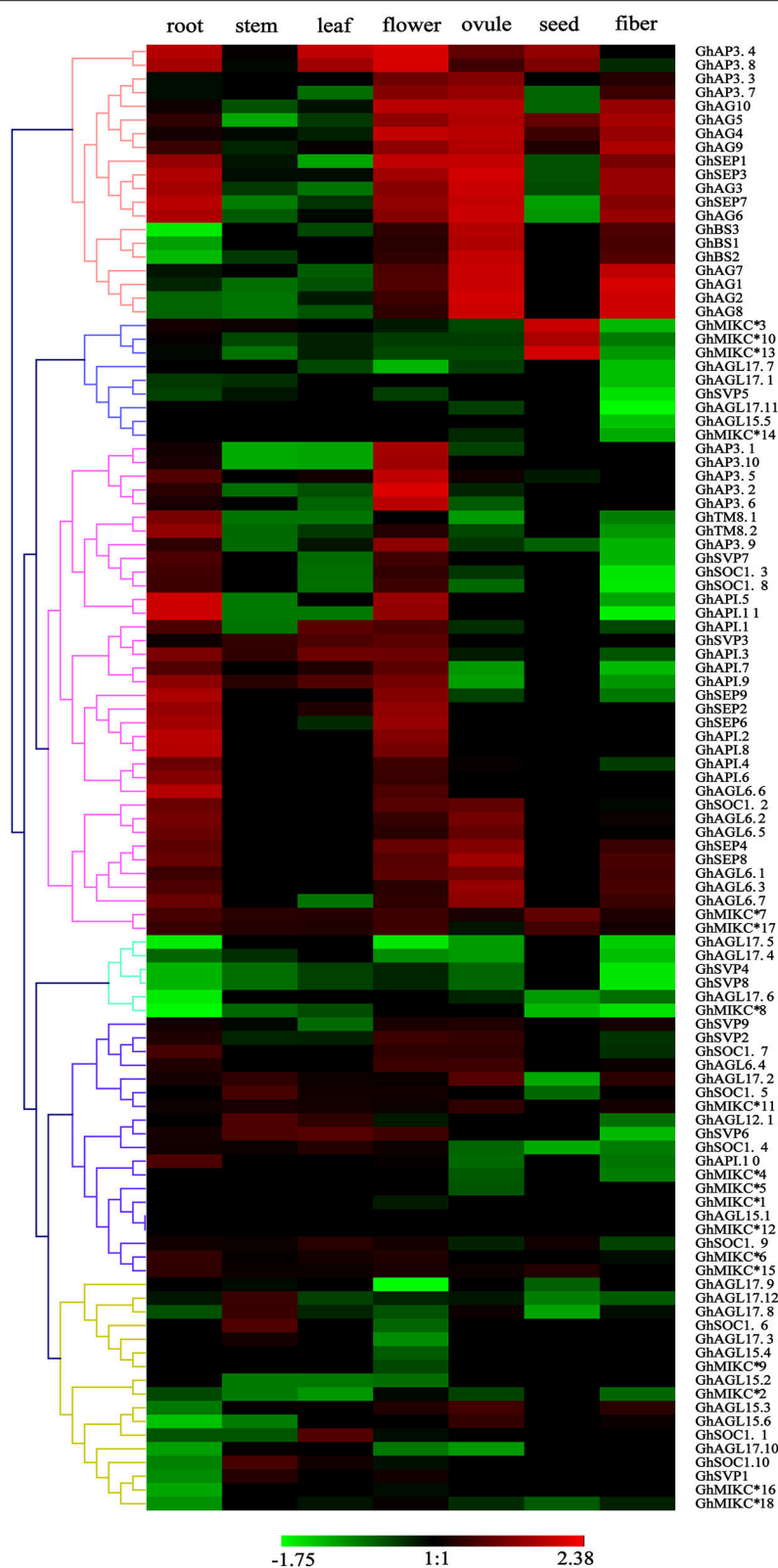


**FIGURE 3 | (A).** Phylogenetic relationships and **(B).** Conserved motifs of GhMIKC proteins. The motif compositions were determined using MEME. Motif 1 contains MADS domain, Motifs 2, 4, 5, 8, and 9 contain K-box domains.



SVP subfamily was expressed especially in flowers. Additionally, some SEP members (*GhSEP1*, *GhSEP3*, *GhSEP4*, *GhSEP7*, and *GhSEP8*), and the AG and BS subgroups, were highly expressed in reproductive organs (ovules and fibers). Simultaneously, SEP, AP1 and five of the AGL6 genes (*GhAGL6.1*, *GhAGL6.2*, *GhAGL6.5*, *GhAGL6.6*, and *GhAGL6.7*) were also detected in roots. Interestingly, the SOC family displayed diverse expression profiles. *GhSOC1.3* and *GhSOC1.7* had high expression levels

in roots. In addition, *GhSOC1.8* was mainly expressed in flower, while *GhSOC1.1*, *GhSOC1.4* and *GhSOC1.9* were highly expressed in leaves. *GhSOC1.6* and *GhSOC1.10* were exclusively and highly expressed in stems. The *GhAGL15* (*GhAGL15.2*, *GhAGL15.4*, and *GhAGL15.5*) and *GhAGL17* (*GhAGL17.7*, *GhAGL17.9*, and *GhAGL17.11*) subfamilies were relatively highly expressed in roots, and *GhAGL17.2*, *GhAGL17.8*, and *GhAGL17.12* were expressed in flowers. Four of the *GhMIKC*<sup>\*</sup>



**FIGURE 5 |** Heat map showing the hierarchical clustering of expression levels of *Gossypium hirsutum* L. MIKC genes in seven different tissues. The relative gene expression data was normalized. Gene names are displayed to the right of each row. Cluster analyses of gene expression levels with different color scales are displayed at the bottom.

genes (*GhMIKC\*6*, *GhMIKC\*7*, *GhMIKC\*17*, and *GhMIKC\*18*) had high expression levels in flowers, and *GhMIKC\*7* and *GhMIKC\*17* were also highly expressed in seeds.

ABCDE model genes regulate the formation of five floral organs in *Arabidopsis* (Sánchez-Fernández et al., 2001; Dietrich et al., 2009; Kuromori, 2010). To validate the participation of MIKC genes in regulating flowering, we selected 16 of ABCDE model orthologous genes to test their expression in five parts of floral organs (sepal, petal, stamen, carpel, and ovule) by qRT-PCR in *G. hirsutum* (Figure 6). *GhAP1.4* and *GhAP1.11* (*A* class) showed high expression levels in sepals, petals, and carpel. Differently, *GhAP1.8* was preferentially expressed in sepal. *GhAP3.5*, *GhAP3.6*, and *GhAP3.8* of the *AP3* subfamily, belonging to *B* class, were expressed in petals and stamens. *GhAG4* of the *C* class displayed the highest expression level in stamen. *GhAG7* and *GhAG8* of the *D* class had higher expression levels in carpel and ovules. *GhSEP1*, *GhSEP4*, and *GhSEP6* (*E* class) were expressed in four different floral organs. *GhBS2* and *GhBS3* (*B* sister class) were mainly expressed in carpel and ovules. *SOC1* accelerates the flowering time, and thus, it is involved in the promotion of floral organ formation. Therefore, high expression levels of *GhSOC1.2* and *GhSOC1.8* were detected in sepals, stamens and carpel. These results were consistent with the ABCDE model.

## Overexpression of the *GhAGL17.9* Gene in *Arabidopsis*

AGL17 is the biggest subgroup (Figure 1B). To further investigate the role of the *GhAGL17* subfamily in plant growth and development, we transformed *GhAGL17.9* into *Arabidopsis* (Columbia-0) driven by the *CAULIFLOWER MOSAIC VIRUS* (CaMV) 35S promoter. We identified 12 T<sub>3</sub> generation transgenic lines that showed an early flowering phenotype. QRT-PCR results confirmed that *GhAGL17.9* was overexpressed in transgenic lines L1 and L3 (Figure 7). Meanwhile, the numbers of rosette leaves were significantly decreased compared with WT (Table 2). To explore the molecular mechanisms that impact the flowering time in transgenic lines, qRT-PCR was used to detect the expression of flowering-related genes in transgenic lines. *LFY* is a flowering integration promoting factor, *AGL17* can positively regulate the expression of *LFY* gene (Han et al., 2008), and *CO* is a photoperiod pathway regulator, *AGL17* acts downstream of *CO* (Han et al., 2008). As shown in Figure 7, the expression levels of *LFY* gene in lines 35S-L1 and 35S-L3 were three times higher than in the wild type. *CO* gene expression was not significantly increased. *SOC1* is a flowering promoter that regulates different signals of the flowering pathways (Lee and Lee, 2010; Ding et al., 2013). The up-regulation of *SOC1* activates downstream targets, including *LFY* and promotes flowering in *Arabidopsis* (Schönrock et al., 2006; Lee et al., 2008). Approximate four-fold increases in *SOC1* expression levels were observed in two transgenic lines.

## DISCUSSION

Cotton, as an oil crop, plays an important role in agriculture and industry all around the world (Houhoula et al., 2003; Waheed

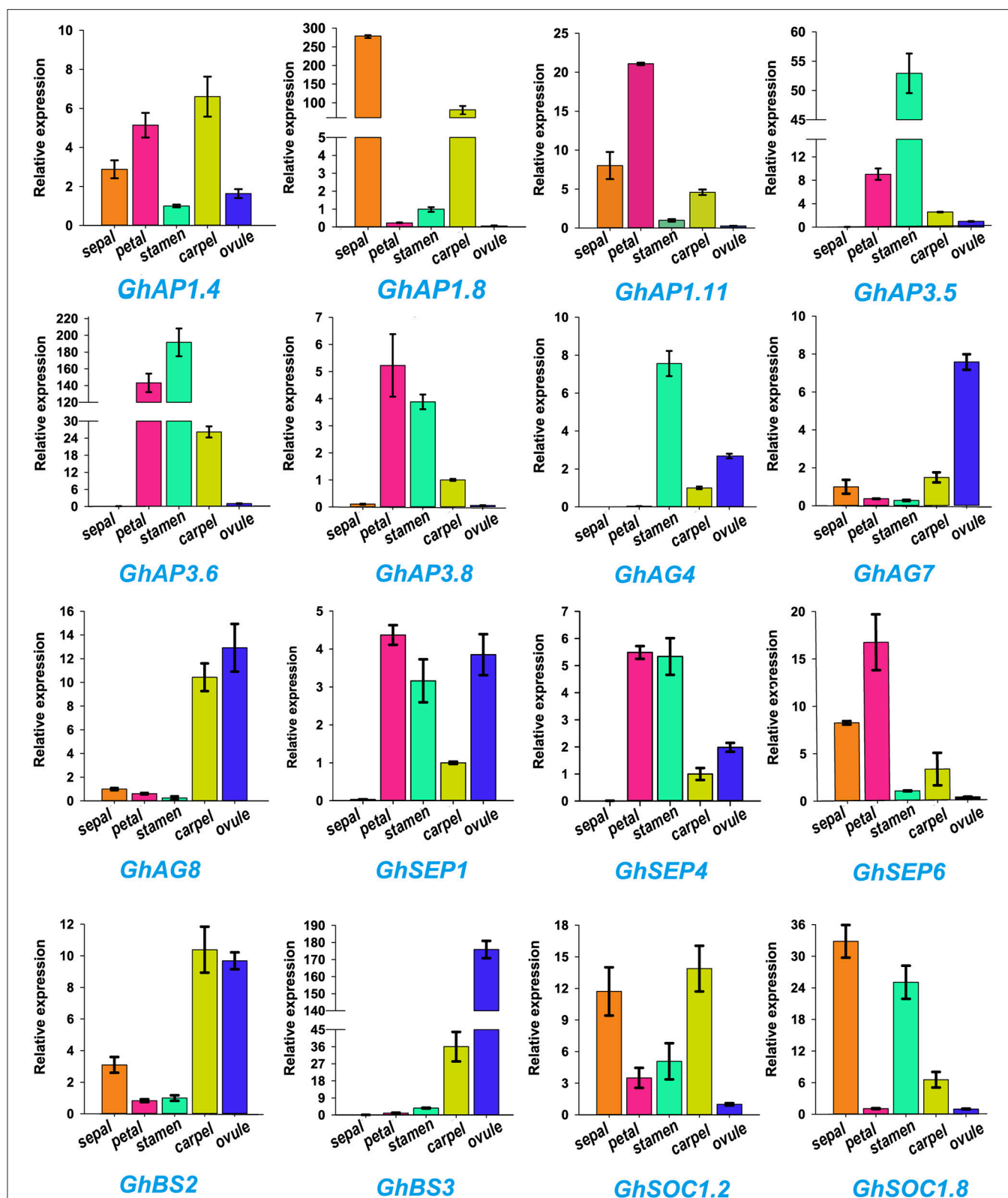
et al., 2010; Mujeli et al., 2016). Floral organs developments affect the yield and quality of cotton seed. The MIKC family members are plant-specific transcription factors containing MADS and K-box domains, and play crucial roles in plant seed development and floral identity (Nesi et al., 2002; De Folter et al., 2006; Mondragon-Palomino and Theissen, 2011). Many MIKC homologs have been analyzed in many plants, including *Arabidopsis*, *O. sativa*, *P. tremula*, *Z. mays*, *S. bicolor*, *B. rapa*, and *R. sativus* (Parenicová et al., 2003; Leseberg et al., 2006; Arora et al., 2007; Zhao et al., 2010; Duan et al., 2015; Li et al., 2016). However, the characterization and functional analysis of the MIKC family has not been performed in *G. hirsutum*, an allotetraploid species. In this study, we performed a comprehensive analysis of *GhMIKCs*, which included investigating chromosomal locations, phylogenetic relationships, gene structures, conserved motifs, and expression profiles in different tissues.

## Overall Summary of the MIKC Family in *Gossypium hirsutum* L

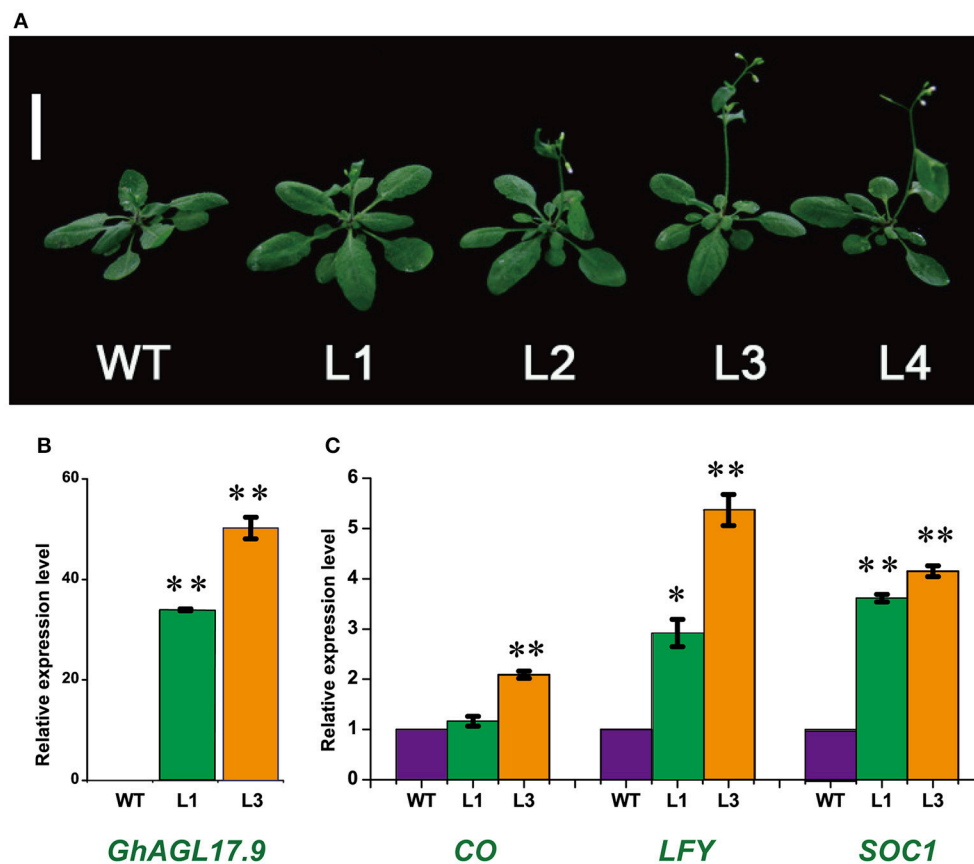
In total, 110 MIKC genes were identified based on *G. hirsutum* genome sequences. Based on phylogenetic relationships with *Arabidopsis* and *O. sativa* orthologs (Figure S1), the *G. hirsutum* type II MADS family (MIKC<sup>C</sup>) was divided into 13 subfamilies (Figure 1C). Interestingly, an *FLC* subfamily was not identified in the *G. hirsutum* genome. Similar results were found in *O. sativa*, *C. sativus*, *Z. mays*, and *S. bicolor* genomes as well (Arora et al., 2007; Zhao et al., 2010; Hu and Liu, 2012). The *FLC* genes are involved in controlling flowering time through the vernalization and autonomous pathways (Helliwell et al., 2006, 2011; Greb et al., 2007). Vernalization is not required for flowering in *O. sativa*, *C. sativus*, *Z. mays*, and *S. bicolor* (Arora et al., 2007; Zhao et al., 2010; Hu and Liu, 2012). Thus, vernalization might not be essential for cotton flowering as well. In addition, we found that in most subgroups, the numbers of proteins in *G. hirsutum* were not doubled, compared with in the diploids *Arabidopsis* and *O. sativa*. This implied that gene duplication could give rise to the amplification of MIKC subfamily genes in a variety of forms (Flagel et al., 2008; Hargreaves et al., 2014). As previously reported, multiple duplications and diversifications in the different clades of different species cause different evolutionary constraints (Lynch and Conery, 2000; Flagel and Wendel, 2009; Airoidi and Davies, 2012).

Chromosomal assignments indicated that the gene locations were equally divided among four pairs of chromosomes (At-chr6 and Dt-chr6, At-chr7 and Dt-chr7, At-chr10 and Dt-chr10, and At-chr13 and Dt-chr13) in A as well as in D genome (Figure 4). However, five D-genome chromosomes (Dt-chr2, Dt-chr4, Dt-chr9, Dt-chr11, and Dt-chr12) contained more genes compared with the corresponding A-genome chromosomes (At-chr2, At-chr4, At-chr9, At-chr11, and At-chr12). Additionally, large numbers of MIKC genes were located on the last three chromosomes (chr11, chr12, and chr13) of both genomes. This could indicate that the current phenomena were derived from differential rates of genomic evolution and inter-genomic hereditary information transfer (Paterson et al., 2000; Wendel and Cronn, 2003).





**FIGURE 6 |** Expression profiles of 16 *Gossypium hirsutum* L. MIKC genes in five different tissues (sepal, petal, stamen, carpel, and ovule) as determined by qRT-PCR. The relative expression levels are shown against the reference gene *His3*. Error bars represent the standard deviations of three independent experiments.



**FIGURE 7 | Phenotypes of transgenic *Arabidopsis* plants overexpressing *GhAGL17.9* under the *Cauliflower mosaic virus* (CaMV) 35S promoter. (A). Morphology of wild type (WT) and transgenic seedlings after 22 days of growth. Bar = 2 cm. (B). A qRT-PCR analysis of *GhAGL17.9* overexpression in WT and transgenic *Arabidopsis*. Significant differences compared with WT (*t*-test):\*\*,  $P < 0.01$ . (C). Expression levels of *SOC1*, *CO*, and *LFY* as determined by qRT-PCR in WT and *GhAGL17.9*-overexpression plants. *Actin2* was used as the internal control. Error bars represent the standard deviations of three independent experiments. Significant differences compared with WT (*t*-test):\*,  $P < 0.05$ ;\*\*,  $P < 0.01$ .**

**TABLE 2 | Flowering time and the leaf numbers of rosette in WT and p35S::GhAGL17.9 plants.**

Genotype	Days to the first open flower	Rosette leaf number	<i>n</i>
WT	25.37 ± 0.61	8.2 ± 0.94	20
L1	24.29 ± 0.55**	7.33 ± 0.9*	20
L2	23.72 ± 0.46**	5.82 ± 0.73**	18
L3	23.83 ± 0.72**	5.88 ± 1.17**	21
L4	23.81 ± 0.782**	5.65 ± 0.7**	16

\*Represents a significant difference from wild type (*t*-test,  $p < 0.05$ );

\*\*Represents a significant difference from wild type (*t*-test,  $p < 0.01$ );

Data are presented as the mean ± SD;

Plants were grown under long-day conditions (16 h of light/8 h of dark).

## Expression Profiles of MIKC Genes in *Gossypium hirsutum* L

Global expression patterns analyses in seven different tissues showed that the *API*, *AP3*, *AG*, *SEP*, and *BS* subfamilies were almost all expressed in the flower development stage (Figure 6). Floral organ identities and flower meristem are

regulated by five kinds of genetic functional genes (A-B-C-D-E) during flower development, from sepals to ovules (Díaz-Riquelme et al., 2009; Na et al., 2014). A qRT-PCR analysis showed the expression patterns of the orthologous genes of the ABCDE model in flower organogenesis (Figure 6), which were consistent with previous findings in *Arabidopsis* (Ó'Maoiléidigh et al., 2014; Xie et al., 2015). Further, the *API* subgroup of *A* class genes were not only expressed in sepals and petals, but also exhibited carpel expression profiles. Before and after pollination, the *API*-like gene may aid in the carpel development in Orchidaceae, which triggered ovary development (Mondragon-Palomino and Theissen, 2011; Aciri-Nunes-Miranda and Mondragón-Palomino, 2014). Thus, *API* subgroup genes may have similar expression patterns in Orchidaceae and allotetraploid cotton. A few *GhMIKC\** genes were highly expressed in flowers and seeds, which was in accordance with previous results in *Arabidopsis* (Verelst et al., 2007) and *O. sativa* (Liu et al., 2013). These results indicated that the expression profiles of *MIKC\** genes were involved in functional redundancy and conservation in the process of *G. hirsutum* evolution.

## Role of the *GhAGL17* Gene in Flowering

In *Arabidopsis*, *AGL17* acts as a novel flowering promoter, which is involved in the photoperiod pathway. Under long-day conditions, the overexpression of *AtAGL17* causes early flowering (Han et al., 2008). As the largest subgroup of the *GhMIKC<sup>C</sup>* family, one member of the *AGL17s*, *GhAGL17.9*, was overexpressed in *Arabidopsis* to explore its biological functions. The transgenic lines displayed earlier flowering than wild type (Figure 7). The expression levels of the related positive marker genes, especially *LFY* and *SOC1*, which are involved in regulating the flowering process, were higher in p35S::*GhAGL17.9* lines than in wild type. *LFY* overexpression can prematurely cause plant development and accelerate blossoming processes (Nilsson et al., 1998; Dornelas and Amaral, 2004). *AGL17* targets *LFY* to promote flowering (Han et al., 2008). *SOC1* encodes a MIKC protein, a floral pathway integrator, which is regulated by a variety of flower signaling pathways (Lee et al., 2000; Wang et al., 2009; Ding et al., 2013). However, the relationship between *AGL17* and *SOC1* in flowering is not clear, which remains to be functionally explored further in the future.

## CONCLUSIONS

In this study, 110 MIKC genes were first identified in the *G. hirsutum* genome. The family was divided into 13 subgroups based on a phylogenetic tree, exon/intron structures, and the distributions of conserved motifs. Chromosomal locations of MIKC gene family members were also determined. Finally, the expression patterns of *GhMIKCs* were explored using transcriptome sequencing and qRT-PCR, which revealed the expression levels at different developmental stages. Most MIKC<sup>C</sup>

genes were highly expressed in the floral organs, which was consistent with the ABCDE model. The overexpression of *GhAGL17.9* in *Arabidopsis* resulted in early flowering through the upregulated expression of *SOC1*, *CO*, and *LFY*, which suggested that *GhMIKCs* play vital roles in cotton flowering. Our work provides functional insights into the roles of *GhMIKC* genes in cotton flowering.

## AUTHOR CONTRIBUTIONS

ZuY and FL conceived and designed the experiments. ZR and DY performed the experiments. ZhY conducted the phylogeny analysis. CL and LL prepared the materials. HZ and QC analyzed the data. ZR and ZuY wrote the paper. GQ, YL, JL, ZL, and LW helped to revise the paper. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 31501345), Zhengzhou Science and Technology Program (153PXXCY180) and Young Elite Scientist Sponsorship Program by CAST. We thank Peng Huo (Zhengzhou Research Center, Institute of Cotton Research of CAAS, Zhengzhou) for technical assistance.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00384/full#supplementary-material>

## REFERENCES

- Aciri-Nunes-Miranda, R., and Mondragón-Palomino, M. (2014). Expression of paralogous SEP-, FUL-, AG- and STK-like MADS-box genes in wild-type and peloric *Phalaenopsis* flowers. *Front. Plant Sci.* 5:76. doi: 10.3389/fpls.2014.00076
- Airoldi, C. A., and Davies, B. (2012). Gene Duplication and the evolution of plant MADS-box transcription factors. *J. Genet. Genomics* 39, 157–165. doi: 10.1016/j.jgg.2012.02.008
- Alhassan, Y., Kumar, N., Bugaje, I. M., Pali, H. S., and Kathkar, P. (2014). Co-solvents transesterification of cotton seed oil into biodiesel: effects of reaction conditions on quality of fatty acids methyl esters. *Energy Conversion Manage.* 84, 640–648. doi: 10.1016/j.enconman.2014.04.080
- Alvarez-Buylla, E. R., Pelaz, S., Liljegren, S. J., Gold, S. E., Burgeff, C., Ditta, G. S., et al. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5328–5333. doi: 10.1073/pnas.97.10.5328
- Arora, R., Agarwal, P., Ray, S., Singh, A. K., Singh, V. P., Tyagi, A. K., et al. (2007). MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics* 8:242. doi: 10.1186/1471-2164-8-242
- Becker, A., and Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* 29, 464–489. doi: 10.1016/S1055-7903(03)00207-0
- Becker, A., Winter, K. U., Meyer, B., Saedler, H., and Theissen, G. (2000). MADS-box gene diversity in seed plants 300 million years ago. *Mol. Biol. Evol.* 17, 1425–1434. doi: 10.1093/oxfordjournals.molbev.a026243
- Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (1991). Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* 112, 1–20.
- Carlsson, A. S. (2009). Plant oils as feedstock alternatives to petroleum - A short survey of potential oil crop platforms. *Biochimie* 91, 665. doi: 10.1016/j.biochi.2009.03.021
- Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J. Cell Mol. Biol.* 16, 735–743.
- Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31–37. doi: 10.1038/353031a0
- De Folter, S., Shchennikova, A. V., Franken, J., Busscher, M., Baskar, R., Grossniklaus, U., et al. (2006). A B sister MADS-box gene involved in ovule and seed development in petunia and *Arabidopsis*. *Plant J.* 47, 934–946. doi: 10.1111/j.1365-3113.2006.02846.x
- Díaz-Riquelme, J., Lijavetzky, D., Martínez-Zapater, J. M., and Carmona, M. J. (2009). Genome-wide analysis of MIKCC-type MADS box genes in grapevine. *Plant Physiol.* 149, 354–369. doi: 10.1104/pp.108.131052
- Dietrich, D., Schmuths, H., De Marcos Lousa, C., Baldwin, J. M., Baldwin, S. A., Baker, A., et al. (2009). Mutations in the *Arabidopsis* Peroxisomal ABC transporter COMATOSE allow differentiation between multiple functions in planta: insights from an Allelic Series. *Mol. Biol. Cell* 20, 530–543. doi: 10.1091/mbc.E08-07-0745
- Ding, L., Wang, Y., and Yu, H. (2013). Overexpression of DOSOC1, an ortholog of *Arabidopsis* SOC1, promotes flowering in the Orchid *dendrobium chao parya* smile. *Plant Cell Physiol.* 54, 595–608. doi: 10.1093/pcp/pct026

- Dornelas, M. C., and Amaral, W. A. N. D. (2004). EgLFY, the *Eucalyptus grandis* homolog of the Arabidopsis gene LEAFY is expressed in reproductive and vegetative tissues. *Braz. J. Plant Physiol.* 16, 105–114. doi: 10.1590/S1677-04202004000200006
- Duan, W., Song, X., Liu, T., Huang, Z., Ren, J., Hou, X., et al. (2015). Genome-wide analysis of the MADS-box gene family in Brassica rapa (Chinese cabbage). *Mol. Genet. Genomics* 290, 239–255. doi: 10.1007/s00438-014-0912-7
- Fan, H. Y., Hu, Y., Tudor, M., and Ma, H. (1997). Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. *Plant J.* 12, 999–1010. doi: 10.1046/j.1365-313X.1997.12050999.x
- Ferrández, C., Liljgren, S. J., and Yanofsky, M. F. (2000). Negative regulation of the SHATTERPROOF genes by FRUITFULL during Arabidopsis fruit development. *Science* 289, 436–438. doi: 10.1126/science.289.5478.436
- Flagel, L., Udall, J., Dan, N., and Wendel, J. (2008). Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol.* 6:16. doi: 10.1186/1741-7007-6-16
- Flagel, L. E., and Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytol.* 183, 557–564. doi: 10.1111/j.1469-8137.2009.02923.x
- Greb, T., Mylne, J. S., Crevillen, P., Geraldo, N., An, H., Gendall, A. R., et al. (2007). The PHD finger protein VRN5 functions in the epigenetic silencing of Arabidopsis FLC. *Curr. Biol.* 17, 73–78. doi: 10.1016/j.cub.2006.11.052
- Guo, Y., Zhu, Q., Zheng, S., and Li, M. (2007). Cloning of a MADS Box Gene (GhMADS3) from cotton and analysis of its homeotic role in transgenic tobacco. *Acta Genet. Sin.* 34, 527–535. doi: 10.1016/s1673-8527(07)60058-7
- Han, P., García-Ponce, B., Fonseca-Salazar, G., Alvarez-Buylla, E. R., and Yu, H. (2008). AGAMOUS-LIKE 17, a novel flowering promoter, acts in a FT-independent photoperiod pathway. *Plant J.* 55, 253–265. doi: 10.1111/j.1365-313X.2008.03499.x
- Hargreaves, A. D., Swain, M. T., Hegarty, M. J., Logan, D. W., and Mulley, J. F. (2014). Restriction and recruitment—gene duplication and the origin and evolution of snake venom toxins. *Genome Biol. Evol.* 6, 2088. doi: 10.1093/gbe/evu166
- Hartmann, U., Höhmann, S., Nettesheim, K., Wisman, E., Saedler, H., and Huijser, P. (2000). Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis. *Plant J.* 21, 351–360. doi: 10.1046/j.1365-313x.2000.00682.x
- Helliwell, C. A., Robertson, M., Finnegan, E. J., Buzas, D. M., and Dennis, E. S. (2011). Vernalization-repression of Arabidopsis FLC requires promoter sequences but not antisense transcripts. *PLoS ONE* 6, 240–247. doi: 10.1371/journal.pone.0021513
- Helliwell, C. A., Wood, C. C., Robertson, M., Peacock, W., and Dennis, E. S. (2006). The Arabidopsis FLC protein interacts directly in vivo with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex. *Plant J.* 46, 183–192. doi: 10.1111/j.1365-313X.2006.02686.x
- Hepworth, S. R., Valverde, F., Ravenscroft, D., Mouradov, A., and Coupland, G. (2002). Antagonistic regulation of flowering-time gene SOC1 by CONSTANS and FLC via separate promoter motifs. *Embo J.* 21, 4327–4337. doi: 10.1093/emboj/cdf432
- Honma, T., and Goto, K. (2001). Complexes of MADS-BOX proteins are sufficient to convert leaves into floral organs. *Nature* 409, 525–529. doi: 10.1038/35054083
- Houhoula, D. P., Oreopoulou, V., and Tzia, C. (2003). The effect of process time and temperature on the accumulation of polar compounds in cottonseed oil during deep-fat frying. *J. Sci. Food Agric.* 83, 314–319. doi: 10.1002/jsfa.1314
- Hu, L., and Liu, S. (2012). Genome-wide analysis of the MADS-box gene family in cucumber. *Genome* 55, 245–256. doi: 10.1139/g2012-009
- Jiang, S.-C., Pang, C.-Y., Song, M.-Z., Wei, H.-L., Fan, S.-L., and Yu, S.-X. (2014). Analysis of MIKC<sup>C</sup>-Type MADS-box gene family in *Gossypium hirsutum*. *J. Integr. Agric.* 13, 1239–1249. doi: 10.1016/S2095-3119(13)60520-1
- Kaufmann, K., Melzer, R., and Theissen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* 347, 183–198. doi: 10.1016/j.gene.2004.12.014
- Kramer, E. M., and Irish, V. F. (1999). Evolution of genetic mechanisms controlling petal development. *Nature* 399, 144–148. doi: 10.1038/20172
- Kuromori, T. (2010). AbA transport factors found in Arabidopsis ABC transporters. *Plant Signal. Behav.* 5, 1124–1126. doi: 10.4161/psb.5.9.12566
- Lee, H., Suh, S. S., Park, E., Cho, E., Ahn J. H., Kim, S. G., et al. (2000). The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis. *Genes Dev.* 14, 2366–2376. doi: 10.1101/gad.813600
- Lee, J. H., Yoo, S. J., Park, S. H., Hwang, I., Lee, J. S., and Ahn, J. H. (2007). Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. *Genes Dev.* 21, 397–402. doi: 10.1101/gad.1518407
- Lee, J., and Lee, I. (2010). Regulation and function of SOC1, a flowering pathway integrator. *J. Exp. Bot.* 61, 2247–2254. doi: 10.1093/jxb/erq098
- Lee, J., Oh, M., Park, H., and Lee, I. (2008). SOC1 translocated to the nucleus by interaction with AGL24 directly regulates LEAFY. *Plant J. Cell Mol. Biol.* 55, 832–843. doi: 10.1111/j.1365-313X.2008.03552.x
- Leseberg, C. H., Li, A., Kang, H., Duvall, M., and Mao, L. (2006). Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*. *Gene* 378, 84–94. doi: 10.1016/j.gene.2006.05.022
- Li, C., Wang, Y., Xu, L., Nie, S., Chen, Y., Liang, D., et al. (2016). Genome-wide characterization of the MADS-Box gene family in Radish (*Raphanus sativus* L.) and assessment of its roles in flowering and floral organogenesis. *Front. Plant Sci.* 7:1390. doi: 10.3389/fpls.2016.01390
- Liljgren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L., and Yanofsky, M. F. (2000). SHATTERPROOF MADS-box genes control seed dispersal in Arabidopsis. *Nature* 404, 766–770. doi: 10.1038/35008089
- Liu, C., Chen, H., Er, H. L., Soo, H. M., Kumar, P. P., Han, J. H., et al. (2008). Direct interaction of AGL24 and SOC1 integrates flowering signals in Arabidopsis. *Development* 135, 1481–1491. doi: 10.1242/dev.020255
- Liu, Y., Cui, S., Wu, F., Yan, S., Lin, X., Du, X., et al. (2013). Functional conservation of MIKC<sup>C</sup>-Type MADS box genes in Arabidopsis and rice pollen maturation. *Plant Cell* 25, 1288–1303. doi: 10.1105/tpc.113.110049
- Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151
- Ma, H., and Depamphilis, C. (2000). The ABCs of floral evolution. *Cell* 101, 5–8. doi: 10.1016/S0092-8674(00)80618-2
- Ma, H., Yanofsky, M. F., and Meyerowitz, E. M. (1991). AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev.* 5, 484–495. doi: 10.1101/gad.5.3.484
- Malik, T. H., and Ahsan, M. Z. (2016). Review of the Cotton Market in Pakistan and its future prospects. *Oilseeds Fats Crops and Lipids* 23:D606. doi: 10.1051/ocl/2016043
- Messenguy, F., and Dubois, E. (2003). Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Genes* 316, 1–21. doi: 10.1016/S0378-1119(03)00747-9
- Michaelk, D., Deborahl, B., Williamrjr, M., Btodt, C., Fredm, B., Johnr, G., et al. (2010). Fatty acid profiles of cottonseed genotypes from the national cotton variety trials. *J. Cotton Sci.* 14, 64–73.
- Michaels, S. D., and Amasino, R. M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11, 949–956. doi: 10.1105/tpc.11.5.949
- Michaels, S. D., Ditta, G., Gustafson-Brown C., Pelaz S., Yanofsky, M., and Amasino, R. M. (2003). AGL24 acts as a promoter of flowering in Arabidopsis and is positively regulated by vernalization. *Plant J.* 33, 867–874. doi: 10.1046/j.1365-313X.2003.01671.x
- Mondragón-Palomino, M. (2013). Perspectives on MADS-box expression during orchid flower evolution and development. *Front. Plant Sci.* 4:377. doi: 10.3389/fpls.2013.00377
- Mondragon-Palomino, M., and Theissen, G. (2011). Conserved differential expression of paralogous DEFICIENS- and GLOBOSA-like MADS-box genes in the flowers of Orchidaceae: refining the “orchid code”. *Plant J.* 66, 1008–1019. doi: 10.1111/j.1365-313X.2011.04560.x
- Mujeli, M., Kefas, H. M., Shitu, A., and Ayuba, I. (2016). Optimization of biodiesel production from crude cotton seed oil using central composite design. *Am. J. Chem. Biochem. Eng.* 1, 8–14. doi: 10.11648/j.ajcbe.20160101.12
- Na, L., Liu, Y., Ming, Z., Min, J., and Li, H. (2014). Thinking out of the box: MADS-box genes and maize spikelet development. *Afr. J. Biotechnol.* 13, 4673–4679. doi: 10.5897/AJB11.3885
- Nesi, N., Debeaujon, I., Jond, C., Stewart, A. J., Jenkins, G. I., Caboche, M., et al. (2002). The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell* 14, 2463–2479. doi: 10.1105/tpc.004127
- Nilsson, O., Lee, I., Blázquez, M. A., and Weigel, D. (1998). Flowering-time genes modulate the response to LEAFY activity. *Genetics* 150, 403–410.



- Ó'Maoiléidigh, D. S., Graciet, E., and Wellmer, F. (2014). Gene networks controlling *Arabidopsis thaliana* flower development. *New Phytol.* 201, 16–30. doi: 10.1111/nph.12444
- Pang, C. Y., Wang, H., Pang, Y., Xu, C., Jiao, Y., Qin, Y. M., et al. (2010). Comparative proteomics indicates that biosynthesis of pectic precursors is important for cotton fiber and *Arabidopsis* root hair elongation. *Mol. Cell. Proteomics* 9, 2019. doi: 10.1074/mcp.M110.000349
- Parentová, L., de Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., et al. (2003). Molecular and phylogenetic analyses of the complete MADS-Box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell Online* 15, 1538–1551. doi: 10.1105/tpc.011544
- Paterson, A. H., Bowers, J. E., Burow, M. D., Draye, X., Elsik, C. G., Jiang, C. X., et al. (2000). Comparative genomics of plant chromosomes. *Plant Cell* 12, 1523–1540. doi: 10.1105/tpc.12.9.1523
- Pinyopich, A., Ditta, G. S., Savidge, B., Liljgren, S. J., Baumann, E., Wisman, E., et al. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* 424, 85–88. doi: 10.1038/nature01741
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et al. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.* 33, W116–W120. doi: 10.1093/nar/gki442
- Reeves, P. A., He, Y., Schmitz, R. J., Amasino, R. M., Panella, L. W., and Richards, C. M. (2007). Evolutionary conservation of the FLOWERING LOCUS C-mediated vernalization response: evidence from the sugar beet (*Beta vulgaris*). *Genetics* 176, 295–307. doi: 10.1534/genetics.106.069336
- Riechmann, J. L., Krizek, B. A., and Meyerowitz, E. M. (1996). Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc. Natl. Acad. Sci. U.S.A.* 93, 4793–4798. doi: 10.1073/pnas.93.10.4793
- Rounsley, S. D., Ditta, G. S., and Yanofsky, M. F. (1995). Diverse roles for MADS box genes in *Arabidopsis* development. *Plant Cell* 7, 1259–1269. doi: 10.1105/tpc.7.8.1259
- Sturn, A., Quackenbush, J., and Trajanoski, Z. (2002). Cluster analysis of microarray data. *Bioinformatics* 18, 207–208. doi: 10.1093/bioinformatics/18.1.207
- Sánchez-Fernández, R., Davies, T. G., Coleman, J. O., and Rea, P. A. (2001). The *Arabidopsis thaliana* ABC protein superfamily, a complete inventory. *J. Biol. Chem.* 276, 30231–30244. doi: 10.1074/jbc.M103104200
- Sawan, Z. M. (2014). Cottonseed yield and its quality as affected by mineral fertilizers and plant growth retardants. *Agric. Sci.* 05, 186–209. doi: 10.4236/as.2014.53023
- Schönrock, N., Bouveret, R., Leroy, O., Borghi, L., Köhler, C., Grissem, W., et al. (2006). Polycomb-group proteins repress the floral activator AGL19 in the FLC-independent vernalization pathway. *Genes Dev.* 20, 1667–1678. doi: 10.1101/gad.377206
- Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Kröber, S., et al. (2006). The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. *Genes Dev.* 20, 898–912. doi: 10.1101/gad.373506
- Silva, C. S., Puranik, S., Round, A., Brennich, M., Jourdain, A., Parcy, F., et al. (2015). Evolution of the Plant Reproduction Master Regulators, L. F. Y., and the MADS Transcription Factors: The Role of Protein Structure in the Evolutionary Development of the Flower. *Front. Plant Sci.* 6:1193. doi: 10.3389/fpls.2015.01193
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Tapia-López, R., García-Ponce, B., Dubrovsky, J. G., Garay-Arroyo, A., Pérez-Ruiz, R. V., Kim, S. H., et al. (2008). An AGAMOUS-related MADS-box gene, XAL1 (AGL12), regulates root meristem cell proliferation and flowering transition in *Arabidopsis*. *Plant Physiol.* 146, 1182–1192. doi: 10.1104/pp.107.108647
- Theissen, G. (2001). Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol.* 4, 75–85. doi: 10.1016/S1369-5266(00)00139-4
- Theissen, G., Kim, J. T., and Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J. Mol. Evol.* 43, 484–516. doi: 10.1007/BF02337521
- Theissen, G., and Melzer, R. (2007). Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Ann. Bot.* 100, 603–619. doi: 10.1093/aob/mcm143
- Verelst, W., Saedler, H., and Münster, T. (2007). MIKC\* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific *Arabidopsis* promoters. *Plant Physiol.* 143, 447–460. doi: 10.1104/pp.106.089805
- Waheed, A., Rasool, G., and Asghar, A. (2010). Effect of interesterified palm and cottonseed oil blends on cookie quality. *Agric. Biol. J. North Am.* 1, 402–406. doi: 10.5251/abjna.2010.1.3.402.406
- Wang, J. W., Czech, B., and Weigel, D. (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell* 138, 738–749. doi: 10.1016/j.cell.2009.06.014
- Wang, Q., Zhu, Y., Sun, L., Li, L., Jin, S., and Zhang, X. (2016). Transgenic Bt cotton driven by the green tissue-specific promoter shows strong toxicity to lepidopteran pests and lower Bt toxin accumulation in seeds. *Sci. China Life Sci.* 59, 172–182. doi: 10.1007/s11427-015-4920-6
- Wendel, J. F., and Cronn, R. C. (2003). Polyploidy and the evolutionary history of cotton. *Adv. Agron.* 78, 139–186. doi: 10.1016/S0065-2113(02)78004-8
- Wu, D., Yu, S. X., Fan, S. L., Song, M. Z., and Wang, L. N. (2009). Cloning and expression analysis of a MADS-box protein gene (GhMADS-13) from upland cotton. *Genomics Appl. Biol.* 28, 223–228.
- Xie, W., Huang, J., Liu, Y., Rao, J., Luo, D., He, M. (2015). Exploring potential new floral organ morphogenesis genes of *Arabidopsis thaliana* using systems biology approach. *Front. Plant Sci.* 6:829. doi: 10.3389/fpls.2015.00829
- Yang, T., and Zheng, Y. (2016). State and trends of oil crops production in China. *OCOL* 23, D603. doi: 10.1051/ocol/2016046
- Zahn, L. M., Feng, B., and Ma, H. (2006). Beyond the ABC-Model: regulation of floral homeotic genes. *Adv. Bot. Res.* 44, 163–207. doi: 10.1016/S0065-2296(06)44004-0
- Zhang, L., Luo, Y., Hou, Z., He, Z., and Eli, W. (2014). Synthesis of carbonated cotton seed oil and its application as lubricating base oil. *J. Am. Oil Chemists' Soc.* 91, 143–150. doi: 10.1007/s11746-013-2358-1
- Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., et al. (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* 33, 531–537. doi: 10.1038/nbt.3207
- Zhao, Y., Li, X., Chen, W., Peng, X., Cheng, X., Zhu, S., et al. (2010). Whole-genome survey and characterization of MADS-box gene family in maize and sorghum. *Plant Cell Tissue Organ Culture* 105, 159–173. doi: 10.1007/s11240-010-9848-8
- Zhou, Y., Li, B. Y., Li, M., Li, X. J., Zhang, Z. T., Li, Y., et al. (2014). A MADS-box gene is specifically expressed in fibers of cotton (*Gossypium hirsutum*) and influences plant growth of transgenic *Arabidopsis* in a GA-dependent manner. *Plant Physiol. Biochem.* 75, 70–79. doi: 10.1016/j.plaphy.2013.12.003

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Ren, Yu, Yang, Li, Qanmber, Li, Li, Liu, Lu, Wang, Zhang, Chen, Li and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Modeling Allometric Relationships in Leaves of Young Rapeseed (*Brassica napus* L.) Grown at Different Temperature Treatments

Tian Tian<sup>1</sup>, Lingtong Wu<sup>1</sup>, Michael Henke<sup>2</sup>, Basharat Ali<sup>1,3</sup>, Weijun Zhou<sup>1\*</sup> and Gerhard Buck-Sorlin<sup>4\*</sup>

<sup>1</sup> Institute of Crop Science and Zhejiang Key Laboratory of Crop Germplasm, Zhejiang University, Hangzhou, China,

<sup>2</sup> Department of Ecoinformatics, Biometrics and Forest Growth, Georg-August University of Göttingen, Göttingen, Germany,

<sup>3</sup> Institute of Crop Science and Resource Conservation, University of Bonn, Bonn, Germany, <sup>4</sup> IRHS, INRA, AGROCAMPUS OUEST, University of Angers, Beaucouzé, France

## OPEN ACCESS

### Edited by:

Dragana Miladinović,  
Institute of Field and Vegetable Crops,  
Serbia

### Reviewed by:

Najeeb Ullah,  
University of Sydney, Australia  
Ivana Maksimovic,  
University of Novi Sad, Serbia

### \*Correspondence:

Weijun Zhou  
wjzhou@zju.edu.cn  
Gerhard Buck-Sorlin  
gerhard.buck-sorlin@agrocampus-ouest.fr

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 10 December 2016

**Accepted:** 20 February 2017

**Published:** 21 March 2017

### Citation:

Tian T, Wu L, Henke M, Ali B,  
Zhou W and Buck-Sorlin G (2017)  
Modeling Allometric Relationships  
in Leaves of Young Rapeseed  
(*Brassica napus* L.) Grown at Different  
Temperature Treatments.  
Front. Plant Sci. 8:313.  
doi: 10.3389/fpls.2017.00313

Functional-structural plant modeling (FSPM) is a fast and dynamic method to predict plant growth under varying environmental conditions. Temperature is a primary factor affecting the rate of plant development. In the present study, we used three different temperature treatments (10/14°C, 18/22°C, and 26/30°C) to test the effect of temperature on growth and development of rapeseed (*Brassica napus* L.) seedlings. Plants were sampled at regular intervals (every 3 days) to obtain growth data during the length of the experiment (1 month in total). Total leaf dry mass, leaf area, leaf mass per area (LMA), width-length ratio, and the ratio of petiole length to leaf blade length (PBR), were determined and statistically analyzed, and contributed to a morphometric database. LMA under high temperature was significantly smaller than LMA under medium and low temperature, while leaves at high temperature were significantly broader. An FSPM of rapeseed seedlings featuring a growth function used for leaf extension and biomass accumulation was implemented by combining measurement with literature data. The model delivered new insights into growth and development dynamics of winter oilseed rape seedlings. The present version of the model mainly focuses on the growth of plant leaves. However, future extensions of the model could be used in practice to better predict plant growth in spring and potential cold damage of the crop.

**Keywords:** allometry, functional-structural plant model (FSPM), growth function, source-sink relations, temperature, winter oilseed rape, GroIMP

## INTRODUCTION

Winter oilseed rape (*Brassica napus* L.) is an important oilseed and fodder crop worldwide (Zhang and Wu, 2007), not only because of its high oil yield (between 40 and 50% of dry mass of seeds), but also because the oil has a high nutritious value, due to the presence of different kinds of aliphatic acids and vitamins (Yang and Tu, 2003). This makes rapeseed oil a very important and healthy ingredient for human nutrition, with a long history as edible oil in Asia and Europe. In addition, if the content of erucic acid is higher than 50%, the seeds can be used as raw ingredient for the

production of lubricating oil and suspending agents (Yang and Tu, 2003). Sowing of rapeseed crop mostly starts in September or October at latitudes of 30° in Eastern China. Undergoing vernalization throughout winter, rapeseed blooms in April and harvest in May next year.

A number of climatic and nutritional factors are known to influence both rates and daily integrals of photosynthesis and respiration. Temperature is known to be of utmost importance for photosynthesis and respiration rate (Marcelis, 1996). High temperature leads to a decrease in carboxylation rate as it decreases the specificity of Rubisco for CO<sub>2</sub> (Lambers et al., 2008), whereas low temperature affects the sucrose metabolism, which leads to accumulation of phosphorylated intermediates and ultimately reduces inorganic phosphates. Finally, inhibition occurs in ATP and photosynthetic accumulation (Chawade, 2011). Besides those effects on photosynthesis, temperature induces some modifications in the morphology of rapeseed. The stems become thinner and the leaves turn thicker under high temperature (Qaderi et al., 2006, 2010). These effects on rapeseed growth (as the decrease of total leaf area) will lead to differences in photosynthesis. Fresh and dry masses have been considered as being positively correlated with temperature during the seedling stage (Rapacz, 1998). Temperature is thus an important factor for photosynthesis rate and consequently growth of rapeseed plants. In order to avoid a potentially considerable reduction in production, it is therefore, important to be able to predict accurately the phenological stage and the topological and geometrical structure of the plant during this period, in order to better assess recovery and survival rates of the crop, which are pivotal for ultimate yield.

To accelerate breeding and optimize the production of rapeseed, the use of crop models as a tool in rapeseed research has been proposed by a number of researchers (Diepenbrock, 2000; Fourcaud et al., 2008; Xu et al., 2011; Robertson and Lilley, 2016). Models for oilseed rape production have been diversified in objectives and methodology. In some classic crop models, the plant canopy was divided into different layers, where the main processes of biomass production were computed for each layer separately and then integrated for the whole crop (Tang et al., 2007, 2009; da Luz et al., 2012). Some models considered the 3D structure and selected functions of the whole plant (Groer et al., 2007). Other models mainly focused on the accurate description of the rates of photosynthesis, respiration, and biomass production at the leaf and whole-plant scale, with local climate parameters as input (Paul and Driscoll, 1997; Müller and Diepenbrock, 2006; Deligios et al., 2013).

In the present study, we investigated the growth and development of young winter oilseed rape plants under different temperature treatments, with a special consideration of the morphology of leaves of different ranks. On the basis of the data established in this study a functional-structural plant model (FSPM) for young rapeseed plants was devised, representing leaf extension kinetics; data about leaf photosynthesis, respiration, biomass production and allocation at the organ and whole-plant scale were obtained from the literature. FSPM being a modeling paradigm that allows not only an instructive

visualization of growth and development, but also the integration of heterogeneous datasets in a common scheme, the model we created already helped us to better interpret the correlated datasets of our experiments, apart from the conclusions drawn based on the statistical analyses. However, in this model, we were only able to explore leaf development of young plants. In a further extension of the model, we will eventually use it as a visual decision-support tool for rapeseed breeding and production utilization.

The present model is based on a previous rapeseed model by Groer et al. (2007) and the general sink-source-based FSPM-prototype (FSPM-P; prototype model) by Henke et al. (2016). The FSPM-P is a general model based on the 3D structure of the plant including the physiological characteristics (photosynthesis, maintenance respiration, growth respiration, and plant organ biomass accumulation). The FSPM-P being not designed for a specific crop, it can be adapted in principle to any crop plant with C3 photosynthesis. Its main advantage is that it is easy to parameterize, use, and extend (Henke et al., 2016). In the present study, we discuss how this model could be used to predict the kinetics of storage of carbon reserves (starch) during autumn and winter that might be the key to successful reestablishment of the crop in spring.

## MATERIALS AND METHODS

### Setup of the Temperature Treatment

Two experiments were set up, differing in planting date. Both experiments focused on young rapeseed plants. Seeds of the leading commercial cultivar of winter oilseed rape (*Brassica napus* L. cv. ZS 758) were germinated in the dark on moist filter paper at 25°C for 2 days. The seedlings were then transferred to plastic pots (height 20 cm, diameter 15 cm) filled with a fertile soil mixture [available N 200 mg kg<sup>-1</sup>, P<sub>2</sub>O<sub>5</sub> 350 mg kg<sup>-1</sup>, KCl 500 mg kg<sup>-1</sup>, Ca(H<sub>2</sub>PO<sub>4</sub>)<sub>2</sub>·H<sub>2</sub>O 3000 mg kg<sup>-1</sup>, K<sub>2</sub>SO<sub>4</sub> 200 mg kg<sup>-1</sup>, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 500 mg kg<sup>-1</sup>], and cultivated in the greenhouse at a temperature of 10–15°C for 15 days until they had one visible leaf. The potted seedlings were then transferred to a climate chamber (PGX450D, Saifu Instruments, Ningbo, China) and cultivated at three different temperature regimes, combined with a 14-h photoperiod (36 plants per temperature treatment). These temperature regimes correspond to the typical temperature prevalent from September to November at latitudes of 30° in Eastern China<sup>1</sup>. The temperature treatments were: 10/14°C (low temperature, “L”), 18/22°C (medium temperature, “M”), and 26/30°C (high temperature, “H”; night/day temperatures, respectively). Relative humidity was maintained at around 80%. Incident photosynthetically active radiation (PAR) in the temperature chambers was measured above the plant canopy using a Li6400R portable photosynthesis system (LI-COR, Lincoln, OR, USA). It was on average 150 μmol m<sup>-2</sup> s<sup>-1</sup>, coming from 39 wall-mounted fluorescent light tubes (fluorescent phosphor, type SP41), with a total output of 330 μmol m<sup>-2</sup> s<sup>-1</sup>.

<sup>1</sup><http://www.tutiempo.net/en/Climate/Hangzhou/584570.htm>

## Plant Measurements and Analyses

Three plants were chosen randomly every 3 days after putting them into the climate chamber. These plants were photographed, and then sampled destructively for determination of dry weights of each organ. Leaf length, leaf width, petiole length, and leaf area were measured from photographs using ImageJ® 1.41o software (Abramoff et al., 2004). Leaf dry weight was measured after 5 days storage of samples in a drying cabinet at 80°C. Destructive sampling commenced after the plants were put into the temperature chamber and continued for a month. Total leaf number per plant, as well as leaf blade length, leaf blade width, and petiole length were determined for each phytomer (in the sense of Room et al., 1994).

Statistical analyses of total leaf area and dry mass were carried out using the glm procedure of SAS 9.1.3 (SAS® Institute Inc., Cary, NC, USA). The different means of total leaf area and dry mass were based on statistical analyses.

## Leaf Kinetics

Leaf kinetics (i.e., extension of the blade in length over time) was determined from the same samples as in Section “Plant Measurements and Analyses” by measuring the lengths of leaf blades every 3 days during their development. The kinetics of a rapeseed leaf blade can be conveniently described using a sigmoid logistic curve, which describes the current length of a blade as a function of plant age (expressed as accumulated temperature). **Figure 1** describes the leaf length development at different temperature treatments, as a function of accumulated temperature, for the first five leaf ranks. To fit the data, a logistic equation was used with three parameters:

$$y_{r,T} = \frac{y_{m,r,T}}{1 + \left[ \frac{ts}{ts_{0,r,T}} \right]^{b_{r,T}}} \quad (1)$$

where  $y_{m,r,T}$  [cm] is the maximum leaf length at rank  $r$  at temperature treatment  $T$ ;  $ts_{0,r,T}$  [°Cd] is the accumulated temperature at which extension rate is maximal; and  $b_{r,T}$  describes the slope of the curve, for a given rank  $r$  and temperature  $T$  (**Table 1**).

## FSPM of Rapeseed Development

The dynamic FSPM of young rapeseed was developed by modification and reparameterization of the FSPM-P (Henke et al., 2016), a fully functional prototype model written in the language XL (Kniemeyer, 2008), and running on the Growth Grammar-related Interactive Modeling Platform (GroIMP) modeling platform (Kniemeyer et al., 2007). Additionally, elements were used from a previous model of rapeseed development by Groer et al. (2007). Essentially, the present model comprised modules for the main biophysical and physiological processes observable in crop plants: light interception, photosynthesis, assimilate partitioning, growth, and respiration.

Light interception was modeled using the twilight light model provided within the GroIMP platform: This model is based on a Monte Carlo raytracer and allows the computation of the spatial and temporal light distribution for suitable objects (such

as leaves) in a simulated scene [for details see Buck-Sorlin et al. (2011)]. The climate chamber used in reality contained 39 wall-mounted light tubes (fluorescent phosphor, type SP41). As this setup would have been too expensive in computational terms, it was reconstructed in GroIMP using area lights associated with parallelogram objects representing the two lateral walls and the back wall, with a total output of 330  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . Two virtual light sensors were placed inside the chamber, above each one of the two shelves. Simulations reproduced the measured photosynthetic photon flux density (PPFD), 150  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . Virtual lamps, sensors, and the construction elements were specified as geometric objects with optical properties, as described in Buck-Sorlin et al. (2009).

To model leaf photosynthetic rate the model LEAFC3 was used (Nikolov et al., 1995). LEAFC3 is a model of the short-term steady-state fluxes of CO<sub>2</sub>, water vapor, and heat from leaves of C3 plant species, explicitly coupling all major processes involved in photosynthesis (biochemistry of assimilation process, stomatal conductance, and leaf energy balance). The model was parameterized using rapeseed-specific parameters derived from the literature (Liu et al., 2003; Ma et al., 2009).

## Dry Matter Dynamics at Plant Level

To model the temporal dynamics of the integral of dry matter accumulation as a function of accumulated temperature at the plant level, a modified logistic function was used (Li and Ma, 2012) as proposed by Wang (1986). Furthermore, in order to improve the comparison of the dynamics at different temperatures, the input data were normalized, yielding the following equation:

$$y' = \frac{y_m}{1 + e^{a ts'^2 + b ts' + c}} \quad (2)$$

where  $y'$  is the normalized dry mass ( $y/y_{max}$ ),  $ts'$  is the normalized accumulated temperature ( $ts/ts_{max}$ ),  $y_m$ ,  $a$ ,  $b$ , and  $c$  are shape parameters (**Table 2**). Inspection of the curves of the normalized data yielded that for some leaf ranks and temperature treatments a double logistic function with a transition point at an accumulated temperature  $ts'_{tr}$  would be more suitable:

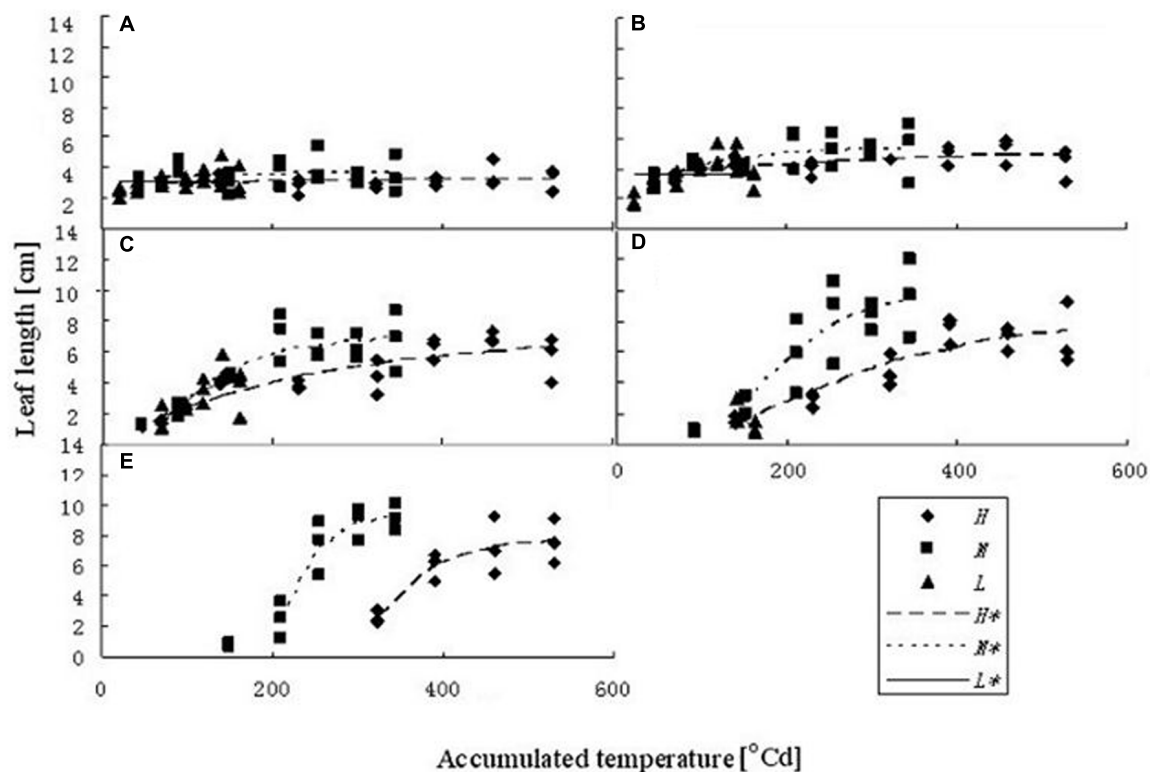
$$y' = \begin{cases} ts' < ts'_{tr} : \frac{y'_{m1}}{1 + e^{a_1 ts'^2 + b_1 ts' + c_1}} \\ ts' \geq ts'_{tr} : \frac{1}{1 + e^{a_h ts'^2 + b_h ts' + c_h}} \end{cases} \quad (3)$$

where  $a_1$ ,  $b_1$ ,  $c_1$ ,  $y'_{m1}$ ,  $a_h$ ,  $b_h$ ,  $c_h$ , and  $y_{mh}$  are the shape parameters (**Table 3**) for the two logistic curves below and above the threshold accumulated temperature  $ts'_{tr}$ , respectively.

In this way, the growth function for each leaf rank under different temperature treatments was obtained. With the exception of the third leaf at the high temperature treatment, which required the double growth function (Equation 3) the expansion kinetics of all leaves could be expressed using Equation 2.

As the plants were grown under conditions of potential production (ample water and nutrients, free of weeds, pests and competition), it was assumed that the measured growth corresponded to the potential growth at the given temperature.





**FIGURE 1 | Leaf length dynamics at three temperature treatments with different leaf ranks. (A–E)** Show the length dynamics for different leaf ranks (Ranks 1–5). *H*, *M*, and *L* represent the data of the leaf length from the experiments under high, medium, and low temperature treatment, respectively. *H\**, *M\**, and *L\** represent the leaf length growth function according to the leaf length data of the high, medium, and low temperature treatment, respectively.

## Leaf Area Dynamics at the Whole-Plant Level

The dynamics of leaf area at the level of the plant as a function of the accumulated temperature was modeled using a logistic curve, with an exponential, linear and maturation phase of growth, an equation with four parameters (i.e., twelve parameters in total, as

the data were fitted for each temperature treatment):

$$LA_t = LA_{0,t} + \frac{LA_{max,t}}{1 + (ts_t - TS_{m,t})^{-b}} \quad (4)$$

Where  $LA_{0,t}$  is the initial leaf area;  $LA_{max,t}$  is the maximum leaf area at a given temperature treatment;  $ts_t$  is the accumulated

**TABLE 1 | Parameters of the logistic curve for three temperature treatments and different leaf ranks in *Brassica napus* seedlings.**

Temperature treatment	Parameter	Leaf rank				
		1	2	3	4	5
<i>H</i>	$ts_0$	152.66	121.77	246.15	272.08	344.69
	$b$	-0.09	-0.32	-1.10	-2.58	-9.88
	$y_m$	6.08	7.99	9.04	8.73	7.74
<i>M</i>	$ts_0$	152.91	41.74	117.05	192.93	228.26
	$b$	-0.11	-1.05	-2.48	-4.05	-11.67
	$y_m$	6.95	5.94	7.36	10.25	9.27
<i>L</i>	$ts_0$	215.52	167.12	12.16	157.92	—
	$b$	48.29	62.99	-4.02	11.01	—
	$y_m$	3.08	3.69	4.46	2.62	—

*H*, *M*, and *L*: high, medium, and low temperature treatments.

**TABLE 2 | Parameters of the two growth functions fitted to the normalized time series data of total dry matter in *B. napus* seedlings grown at three temperature treatments.**

Parameter	Temperature treatment		
	<i>H</i>	<i>M</i>	<i>L</i>
$ym_l$	7.58	1.10	7.57
$a_l$	23.99	0.05	7.90
$b_l$	-16.06	-5.56	-8.23
$c_l$	6.15	3.72	4.70
$ts_{tr}$	0.38	0.76	0.53
$ym_h$	1.00	1.00	1.00
$a_h$	-20.82	-106.04	-2.37
$b_h$	16.49	175.29	-3.99
$c_h$	-2.07	-73.18	2.88

*H*, *M*, and *L*: high, medium, and low temperature treatments.

**TABLE 3 | Parameters of the growth function fitted to the normalized time series data of dry matter in *B. napus* with different leaf ranks at three temperature treatments.**

Temperature treatment	Leaf rank						
	Parameter	1	2	3	4	5	6
<i>H</i>	$ym_t$	0.78	0.93	0.32	16.91	0.96	16.35
	$a_t$	45.49	18.54	−6.63	9.16	−16.11	5.46
	$b_t$	−48.05	−22.77	−29.28	−16.26	4.44	−14.69
	$c_t$	2.79	2.72	5.66	10.04	4.82	11.96
	$ts_{tr}$	—	—	0.61	—	—	—
	$ym_h$	—	—	15.31	—	—	—
	$a_h$	—	—	8.20	—	—	—
	$b_h$	—	—	−16.29	—	—	—
	$c_h$	—	—	10.63	—	—	—
<i>M</i>	$ym_t$	0.86	0.92	0.83	0.95	1.02	14.07
	$a_t$	25.33	0.21	−21.35	7.03	17.91	15.86
	$b_t$	−29.04	−7.69	1.69	−17.70	−40.13	−35.45
	$c_t$	2.23	1.55	2.90	8.10	19.44	22.16
<i>L</i>	$ym_t$	0.90	6.83	38.46	—	—	—
	$a_t$	11.62	4.14	11.42	—	—	—
	$b_t$	−14.74	−5.85	−19.47	—	—	—
	$c_t$	1.82	3.98	12.18	—	—	—

*H*, *M*, and *L*: high, medium, and low temperature treatments.

temperature at a given temperature treatment;  $TS_{m,t}$  is the accumulated temperature at which extension rate is maximum at a given temperature treatment and  $b$  describes the slope of the curve (Table 4).

## Framework of the Rapeseed Model

Some basic parameters used in the rapeseed model are listed in Tables 1–4. We used leaf appearance rates derived from data as (Figure 1), describing the kinetics of leaf appearance, expressed as the number of visible leaves as a function of time (days after transfer into the climate chambers). The daily accumulated temperature increment, as derived from the slope, minus the temperature for growth of rapeseed (5°C) (Tang et al., 2007, 2009) was 23°Cd d<sup>−1</sup>, 15°Cd d<sup>−1</sup>, and 7°Cd d<sup>−1</sup>, at temperature regimes of 26/30°C, 18/22°C, and 10/14°C, respectively. Some derived morphological parameters were calculated, such as the width/length ratio (WLR), the petiole-blade ratio (PBR) and leaf mass per area (LMA) (Figures 2–4). Integration of these elements into the FSPM-P yielded a first model version in which the rough shape of the leaves could be represented (Figures 5, 6). Finally, total dry mass, total leaf area of the whole plant, and leaf mass per area (LMA) of each leaf from our data were used to parameterize the growth curves, which in turn were used to simulate leaf extension (Figures 7–9) and dry matter accumulation kinetics (Figure 10).

## RESULTS

The fitted model parameters for the first five leaf blade ranks and temperature treatments are given in Table 1. The first two true

leaves (except cotyledons) of oilseed rape were usually small, and preformed in the seed. As the first leaf was already present or fully extended by the time of the start of the experiment, the fit of timing ( $x_0$ ) and slope ( $b$ ) parameter of the logistic equation was not very good. A true sigmoid curve was only fitted for leaf ranks of three and above (Figure 1). Furthermore, at low temperature leaves were smaller and grew more slowly, also leaf appearance rate was slower, leading to fewer observed leaves within the observation period (Figure 1).

## Morphological Parameters

The relationship between width and length of the leaves at different temperatures was expressed as WLR. The WLR decreased with leaf rank. Young leaves had a tendency to be narrower than older ones (Figure 2). There was a clear relationship between PBR and leaf rank (Figure 3), changing in a linear fashion with rank. More specifically, PBR increased with increasing rank at the *H* treatment, whereas it decreased at the other two treatments.

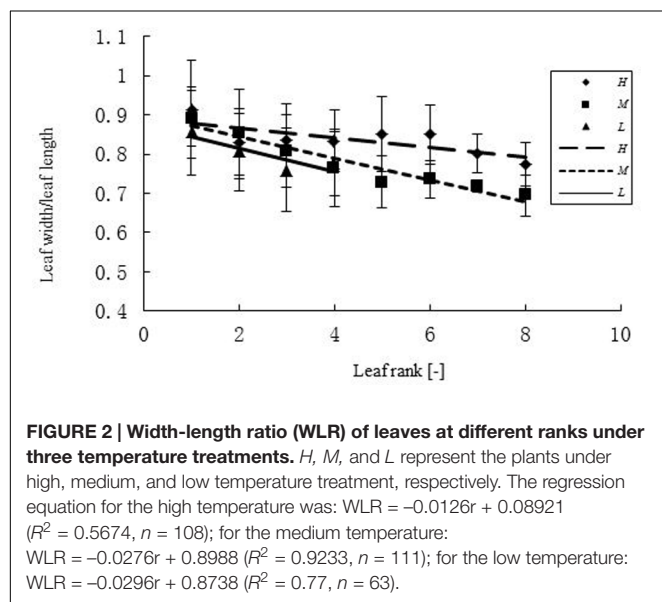
## Total Dry Mass and Total Leaf Area of the Whole Plant

At the first sowing date, the differences in leaf area and leaf dry weight between the *H* and *M* treatment were not significant ( $P = 0.01$ ; Table 5). Only the low temperature treatment was significantly different from the other two, by producing leaves with smaller area and lower dry mass ( $P = 0.01$ ). However, for the second sowing date, all three treatments significantly differed from each other with respect to both parameters ( $P = 0.01$ ; Table 5).

**TABLE 4 | Parameters of the leaf area function in *B. napus* seedlings under three temperature treatments.**

Parameter	Temperature treatment		
	H	M	L
$LA_{0,t}$	17	10	5
$TS_{m,t}$	380	270	135
$LA_{max,t}$	119	184	33
$b$	0.020	0.019	0.025

H, M, and L: high, medium, and low temperature treatments.



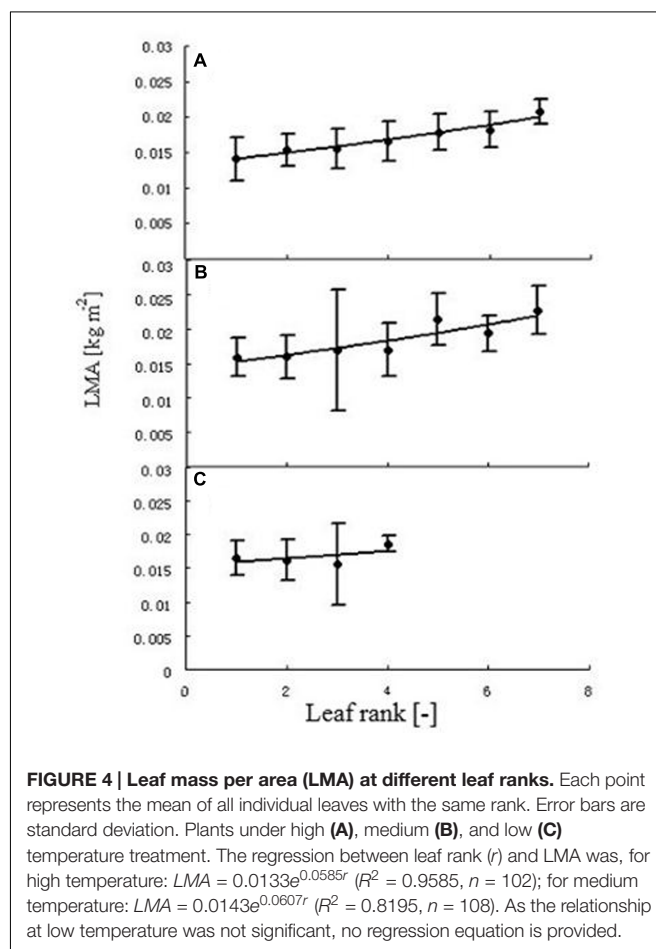
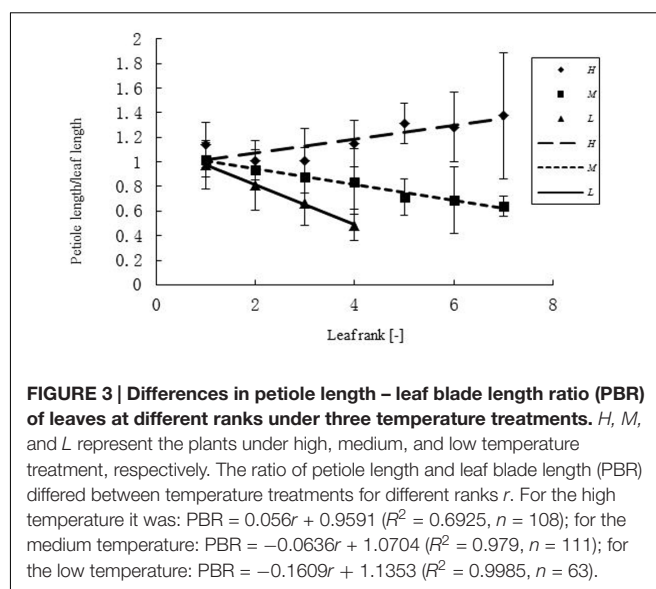
With respect to the duration of treatment (Table 6), it was observed that at the earlier sowing date a significant difference in leaf area occurred between treatment duration of 6 and 9 days, and 12 and 15 days, respectively. A similar pattern was observed for leaf dry weight, but additionally, dry weights differed already between 3 and 6 days duration ( $P = 0.01$ ).

## Leaf Mass Per Area (LMA) of Individual Leaves at Different Treatments

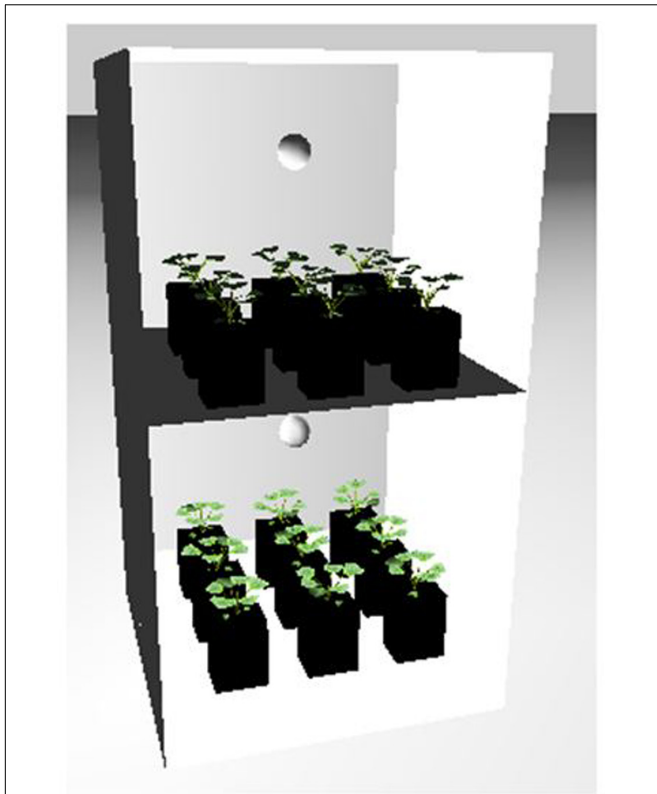
Generally, with the exception of the L treatment, LMA significantly increased with rank (Figure 4). The regression between leaf rank  $r$  and LMA was, for high temperature:  $LMA = 0.0133e^{0.0585r}$  ( $R^2 = 0.9585$ ,  $n = 102$ ), and for medium temperature:  $LMA = 0.0143e^{0.0607r}$  ( $R^2 = 0.8195$ ,  $n = 108$ ). This relationship was not significant for the low temperature treatment.

## Simulation Output

Part of the observations and data given above were used to set up and parameterize the functional-structural model in order to integrate the different results into a spatially explicit form, for the different temperature treatments. Simulations of some developmental stages of young rapeseed plants are shown in Figures 5, 6. As the measured phyllochron was



an input to the model, the accordance between measured and simulated leaf appearance rate naturally was very good (Figure 7). Simulated expansion dynamics of leaf blades of different ranks and at different temperatures are shown in



**FIGURE 5 |** Simulation of the temperature chamber as a 3D model in the Growth Grammar-related Interactive Modeling Platform (GroIMP) modeling environment, with area lights mapped onto the two side walls and the back wall, representing the 39 wall-mounted fluorescent light tubes (door removed for better visibility), and with 2 × 9 plants in pots on the two shelves. For details, see Section “FSPM of Rapeseed Development.”

**Figure 8.** Simulated blade expansion subsided more rapidly and abruptly than in reality (cf. **Figure 1**). The total simulated leaf area at different temperatures, based on the leaf area dynamics function (Equation 4) is shown in **Figure 9**. At the medium temperature treatment simulated leaf area increased faster than under low or high temperature treatment, respectively. Curves were similar among different temperature treatments. However, growth at the medium temperature treatment seemed to be faster than at the high temperature treatment.

## DISCUSSION

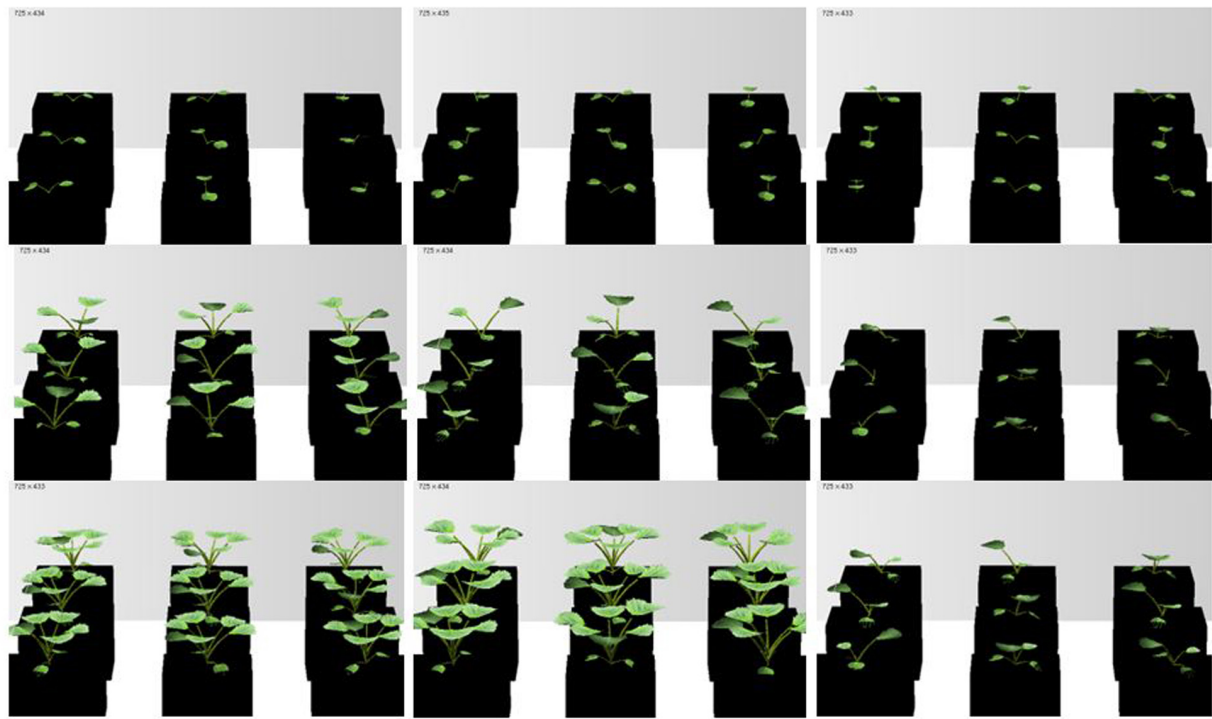
Temperature is a major factor for plant growth rate and daily photosynthesis (Rapacz et al., 2001, 2003; Qaderi et al., 2006, 2010, 2012; Kromdijk and Long, 2016; Zhang et al., 2016). Furthermore, it has also an influence on oil quality (Baux et al., 2008, 2013). In the present study, the effect of temperature on leaf expansion was investigated in oilseed rape. Width-length ratio (WLR) at different temperatures exhibited significant differences only at high temperature (WLR = 0.849). In the *M* and *L* treatments PBR of the upper leaves was small, which means that

the petiole was shorter relative to the blade. Overall, the decrease of PBR with rank will result in a conical leaf arrangement (with upper leaf blades being less distant from the main stem), which could be a plant-level strategy to increase light penetration to the lower leaves. More experimental work will be required to determine the exact reasons for the observed patterns. The PBR is one factor, besides WLR, determining leaf shape. Its change along the main stem can be interpreted in terms of developmental progress from the vegetative to the reproductive stage, with lower-ranked leaves being, in principle, more “vegetative”. The increase in total leaf area and dry mass was lowest under the low temperature treatment. In a study similar to the present one leaf area decreased while leaf blade thickness increased when rapeseed plants were exposed to cold temperatures (Stefanowska et al., 1999). Furthermore, there were no significant differences in leaf area between medium and high temperature at the beginning of juvenile stage. However, when treatment duration was long enough (>6 or >10 days, **Table 6**), leaf area and dry mass at medium temperature were increasing faster than at high temperature. Thus, it can be concluded tentatively that an accumulated temperature increment of 75°Cd is necessary as a phyllochron for the formation of a leaf on the main stem.

Leaf mass per area (LMA) was correlated with environmental and topological factors, such as temperature and leaf rank. LMA was positively correlated with temperature; similar research was also found in other studies (Gausman et al., 1971; Qaderi et al., 2006; Jullien et al., 2009). In the present study, the average LMA under the *H*, *M*, and *L* treatments were 0.01328 kg m<sup>-2</sup>, 0.01766 kg m<sup>-2</sup>, and 0.01639 kg m<sup>-2</sup>, respectively. There was a significant difference between the *H* treatment and the other two treatments. However, *M* and *L* did not differ significantly from each other. A more specific experimental study would be required to obtain a better idea concerning this phenomenon. LMA of upper leaves was higher than that of lower leaves, confirming the findings of Gausman et al. (1971). This effect was not significant for the three or four adjacent ranks in each treatment (**Figure 4**). The absence of a significant difference at the *L* treatment could be explained by the fact that only four leaves per main stem were formed in total.

Plant architecture is of major agronomic importance for the adaptability of a plant for cultivation, and to predict how the plant grows under different environments (Reinhardt and Kuhlemeier, 2002). The impact of plant architecture and morphology on the light climate within the plant canopy and the resulting amount of light locally incident, absorbed by a given assimilating surface, and finally used for photosynthesis plays a major role in plant development as it determines local assimilate availability and thus source strength (Evers et al., 2006). To investigate this complex context and the underlying relationships, we implemented a FSPM as an extension of the FSPM-P model devised by Henke et al. (2016). This model, apart from having a generic description of the sink behavior of young leaves and internodes, also contains provisions to simulate generically photosynthesis on a per-leaf basis, using local light and temperature as an input. Employing a prototype as a departure point permitted us to obtain a usable model relatively rapidly, as the only major steps to be done were parameterization and, eventually, the writing of simple

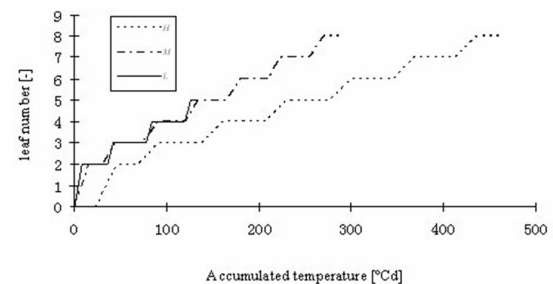




**FIGURE 6 |** Simulated rapeseed seedlings at three temperature treatments: *H*, *M*, and *L* temperature (left to right) and at different times after transfer (days after treatment, dat) to the chamber: 2, 13, and 20 dat (top to bottom).

extensions of rules to accommodate crop-specific processes that were not covered by the general model. The FSPM-P is written in the language XL which is included in the GroIMP platform (Kniemeyer et al., 2007; Kniemeyer, 2008), with specialized tools for plant modeling. GroIMP provides an adequate radiation model that can compute the amount of light absorbed by each plant organ (more generally, objects in a simulated scene; Hemmerling et al., 2008). For more advanced studies of the impact of light quality (e.g., red to far-red ratio) on growth processes, spectral light can be simulated (Henke et al., 2016). The new light model makes use of parallel processing on a graphical processing unit of the computer, thereby accelerating computation enormously (van Antwerpen et al., 2011).

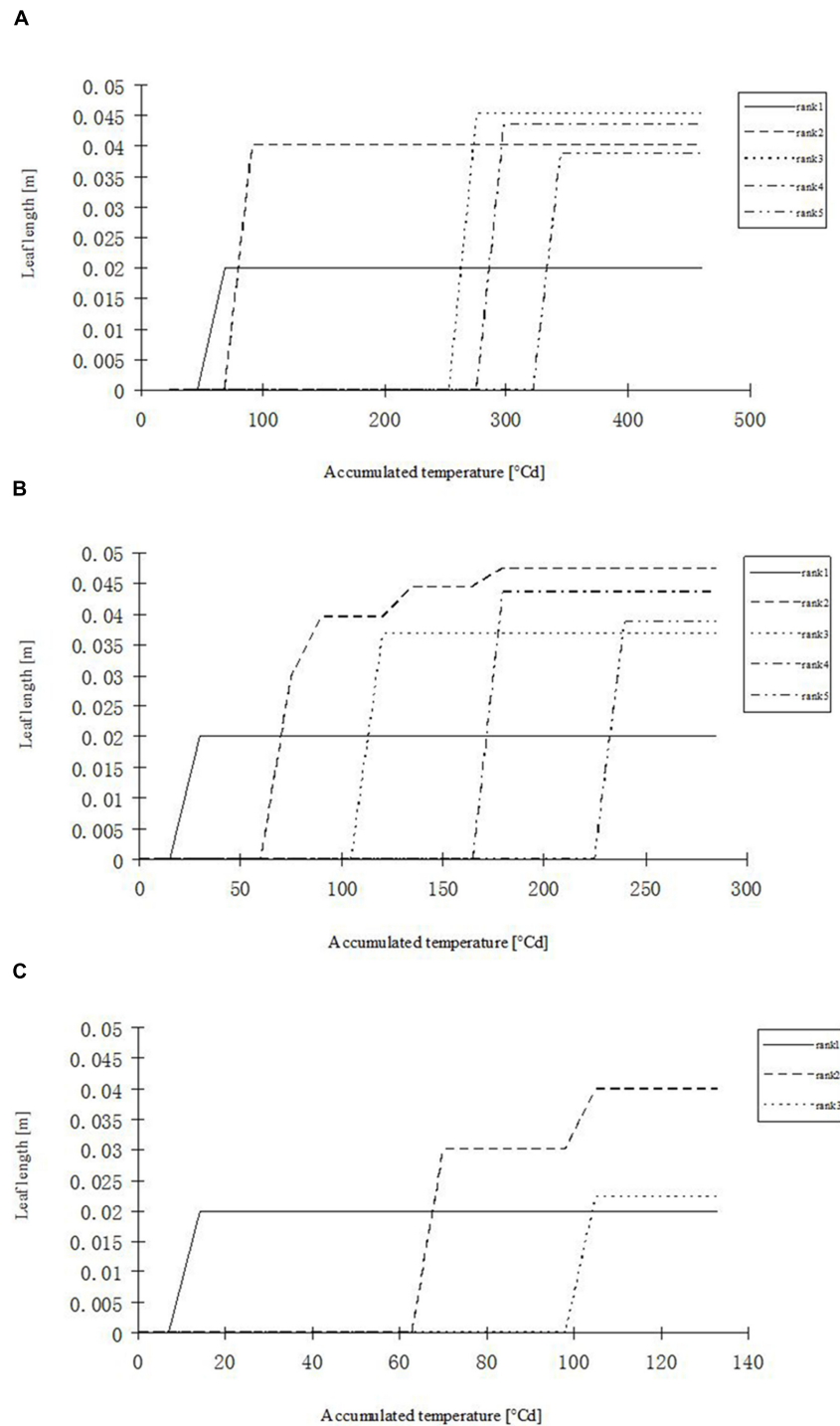
In recent years, models of the main agricultural crops (de Carvalho Lopes and Steidle Neto, 2011) have been developed and their usefulness and importance has been increasingly recognized by the agronomical community (Vos et al., 2010). The most striking advantage of a modeling approach is the time that can be saved when employing a model to predict yield or quality features of a new crop, compared with traditional methods of experimentation and breeding in the field. At the same time, the vivid and realistic visualization of plant development in the case of a 3D model renders teaching much easier and more attractive (Fourcaud et al., 2008). Apart from that, virtual plants or FSPMs, usually integrate a lot of information about the morphology of the crop in question as well as about the nature and spatiotemporal dynamics of physical and biological processes (Buck-Sorlin, 2013). Thus, such models that are capable



**FIGURE 7 |** Simulated leaf number kinetics of rapeseed seedlings at three temperature treatments: *H*, *M*, and *L* signify the plants under high, medium, and low temperature treatment, respectively.

of integrating the dynamics of architectural development with the rate of photosynthesis and respiration are an excellent means to estimate the production of future biomass of a plant (Godin and Sinoquet, 2005).

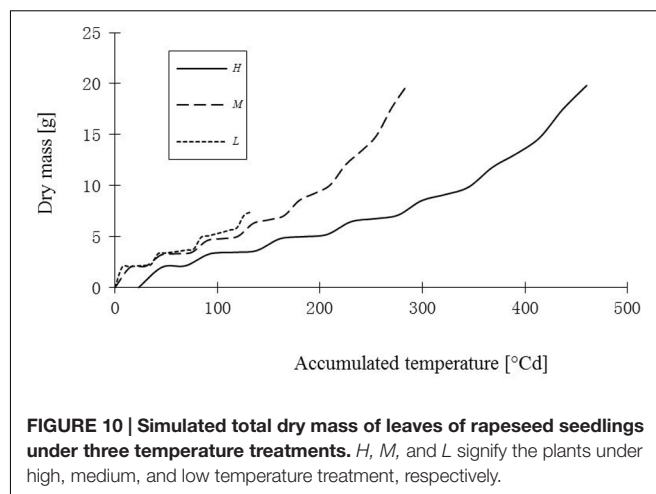
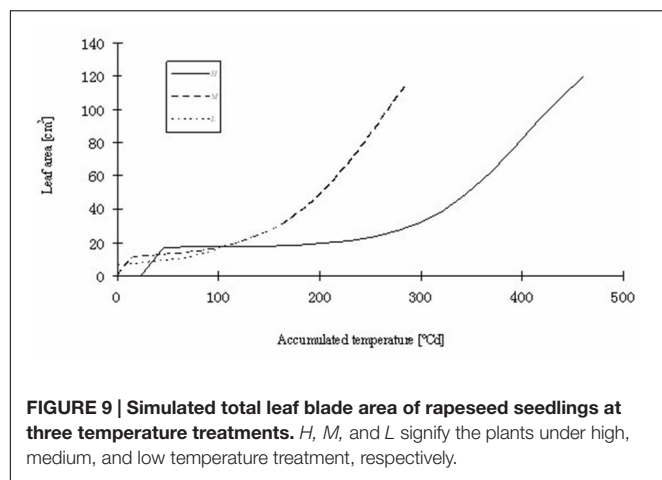
In the field, a rapeseed plant stops growth below 5°C; whereas photosynthesis usually continues well below this threshold temperature for growth (Tang et al., 2007, 2009). This leads to products of photosynthesis accumulating in the plant (as starch), where the extent of accumulation will be a function of photosynthetically active leaf area as well as of the accumulated temperature favorable for photosynthesis (below the threshold for growth yet above the threshold for photosynthesis and above



**FIGURE 8 | Simulated leaf blade length development in rapeseed seedlings of different leaf ranks at three temperature treatments. (A) H temperature treatment, (B) M temperature treatment, (C) L temperature treatment.**

that leading to frost damage). Consequently, regrowth of the plant in spring, as temperatures rise again to favorable values for growth, will be the more rapid the more reserves the plant has been able to accumulate during winter. In a future version

of the model, we will link the rate of regrowth of rapeseed in spring with the amount of stored starch during the winter (Martre et al., 2011). In doing so, an existing photosynthesis module [based on the LEAFC3 model by Nikolov et al. (1995)] will



**TABLE 5 | Total leaf area and dry weight of the whole plant in *B. napus* seedlings under three temperature treatments.**

Parameter	Sowing date	Temperature treatment		
		H	M	L
Leaf area [cm <sup>2</sup> ]	First sowing date	23.63 a*	25.31 a	9.36 b
	Second sowing date	68.27 b	85.41 a	18.16 c
Leaf dry weight [mg]	First sowing date	31.86 a	35.59 a	14.13 b
	Second sowing date	109.50 b	139.49 a	27.54 c

H, M, and L: high, medium, and low temperature treatments.

\* Within each temperature treatment, means followed by the same letter are not significantly different according to the LSD test at  $P \leq 0.01$ . 'a' is the highest and 'c' is the lowest.

be reparameterized from the literature (Leng et al., 2002; Liu et al., 2003; Ma et al., 2009) in order to obtain more accurate values for parameters influencing the rates of photosynthesis and respiration of the young rapeseed. It is possible to compute the total dry mass from an estimate of the amount of carbon fixed by photosynthesis and the amount of carbon lost through maintenance and growth respiration.

The primary aim of the present study was to build a model of the young stages of rapeseed in the leading commercial cultivar (*Brassica napus* L. cv. ZS 758). In order to build a robust and representative model in the future we will need to consider an experimental setup consisting of more plants

and several, contrasting rapeseed cultivars. In addition, as a further step, the current model will be extended to also represent further development from regrowth in spring to maturity (early summer). If successfully parameterized and calibrated such a model could be used to predict harvestable yield, as a function of architectural development at the seedling and juvenile stages. This could in consequence allow the use of such a model as a tool to optimize yield by proposing leaf architecture ideotypes optimized for a given temperature regime in autumn and winter, respectively, an optimal climate dynamics for a given cultivar.

## CONCLUSION

The present study yielded new knowledge about the influence of temperature on leaf expansion in the early development of oilseed rape by combination of observed and literature-derived data. This combination is feasible and useful, as a complete measured dataset is almost never available. Furthermore, embedding of the datasets into an FSPM delivered interesting insights into the interactions of organogenetic and morphogenetic processes with temperature and into source-sink interactions in particular.

**TABLE 6 | Total leaf area and dry weight of the whole plant in *B. napus* seedlings at different temperature treatment durations.**

First sowing date	Parameter	Treatment duration [d]						
		3	6	9	12	15		
	Leaf area [cm <sup>2</sup> ]	13.03 c*	12.21 c	20.28 b	19.46 b	32.17 a		
	Leaf dry weight [mg]	9.22 d	19.10 c	29.18 b	28.77 b	49.68 a		
Second sowing date		3	6	9	12	15	18	21
	Leaf area [cm <sup>2</sup> ]	10.58 d	19.81 d	26.56 cd	47.69 c	82.67 b	101.22 ab	112.44 a
	Leaf dry weight [mg]	17.94 d	34.18 cd	46.1 cd	76.38 c	130.01 b	159.73 ab	180.92 a

\* Within each row, means followed by the same letters are not significantly different according to the LSD test at  $P \leq 0.01$ . 'a' is the highest and 'd' is the lowest.

## AUTHOR CONTRIBUTIONS

GB-S and WZ conceived the study. TT, LW, MH, and BA conducted experiments and analyzed data. MH and GB-S adapted the model. TT, MH, GB-S, and WZ wrote and revised the manuscript. All authors read and approved the manuscript.

## FUNDING

This work was supported by the National High Technology Research and Development Program of China (2013AA103007),

## REFERENCES

- Abramoff, M. D., Magalhaes, P. J., and Ram, S. J. (2004). Image processing with imageJ. *Biophotonics Int.* 11, 36–42.
- Baux, A., Colbach, N., Allirand, J. M., Jullien, A., Ney, B., and Pellet, D. (2013). Insights into temperature effects on the fatty acid composition of oilseed rape varieties. *Eur. J. Agron.* 49, 12–19. doi: 10.1016/j.eja.2013.03.001
- Baux, A., Hebeisen, T., and Pellet, D. (2008). Effects of minimal temperatures on low-linolenic rapeseed oil fatty-acid composition. *Eur. J. Agron.* 29, 102–107. doi: 10.1016/j.eja.2008.04.005
- Buck-Sorlin, G. (2013). “Functional-structural plant modelling,” in *Encyclopedia of Systems Biology*, eds W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota (New York, NY: Springer), 778–781.
- Buck-Sorlin, G., de Visser, P. H. B., Henke, M., Sarlikioti, V., van der Heijden, G. W. A. M., Marcelis, L. M. F., et al. (2011). Towards a functional-structural plant model of cut-rose: simulation of light environment, light absorption, photosynthesis and interference with the plant structure. *Ann. Bot.* 108, 1121–1134. doi: 10.1093/aob/mcr190
- Buck-Sorlin, G. H., Hemmerling, R., Vos, J., and de Visser, P. H. B. (2009). “Modelling of spatial light distribution in the greenhouse: description of the model,” in *Proceedings of the 2009 IEEE 3rd Symposium on Plant Growth Modeling, Simulation, Visualization, and Applications* (Washington, DC: IEEE Computer Society), 79–86.
- Chawade, A. (2011). *Unravelling the Complexity of Cold Acclimation in Plants*. Ph.D. thesis, University of Gothenburg, Gothenburg, 37.
- da Luz, G. L., Medeiros, S. L. P., Tomm, G. O., Bialozor, A., do Amaral, A. D., and Pivoto, D. (2012). Baseline temperature and cycle of canola hybrids. *Cienc. Rural* 42, 1549–1555.
- de Carvalho Lopes, D., and Steidle Neto, A. J. (2011). Simulation models applied to crops with potential for biodiesel production. *Comput. Electron. Agric.* 75, 1–9. doi: 10.1016/j.compag.2010.10.002
- Deligios, P. A., Farci, R., Sulas, L., Hoogenboom, G., and Ledda, L. (2013). Predicting growth and yield of winter rapeseed in a Mediterranean environment: model adaptation at a field scale. *Field Crops Res.* 144, 100–112. doi: 10.1016/j.fcr.2013.01.017
- Diepenbrock, W. (2000). Yield analysis of winter oilseed rape (*Brassica napus* L.): a review. *Field Crops Res.* 67, 35–49. doi: 10.1016/S0378-4290(00)00082-4
- Evers, J. B., Vos, J., Andrieu, B., and Struik, P. C. (2006). Cessation of tillering in spring wheat in relation to light interception and red : far-red ratio. *Ann. Bot.* 97, 649–658. doi: 10.1093/aob/mc1020
- Fourcaud, T., Zhang, X., Stokes, A., Lambers, H., and Körner, C. (2008). Plant growth modelling and applications: the increasing importance of plant architecture in growth models. *Ann. Bot.* 101, 1053–1063. doi: 10.1093/aob/mcn050
- Gausman, H. W., Allen, W. A., Escobar, D. E., Rodriguez, R. R., and Cardenas, R. (1971). Age effects of cotton leaves on light reflectance, transmittance and absorbance and on water content and thickness. *Agron. J.* 63, 465–469.
- Godin, C., and Sinoquet, H. (2005). Functional-structural plant modelling. *New Phytol.* 166, 705–708. doi: 10.1111/j.1469-8137.2005.01445.x
- Groer, C., Kniemeyer, O., Hemmerling, R., Kurth, W., Becker, H., and Buck-Sorlin, G. (2007). “A dynamic 3D model of rape (*Brassica napus* L.) computing yield components under variable nitrogen fertilisation regimes,” in *Proceedings of the 5th International Workshop on Functional-Structural Plant Models, FSPM '07*, Napier.
- Hemmerling, R., Kniemeyer, O., Lanwert, D., Kurth, W., and Buck-Sorlin, G. (2008). The rule-based language XL and the modelling environment GroIMP illustrated with simulated tree competition. *Funct. Plant Biol.* 35, 739–750. doi: 10.1071/FP08052
- Henke, M., Kurth, W., and Buck-Sorlin, G. H. (2016). FSPM-P: towards a general functional-structural plant model for efficient model development. *Front. Comput. Sci.* 10, 1103–1117. doi: 10.1007/s11704-015-4472-8
- Jullien, A., Allirand, J. M., Mathieu, A., Andrieu, B., and Ney, B. (2009). Variations in leaf mass per area according to N nutrition, plant age, and leaf position reflect ontogenetic plasticity in winter oilseed rape (*Brassica napus* L.). *Field Crops Res.* 114, 188–197. doi: 10.1071/FP08052
- Kniemeyer, O. (2008). *Design and Implementation of a Graph Grammar Based Language for Functional-Structural Plant Modelling*. Ph.D. thesis, BTU Cottbus, Senftenberg, 438.
- Kniemeyer, O., Buck-Sorlin, G., and Kurth, W. (2007). “GroIMP as a platform for functional-structural modelling of plants,” in *Functional-Structural Plant Modelling in Crop Production*, eds J. Vos, L. F. M. Marcelis, P. H. B. de Visser, P. C. Struik, and J. B. Evers (Dordrecht: Springer Science+Business Media BV), 43–52.
- Kromdijk, J., and Long, S. P. (2016). One crop breeding cycle from starvation? How engineering crop photosynthesis for rising CO<sub>2</sub> and temperature could be one important route to alleviation. *Proc. R. Soc. B* 283, 20152578. doi: 10.1098/rspb.2015.2578
- Lambers, H., Chapin, I. I. F. S., and Pons, T. L. (2008). *Plant Physiological Ecology*. New York, NY: Springer-Verlag.
- Leng, S. H., Xia, J. F., Hu, Z. Z., and Chen, X. L. (2002). Studies on photosynthetic characteristics of rapeseed leaves. *Chin. J. Oil Crop Sci.* 24, 10–13.
- Li, J. Y., and Ma, Y. Q. (2012). The research on application of plant growth equation based on L-system study. *Image Process. Multimed. Technol.* 6, 25–27.
- Liu, T. M., Cao, C. G., Huang, Y., Liu, F. C., Liu, T. F., and Liu, A. G. (2003). A simulation model of photosynthetic production and dry matter accumulation in rapeseed. *J. Huazhong Agric. Univ.* 22, 533–537.
- Ma, N., Liu, D., Zhang, C. L., Li, J., and Li, G. M. (2009). Regulation effects of exogenous hormones on growth and photosynthesis and yield of rapeseed (*Brassica napus* L.) after Frozen. *Acta Agron. Sin.* 35, 1336–1343. doi: 10.3724/SP.J.1006.2009.01336
- Marcelis, L. F. M. (1996). Sink strength as a determinant of dry matter partitioning in the whole plant. *J. Exp. Bot.* 47(Suppl. 1), 1281–1291. doi: 10.1093/jxb/47.Special\_Issue.1281
- Martre, P., Bertin, N., Salon, C., and Génard, M. (2011). Modelling the size and composition of fruit, grain and seed by process-based simulation models. *New Phytol.* 191, 601–618. doi: 10.1111/j.1469-8137.2011.03747.x
- Müller, J., and Diepenbrock, W. (2006). Measurement and modelling of gas exchange of leaves and pods of oilseed rape. *Agr. For. Meteorol.* 139, 307–322. doi: 10.1016/j.agrformet.2006.07.010
- Nikolov, N. T., Massman, W. J., and Schoettle, A. W. (1995). Coupling biochemical and biophysical processes at the leaf level: an equilibrium photosynthesis model for leaves of C<sub>3</sub> plants. *Ecol. Model.* 80, 205–235. doi: 10.1016/0304-3800(94)00072-P



- Paul, M. J., and Driscoll, S. P. (1997). Sugar repression of photosynthesis: the role of carbohydrates in signalling nitrogen deficiency through source:sink imbalance. *Plant Cell Environ.* 20, 110–116. doi: 10.1046/j.1365-3040.1997.d01-17.x
- Qaderi, M. M., Basraon, N. K., Chinnappa, C. C., and Reid, D. M. (2010). Combined effects of temperature, ultraviolet-B radiation, and watering regime on growth and physiological processes in canola (*Brassica napus*) seedlings. *Int. J. Plant Sci.* 171, 466–481. doi: 10.1086/652389
- Qaderi, M. M., Kurepin, L. V., and Reid, D. M. (2006). Growth and physiological responses of canola (*Brassica napus* L.) to three components of global climate change: temperature, carbon dioxide and drought. *Physiol. Plant.* 128, 710–721. doi: 10.1111/j.1399-3054.2006.00804.x
- Qaderi, M. M., Kurepin, L. V., and Reid, D. M. (2012). Effects of temperature and watering regime on growth, gas exchange and abscisic acid content of canola (*Brassica napus*) seedlings. *Environ. Exp. Bot.* 75, 107–113. doi: 10.1016/j.envexpbot.2011.09.003
- Rapacz, M. (1998). The after-effects of temperature and irradiance during early growth of winter oilseed rape (*Brassica napus* L. var. *oleifera* cv 'Gorczański') seedlings on the progress of their cold acclimation. *Acta Physiol. Plant.* 20, 73–78. doi: 10.1007/s11738-998-0046-9
- Rapacz, M., Tokarz, K., and Janowiak, F. (2001). The initiation of elongation growth during long-term low-temperature stay of spring-type oilseed rape may trigger loss of frost resistance and changes in photosynthetic apparatus. *Plant Sci.* 161, 221–230. doi: 10.1016/S0168-9452(00)00341-1
- Rapacz, M., Waligorski, P., and Janowiak, F. (2003). ABA and gibberellin-like substances during prehardening, cold acclimation, de- and reacclimation of oilseed rape. *Acta Physiol. Plant.* 25, 151–161. doi: 10.1007/s11738-003-0048-6
- Reinhardt, D., and Kuhlemeier, C. (2002). Plant architecture. *EMBO Rep.* 3, 846–851. doi: 10.1093/embo-reports/kvf177
- Robertson, M. J., and Lilley, J. M. (2016). Simulation of growth, development and yield of canola (*Brassica napus*) in APSIM. *Crop Pasture Sci.* 67, 332–344. doi: 10.1071/CP15267
- Room, P. M., Maillette, L., and Hanan, J. S. (1994). Module and metamer dynamics and virtual plants. *Adv. Ecol. Res.* 25, 105–157. doi: 10.1016/S0065-2504(08)60214-7
- Stefanowska, M., Kuraś, M., Kubacka-Zebalska, M., and Kacperska, A. (1999). Low temperature affects pattern of leaf growth and structure of cell walls in winter oilseed rape (*Brassica napus* L., var. *oleifera* L.). *Ann. Bot.* 84, 313–319. doi: 10.1006/anbo.1999.0924
- Tang, L., Zhu, Y., Liu, X. J., Tian, Y. C., Yao, X., and Cao, W. X. (2009). Simulation model for photosynthetic production in oilseed rape. *Pedosphere* 19, 700–710.
- Tang, L., Zhu, Y., Sun, X. F., and Cao, W. X. (2007). Dynamic simulation model for photosynthesis and dry matter accumulation in rapeseed. *Acta Agron. Sin.* 33, 189–195.
- van Antwerpen, D., van der Heijden, G. W. A. M., Marcelis, L. F. M., de Visser, P. B. M., Buck-Sorlin, G., and Jansen, E. (2011). “High performance spectral light transport model for agricultural applications,” in *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS Conference on High Performance Graphics 2011*, eds C. Dachsbacher, W. Mark, and J. Pantaleoni (Vancouver, CO: Eurographics Association).
- Vos, J., Evers, J. B., Buck-Sorlin, G. H., Andrieu, B., Chelle, M., and de Visser, P. H. B. (2010). Functional-structural plant modelling: a new versatile tool in crop science. *J. Exp. Bot.* 61, 2101–2115. doi: 10.1093/jxb/erp345
- Wang, X. L. (1986). How to use the Logistic equation in crop dry matter accumulation. *Chin. J. Agrometeorol.* 1, 14–18.
- Xu, L., Henke, M., Zhu, J., Kurth, W., and Buck-Sorlin, G. (2011). A functional-structural model of rice linking quantitative genetic information with morphological development and physiological processes. *Ann. Bot.* 107, 817–828. doi: 10.1093/aob/mcq264
- Yang, W. Y., and Tu, N. M. (2003). *The Theory of Crop Cultivation*. Beijing: Chinese Agricultural Press.
- Zhang, B., Jia, D., Gao, Z., Dong, Q., and He, L. (2016). Physiological responses to low temperature in spring and winter wheat varieties. *J. Sci. Food Agric.* 96, 1967–1973. doi: 10.1002/jsfa.7306
- Zhang, C. L., and Wu, X. M. (2007). “Rapeseed production, cultivation and research in China,” in *Proceedings of the 12th International Rapeseed Congress*, Wuhan, 7–17.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Tian, Wu, Henke, Ali, Zhou and Buck-Sorlin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

**Visit us:** [www.frontiersin.org](http://www.frontiersin.org)

**Contact us:** [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

[@frontiersin](https://twitter.com/frontiersin)



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership