BREEDING INNOVATIONS IN UNDERUTILIZED TEMPERATE FRUIT TREES

EDITED BY: Giuseppe Ferrara, Agata Gadaleta, Malli Aradhya, Inaki Hormaza and Maria Luisa Badenes

PUBLISHED IN: Frontiers in Plant Science

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-056-7
DOI 10.3389/978-2-88974-056-7

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: frontiersin.org/about/contact

BREEDING INNOVATIONS IN UNDERUTILIZED TEMPERATE FRUIT TREES

Topic Editors:
Giuseppe Ferrara, University of Bari Aldo Moro, Italy
Agata Gadaleta, University of Bari Aldo Moro, Italy
Malli Aradhya, Agricultural Research Service, United States Department of Agriculture, United States
Inaki Hormaza, La Mayora Experimental Station, Spanish National Research Council (CSIC), Spain
Maria Luisa Badenes, Instituto Valenciano de Investigaciones Agrarias, Spain

Citation: Ferrara, G., Gadaleta, A., Aradhya, M., Hormaza, I., Badenes, M. L., eds. (2022). Breeding Innovations in Underutilized Temperate Fruit Trees. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-056-7

Table of Contents

05 ***Editorial: Breeding Innovations in Underutilized Temperate Fruit Trees***
Giuseppe Ferrara, Agata Gadaleta, Malli Aradhya, J. Iñaki Hormaza and Maria Luisa Badenes

07 ***Fruit Development in* Ficus carica *L.: Morphological and Genetic Approaches to Fig Buds for an Evolution From Monoecy Toward Dioecy***
Ilaria Marcotuli, Andrea Mazzeo, Pasqualina Colasuonno, Roberto Terzano, Domenica Nigro, Carlo Porfido, Annalisa Tarantino, Riccardo Aiese Cigliano, Walter Sanseverino, Agata Gadaleta and Giuseppe Ferrara

21 ***Homozygosity Mapping Reveals Population History and Trait Architecture in Self-Incompatible Pear* (Pyrus *spp.*)**
Satish Kumar, Cecilia Hong Deng, Martin Hunt, Chris Kirk, Claudia Wiedow, Daryl Rowan, Jun Wu and Lester Brewer

33 ***Resistance to '*Candidatus *Liberibacter asiaticus,' the Huanglongbing Associated Bacterium, in Sexually and/or Graft-Compatible* Citrus *Relatives***
Mônica N. Alves, Silvio A. Lopes, Laudecir L. Raiol-Junior, Nelson A. Wulff, Eduardo A. Girardi, Patrick Ollitrault and Leandro Peña

49 ***Genome-Wide Association Study Reveals Candidate Genes Involved in Fruit Trait Variation in Persian Walnut* (Juglans regia *L.*)**
Anthony Bernard, Julie Crabier, Armel S. L. Donkpegan, Annarita Marrano, Fabrice Lheureux and Elisabeth Dirlewanger

70 ***Pod Morphology, Primary and Secondary Metabolite Profiles in Non-grafted and Grafted Carob Germplasm Are Configured by Agro-Environmental Zone, Genotype, and Growing Season***
Angelos C. Kyratzis, Chrystalla Antoniou, Lambros C. Papayiannis, Giulia Graziani, Youssef Rouphael and Marios C. Kyriacou

89 ***Construction of a High-Density Genetic Map of* Acca sellowiana *(Berg.) Burret, an Outcrossing Species, Based on Two Connected Mapping Populations***
Marianella Quezada, Rodrigo Rampazo Amadeu, Beatriz Vignale, Danilo Cabrera, Clara Pritsch and Antonio Augusto Franco Garcia

104 ***Fine Mapping of the "black" Peel Color in Pomegranate* (Punica granatum *L.*) *Strongly Suggests That a Mutation in the* Anthocyanidin Reductase (ANR) *Gene Is Responsible for the Trait***
Taly Trainin, Rotem Harel-Beja, Irit Bar-Ya'akov, Zohar Ben-Simhon, Rami Yahalomi, Hamutal Borochov-Neori, Ron Ophir, Amir Sherman, Adi Doron-Faigenboim and Doron Holland

121 ***Comprehensive Characterization and Validation of Chromosome-Specific Highly Polymorphic SSR Markers From Pomegranate* (Punica granatum *L.*) *cv. Tunisia Genome***
Prakash Goudappa Patil, Nripendra Vikram Singh, Abhishek Bohra, Keelara Puttaswamy Raghavendra, Rushikesh Mane, Dhananjay M. Mundewadikar, Karuppannan Dhinesh Babu and Jyotsana Sharma

136 *Analysis of the Segregation Distortion of* FcRAN1 *Genotypes Based on Whole-Genome Resequencing of Fig* (Ficus carica *L.*) *Breeding Parents*
Hidetoshi Ikegami, Kenta Shirasawa, Hiroshi Yakushiji, Shiori Yabe, Masaru Sato, Takeshi Hayashi, Kosuke Tashiro and Hitoshi Nogata

144 AGAMOUS *Gene as a New Sex-Identification Marker in Fig* (Ficus carica *L.*) *Is More Efficient Than RAN1*
Xu Wang, Miaoyu Song, Moshe A. Flaishman, Shangwu Chen and Huiqin Ma

frontiers in Plant Science

EDITORIAL
published: 24 November 2021
doi: 10.3389/fpls.2021.799233

Editorial: Breeding Innovations in Underutilized Temperate Fruit Trees

Giuseppe Ferrara[1]*, *Agata Gadaleta*[2], *Malli Aradhya*[3], *J. Iñaki Hormaza*[4] and *Maria Luisa Badenes*[5]

[1] Department of Plant, Soil and Food Science, University of Bari 'Aldo Moro', Bari, Italy, [2] Department of Environmental and Territorial Sciences, University of Bari 'Aldo Moro', Bari, Italy, [3] Agricultural Research Service, United States Department of Agriculture, Washington, DC, United States, [4] Subtropical Fruit Crops Department, IHSM La Mayora – CSIC – UMA, Málaga, Spain, [5] Department of Citriculture and Crop Production, Instituto Valenciano de Investigaciones Agrarias, Moncada, Spain

Keywords: fig (*Ficus carica* L.), pomegranate (*Punica granatum* L.), walnut (*Juglans regia L.*), carob (*Ceratonia siliqua* L.), pear, feijoa (*Acca sellowiana Berg.*), *Citrinae*

Editorial on the Research Topic

Breeding Innovations in Underutilized Temperate Fruit Trees

The recent growing interest in minor species (i.e., fig, pomegranate, feijoa, etc.) has recently driven new research on breeding and genetics to address producer and consumer traits. Since these species have received little attention from the scientific community, they were less improved *via* conventional breeding, and lacked detailed genomic information on important traits. This lack of data, together with a general poor genetic knowledge of these species, has limited a wider cultivation of varieties with improved characteristics.

For these reasons, and with the objective to increase the interest of scientists, farmers, and consumers for these fruits, this Research Topic "*Breeding Innovations in Underutilized Temperate Fruit Trees*" comprises biochemical, morphological, and genetic studies on some minor species regarding fruit trait variation, resistance, evolution, or sex determination.

In this context, Marcotuli et al. examined the mechanisms behind the bud evolution toward breba or main crop in fig (*Ficus carica* L.), since this aspect remains unclear. The X-ray images of buds showed a great structural similarity between breba and the main crop during the initial stages of development, but breba inflorescence differentiation was completed in two seasons whereas that of main crop started at the end of winter and was completed within 2–3 months in the same season. The higher expression of floral homeotic protein AGAMOUS in breba compared to the main crop may indicate a role of these fruits on staminate flowers' production for pollination of the main crop, as profichi act in the caprifig. Within the same species (*Ficus carica* L.) and for sexual determination, Ikegami et al. analyzed the *FcRAN1* gene (during a breeding program for the selection of female plants) strongly associated with the sex phenotype. A male-biased segregation ratio distortion was obtained in 12 F1 populations, suggesting some genetic factor(s) affecting it. A comparison between the annotated genes and the genes required for normal embryo or gametophyte development and function identified several candidate genes responsible for the segregation distortion in fig. Following the same topic, Wang et al. hypothesized an early sex-identification method to improve breeding efficiency. The use of a deletion as a forward primer, a newly established AG-Marker, was as accurate as the RAN1-Marker, and provided the identification of male plants, giving new clues to understanding *Ficus* sex determination. Moving toward another attracting species, i.e., pomegranate (*Punica granatum* L.), Trainin et al. investigated the black peel color of some pomegranate varieties. Biochemical analysis revealed that delphinidin is highly abundant in the peel of black varieties and the pattern of anthocyanin accumulation is different from that of other pomegranates with red or pink colors of the peel. Genetic analysis of an F2

OPEN ACCESS

Edited and reviewed by:
Athanassios Molassiotis, Aristotle University of Thessaloniki, Greece

***Correspondence:**
Giuseppe Ferrara
giuseppe.ferrara@uniba.it

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

Received: *21 October 2021*
Accepted: *27 October 2021*
Published: *24 November 2021*

Citation:
Ferrara G, Gadaleta A, Aradhya M, Hormaza JI and Badenes ML (2021) Editorial: Breeding Innovations in Underutilized Temperate Fruit Trees. Front. Plant Sci. 12:799233. doi: 10.3389/fpls.2021.799233

population segregating for the black phenotype revealed that it is determined by a single recessive gene. Pomegranate was also studied by Goudappa Patil et al. with regards to the SSR of "Tunisia" pomegranate variety. There was a positive trend in chromosome length and the SSR abundance, as marker density, enhanced with a shorter chromosome length. Examination of the distribution of SSR motif types within a chromosome suggested the abundance of hexanucleotide repeats in each chromosome followed by dinucleotides. A comprehensive set of highly polymorphic genome-wide SSRs was successively developed and tested. These chromosome-specific SSRs could serve as a powerful genomic tool to leverage future genetic studies, germplasm management, and genomics-assisted breeding in pomegranate varieties. Some evolutionary aspects of pear were investigated by Kumar et al. who put a light on runs of homozygosity (ROH) in self-incompatible plants, in particular Asian pears, European pears, and interspecific hybrids using genotyping-by-sequencing. The observed ROH patterns suggested that systematic breeding of European pears would have started earlier than Asian pears. Fruit trait variation in Persian walnut (*Juglans regia* L.) was addressed by Bernard et al. who conducted a genome-wide association study (GWAS) using multi-locus models in a panel of 170 accessions of *J. regia* to elucidate the genetic determinants of fruit quality traits in walnut toward the breeding of new varieties. The authors proposed several candidate genes involved in nut characteristics, such as a gene coding for a beta-galactosidase linked to several size-related traits and known to also be involved in fruit development in other species. With regards of fruit traits, Kyratzis et al. investigated the germplasm of an ancient species, the carob (*Ceratonia siliqua* L.), on the island of Cyprus. The domestic germplasm varies both in terms of pod morphology and composition, reflecting the genetic and physiological characteristics of both grafted and non-grafted accessions, and possibly the impact of agro-environmental conditions. Morphological traits, such as seeds-to-pod weight ratio, pod width, and thickness, were principally under genetic control. Contrarily, chemical compounds, particularly total phenolic content, including condensed tannins, *in vitro* antioxidant capacity, and to a lesser extent gallic acid, organic acids, sugars (glucose and fructose), and minerals were more under agro-environmental control. In the Southern Hemisphere, Quezada et al. worked on feijoa (*Acca sellowiana* Berg.), a fruit tree species native to Uruguay and Brazil. A high-density composite genetic linkage map of feijoa was constructed using two genetically populations. Genotyping by sequencing (GBS) approach was successfully applied for developing single nucleotide polymorphism (SNP) markers. They used both the reference genome of the closely related species *Eucalyptus grandis* and a *de novo* pipeline to construct a composite map. A novel approach for the construction of composite maps where the meiosis information of individuals of two connected populations is captured in a single estimator is described. The topic of resistance was carried out by Alves et al. in order to find sources of genetic resistance to Huanglongbing (HLB)-associated "*Candidatus Liberibacter asiaticus*" (Las), one of the most destructive diseases of citrus. Some genotypes from subtribe *Citrinae*, sexually incompatible but graft-compatible with Citrus, may provide new rootstocks able to restrict bacterial titer in the canopy. Authors tested for Las resistance a wide collection of graft-compatible *Citrinae* species using an aggressive and consistent challenge-inoculation and evaluation procedure. *Eremocitrus glauca* and Papua/New Guinea Microcitrus species as well as their hybrids resulted in full resistance, opening the way for using these underutilized genotypes as Las resistance sources in breeding programs.

AUTHOR CONTRIBUTIONS

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

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 21 August 2020
doi: 10.3389/fpls.2020.01208

Fruit Development in *Ficus carica* L.: Morphological and Genetic Approaches to Fig Buds for an Evolution From Monoecy Toward Dioecy

Ilaria Marcotuli[1], Andrea Mazzeo[2], Pasqualina Colasuonno[1], Roberto Terzano[2], Domenica Nigro[2], Carlo Porfido[2], Annalisa Tarantino[3], Riccardo Aiese Cigliano[4], Walter Sanseverino[4], Agata Gadaleta[1] and Giuseppe Ferrara[2]**

[1] Department of Environmental and Territorial Sciences, University of Bari "Aldo Moro", Bari, Italy, [2] Department of Soil, Plant and Food Sciences, University of Bari "Aldo Moro", Bari, Italy, [3] Department of Agriculture, Food and Environmental Sciences, University of Foggia, Foggia, Italy, [4] Sequentia Biotech SL, Barcelona, Spain

OPEN ACCESS

Edited by:
Inaki Hormaza,
Institute of Subtropical and Mediterranean Horticulture La Mayora, Spain

Reviewed by:
Tommaso Giordani,
University of Pisa, Italy
Aureliano Bombarely,
University of Milan, Italy

***Correspondence:**
Agata Gadaleta
agata.gadaleta@uniba.it
Giuseppe Ferrara
giuseppe.ferrara@uniba.it

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

Received: 02 April 2020
Accepted: 24 July 2020
Published: 21 August 2020

Citation:
Marcotuli I, Mazzeo A, Colasuonno P, Terzano R, Nigro D, Porfido C, Tarantino A, Aiese Cigliano R, Sanseverino W, Gadaleta A and Ferrara G (2020) Fruit Development in Ficus carica L.: Morphological and Genetic Approaches to Fig Buds for an Evolution From Monoecy Toward Dioecy. Front. Plant Sci. 11:1208. doi: 10.3389/fpls.2020.01208

The mechanism behind the bud evolution towards breba or main crop in *Ficus carica L.* is uncertain. Anatomical and genetic studies may put a light on the possible similarities/differences between the two types of fruits. For this reason, we collected complimentary data from anatomical, X-ray imaging, and genetic techniques. The RNA seq together with structural genome annotation allowed the prediction of 34,629 known genes and 938 novel protein-coding genes. Transcriptome analysis of genes during bud differentiation revealed differentially expressed genes in two fig varieties (Dottato and Petrelli) and in breba and main crop. We chose Dottato and Petrelli because the first variety does not require pollination to set main crop and the latter does; moreover, Petrelli yields many brebas whereas Dottato few. Of the 1,615 and 1,904 loci expressed in Dottato and Petrelli, specifically in breba or main crop, respectively, only 256 genes appeared to be transcripts in both varieties. The buds of the two fig varieties were observed under optical microscope and using 3D X-ray tomography, highlighting differences mainly related to the stage of development. The X-ray images of buds showed a great structural similarity between breba and main crop during the initial stages of development. Analysis at the microscope indicated that inflorescence differentiation of breba was split in two seasons whereas that of main crop started at the end of winter of season 2 and was completed within 2 to 3 months. The higher expression of floral homeotic protein AGAMOUS in breba with respect to main crop, since this protein is required for normal development of stamens and carpels in the flower, may indicate an original role of these fruits for staminate flowers production for pollination of the main crop, as profichi in the caprifig. Several genes related to auxin (auxin efflux carrier, auxin response factor, auxin binding protein, auxin responsive protein) and to GA synthesis (GA20ox) were highly expressed in brebas with respect to main crop for the development of this parthenocarpic fruit.

Keywords: *Ficus carica* L., breba, main crop, RNA seq, transcriptome analysis, 3D X-ray tomography, anatomical analysis

INTRODUCTION

Ficus carica L. is one of the ≅700 species of the *Ficus* genus in the *Moraceae* family (Datwyler and Weiblen, 2004; Flaishman et al., 2008). Fig was a popular fruit in the diet of Roman people, as supported by many relics found in cities such as Pompeii, Oplontis, and in different areas of the Roman Empire, from Africa to Europe. Cato, Pliny the Elder, and Columella described horticultural practices (for example, tillage and fertilization) and varieties (African, Winter, Tiburtine, Pompeian, Herculanean, Saguntine), but they did not specifically describe the breba or first crop (generally ripening in May-June) and the main crop (ripening in July-September). The fig fruit is a composite type of fruit and consists of a receptacle tissue enclosing hundreds of individual pedicellate pistillate flowers developing into drupelets. The fertilized drupelets contain small formed seeds with respect to the empty drupelets of the parthenocarpic (persistent) fruits. The composite fruit developing from this inflorescence is called syconium. The first crop is called breba, ripening in May-June in the Northern Hemisphere, with fruits generally larger than the main crop and persistent. The second crop is the main crop (forniti) with a possible third crop harvested in autumn; the second crop may or may not be pollinated (Storey, 1977; Stover et al., 2007; Flaishman et al., 2008; Ferrara et al., 2016). Botanically, the fig is a gynodioecious species, and female trees (*Ficus carica sativa* L.) need to be pollinated, or "caprificated", by the male trees (*Ficus carica caprificus* L.) to set the main crop. In the process of caprification, wasps (*Blastophaga psenes* L.) passively carry pollen from male flowers, located in the inedible profichi fruits borne on caprifig trees (hermaphroditic, with male and short-styled female flowers in the profichi fruits), to the long-styled female flowers of the edible figs (in the main crop). Apart from naturally occurring pollination where caprifig trees are present, caprification by growers has been practiced in the Mediterranean area for centuries, as documented by Aristotle, Theophrastus, and Pliny the Elder (Vallese, 1909; Siniscalchi, 1911; Condit, 1955). In brief, growers placed some profichi fruits, containing pollen and pollinators, on female trees to facilitate pollination of female flowers in edible figs. Collecting profichi and putting them in trees is laborious and costly, so this process has fallen out of favor and varieties have been selected that do not require caprification. The eradication of caprifigs has led to a simplification/erosion of fig biodiversity, in particular of varieties requiring pollination.

The reproductive biology of fig is very complex, since there are fig varieties which need pollination and varieties that do not. As a consequence, edible fig varieties have been grouped in three main groups: 1) the Common type, with trees developing fruits parthenocarpically (persistent), either brebas or main crop; 2) the Smyrna type, non-parthenocarpic, which requires pollination to bear the main crop and do not usually produce brebas; 3) the San Pedro type, bearing brebas parthenocarpically and the main crop generally after caprification or even parthenocarpically (Storey, 1977; Ferrara et al., 2016). In some varieties, the main crop could be divided in two sub-groups, the (summer) main crop (ripening in July-September) and the late main crop ripening in autumn and borne on the trees up to December. There are varieties which produce only the main crop ripening very late (cv. Natalino) and eaten almost at Christmas time.

The male fig (caprifig) can have three fruits similarly to the female fig: 1) the profichi in spring-beginning summer; 2) the mammoni in summer-autumn; and 3) the mamme during autumn-winter. Either the presence or the amount of the three types of fruits may vary consistently among varieties of caprifig. The male flowers for the caprification of the main crop in the edible fig are present in the profichi of the caprifig (**Figure S1**). Female flowers are either short-styled or long-styled, with the first type located in the fruits of the caprifig and the second type in the fruits of the edible fig (**Figure S2**). All these flowers are inserted on a receptacle forming the syconium, which is palatable in the edible fig, and generally non-palatable in the caprifig (caprifig means fig for the goats and not for human beings). Short-styled flowers are used by the wasps for oviposition in the ovary whereas the long-styled ones are not efficient for the wasps because the style is longer than the ovipositor.

All the fig buds are developed on the same shoot (**Figure S3**), i.e., fruit bud, vegetative bud and mixed bud. In particular, the apical/lateral bud is a mixed bud which in spring originates a shoot carrying axillary fruits and/or buds developing in the current and/or successive season all the crops. In the common fig, the basal portion of the shoot is generally occupied by the main crop and the distal by the brebas (sometimes with the late ripening main crop in the middle), but some overlapping situations may occur. In the caprifig in the basal portion, there are the mammoni, then the mamme and finally the profichi in the distal portion, when all these types of fruits are present.

This "difference" of crops lead to the distinction of the varieties in uniferous (only one crop, main crop), biferous (two crops, breba, and main), and even triferous (breba, summer, and late main crop) (Ferrara et al., 2017; Marcotuli et al., 2019). *It is unclear why some fruit buds on the shoot may develop in summer for the main crop, or in autumn-winter for the late main crop or even in the following year (breba) since they originated from the same mixed bud bursting in the current season.*

The aim of the present work was to study the buds of two fig varieties, with different crop types, applying a multidisciplinary approach including anatomical, X-ray imaging, and genetic techniques, in order to better understand the mechanism behind the development towards breba or main (summer or winter) crop.

MATERIALS AND METHODS

Plant Materials and Total RNA Extraction

Two fig varieties Dottato and Petrelli grown at the fig repository at the "P. Martucci" experimental station in Valenzano (Bari) of University of Bari "Aldo Moro" Department of Soil, Plant, and Food Science – Fruit Tree Unit, were used in this work. In particular, Dottato belongs to the Common type, which does not require pollination, and produces mainly main crop (rarely brebas) and the Petrelli belongs to the San Pedro type and produces mainly brebas and sometimes few main crop fruits.

In 2016 and 2017, the two varieties were investigated for the caprification response whereas in the second year several approaches were undertaken to deeply study the buds. Moreover, total RNA from fruit buds of Petrelli and Dottato harvested in 2017 at two different time point, April and July (according to the development of the fruits, breba, and main crop respectively), was extracted according to the RNeasy Plant Mini Kit (QIAGEN®) instructions; for each stage and variety three different biological replicas were used, and for each of its three technical replicates. Considering that the standard deviation among the replicates was not significant, we used the mean for the subsequent analysis. RNA quality and quantity were assessed by Nanodrop 2000 (Thermo Scientific, USA) and checked on 1.5% agarose gel. RNA integrity was evaluated with Bioanalyzer 2100 and TapeStation 4200, only samples with a RIN higher than 8 were used for sequencing. After library construction, with fragment size of 275 bp and standard deviation of 166 bp, using a TruSeq Standard mRNA kit (Illumina USA), RNA sequencing was performed on a HiSeq 2000 Illumina system using a paired-end sequencing technique (2x100 bp).

Genome Structural and Functional Annotation

The reference genome sequence of *F. carica* (Mori et al., 2017) was used as a basis for all the bioinformatics analyses, and it was downloaded from NCBI (accession number GCA_002002945.1). As a first step, a structural genome annotation was performed following the pipeline showed in **Figure S4**. Briefly, RNA-seq reads produced from this study were trimmed using BBDuk v35.85 (Bushnell et al., 2017) setting a minimum base quality of 20 and a minimum read length of 35 bp. Reads were then mapped against the reference genome with STAR v2.7.3a (Dobin et al., 2012) in double pass mode and the following parameters "–alignEndsType Local –alignEndsProtrude 20 ConcordantPair". The resulting BAM files were used as input for Trinity v2.4.0 (Grabherr et al., 2011) transcriptome assembler in genome guided mode and using the following options "–jaccard_clip –min_kmer_cov 2 –SS_lib_type RF –genome_guided_min_reads_per_partition 10". The raw assembly was processed with CD-HIT-EST v4.8.1 (Fu et al., 2012) to remove redundant sequences with the options "-r 0 -g 1" and then the assembly was evaluated with Transrate v1.0.3 (Smith-Unna et al., 2016) and BUSCOv3 (Simão et al., 2015).

The assembled transcript and the transcriptome sequences published in Solorzano Zambrano (Solorzano Zambrano et al., 2017) were used with the Maker2 v3.0.0 (Cantarel et al., 2008) pipeline to create a first annotation based only on expressed sequences. The obtained annotation was used to perform a training of an Augustus v3.3 (Stanke and Morgenstern, 2005) model which was then used with Maker2 to perform a new annotation iteration. The new obtained GTF was used to refine the training Augustus. This process was iterated 4 times before obtaining the final annotation. The newly obtained genome annotation was compared with the official one, to avoid loss of information the genes present in the official annotation were always maintained, in case of 1-to-1 overlap the official gene structure was maintained. In case of 1-to-many overlaps between our annotation and the NCBI annotation, a more in-depth analysis was performed: the corresponding proteins were BLASTed against TrEMBL and UniProt Plantae databases (April 2018) (minimum e-value 0.001) and the best proteins/genes were selected as the ones showing the highest alignment length. The predicted protein sequences were annotated with the AHRD v3.0 (https://github.com/groupschoof/AHRD) pipeline after performing a BLASTp against SwissProt and TrEMBL Plant sequences (April 2018).

Quantification and Differential Gene Expression Analysis

Gene expression values were quantified for each sample with FeatureCounts v2.0 (Liao et al., 2013) together with the new genome annotation in order to calculate gene expression values as raw read counts. In addition, normalized TMM (Trimmed Mean of M-values) and FPKM (Fragments Per Kilobase per Million Mapped Fragment) values were calculated for all the genes.

The HTSFilter (Rau et al., 2013) package was chosen to remove the not expressed genes and the ones showing too much variability and implemented with a filtering procedure for replicated transcriptome sequencing data based on a Jaccard similarity index. The filter was applied to the different experimental conditions in order to identify and remove genes that appear to generate an uninformative signal. The TMM normalization was used for this step and the genes with TMM values lower than 7 were removed.

All the statistical analyses were performed with R with the packages HTSFilter (Rau et al., 2013) and edgeR (Robinson et al., 2009). Genes were considered statistically differentially expressed with an FDR value <=0.05.

Identification of Differentially Expressed Genes

The sequences of genes involved in the complex system of flowering and fruit maturation reported in literature were downloaded from GenBank (http://www.ncbi.nlm.nih.gov/Genbank/) and *Morus notabilis* Genome Database (http://morus.swu.edu.cn/morusdb/). All the sequences obtained were used as queries to blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against the yielded sequences from the *Ficus carica* RNA-seq (e-value threshold ≤E-10 and identity percentage higher than 80%). To determine the role in fruit ripening, differential expression analysis was carried out comparing the two varieties and the two different time points (fruits).

Caprification Trial

Petrelli is a San Pedro-type variety defined as biferous since it needs caprification for the main crop, whereas Dottato, which is a Common-type fig defined as uniferous, does not need caprification. The trial was conducted in the fig repository of the Fruit Tree Unit - Department of Soil, Plant, and Food Science - University of Bari "Aldo Moro." In this repository, trees are trained to vase and spaced at 6 × 6 m. The orchard is subjected to the cultural practices commonly adopted for fig trees in the area and

weed management is done through mowing (no use of herbicides at all). The experimental plot consisted of 3 treatments with 3 biological replicates (trees) each: 1) open pollination of the fruits (T1); fruits covered with net without pollination, i.e., parthenocarpic (T2); fruit covered with net with hand pollination (T3). The shoots bearing fruits were covered with nets to prevent pollination by *Blastophaga psenes* (**Figure S5**). Hand-pollination was conducted every 4 to 5 days each year (2016–2017) by using a syringe with a thin needle. Pollen grains of the profichi were dispersed in solution containing 2% of sucrose and were injected in the fruits through the ostiole when ready to be pollinated (**Figure S6**). All the fruits were counted prior to bagging and successively at the end of the trial for calculating the percentage of fruit-set. At harvest, on the fruits that set, quantitative and qualitative measurements were made including size, weight, pH, total soluble solids (TSS, as °Brix), and titratable acidity (TA).

Analysis of variance (ANOVA) was performed with XLSTAT-Pro software (Addinsoft, Paris, France) at the 0.05 P level. The assumptions of variance were verified with the Levene test (homogeneity of variance) and the Lillefors and Shapiro-Wilk tests (normal distribution). The mean values obtained for the different factors were statistically separated by using the REGWQ test.

Microscopic Bud Analysis

In 2017, 30 fruit buds and 30 mixed buds taken from the current year shoots were put in FAA (90% ethanol 50%, 5% acetic acid, and 5% formaldehyde) for 72 h. Buds were successively washed in distilled water and dehydrated in alcoholic solutions and finally included in metacrylate. The included buds were cross-sectioned by using a steel blade microtome (LKB Bromma 11800 Pyramitome). The sections of the buds were put on the slides and observed at the microscope (Nikon H550L).

In 2018, 10 fruit buds and 10 mixed buds were taken from the shoots every 3 weeks from summer (July) until end of winter (February) from each variety. External scales were carefully removed from each bud and sections (0.04–0.08 mm thick) were cut with a hand microtome and stained with 0.05% (w/v) toluidine blue. Images were acquired under a Nikon H550L light microscope (Nikon, Tokyo, Japan) equipped with a digital camera (Canon, Tokyo, Japan).

In order to better investigate the development of the different buds we set up a small experiment in a growth chamber at the end of the summer. Potted fig plants were placed in the growth chamber with controlled light (16/8) and temperature (26/22 °C). Bud burst of both brebas and apical buds was monitored for the successive months.

X-Ray Analyses

High resolution micro X-ray computed tomography (μCT) analysis of fig buds was carried out at the MicroXRayLab of University of Bari "Aldo Moro." Since sample dehydration during the scan strongly affects tomography data resulting in image distortions, fig buds were preliminary dehydrated with ethanol solutions with increasing concentration (up to 80% ethanol) as reported by Glauert (Glauert, 1974).

Each fig bud sample was glued at the tip of a wooden toothpick and analyzed with a SkyScan 1272 μCT (Bruker Gmbh, Germany), equipped with a W microfocus X-ray source operating at 40 kV and 250 mA. For all samples, a rotation step of 0.1 deg (0–180 deg) and an exposure of 1630 ms per radiography was set, for a total time of about 5-hours acquisition per sample. Image resolution (pixel size) varied between 1.6 and 2.1 μm, depending on sample shape and dimension. Flat field correction, frame averaging (3) and random movement (10) were also applied to improve the accuracy of the results.

After acquisition, image reconstruction was carried out using the software NRecon (version 1.6.10.4, InstaRecon®), while for the 3D rendering the software DataViewer (version 1.5.2.4, Bruker microCT®) and CTvox (version 3.1.1 r1191, Bruker microCT®) were jointly used.

RESULTS

Genome Structural and Functional Annotation

The genome of *F. carica* was published in 2017 (Mori et al., 2017), which reported 36,138 gene models, 69% of which were functionally annotated. Own transcriptomics data and publicly available data were used to update the annotation with variety specific expressed genes that might not have been represented in the official one. In order to do so, the approach described in **Figure S4** was adopted. Briefly, a reference guided transcriptome assembly produced 50,866 transcripts with a mean length of 875.16 bp. The dataset used for the assembly included 167 million of paired-end 100 bp reads from six samples. Analysis performed with the Viridiplantae dataset of BUSCO conserved genes found that the assembled transcriptome included 76.6% of complete genes (i.e., single copy and full length) (**Figure S7**). The assembled transcripts, together with the transcripts from the official annotation and a *de novo* assembly produced by Solorzano Zambrano (Solorzano Zambrano et al., 2017; Usai et al., 2020) were used with the Maker2 pipeline to produce a pre-annotation that was then used to create an Augustus gene model. The model was refined 4 times before the genome annotation was obtained, which contained 35,567 genes. About 34,600 gene models were in common between the official annotation and our version. In addition, 938 new genes were annotated and 1,509 genes from the Mori et al. (2017) annotation were removed since they were gene fusions or gene fragments (see *Methods*). The BUSCO analysis (**Figure S7**) revealed that 79% of genes were complete in the new annotation, 68.8% were also in single copy whereas 10.3% were duplicated. On the other hand, the official annotation contained 67.3% of single copy genes and 10.2% of duplicated genes. In addition, the new annotation contained 6.18% of fragmented genes against the 6.45% of the reference. Finally, the percentage of missing genes was 14.5% in the new annotation and 15.9% in the reference. A description was assigned to 25,910 genes (72.84% of the total) whereas a Gene Ontology annotation was assigned to 24,987 genes (70.25% of the total).

Differential Gene Expression Analysis

A total of 35,567 (34,629 known and 938 novel genes) protein-coding genes were predicted in the genome assembly. Among them, 21,762 and 21,441 were expressed in Dottato fruits in brebas and main crop, respectively, while fruits of Petrelli had 21,801 and 21,852 genes expressed in brebas and main crop, respectively (**Figure S8**). Transcripts with expression values of zero in all the samples were not considered into the analysis. Among the total of expressed loci, in breba 1,072 and 1,111 loci were specific for Dottato and Petrelli, respectively, whereas 20,690 genes were expressed in both varieties. In the main crop, instead, 992 and 1,403 genes were detected in fruits of Dottato and Petrelli, respectively. Both varieties shared 20,449 expressed loci (**Figure S9**).

Within the 1,615 and 1,904 loci expressed in Dottato and Petrelli specifically in breba or main crop, respectively, only 256 genes appeared to be transcripts in both varieties. The other loci were clearly specific for each variety and in particular 1,359 loci for Dottato and 1,648 genes for Petrelli (**Figure S9**).

Two databases (KEGG and GO) were used to annotate all unigenes with comprehensive gene function information. A total of 7,056 genes were successfully annotated using KEGG database, representing the 19.8% of the total. The KEGG database identified the putative biological pathways of genes (**Table 1**) with a total 130 genes expressed in Dottato (66 loci in breba and 64 loci in main crop) and 203 transcript specific in Petrelli (95 loci in breba and 108 in main crop). With respect to the KEGG pathway distributions, SAUR (Small Auxin Up RNAs) family proteins, followed by transcription factor MYB and interleukin-1 receptor-associated kinase 4 were the most common. The signaling molecules and the sugar metabolizing enzymes had a low gene count (two or three genes per pathway).

The GO database assigned a predicted function to 24,987 genes (70.2%), including biological process, cellular component and molecular function. Among the annotated genes, 9,611 (27%) were found differentially expressed, and specifically 4,664 (13.1%) were up-regulated, while 4,947 (13.9%) were down-regulated in both Dottato and Petrelli. As a following step, a deferential expression analysis using separately the two groups of the up- and down-regulated genes was performed comparing Dottato vs. Petrelli. The evaluation underlined how 15 genes were up-regulated of which 10 genes were responsible of molecular function, 2 for cellular component, and 3 coding for biological processes (**Figure S10A**), with the signal transduction pathway (3–5 genes) the most represented. Analyzing the down-regulated genes, 20 were differentially expressed in the two varieties, and in details 12 corresponded to molecular function, 1 for cellular component, and 7 to different biological processes. Even in the down-regulated transcripts, the signal transduction pathway genes (2–3 genes) were the most common (**Figure S10B**).

TABLE 1 | Comparative analysis of differentially expressed genes between the two stages of bud development in Dottato and Petrelli, the number of genes annotated in KEGG database and its specific description.

Variety	No. of KEGG annotated genes	Most represented KEGG pathways	
		No. of loci	Description
Dottato (Breba)	66	2	Ferric-chelate reductase
		2	Glycerol-3-phosphate acyltransferase
		2	Solute carrier family 15 (peptide/ histidine transporter), member 3/4
		3	Beta-glucosidase
		3	Laccase
		5	SAUR family protein
Petrelli (Breba)	95	2	Galacturan 1,4-alpha-galacturonidase
		2	Interleukin-1 receptor-associated kinase 4
		2	Disease resistance protein RPM1
		2	HSP20 family protein
		2	Omega-hydroxypalmitate O-feruloyl transferase
		2	Cytochrome P450 family 714 subfamily A polypeptide 1
		3	Peroxidase
		3	Pectinesterase
		4	Transcription factor MYB, plant
		10	SAUR family protein
Dottato (Main crop)	64	2	Galacturan 1,4-alpha-galacturonidase
		4	SAUR family protein
		5	Transcription factor MYB, plant
Petrelli (Main crop)	108	2	Endoglucanase
		2	Polygalacturonase
		2	Laccase
		2	Indole-3-pyruvate monooxygenase
		2	Calcium-binding protein CML
		2	ABA-responsive element binding factor
		2	SAUR family protein
		2	Solute carrier family 15 (peptide/ histidine transporter), member 3/4
		2	Cytokinin riboside 5′-monophosphate phosphoribohydrolase
		3	Nuclear transcription Y subunit beta
		3	MADS-box transcription enhancer factor 2A
		4	Peroxidase
		4	Transcription factor MYB, plant
		5	Interleukin-1 receptor-associated kinase 4

Transcriptome Analysis of Flowering Genes During Bud Differentiation

To find the main genes potentially involved in the flowering process and in the bud differentiation of main crop or breba, we used our transcriptome as a query to blast and select candidate genes.

Gene families including sequences of *MADS-box* (comprising *Agamous*, *AG*, and *Defieciens*, *DEF*, genes), flowering time control protein (*FCA*, *FPA*, and *FY*) and *Apetala* (*AP*), which are responsible of the flowering, and the hormones, such as auxin (*AUX*), gibberellin (*GA*), ethylene (*ETH*) and indole-3-acetic acid (*IAA*), which regulate the plant growth, were identified in NCBI and *Morus* (phylogenetically closest genus belonging to the *Moraceae* family) database and the expression levels of the

correspondent transcripts were compared in both Dottato and Petrelli fruit types.

The blast analysis against the *Ficus carica* annotated genes and the homologue sequence from *Morus* database, allowed the identification of 151 loci associated with genes for either flowering or hormones for plant growth. Among those, 5 loci were obtained from the *Ficus* genome, instead 146 genes were detected through *Morus* database. In particular, the most represented classes were auxin and ethylene gene families with 54 and 45 transcripts, respectively, followed by gibberellin (16 genes) (**Table S1**).

To explore the flowering genes, the expression in the two crops (breba and main crop) of all genes obtained from this analysis was examined and compared between the two varieties. The expression patterns of some of the genes following the auxin, ethylene, and gibberellin groups significantly differed between Dottato and Petrelli. Looking at the genes primary implicated in the flowering process eighteen genes were identified. Among them, the results showed high variability not only between but also in the same varieties for breba and main crop. In particular flowering locus T (*FcFT*), flowering time control protein (*FCA*), and flowering time control protein (*FPA*) had the highest expression level in the main crop of Petrelli; floral homeotic protein AGAMOUS (*AG*), floral homeotic protein APETALA (*AP2*) and flowering time control protein (*FY*) were more expressed in breba of Dottato; 1-aminocyclopropane-1-carboxylate oxidase-like 3 (*ACO3*) and floral homeotic protein APETALA (*AP1*) had high values of expression in main crop of Dottato and breba of Petrelli, respectively (**Table 2**).

No detectable variation in the expression pattern was observed for the following genes between breba and main crop and between the two varieties: 1-aminocyclopropane-1-carboxylate oxidase 2 (*ACO2*), 1-aminocyclopropane-1-carboxylate synthase 4 (*ACS4*), 1-aminocyclopropane-1-carboxylate synthase-like 1L (*ACS1L*), two of the *AP2* and Floral homeotic protein DEFICIENS (*DEF*) (**Table 2**).

Caprification

In the year 2016, fruit-set was variable as a function of the flights of *Blastophaga psenes* for both the climatic conditions and the number of individuals in the fruits. In the case of open pollination, in Petrelli fruit-set was 66.7% and in Dottato 90.9%. Similar values were obtained by hand-pollination under the net, with 57.1 and 95.0% for Petrelli and Dottato, respectively (**Table 3**). When pollination was prevented to obtain parthenocarpic

TABLE 3 | Effects of open, hand and no pollination (parthenocarpic) on fruit-set (%) of Petrelli and Dottato for 2016 and 2017.

Treatment	2016		2017	
	Petrelli	Dottato	Petrelli	Dottato
Open pollination	66.7a	90.9	81.5a	100
Parthenocarpic	12.0c	93.4	28.6b	100
Hand pollination	57.1b	95.0	86.4a	100

Small letters indicate significant differences ($P < 0.05$) between treatments within each year and variety according to REGQW test.

TABLE 2 | Expression patterns of the genes primary implicated in the flowering process significantly different between Dottato and Petrelli.

Gene	Enzyme	Source		Ficus Locus	Expression level				Differential expression
		Specie	Accession number		DB	PB	DM	PM	
		Ficus carica							
FcFT	flowering locus T		AB594722.1	s01389g28557	12.1	1.5	2.1	30.3	–
ACO2	1-aminocyclopropane-1-carboxylate oxidase 2		KP892660.1	s00021g02797	0.1	0.1	0.0	0.3	–
ACS4	1-aminocyclopropane-1-carboxylate synthase 4		KP892659.1	s00001g00020	3.0	2.3	3.1	8.1	–
ACO3	1-aminocyclopropane-1-carboxylate oxidase-like 3		KP892661.1	s00803g24021	11.3	3.4	34.6	18.8	–
ACS1L	1-aminocyclopropane-1-carboxylate synthase-like 1L		KP892658.1	s00311g15561	0.4	0.0	0.4	3.7	–
		Morus notabilis							
AG	Floral homeotic protein AGAMOUS		EXC21999.1	s00824g24272	113.0	78.6	22.9	31.2	Up
			EXC21999.1	s00026g03239	145.3	116.3	43.6	26.9	Up
AP1	Floral homeotic protein APETALA 1		EXB44879.1	s00016g02309	661.8	823.9	353.3	472.1	–
AP2	Floral homeotic protein APETALA 2		EXB30380.1	s00088g07632	0.2	0.3	0.2	0.4	–
			EXB30380.1	s00991g25958	0.4	0.6	0.2	0.0	–
			EXC24730.1	s00433g18352	118.1	94.2	36.1	46.5	Up
			EXB84815.1	s00019g02645	123.5	121.7	54.6	33.5	Up
DEF	Floral homeotic protein DEFICIENS		EXC35487.1	s00151g10560	0.1	0.1	0.0	0.0	Up
			EXB48382.1	s00042g04637	6.0	6.9	1.8	1.1	Up
FCA	Flowering time control protein FCA		EXC27503.1	augustus_masked-BDEM01000163.1-processed-gene-1.14	43.3	38.0	41.6	58.7	–
FPA	Flowering time control protein FPA		EXB23115.1	s00987g25910	12.3	9.4	16.6	20.7	Down
			EXC35026.1	s00385g17337	12.5	11.9	14.6	15.4	–
FY	Flowering time control protein FY		EXB62656.1	augustus_masked-BDEM01001238.1-processed-gene-0.1	26.8	19.9	23.2	24.5	–

DB, Dottato Breba; DM, Dottato Main crop; PB, Petrelli Breba; PM, Petrelli Main crop.

fruits, fruit-set in Petrelli was only 12.0% but fruit set in Dottato was not affected (93.4%).

In 2017 data were different in the number but not in the trend, maybe because more caprifigs were used for releasing the pollen grains. In particular, fruit-set was always 100% in Dottato, for all the treatments. In Petrelli, fruit-set was higher than the previous year, with 81.5 and 86.4% for open and hand pollination, respectively, but was only 28.6% in non-pollinated (parthenocarpic) fruits (**Table 3**).

Quantitative and qualitative traits showed differences between the pollinated and non-pollinated fruits (**Table 4**). Hand-pollinated Petrelli and Dottato fruits were heavier than open pollinated or parthenocarpic fruits (**Table 4**). Differences arose also for size, with hand pollinated Dottato fruits higher than fruits from the other treatments; hand pollinated Petrelli fruits resulted also larger than fruits of the other two treatments (**Table 4**). Hand pollinated fruits of Dottato ripened earlier than non-pollinated fruits (5–6 days) with higher total soluble solids (TSS) than fruits from other treatments; but TA and pH of hand-pollinated fruits were lower and higher, respectively, than parthenocarpic fruits. The pulp of hand pollinated Dottato fruits was less juicy and with an intense red color and with a large number of drupelets, whereas in the open pollinated and parthenocarpic fruits the color of the pulp was lighter (**Figure S11**). In the case of Petrelli fruits, the lower titratable acidity (TA) was noticed for hand pollinated fruits which also had a higher pH with respect to open pollinated and parthenocarpic fruits (**Table 4**). The taste of the three types of fruits was almost identical.

TABLE 4 | Effects of open, hand, and no pollination (parthenocarpic) on fruit quality of fruits of Petrelli and Dottato.

Treatment	Weight (g)		Height (mm)		Width (mm)		TSS (°Brix)		TA (g/kg)		pH	
	Petrelli	Dottato	Petrelli	Dottato	Petrelli	Dottato	Petrelli	Dottato	Petrelli	Dottato	Petrelli	Dottato
Open Pollination	57.7b	58.7b	45.8	52.9b	47.3b	49.5	19.8	23.3b	1.26a	1.22c	5.47b	5.80a
Parthenocarpic	60.4b	55.0b	47.6	50.8b	49.7b	48.4	19.6	22.3c	1.22a	2.14a	5.47b	5.00c
Hand pollination	80.5a	64.4a	48.8	61.7a	55.1a	50.1	19.1	24.7a	1.10b	1.70b	5.57a	5.28b

Small letters indicate significant differences (P <0.05) between treatments within each year and variety according to REGQW test.

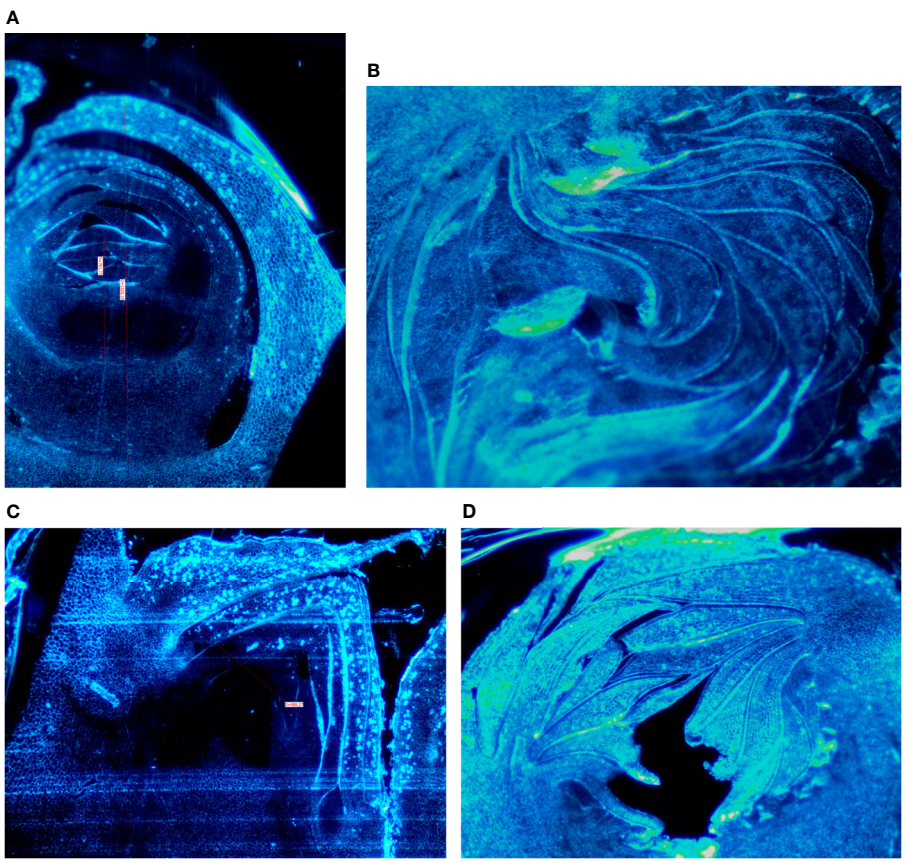

FIGURE 1 | Section of fruit buds of Petrelli in distal nodes **(A)** and basal ones **(B)**. Section of fruit buds of Dottato in distal nodes **(C)** and basal ones **(D)**.

Microscopic Buds Analysis

Buds taken from the shoots showed some differences mainly related to the stage of the development, less developed in the distal nodes with respect to the more developed buds at the basal nodes of the current year shoot. In Petrelli, the fruit buds in the distal nodes showed small, elongated and curved ostiolar scales (**Figure 1A**), whereas in the basal fruit buds the ostiolar scales were much larger with cells more developed (presence of florets) in the cavity of the syconium (**Figure 1B**). In Dottato, differences in the development of the fruit buds were also evident; in the smaller fruit buds at the apical position the ostiolar scales were still underdeveloped, and the florets were almost invisible (**Figure 1C**), whereas at the basal nodes buds had ostiolar scales completely developed, and the structures of the florets resulted also visible (**Figure 1D**).

The evolution of brebas buds from the dormant season to summer is clearly visible for both Dottato (**Figure S12**) and Petrelli (**Figure S13**). The greatest development of the brebas of Dottato was between March-April. Similar growth was observed for Petrelli, with another smaller size increase in May-June (**Figures S12** and **S13**).

Microscopic analyses of buds from July until February showed that differentiation of flat receptacles had not occurred at apical/lateral mixed buds until December (**Figure 2A**). Apical/lateral buds showed small undifferentiated inflorescences (main crop) at the basal-nodes of the meristem but not at the apical nodes by the end of January (**Figure 2B**). Conversely, the brebas had already formed florets (not completely differentiated) in July, which were characterized by masses of cylindrical pistillate primordia, and remained at this stage from summer until the end of winter before bud burst (**Figure 2C**). To our knowledge, this was the first time such a long survey of bud development in *Ficus carica* was conducted either on apical/lateral buds or fruit buds. The sections of fruit and mixed buds of Dottato and Petrelli are shown from July to February in **Figures S14** and **S15**, respectively.

X-Ray Analyses

The X-ray images of Dottato showed a great structural similarity between the breba (**Figure 3**x) and the main crop (**Figure 3**y). The two fruits were taken at the same size and, as shown by the microanalyses, the two structures are very similar. The ostiolar scales are well developed (**Figure 3**x, y a) and completed their development before the final growth of both the receptacle and the florets. The fruit bud of the breba is covered (**Figure 3**x b) with several bud scales, since it is an overwintering bud, whereas the fruit bud of the main crop (**Figure 3**y b) develops on the current season's shoot. In the case of Petrelli, **Figures 3**w–z show the two structures which differentiate either for the bud scales, more abundant in the case of the bud of the breba (**Figure 3**w) or for the more elongated shape of the main crop (**Figure 3**z). Ostiolar scales and the florets in the receptacle are almost identical in the two crops, which make them almost undistinguishable from an anatomical point of view in the first developmental stages.

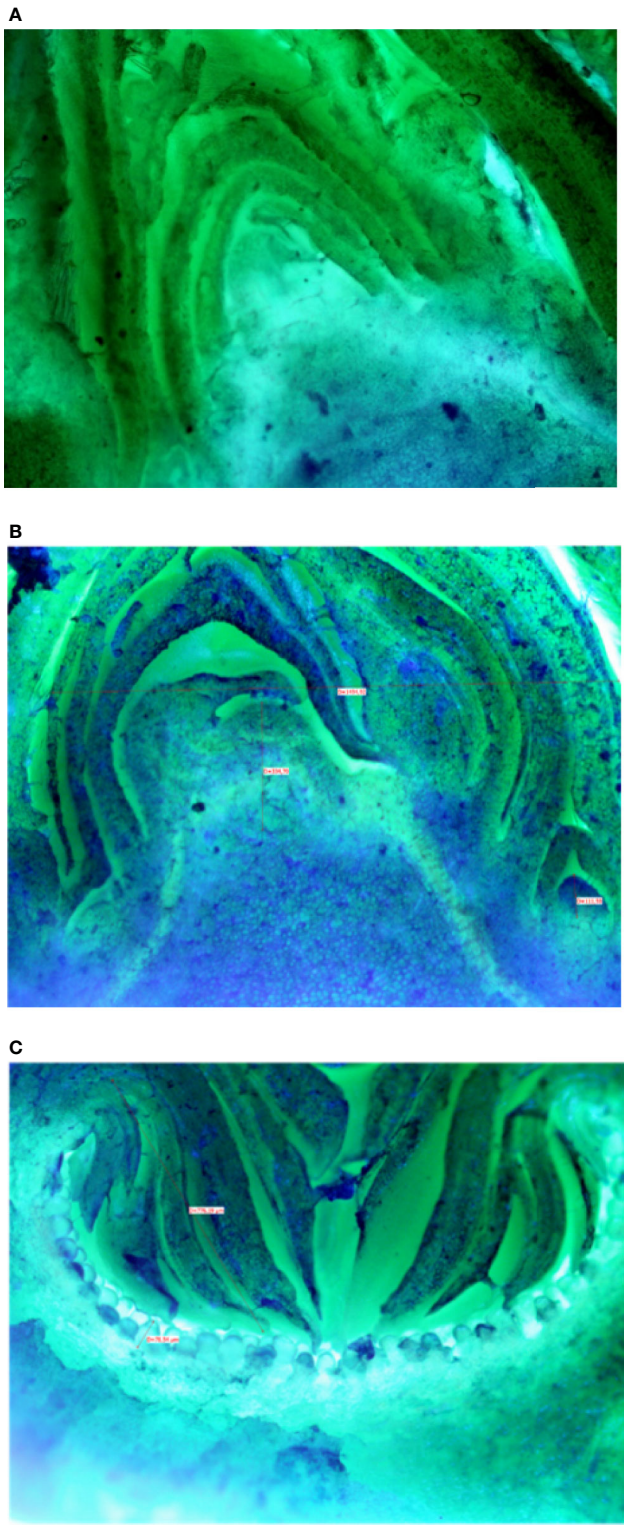

FIGURE 2 | Sections of: the apical bud (mixed) with no signs of inflorescence differentiation in December **(A)**; the apical bud (mixed) with small-undifferentiated inflorescence at the basal node **(B)** in February; a fruit bud (breba) already differentiated **(C)** with florets in the receptacle in summer (August-September).

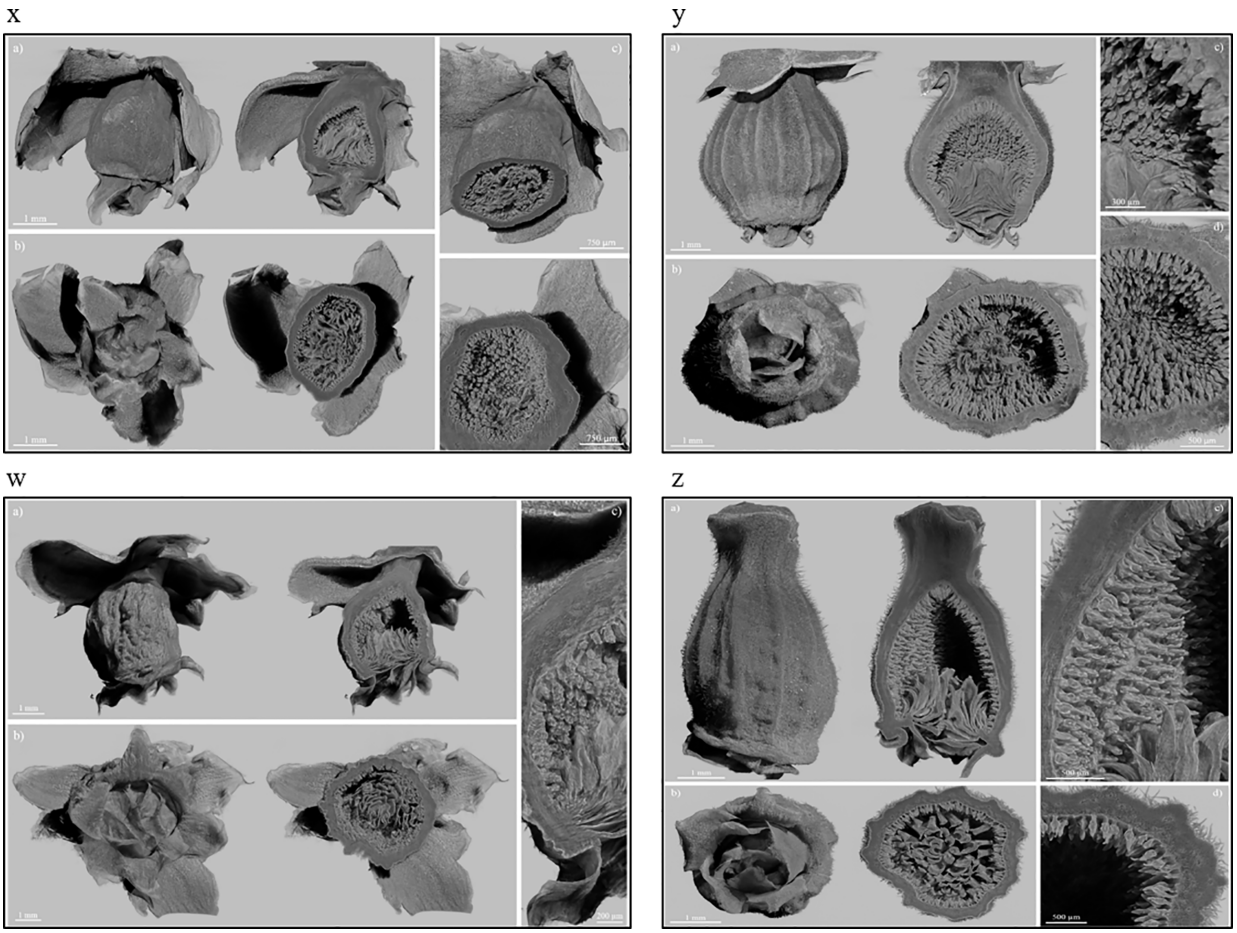

FIGURE 3 | Fruits of Dottato, breba **(x)** and main crop **(y)**, and Petrelli, breba **(w)** and main crop **(z)**, as obtained by X-ray analysis. For each panel, subfigures a) and b) show the frontal and bottom views, respectively, with related internal cross sections; subfigures c) and d) show internal details.

DISCUSSION

Molecular Analyses

In the current study, genes involved in the natural process of flowering and flower differentiation in breba or main crop of fig were investigated. The expression of fig flowering-related and hormone genes was examined to understand their functions in fig fruits and further the regulation mechanisms in this unique species. In particular, we were able to identify twelve genes expressed in the two crops (breba and main crop) and between the two varieties analysed (Dottato and Petrelli) involved in the formation of organs of the fig inflorescence (sepals, petals, stamens, ovary). Among those genes, some of them showed a differential expression pattern underlining the putative role of each of them in flowering process. In particular, the *flowering locus T* gene, which showed high expression level in the main crop of Petrelli, has been identified in other species (e.g., *Arabidopsis thaliana* and cucurbits) as a key gene in the regulation of flowering, in relation to day length changes, (Turck et al., 2008) and floral induction (Lin et al., 2007). Several reports have identified other genes related to flowering, including the *APETALA1* gene (which appeared to be differentially expressed in our analysis) in activating the production of reproductive organs (Abe et al., 2005; Teper-Bamnolker and Samach, 2005; Wigge et al., 2005; Amasino and Michaels, 2010). Similar mechanism has been proposed for the *APETALA2* gene (*AP2*), which in *Arabidopsis* promotes early floral meristem identity (Jofuku et al., 1994) and is required for the transition of an inflorescence meristem into a floral meristem (Drews et al., 1991). In addition, AP2 takes part in the specification of floral identity in *Arabidopsis* plant (Drews et al., 1991; Krogan et al., 2012).

A large number of genes following the flowering time control family (including *FCA*, *FPA*, and *FY* identified in our work) have been identified in grapevine (Kamal et al., 2019), as the main genes involved in the high complex flower development and long duration of bud initiation, and in *Arabidopsis thaliana*, in which promote the transition of the vegetative meristem to reproductive development.

In our analysis, we detected, also, the AGAMOUS gene, which appeared to be highly expressed in Dottato compared to the Petrelli, especially in the brebas (parthenocarpic fruits). In

Arabidopsis this is a transcription factor involved in the control of the organ identity at the early stages of flowers development and interacts genetically with the other homeotic genes such as APETALA2 (Yanofsky et al., 1990; Bowman et al., 1991a; Bowman et al., 1991b; Drews et al., 1991). APETALA2 resulted also upregulated in Dottato with respect to Petrelli, and in general, AGAMOUS, APETALA1, and APETALA2 resulted more expressed in brebas than in the main crop.

Interestingly, we detected a differential expression level of some *1-aminocyclopropane-1-carboxylate oxidase* (*ACO*) genes, with the *ACO3* the most interesting, showing high expression amount in the main crop of Dottato. Apparently the ACO family is not directly related with the flowering process, but catalyzes the final step of ethylene biosynthesis (Houben and Van De Poel, 2019), a pathway usually connected to climacteric fruit ripening and senescence and one of the primary targets of biotechnology for increasing the shelf life of vegetables and fruits. Nevertheless, some studies have demonstrated that *ACO* gene regulation can affect flower formation and development (Houben and Van De Poel, 2019). In detail, the silencing of *ACO* genes can delay flower senescence and abscission in petunia, carnation, and torenia (Savin et al., 1995; Aida et al., 1998; Huang et al., 2007; Tan et al., 2014). Additionally, a mutation of the *ACO2* gene in cucumber affected the sex determination, producing only male flowers (Chen et al., 2016).

Ikegami (Ikegami et al., 2013) reported that FcFT1 mRNA levels were higher in basal nodes (main crop) with respect to distal and younger nodes (breba) thus suggesting a partial correlation between inflorescence differentiation and FcFT1 expression. We observed a higher expression of FT in the main crop of Petrelli with respect to the other fruits. Receptacles with florets differentiated at the same time as FcFT1 expression levels begin to increase, supporting a relationship between FcFT1 continuous expression and flowering as well as fruiting (Ikegami et al., 2013). FcFT1 expression activated by light is indispensable to fruit bearing, including inflorescence differentiation, and the particular two crops of fig (main and breba) in two distinct seasons are probably due to the long-term stable expression of FcFT1 (Ikegami et al., 2013).

We found a higher expression of several genes related to auxin (auxin efflux carrier, auxin response factor, auxin binding protein, auxin responsive protein) and to GA synthesis (GA20ox) in brebas with respect to main crop. Auxin and gibberellin content was higher in brebas than in the main crop of San Pedro cv. King (Lodhi et al., 1969; Chai et al., 2017) to support the growth of parthenocarpic fruits and expression of GA and auxin-biosynthesis gene is repressed in main crop of San Pedro (seeded fruit) with respect to breba, whereas ABA and ethylene-biosynthesis genes are enhanced (Chai et al., 2017). In particular, genes related to GA synthesis (GA20ox) were upregulated in brebas of San Pedro fig at the stages of anthesis and post-anthesis whereas ABA genes were highly expressed in the main crop fruits (Chai et al., 2017).

Caprification

When Petrelli fruits were caprificated (open or hand pollination) fruit-set was always higher than in non caprificated fruits. In contrast, Dottato had high fruit-set regardless of how it was pollinated. Better results (higher fruit-set) were achieved in the second season when pollen was collected from more caprifigs. Important characteristics of the caprifigs (profichi crop) such as fruit size, fruit number per shoot, amount of pollen production and pollen viability can affect the fruit-set and successive quality of fruits; thus, caprification with different types of caprifigs will give different results in terms of fruit-set, yield, and quality as indicated by the data of the second season.

The lower fruit-set of 2016 for the open pollination treatment may be due to the lack of caprification or to the deficiency of caprification caused by weather conditions such as temperature and wind, which can reduce pollinator activity (Oukabli et al., 2003). The differences in hand pollinated fruits collected in 2016 and 2017 are the consequence of the pollen sources used in the two years. In Tunisia, different caprifigs affected both the fruit-set and qualitative parameters such as fruit size, skin thickness, flesh thickness, seed number, TA, and TSS (Gaaliche et al., 2011c). Similarly, in Iran pollen of different caprifigs had a significant effect on fruit length, TSS, ostiole diameter, and percentage of seed germination (Rahemi and Jafari, 2008). Another study conducted in Iran on the cvs Payves and Sabz with three different caprifigs (Avgeizi, Sarbasteh, and Kouhi) demonstrated that the pollen source had significant effect on fruit length, skin color, total soluble solids, total phenolics, total flavonoids, and total anthocyanins (Pourghayoumi et al., 2012). The caprification intensity and more than one type of caprifig in the orchard can either extend the caprification period or affect quality and yield of the main crop (Mars et al., 2017). The differences between hand-pollinated and open-pollinated fruits could be ascribed to both the amount and intensity of pollination. In the case of open-pollinated fruits, the profichi fruits of the three types of caprifigs in the fig orchard were left hanging on the caprifigs, whereas when hand-pollination was accomplished, pollen grains from the profichi of the three caprifigs, were collected and added in the sucrose mixture injected with the syringe directly in the fruit cavity. The mixture of pollen grains used for hand-pollination may have enhanced some characteristics of the fig fruits with respect of the pollen grains carried by the wasps from the profichi fruits in the orchard.

In Turkey, other authors reported that pollinated fruits were distinguished by superior quality (higher TSS, firmness, and more intense external and internal color) and could be better stored than parthenocarpic ones (Aksoy et al., 2003). In our case, better results were obtained with hand pollination of Dottato fruits which ripened earlier and had a red colored pulp. In general, pollination affected the quality parameters of the fruits in particular fruit size and weight (Gaaliche et al., 2011a) but, being pollinated or not, figs develop almost the same concentrations of aroma when ripe and non-pollinated fruits tend to develop more ketones and alcohol compounds (Trad et al., 2012). However, some compounds responsible for pleasant aromas and flavors degenerate when figs reach maturity without being pollinated; that was noticed in the case of butyl and hexyl acetate and beta-ionone in Bouhouli fruits (Trad et al., 2012).

In the case of Brown Turkey figs, ripening pollinated fruit differed from parthenocarpic ones in their shape, which was round for the former and pear-shaped for the latter (Rosianski et al., 2016). Similarly to our results for Dottato (common type), pollinated Brown Turkey figs were heavier than parthenocarpic figs and also larger (Rosianski et al., 2016), whereas Dottato pollinated fruits resulted higher and heavier than parthenocarpic ones in our study. The different color of the pulp between pollinated and parthenocarpic figs was also clearly observed for the inner inflorescence of Brown Turkey throughout fruit development. In particular, in parthenocapric fruits anthocyanin level reached the values of pollinated fruits in much later stages (Rosianski et al., 2016). In general, pollinated ripe fig fruits are much better than parthenocarpic fruits in growth, width, weight, firmness and taste qualities (Rosianski et al., 2016).

Buds Analyses

The analyses of the buds at the beginning of the season, when buds were small and in the growing stage, showed significant differences in the development of the basal buds with respect to the distal buds. The basal buds are devoted to develop the main crop, the middle buds main (late) crop and the distal nodes bear brebas. The different development of the buds can be explained by their different evolution, development in the current season for the basal buds/main crop and development in the successive season for the distal buds/brebas. We observed that from July onward, the distal buds had already developed florets in the receptacle and remained in a dormant stage during autumn-winter time. Apical/lateral mixed buds showed an undifferentiated apex, thus suggesting the differentiation of the inflorescences (breba and main) occurs in the successive growing season. However, at the lower nodes of the apical/lateral buds the presence of undifferentiated primordia was visible at the end of winter (January-February), but no signs of primordia were visible in the middle-distal nodes. It is clear that the inflorescence formation and differentiation process occurs sequentially, from the basal nodes (main crop) toward the middle (main crop-brebas) and distal ones (brebas). The formation of the primordia (main crop) in the basal nodes of the meristem started at the end of winter when temperature arose and successively proceeded during the season to the distal nodes.

With regards of the potted fig plants in the growth chamber, the controlled climatic conditions did not stimulate the development of the distal fruit buds (brebas) which stayed dormant for the successive months. Conversely, light and temperature induced the bud burst of either vegetative or mixed buds with the latter developing main crops on the growing shoots (unpublished data). Environmental factors (light, temperature) were not the limiting factors for fruit buds to burst but probably some endogenous factors (hormones, amino acids) or a limited cold requirement could play a major role for the activation of genes involved in fruit buds burst. The fruit buds of the distal nodes are covered with scales to pass dormancy and protect from winter injury, so they are prone to overwinter instead of developing in the current season although light and temperature were not a limiting factor in the growth chamber. Chill requirements for these fruit buds, although low, were not probably satisfied, and they did not burst. Vegetative and mixed buds reacted to the climatic conditions and burst with the formation of new shoots and developing main crops. In summer in the field, all the buds on the current year shoot are differentiated but some of them enlarge and grow as main crop and others do not and will develop (or drop or become latent) in the successive year as brebas.

The most relevant fruit tree species such as sweet cherry, apple, peach but also grape have flower (mixed) buds which partially differentiate during the summer of the previous year (such as the fruit buds of brebas) and the process is completed at the beginning of the successive season (spring) for a two-year process (Koutinas et al., 2010). In the case of figs, the distal fruits buds (brebas) differentiate in two seasons (spring-summer of year 1 and spring of year 2, such as the inflorescence of grape) whereas inflorescence of the main crop differentiates in the same year of formation since at the end of winter only rudimentary primordia are visible in the basal nodes of the apical/lateral mixed buds. In this study, we showed the differentiation started at the basal nodes at the beginning of the growing season and proceeded in the successive nodes during the season. The singularity of the fig is that only some differentiated inflorescences develop in the season (main crop) whereas others, start to differentiate (florets present), but do not continue to grow and develop during the current season. Molecular analyses also indicated a higher expression level of APETALA1 and APETALA2, AGAMOUS genes involved in flower organ formation, in brebas of both Dottato and Petrelli with respect to the main crop. The higher expression of AGAMOUS (required for stamen development) in breba may suggest a possible presence of staminate flowers at the ostiole of brebas in the origin of fig during its evolution. Higher expression of auxin and gibberellin genes in brebas may confirm this hypothesis, since brebas are parthenocarpic (persistent) fruits with no seeds and hormones are necessary for development and ripening of fruits. When the fig was monoecious, the growth of brebas was probably necessary to allow pollen dispersal of male flowers by the wasps, as in the profichi of the caprifig. AGL11 highly expressed in brebas and upregulated in Dottato is a possible candidate gene for parthenocarpy of fig fruits (both breba and main crop) as for seedlessness in grape (Mejía et al., 2011). Moreover, the ethylene-synthesis gene ACO3 was more expressed in main crop fruits with respect to breba fruits for the ripening of climateric fruits.

A study conducted on five varieties grown in Tunisia showed a growth model with two main vegetative growth flushes, with the first much longer than the second (Gaaliche et al., 2011b). The second flush, generally occurring at the end of summer, is strongly influenced by the climatic conditions, but it never lasts after the end of the summer (Gaaliche et al., 2011b). The buds of brebas are located in the distal portion of the shoot that develop late in the growing season, with respect to the portion where main crops are borne, and these buds may be more susceptible to erratic climatic conditions at the end of summer. In a three-year study, the main crop amount exerted a strong influence on the fruiting intensity of the successive year; in particular, a heavy main crop load decreased the number of brebas in the following year, while a heavy breba crop load reduced the main crop load of the current season (Gaaliche et al., 2011b). In the case of heavy

main crop, the number of fruit buds (brebas) which remain dormant is clearly lower since almost all buds on the current shoot developed during the season. Conversely, the higher number of fruit buds (brebas) retained on 1-year shoots and developing in the successive season will compete with both the current year shoot growth, and the main crop fruits thus reducing the yield of the main crop and the shoot length. The behavior is an alternate bearing as in many tree fruit species to balance between vegetative and reproductive activity.

In this study, in apical (mixed) buds before bud burst (from July of year 1 until March of year 2) we did not observe inflorescence primordia (pistillate florets) but only a rudimentary undifferentiated structure until the end of winter. The mixed bud will differentiate first inflorescences (lower nodes, main crop) in spring of year 2. The flower primordia will completely differentiate in spring-summer both for the main crop and the brebas on all the nodes of the new shoot, since florets are already present in the fruit buds (brebas) in July in the distal nodes of the shoot (**Figures S14** and **S15**).

Moreover, X-ray images showed brebas and main crop were very morphologically similar at the first stages of development with the difference only in the number and thickness of bud scales. The two types of crops are almost undistinguishable at the very beginning of the structure enlargement. It seems that the buds of the brebas are prone to pass a dormant period (number and thickness of scales), whereas the fruits of main crop have lighter scales because have to develop in the current season but may ripe very late in the season (autumn-winter) like the mamme fruits of the caprifig.

Phylogenetic studies have shown that the common fig has a monoecious ancestor (Machado et al., 2001) and successively evolved in a gynodioecious species. The presence of these dormant fruit buds evolving in brebas at the beginning of the season may be a relic of the ancient monoecy, with the wasps entering the different inflorescences (main summer crop, main late crop, breba) during the year to lay the eggs, as they nowadays do in the caprifigs. In fact, the primitive condition in the mutualism fig-wasp, dating back ca. 90 million years ago, was the monoecius breeding system in the fig and the passive pollination in the wasp which was useful for either the fig (seeds) and the wasp (offsprings) (Machado et al., 2001). The particular bud evolution of the fig for main crop (in the current season) and breba (in the successive season) may be because the mutualistic evolution with the wasp in order to host the different generations to both pollinate and ovoposit as happening in the caprifig.

Dioecy was considered as an adaptation to seasonal climates in order to support the *Ficus Blastophaga* symbiosis (Kjellberg et al., 1987). However, Kerdelhué and Rasplus (Kerdelhué and Rasplus, 1996) proposed an alternative evolutionary scenario in which dioecy would have appeared under the selective pressure of non pollinating fig wasps on the mutualism; this would have led to a reduction of ovary layers in the monoecius fig with the development of a higher number of long-styled (and short-styled) flowers in monoecious trees which successively would have evolved towards dioecy, with male and female trees carrying different types of flowers. The higher expression of floral homeotic protein AGAMOUS in breba with respect to main crop may indicate an original role of these fruits (at monoecious stage) for staminate flowers production, as profichi in the caprifig, since this protein is required for normal development of stamens and carpels in the flower. We suggest that the monoecious ancestor of fig had probably male flowers in the brebas for the pollination of main crop. The development of long-styled pistillate flowers (seeds production), loss of staminate flowers in some trees (female) and an increase of short-styled pistillate flowers (gall production) in other trees (male) were the successive steps for the complete evolution towards dioecy. The colonization of temperate environments from tropical ones was probably a consequence of the dioecy more than a cause (Kerdelhué and Rasplus, 1996). The symbiosis between fig and wasps is also under chemical signals used by the figs to attract the wasps into the ostiole (Souza et al., 2015; Mars et al., 2017) as a possible consequence of the evolution towards dioecy in order to make the volatile chemical signals of seed figs (female) identical to those of gall figs (male) to allow pollination (Borges et al., 2008). The monoecious fig in symbiosis with the wasp had two to three types of fruits for the different generations of the wasp to complete the biological cycle. During this evolution towards monoecy, maybe breba inflorescences also had male flowers close to the ostiole to pollinate the main crop to get either seeds for the fig and galls for the wasp. The late main crop maybe acted as overwintering fruit for the wasps as for the mamme in the caprifig and in spring the cycle started again. In an intermediate evolution caprifig fruits were possibly edible, as nowadays may happen for few varieties. After evolution in male and female trees, the fruits became less palatable in the caprifig (offsprings of the wasp) and more palatable and attractive in the edible fig for the production of seeds to be dispersed by animals eating the fruits. In this evolution towards dioecy, the volatile compounds emitted from the male fig were similar to those of the female fig in order to attract the wasps both for seeds and offsprings as a perfect mutualism (Soler et al., 2012).

CONCLUSIONS

This two-year study, using biological, morphological, anatomical, and genetic approaches showed the positive effect of caprification on qualitative and quantitative parameters of the main crop. Moreover, X-ray images and sections of both brebas and main crop showed a very similar structure at the first stages of development. However, the brebas (fruit bud) differentiated in two seasons passing the dormancy period whereas the main crop (mixed bud) differentiated and developed in the same season. The mixed bud started to develop undifferentiated inflorescence primordia at the end of the winter (with increasing temperatures) which will differentiate almost complete inflorescences within 2 to 3 months both for breba and main crop. The main crop completed its development during the season whereas the breba stopped in summer and developed in the successive spring. Molecular analysis identified a set of key genes involved in the flowering process, fruit development and ripening differentially expressed in

both varieties and types of fruit (breba and main crop). This behaviour may be related to the mutualistic coevolution with the pollinating wasp in order to have different fruits to host multiple generations of wasps all year round. A relic of the original monoecy of the species when caprifig and edible fig were probably the same thing, with staminate flowers possibly present in brebas for the pollination of the main crop.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available and can be found in NCBI. BioProject PRJNA623468.

AUTHOR CONTRIBUTIONS

AG and GF design of the work. IM, PC, AM, RT, DN, CP, AT, RA, WS, and GF carried out the analysis. IM, PC, RA, WS, AG, and GF carried out the interpretation of data. IM, PC, AG, and GF have drafted the work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01208/full#supplementary-material

REFERENCES

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., et al. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* 309, 1052–1056. doi: 10.1126/science.1115983

Aida, R., Yoshida, T., Ichimura, K., Goto, R., and Shibata, M. (1998). Extension of flower longevity in transgenic torenia plants incorporating ACC oxidase transgene. *Plant Sci.* 138, 91–101. doi: 10.1016/S0168-9452(98)00139-3

Aksoy, U., Balci, B., Can, H. Z., and Hepaksoy, B. (2003). Some significant results of the research-work in Turkey on fig. *Acta Hortic.* 605, 173–180. doi: 10.17660/ActaHortic.2003.605.26

Amasino, R. M., and Michaels, S. D. (2010). The timing of flowering. *Plant Physiol.* 154, 516–520. doi: 10.1104/pp.110.161653

Borges, R. M., Bessière, J.-M., and Hossaert-Mckey, M. (2008). The chemical ecology of seed dispersal in monoecious and dioecious figs. *Funct. Ecol.* 22, 484–493. doi: 10.1111/j.1365-2435.2008.01383.x

Bowman, J. L., Drews, G. N., and Meyerowitz, E. M. (1991a). Expression of the Arabidopsis floral homeotic gene AGAMOUS is restricted to specific cell types late in flower development. *Plant Cell* 3, 749–758. doi: 10.1105/tpc.3.8.749

Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (1991b). Genetic interactions among floral homeotic genes of Arabidopsis. *Development* 112, 1–20.

Bushnell, B., Rood, J., and Singer, E. (2017). BBMerge – Accurate paired shotgun read merging *via* overlap. *PloS One* 12, e0185056. doi: 10.1371/journal.pone.0185056

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., et al. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18, 188–196. doi: 10.1101/gr.6743907

Chai, L., Wang, Z., Chai, P., Chen, S., and Ma, H. (2017). Transcriptome analysis of San Pedro-type fig (Ficus carica L.) parthenocarpic breba and non-parthenocarpic main crop reveals divergent phytohormone-related gene expression. *Tree Genet. Genomes* 13, 83. doi: 10.1007/s11295-017-1166-4

Chen, H., Sun, J., Li, S., Cui, Q., Zhang, H., Xin, F., et al. (2016). An ACC oxidase gene essential for cucumber carpel development. *Mol. Plant* 9, 1315–1327. doi: 10.1016/j.molp.2016.06.018

Condit, I. J. (1955). Fig varieties: a monograph. *Hilgardia* 23, 323–539. doi: 10.3733/hilg.v23n11p323

Datwyler, S. L., and Weiblen, G. D. (2004). On the origin of the fig: phylogenetic relationships of Moraceae from ndhF sequences. *Am. J. Bot.* 91, 767–777. doi: 10.3732/ajb.91.5.767

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2012). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Drews, G. N., Bowman, J. L., and Meyerowitz, E. M. (1991). Negative regulation of the Arabidopsis homeotic gene AGAMOUS by the APETALA2 product. *Cell* 65, 991–1002. doi: 10.1016/0092-8674(91)90551-9

Ferrara, G., Mazzeo, A., Pacucci, C., Matarrese, A. M. S., Tarantino, A., Crisosto, C., et al. (2016). Characterization of edible fig germplasm from Puglia, southeastern Italy: is the distinction of three fig types (Smyrna, San Pedro and Common) still valid? *Sci. Hortic.* 205, 52–58. doi: 10.1016/j.scienta.2016.04.016

Ferrara, G., Mazzeo, A., Gallotta, A., Pacucci, C., Matarrese, A. M. S., Tarantino, A., et al. (2017). Fruit-set and SSR markers of fig cultivars from Puglia region, Southeastern Italy. *Acta Hortic.* 1173, 39–44. doi: 10.17660/ActaHortic.2017.1173.7

Flaishman, M., Rodov, V., and Stover, E. (2008). The fig: botany, horticulture and breeding. *Hortic. Rev.* 34, 96. doi: 10.1002/9780470380147.ch2

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Gaaliche, B., Hfaiedh, L., Trad, M., and Mars, M. (2011a). Caprification efficiency of three Tunisian fig (Ficus carica L.) cultivars. *J. Nat. Prod. Plant Res.* 1, 20–25.

Gaaliche, B., Lauri, P. E., Trad, M., Costes, E., and Mars, M. (2011b). Interactions between vegetative and generative growth and between crop generations in fig tree (Ficus carica L.). *Sci. Hortic.* 131, 22–28. doi: 10.1016/j.scienta.2011.09.022

Gaaliche, B., Trad, M., and Mars, M. (2011c). Effect of pollination intensity, frequency and pollen source on fig (Ficus carica L.) productivity and fruit quality. *Sci. Hortic.* 130, 737–742. doi: 10.1016/j.scienta.2011.08.032

Glauert, A. M. (1974). *Fixation, Dehydration and Embedding of Biological Specimens* (Amsterdam: North-Holland).

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Houben, M., and Van De Poel, B. (2019). 1-Aminocyclopropane-1-Carboxylic Acid Oxidase (ACO): the enzyme that makes the plant hormone ethylene. *Front. Plant Sci.* 10, 695. doi: 10.3389/fpls.2019.00695

Huang, L.-C., Lai, U.-L., Yang, S.-F., Chu, M.-J., Kuo, C.-I., and Tsai, M.-F. (2007). Delayed flower senescence of Petunia hybrida plants transformed with antisense broccoli ACC synthase and ACC oxidase genes. *Postharvest Biol. Technol.* 46, 47–53. doi: 10.1016/j.postharvbio.2007.03.015

Ikegami, H., Nogata, H., Inoue, Y., Himeno, S., Yakushiji, H., Hirata, C., et al. (2013). Expression of FcFT1, a FLOWERING LOCUS T-like gene, is regulated by light and associated with inflorescence differentiation in fig (Ficus carica L.). *BMC Plant Biol.* 13, 216. doi: 10.1186/1471-2229-13-216

Jofuku, K., Denboer, B. G. W., Vanmontagu, M., and Okamuro, J. K. (1994). Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *Plant Cell* 6, 1211–1225. doi: 10.1105/tpc.6.9.1211

Kamal, N., Ochßner, I., Schwandner, A., Viehöver, P., Hausmann, L., Töpfer, R., et al. (2019). Characterization of genes and alleles involved in the control of flowering time in grapevine. *PloS One* 14, e0214703. doi: 10.1371/journal.pone.0214703

Kerdelhué, C., and Rasplus, J. Y. (1996). The evolution of dioecy among Ficus (Moraceae): an alternative hypothesis involving non-pollinating fig wasp pressure on the Fig-Pollinator Mutualism. *Oikos* 77, 163–166. doi: 10.2307/3545597

Kjellberg, F., Gouyon, P. H., Ibrahim, M., Raymond, M., and Valdeyron, G. (1987). The stability of the symbiosis between Dioecious figs and their pollinators: a study of Ficus carica L. and Blastophaga psenes L. *Evolution* 41, 693–704. doi: 10.1111/j.1558-5646.1987.tb05846.x

Koutinas, N., Pepelyankov, G., and Lichev, V. (2010). Flower induction and flower bud development in apple and sweet cherry. *Biotechnol. Biotechnol. Equip.* 24, 1549–1558. doi: 10.2478/V10133-010-0003-9

Krogan, N. T., Hogan, K., and Long, J. A. (2012). APETALA2 negatively regulates multiple floral organ identity genes in *Arabidopsis* by recruiting the co-repressor TOPLESS and the histone deacetylase HDA19. *Development* 139, 4180–4190. doi: 10.1242/dev.085407

Liao, Y., Smyth, G. K., and Shi, W. (2013). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Lin, M.-K., Belanger, H., Lee, Y.-J., Varkonyi-Gasic, E., Taoka, K.-I., Miura, E., et al. (2007). FLOWERING LOCUS T protein may act as the long-distance florigenic signal in the Cucurbits. *Plant Cell* 19, 1488–1506. doi: 10.1105/tpc.107.051920

Lodhi, F., Bradley, M. V., and Crane, J. C. (1969). Auxins and Gibberellin-like substances in Parthenocarpic and non-parthenocarpic Syconia of Ficus carica L., cv. King. *Plant Physiol.* 44, 555–561. doi: 10.1104/pp.44.4.555

Machado, C. A., Jousselin, E., Kjellberg, F., Compton, S. G., and Herre, E. A. (2001). Phylogenetic relationships, historical biogeography and character evolution of fig-pollinating wasps. *Proc. Biol. Sci.* 268, 685–694. doi: 10.1098/rspb.2000.1418

Marcotuli, I., Mazzeo, A., Nigro, D., Giove, S., Giancaspro, A., Colasuonno, P., et al. (2019). Analysis of genetic diversity of *Ficus carica* L. (Moraceae) collection using simple sequence repeat (SSR) markers. *Acta Sci. Polonorum Hortorum Cultus* 18, 93–109. doi: 10.24326/asphc.2019.4.9

Mars, M., Trad, M., and Gaaliche, B. (2017). The unique fig caprification system and its effects on productivity and fruit characteristics. *Acta Hortic.* 1173, 127–136. doi: 10.17660/ActaHortic.2017.1173.22

Mejía, N., Soto, B., Guerrero, M., Casanueva, X., Houel, C., and Miccono, M. (2011). Molecular, genetic and transcriptional evidence for a role of VvAGL11 in stenospermocarpic seedlessness in grapevine. *BMC Plant Biol.* 11, 57. doi: 10.1186/1471-2229-11-57

Mori, K., Shirasawa, K., Nogata, H., Hirata, C., Tashiro, K., Habu, T., et al. (2017). Identification of RAN1 orthologue associated with sex determination through whole genome sequencing analysis in fig (Ficus carica L.). *Sci. Rep.* 7, 41124. doi: 10.1038/srep41124

Oukabli, A., Mamouni, A., Laghezali, M., Ater, M., Roger, J. P., and Khadari, B. (2003). Local caprifig tree characterization and analysis of interest for pollination *Acta Hortic.* 605, 61–64. doi: 10.17660/ActaHortic.2003.605.7

Pourghayoumi, M., Bakhshi, D., Rahemi, M., and Jafari, M. (2012). Effect of pollen source on quantitative and qualitative characteristics of dried figs (Ficus carica L.) cvs 'Payves' and 'Sabz' in Kazerun – Iran. *Sci. Hortic.* 147, 98–104. doi: 10.1016/j.scienta.2012.08.026

Rahemi, M., and Jafari, M. (2008). Effect of caprifig type on quantity and quality of estahban dried fig ficus carica Cv. SABZ. *Acta Hortic.* 798, 249–252. doi: 10.17660/ActaHortic.2008.798.35

Rau, A., Gallopin, M., Celeux, G., and Jaffrézic, F. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics* 29, 2146–2152. doi: 10.1093/bioinformatics/btt350

Robinson, M. D., Mccarthy, D. J., and Smyth, G. K. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Rosianski, Y., Freiman, Z. E., Cochavi, S. M., Yablovitz, Z., Kerem, Z., and Flaishman, M. A. (2016). Advanced analysis of developmental and ripening characteristics of pollinated common-type fig (Ficus carica L.). *Sci. Hortic.* 198, 98–106. doi: 10.1016/j.scienta.2015.11.027

Savin, K. W., Baudinette, S. C., Graham, M. W., Michael, M. Z., Nugent, G. D., Lu, C.-Y., et al. (1995). Antisense ACC Oxidase RNA delays carnation petal senescence *HortScience* 30, 970–972. doi: 10.21273/HORTSCI.30.5.970

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Siniscalchi, A. (1911). La coltivazione del fico nel Cilento. *Bol. Arbor. Ital.* 7, 25–54.

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M., and Kelly, S. (2016). TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. *Genome Res.* 26, 1134–1144. doi: 10.1101/gr.196469.115

Soler, C. C. L., Proffit, M., Bessière, J.-M., Hossaert-Mckey, M., and Schatz, B. (2012). Evidence for intersexual chemical mimicry in a dioecious plant. *Ecol. Lett.* 15, 978–985. doi: 10.1111/j.1461-0248.2012.01818.x

Solorzano Zambrano, L., Usai, G., Vangelisti, A., Mascagni, F., Giordani, T., Bernardi, R., et al. (2017). Cultivar-specific transcriptome prediction and annotation in Ficus carica L. *Genomics Data* 13, 64–66. doi: 10.1016/j.gdata.2017.07.005

Souza, C. D., Pereira, R. A. S., Marinho, C. R., Kjellberg, F., and Teixeira, S. P. (2015). Diversity of fig glands is associated with nursery mutualism in fig trees. *Am. J. Bot.* 102, 1564–1577. doi: 10.3732/ajb.1500279

Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33, W465–W467. doi: 10.1093/nar/gki458

Storey, W. B. (1977). *The Fig: Its Biology, History, Culture, and Utilization* (Riverside, CA: Jurupa Mountains Cultural Center).

Stover, E., Aradhya, M., Ferguson, L., and Crisosto, C. H. (2007). The fig: overview of an ancient fruit. *HortScience* 42, 1083–1087. doi: 10.21273/HORTSCI.42.5.1083

Tan, Y., Liu, J., Huang, F., Guan, J., Zhong, S., Tang, N., et al. (2014). PhGRL2 protein, interacting with PhACO1, is involved in flower senescence in the petunia. *Mol. Plant* 7, 1384–1387. doi: 10.1093/mp/ssu024

Teper-Bamnolker, P., and Samach, A. (2005). The flowering integrator FT regulates *SEPALLATA3* and *FRUITFULL* accumulation in *Arabidopsis* leaves. *Plant Cell* 17, 2661–2675. doi: 10.1105/tpc.105.035766

Trad, M., Ginies, C., Gaaliche, B., Renard, C. M. G. C., and Mars, M. (2012). Does pollination affect aroma development in ripened fig [Ficus carica L.] fruit? *Sci. Hortic.* 134, 93–99. doi: 10.1016/j.scienta.2011.11.004

Turck, F., Fornara, F., and Coupland, G. (2008). Regulation and identity of Florigen: FLOWERING LOCUS T moves center stage. *Annu. Rev. Plant Biol.* 59, 573–594. doi: 10.1146/annurev.arplant.59.032607.092755

Usai, G., Mascagni, F., Giordani, T., Vangelisti, A., Bosi, E., and Zuccolo, A. (2020). Epigenetic patterns within the haplotype phased fig (Ficus carica L.) genome. *Plant J.* 102 (3), 600–614. doi: 10.1111/tpj.14635

Vallese, F. (1909). *Il fico* (Catania, Italy: F. Battiato).

Wigge, P. A., Kim, M. C., Jaeger, K. E., Busch, W., Schmid, M., Lohmann, J. U., et al. (2005). Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science* 309, 1056–1059. doi: 10.1126/science.1114358

Yanofsky, M. F., Ma, H., Bowman, J. L., Drews, G. N., Feldmann, K. A., and Meyerowitz, E. M. (1990). The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. *Nature* 346, 35–39. doi: 10.1038/346035a0

Conflict of Interest: RA and WS were employed by the company *Sequentia Biotech SL*.

The remaining 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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 05 January 2021
doi: 10.3389/fpls.2020.590846

Homozygosity Mapping Reveals Population History and Trait Architecture in Self-Incompatible Pear (*Pyrus* spp.)

Satish Kumar[1], Cecilia Hong Deng[2], Martin Hunt[3], Chris Kirk[3], Claudia Wiedow[3], Daryl Rowan[3], Jun Wu[4] and Lester Brewer[5]*

[1] Hawke's Bay Research Centre, The New Zealand Institute for Plant and Food Research Limited, Havelock North, New Zealand, [2] Mount Albert Research Centre, The New Zealand Institute for Plant and Food Research Limited, Auckland, New Zealand, [3] Palmerston North Research Centre, The New Zealand Institute for Plant and Food Research Limited, Palmerston North, New Zealand, [4] Centre of Pear Engineering Technology Research, Nanjing Agricultural University, Nanjing, China, [5] Motueka Research Centre, The New Zealand Institute for Plant and Food Research Limited, Motueka, New Zealand

Runs of homozygosity (ROH) have been widely used to study population history and trait architecture in humans and livestock species, but their application in self-incompatible plants has not been reported. The distributions of ROH in 199 accessions representing Asian pears (45), European pears (109), and interspecific hybrids (45) were investigated using genotyping-by-sequencing in this study. Fruit phenotypes including fruit weight, firmness, Brix, titratable acidity, and flavor volatiles were measured for genotype-phenotype analyses. The average number of ROH and the average total genomic length of ROH were 6 and 11 Mb, respectively, in Asian accessions, and 13 and 30 Mb, respectively, in European accessions. Significant associations between genomic inbreeding coefficients (F_{ROH}) and phenotypes were observed for 23 out of 32 traits analyzed. An overlap between ROH islands and significant markers from genome-wide association analyses was observed. Previously published quantitative trait loci for fruit traits and disease resistances also overlapped with some of the ROH islands. A prominent ROH island at the bottom of linkage group 17 overlapped with a recombination-supressed genomic region harboring the self-incompatibility locus. The observed ROH patterns suggested that systematic breeding of European pears would have started earlier than of Asian pears. Our research suggest that F_{ROH} would serve as a novel tool for managing inbreeding in gene-banks of self-incompatible plant species. ROH mapping provides a complementary strategy to unravel the genetic architecture of complex traits, and to evaluate differential selection in outbred plants. This seminal work would provide foundation for the ROH research in self-incompatible plants.

Keywords: ***Pyrus*, runs of homozygosity, signatures of selection, genotyping-by-sequencing, inbreeding, germplasm, quantitative trait loci, genome wide association study**

OPEN ACCESS

Edited by:
Agata Gadaleta, University of Bari Aldo Moro, Italy

Reviewed by:
Lorenzo León, IFAPA Centro Alameda del Obispo, Spain
Markus Neuditschko, Agroscope, Switzerland

****Correspondence:***
Satish Kumar
satish.kumar@plantandfood.co.nz

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

Received: *03 August 2020*
Accepted: *04 December 2020*
Published: *05 January 2021*

Citation:
Kumar S, Deng CH, Hunt M, Kirk C, Wiedow C, Rowan D, Wu J and Brewer L (2021) Homozygosity Mapping Reveals Population History and Trait Architecture in Self-Incompatible Pear (Pyrus spp.). Front. Plant Sci. 11:590846. doi: 10.3389/fpls.2020.590846

INTRODUCTION

Systematic genetic improvement of outbred plants for economically important traits such as yield, consumer acceptance and nutritional value, has led to the loss of genetic diversity within and among accessions in domesticated gene pools (Van de Wouw et al., 2010; Smýkal et al., 2018). Mating among related individuals would lead to inbreeding, which increases the level of homozygosity and

reduces recombination frequency in the genome (Charlesworth, 2003). In addition to inbreeding, long tracts of consecutive homozygous segments in the genome can arise through mechanisms such as natural and artificial selection, genetic drift and population bottlenecks (Ceballos et al., 2018). Runs of homozygosity (ROH), first described by Gibson et al. (2006), are successive homozygous segments of the genome where the two haplotypes inherited from the parents are identical-by-descent. Consanguinity would result in long ROH, whereas larger populations have fewer, shorter ROH (Ceballos et al., 2018).

As recombination interrupts long chromosome segments over time, the length of ROH segment depends in part on the number of generations since the parents shared an ancestor in common (Curik et al., 2014). In an inbred population we would expect to see longer homozygous segments than in outbred populations. Long ROH could still be observed in outbred accessions, perhaps due to unusual mutation and recombination suppression at certain genomic locations. The shorter ROH would indicate the presence of more ancient relatedness which is unaccounted for in the absence of the individual's historic pedigree record. Therefore, the extent and frequency of ROH could reveal population history of a species, such as inbreeding, change of population size, and admixture (Ceballos et al., 2018; Clark et al., 2019). ROH mapping also allows a comparison of the degree of homozygosity among populations with varying degrees of isolation and inbreeding (Kirin et al., 2010).

With the development of cost-effective genome sequencing technologies, large numbers of single nucleotide polymorphisms (SNPs) can be generated at a relatively low price. This facilitates ROH analysis to capture the genomic regions contributing to inbreeding, and thus to assess the breeding history and to identify the genetic components for trait selection. ROH were first recorded in humans by using 8,000 short tandem-repeat polymorphisms (Broman and Weber, 1999). Using ∼700,000 SNPs, Gibson et al. (2006) reported the widespread occurrence of ROH in humans and revealed the harmful effects of recessive deleterious variants present in the ROH regions. Clark et al. (2019) showed genomic inbreeding coefficients (F_{ROH}) derived from ROH, were significantly associated with deleterious effects in humans.

The ROH patterns were shown to differ markedly among cattle breeds (Purfield et al., 2012), and the genomic regions with significant excesses of ROH (termed as ROH islands; Curik et al., 2014) were reported to be associated with signatures of positive selection in horses (Metzger et al., 2015; Grilz-Seger et al., 2019). Beynon et al. (2015) used ROH to reveal population history and structure in a sheep population. ROH analysis was shown to be in agreement with other approaches (e.g., genome-wide association (GWA); haplotype analysis and signatures of selection) to identify the SLICK hair locus in cattle (Huson et al., 2014). Biscarini et al. (2014) also showed agreement between ROH-based and GWA methods to identify quantitative trait loci (QTLs) in farm animals. ROH-guided analyses have also been shown to be a reliable tool for the design of mating schemes to minimize inbreeding (Toro and Varona, 2010; Biscarini et al., 2014).

Inbreeding usually results in the loss of vigor and reduced reproductive fitness of offspring in outbred plant species (Moore and Janick, 1975; Angeloni et al., 2011). Lander and Botstein (1987) suggested that the deleterious recessive variants can be identified in inbred individuals by the presence of long homozygous regions. In the process of evolution and the development of new cultivars under the influence of different mating systems, directional selection, different population sizes and development histories would generate unique ROH distribution patterns in the plant genome; therefore, the number, length, distribution and frequency of ROH in plant genomes would provide rich genetic background information, such as population histories and inbreeding levels.

Despite many studies in humans and livestock populations, the use of ROH to infer inbreeding, population history and trait architecture has apparently not been explored in outbred plant species. The availability of reference genomes and cost-effective genotyping technologies provide an excellent opportunity to evaluate the use of ROH, which still appears an unexplored research field in outbred plants. Pear (*Pyrus* spp.), which exhibits gametophytic self-incompatibility, is among the important temperate fruit tree species, with at least 3,000 years of cultivation history. The genus *Pyrus* is believed to have originated in the mountainous regions of western China (Wu et al., 2018). Pear is commercially grown in more than 50 countries in different geographical regions, but *Pyrus communis* is the predominant species cultivated in Europe, and the major cultivated species in Asia include *P. pyrifolia*, *P.* bretschneideri, *P. sinkiangensis,* and *P. ussuriensis*. Asian pears display a crisp texture, while the European pear is well known for buttery and juicy texture. Various pear breeding programs use interspecific hybrids to develop cultivars with novel combinations of texture and flavor (Brewer and Palmer, 2010).

Natural and artificial selection, as well as independent evolution, has resulted in *Pyrus* species that differ extensively especially in their fruit characteristics (Wu et al., 2018). Different pear species could conceptually be considered as subpopulations, so investigation of ROH patterns would provide insight into their disparate histories. Here we use *Pyrus* as an example to demonstrate application of the ROH concept to investigate population history and trait architecture in self-incompatible outbred plant species.

MATERIALS AND METHODS

Plant Material, Phenotyping, and Genotyping

Accessions of the European and Asian pear species were imported into New Zealand to initiate an interspecific hybrid breeding program in 1983 (Brewer and Palmer, 2010). The successive generations of hybrids were mainly developed from crosses among a few selected hybrids from the previous generation. The imported accessions, as well as the selections from the interspecific hybrid program were propagated over a number of years and planted in duplicate at the Plant and Food Research (PFR) Pear Repository for further assessment and long-term

conservation. All trees received standard orchard management for nutrition, pesticide, and irrigation. Six fruit from each plant in the repository were harvested over two consecutive years 2014 and 2015. An average value of six fruits was used to represent each phenotype of each accession.

For the purpose of this study, a total of 199 accessions, including 46 representing Asian species (36 *P. pyrifolia*, 10 *P.* × *bretschneideri*), 108 of European pear (*P. communis*), and 45 hybrids between Asian and European species (**Table 1**) were sampled. Protocol for fruit harvesting and assessment were as reported earlier (Kumar et al., 2017); six fruit from each accessions were stored for 28 days at 3°C, then a further 1 day at 20°C before evaluation. Skin russet coverage (RUS), sensory flavor intensity (FINT) and skin bitterness (BIT) were scored on intensity scales where 0 = none and 9 = highest. Scuffing (SCUF) was also rated on a 0–9 scale (0 = no darkening; 9 = solid brown or black coloration) after each fruit was firmly rubbed across the cup of a molded pulp fibreboard fruit packing tray and assessed after 2 h. Fruit shape index (SHAP) was measured using a two dimensional shape chart and fruit weight (AVFW) was recorded as the average weight of the six fruit. Fruit firmness (FF) was determined on opposite sides of each fruit after peel removal using a Fruit Texture Analyzer (GÜSS) fitted with an 11 mm diameter probe tip. Soluble solids concentration (SSC) for each fruit was measured, with the juice expressed during the firmness probe, using a digital refractometer (Atago PR-32). Bulked juice from the cortical flesh of the sample fruit was used to measure titratable acidity (TA) using an automatic acid titrator (Metrohm 716 DMS) and the percentage of malic acid in fruit juice was recorded.

The flavor volatile analysis procedure using GC-MS, was used as described by Rowan et al. (2009), except that fruit was placed in 4-L unused commercial metal paint cans rather than glass jars. Sample size varied from 300 to 1,000 g. Volatiles were collected onto Tenax-TA using an air flow of 55 ml/min for 2 h. After volatile collection, the absorbent traps were eluted with diethyl ether (2 × 1 mL) containing tetradecane at 10 nL mL^{-1} into pre-weighed 4 mL glass vials at a flow rate of 2 mL min^{-1}. Samples were stored at −20°C before analysis using a Waters GCT GC-MS/Agilent 6890N GC equipped with an Optic 3 injector. Volatiles were identified based on their retention indices and by comparison with commercial mass spectral databases and authentic compounds. Generally, base peak intensities were used to aid automated peak identification and integration using Waters QuanLynx software. Fruit volume, and hence surface area, was calculated, and volatile concentrations are reported as ng tetradecane (m/z 57) equivalents released cm^{-2} fruit surface area per hour.

Young leaves were collected in spring 2013 for DNA extraction. Protocols for DNA extraction, genotyping-by-sequencing (GBS; Elshire et al., 2011) library preparation were the same as those reported earlier by Kumar et al. (2017). Briefly, GBS libraries were multiplexed into 5 pools, with 36–55 libraries per pool, for NGS sequencing on the Illumina HiSeq2000 platform and the sequence data were analyzed using TASSEL (Bradbury et al., 2007). The fastq file were mapped to the *P.* × *bretschneideri* (cultivar "Suli") (Xue et al., 2018). SNPs

TABLE 1 | List of pear accessions and their *Pyrus* species group.

CULTIVAR/ Selection	Species	CULTIVAR/ Selection	Species
3189	*P. communis*	NOUVEAU POITER	*P. communis*
2-301	*P. communis*	OLD HOME	*P. communis*
6/23/94	*P. communis*	OTTAWA-291	*P. communis*
6-31-100	*P. communis*	OVID	*P. communis*
6-31-68	*P. communis*	C01	*P. communis*
ANGELYS	*P. communis*	C02	*P. communis*
Aurora	*P. communis*	C03	*P. communis*
AUTUMN BERGAM	*P. communis*	C04	*P. communis*
BEURRE BOSC	*P. communis*	C05	*P. communis*
BEURRE CAPIAMONT	*P. communis*	C06	*P. communis*
BEURRE EASTER	*P. communis*	C07	*P. communis*
BEURRE HARDY	*P. communis*	C08	*P. communis*
BROCKWORTH	*P. communis*	C09	*P. communis*
BROWN BEURRE	*P. communis*	C10	*P. communis*
BUTIRRA PRECOCE MORRETINI	*P. communis*	C11	*P. communis*
BUTIRRA ROSATA MORRETINI	*P. communis*	C12	*P. communis*
CALIFORNIA	*P. communis*	C13	*P. communis*
CARMEN	*P. communis*	C14	*P. communis*
COLETTE	*P. communis*	C15	*P. communis*
CONCORDE	*P. communis*	C16	*P. communis*
CRIMSON GEM COMICE	*P. communis*	C19	*P. communis*
D'Incontinue	*P. communis*	C20	*P. communis*
DOYENNE DU COMICE	*P. communis*	C24	*P. communis*
ELDORADO	*P. communis*	C25	*P. communis*
Elizabeth Cole	*P. communis*	C26	*P. communis*
FERTILITY	*P. communis*	C27	*P. communis*
FLEMISH BEAUTY	*P. communis*	P327-57	*P. communis*
FLORIDA HOME	*P. communis*	PACKHAM'S TRIUMPH	*P. communis*
Gleau Morceau	*P. communis*	PASSA CRASSANA	*P. communis*
GOLDEN RUSSET BOSC	*P. communis*	PATRICK BARRY	*P. communis*
GORHAM	*P. communis*	PATTEN	*P. communis*
GRAND CHAMPION	*P. communis*	Peamy	*P. communis*
HARROW DELIGHT	*P. communis*	PIERRE CORNEILLE	*P. communis*
HIGHLAND	*P. communis*	PRESIDENT D'OSMOND	*P. communis*
HOSKINGS	*P. communis*	President Heron	*P. communis*
HOWELL	*P. communis*	PRINCESS	*P. communis*
HW606 (Harovin Sundown)	*P. communis*	PT AV-63-2076	*P. communis*
JUMBO (STARKS)	*P. communis*	RED ANJOU	*P. communis*
Jupp	*P. communis*	RED SENSATION BARTLETT	*P. communis*

(Continued)

TABLE 1 | Continued

CULTIVAR/ Selection	Species	CULTIVAR/ Selection	Species
LOUISE BON DE JERSEY	*P. communis*	REIMER RED	*P. communis*
Margeurite Marrilat	*P. communis*	ROGUE RED	*P. communis*
MAX RED BARTLETT	*P. communis*	ROSEMARIE	*P. communis*
Moders	*P. communis*	Ruby	*P. communis*
MOONGLOW	*P. communis*	RX359	*P. communis*
Nellie	*P. communis*	RX529	*P. communis*
New York	*P. communis*	RX810	*P. communis*
SIERRA	*P. communis*	Shingo	*P. pyrifolia*
SILVERBELL	*P. communis*	SHINKO	*P. pyrifolia*
STARKING DELICIOUS	*P. communis*	SHINSUI	*P. pyrifolia*
STARKRIMSON	*P. communis*	SUISEI	*P. pyrifolia*
SUPER COMICE	*P. communis*	TAMA	*P. pyrifolia*
SWISS BARTLETT	*P. communis*	WASEAKA	*P. pyrifolia*
TAYLORS GOLD	*P. communis*	Yasato	*P. pyrifolia*
TENN	*P. communis*	B01	*P. bretschneideri*
TN09-46	*P. communis*	B02	*P. bretschneideri*
TOSCA	*P. communis*	B03	*P. bretschneideri*
US307	*P. communis*	PINGGUOLI	*P. bretschneideri*
US56112/46	*P. communis*	QIYUESU	*P. bretschneideri*
UVEDALES ST GERMAINE	*P. communis*	TSULI	*P. bretschneideri*
Velvetine	*P. communis*	XINYALI	*P. bretschneideri*
WINTER NELIS	*P. communis*	XUEHUALI	*P. bretschneideri*
WORDEN SECKLE	*P. communis*	YALI	*P. bretschneideri*
Cangxili	*P. pyrifolia*	Crispie	Hybrid
Choju	*P. pyrifolia*	HWA HONG	Hybrid
CHOJURO	*P. pyrifolia*	MAXIE	Hybrid
DAN BAE	*P. pyrifolia*	H01	Hybrid
DOITSU	*P. pyrifolia*	H02	Hybrid
GION	*P. pyrifolia*	H03	Hybrid
Gold Nijisseiki	*P. pyrifolia*	H04	Hybrid
HAKKO	*P. pyrifolia*	H05	Hybrid
HEISHI	*P. pyrifolia*	H06	Hybrid
HOKUSEI	*P. pyrifolia*	H07	Hybrid
Hougetsu	*P. pyrifolia*	H08	Hybrid
IMAMURA AKI	*P. pyrifolia*	H09	Hybrid
NIITAKA	*P. pyrifolia*	H10	Hybrid
P01	*P. pyrifolia*	H11	Hybrid
P02	*P. pyrifolia*	H12	Hybrid
P03	*P. pyrifolia*	H13	Hybrid
P04	*P. pyrifolia*	H14	Hybrid
P05	*P. pyrifolia*	H15	Hybrid
P07	*P. pyrifolia*	H16	Hybrid
P08	*P. pyrifolia*	H17	Hybrid
P09	*P. pyrifolia*	H18	Hybrid
P10	*P. pyrifolia*	H19	Hybrid
P11	*P. pyrifolia*	H20	Hybrid
P12	*P. pyrifolia*	H21	Hybrid
P13	*P. pyrifolia*	H22	Hybrid
P14	*P. pyrifolia*	H23	Hybrid
P15	*P. pyrifolia*	H27	Hybrid
P16	*P. pyrifolia*	H31	Hybrid
P17	*P. pyrifolia*	H32	Hybrid
Red Hosui	*P. pyrifolia*	H33	Hybrid
		H34	Hybrid
		H35	Hybrid
		H36	Hybrid
		H37	Hybrid
		H38	Hybrid
		H40	Hybrid
		H41	Hybrid
		H42	Hybrid
		H45	Hybrid
		H46	Hybrid
		H47	Hybrid
		H48	Hybrid
		H49	Hybrid
		H50	Hybrid
		H51	Hybrid

(Continued)

with minor allele frequency (MAF) < 0.05, and missing data frequency $> 20\%$ were dropped.

Measurement of ROH

A procedure to discover ROH in PLINK software (Purcell et al., 2007) using a sliding-window approach along the genome is depicted in **Figure 1**. Briefly, a window of pre-determined number of SNPs was examined for homozygosity (allowing pre-determined number of heterozygous and missing calls) and then, for each SNP, the proportion of "homozygous" windows that overlap that position was calculated. ROH segments were then called based on a threshold for the average (Bjelland et al., 2013). To minimize the number of ROH that occurred by chance, the minimum number of SNPs that constituted a ROH was calculated following Lencz et al. (2007):

$$l = \frac{\log_e \frac{\alpha}{n_s n_i}}{\log_e (1 - het)} \quad (1)$$

where α is the percentage of false-positive ROH (set to 0.05), n_s is the number of SNPs per individual, n_i is the number of individuals and *het* is the mean SNP heterozygosity across all SNPs. Following Equation 1, the minimum number of SNPs constituting an ROH was set to 48 in this study. ROH segments were determined using PLINK v.1.7 based on the following settings: one heterozygous genotype and two missing SNP were allowed per window of 48 SNPs; minimum SNP density was set to one SNP per 50 kb, with a maximum gap between consecutive SNPs was set to 1 Mb to avoid low SNP density affecting ROH length; a minimum ROH length of 500 kb. The adjacent SNPs having a proportion of ROH occurrences over the adopted threshold formed ROH islands.

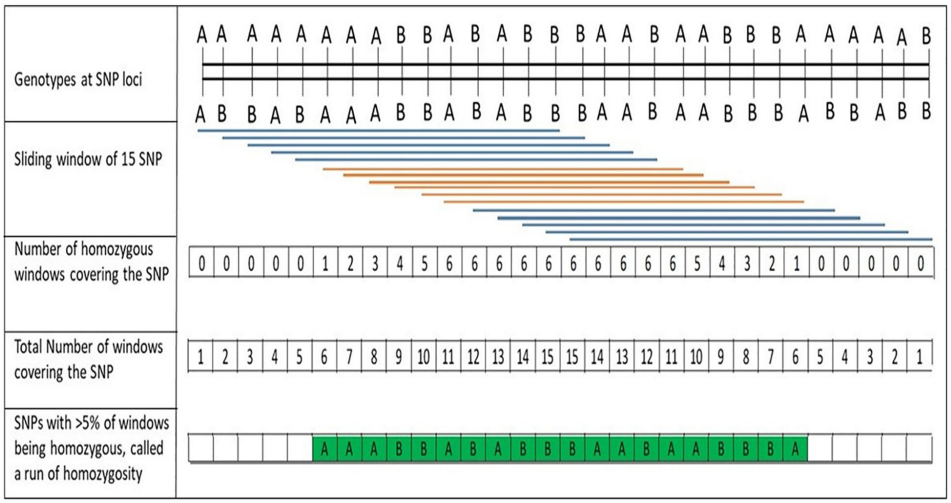

FIGURE 1 | Description of the process for discovery of runs of homozygosity (ROH) using a sliding window of single nucleotide polymorphisms (SNPs) along the chromosome, as implemented with PLINK software (Purcell et al., 2007; Bjelland et al., 2013).

In this study, putative ROH islands were determined based on overlapping ROH regions, shared by at least 15% of studied accessions. The adjacent SNPs were merged into genomic regions corresponding to ROH islands.

Measures of Genomic Inbreeding

For each accession, three estimates of the genomic inbreeding coefficient (F) were calculated, F_{ROH}, F_{SNP}, and F_{GRM}. F_{ROH} is the fraction of each genome in ROH > 0.5 Mb. For example, in a sample for which n ROH of length l_i (in Mb) were identified, then F_{ROH} was calculated as:

$$F_{\mathrm{ROH}} = \frac{1}{L}\sum_{i}^{n} l_i \tag{2}$$

where L represents the genome length. F_{SNP}, which is a method of moment based measure of inbreeding in the most recent generation (Clark et al., 2019), was estimated as follows using PLINK software:

$$F_{\mathrm{SNP}} = \frac{O(HOM) - E(HOM)}{N - E(HOM)} \tag{3}$$

where O(HOM) is the observed number of homozygous SNPs, E(HOM) is the expected number of homozygous SNPs, and N is the total number of genotyped SNPs. F_{GRM}, a genomic relationship-based inbreeding coefficient, was calculated using the method described by VanRaden et al. (2011):

$$G = \frac{ZZ'}{2\sum p(1-p)} \tag{4}$$

where $\boldsymbol{Z}$ is an $n \times m$ matrix (n = number of individuals, m = number of SNP loci) representing genotypes at each locus. The coefficient of the ith column of the $\boldsymbol{Z}$ matrix are ($0–2p_i$), ($1–2p_i$), and ($2–2p_i$) for genotypes AA, AB, and BB, respectively, p_i is the allele frequency of allele A at the ith SNP. $\boldsymbol{G}$ was calculated using $p = 0.5$, which is the same as the method used by the USDA-ARS Animal Improvement Programs (VanRaden et al., 2011). The values on the diagonal of $\boldsymbol{G}$ denote the relationship of an accession to itself, or its genomic inbreeding coefficient (F_{GRM}).

Effect Size Estimates for Quantitative Traits

For each trait, the phenotypes were modeled in two steps. First, a mixed linear model (MLM) was fitted accounting for fixed effects and random effects:

$$y = Xb + Zu + e \tag{5}$$

where y is a vector of measured trait values, b is a vector of unknown fixed covariate effects (e.g., overall mean, year effect), X and Z are the known design matrices for the fixed and random effects, respectively; u is an unknown vector of additive genetic effects with a normal distribution $N(0, \sigma_A{}^2 G)$, where G is the genomic relationship matrix (GRM); and e is an unknown vector of residuals. In the second step, estimates of random additive effects (u') from Equation 5 were regressed on F_{ROH} as follows:

$$u' = \mu + \beta * F_{\mathrm{ROH}} + \varepsilon \tag{6}$$

where μ is the overall mean, β is the unknown scalar effect of F_{ROH} on the trait, F_{ROH} is a known vector of individual F_{ROH}, and ε is an unknown vector of residuals.

Marker-trait genome-wide association (GWA) analysis were also conducted for each trait using unified MLM as implemented in R package GAPIT (Lipka et al., 2012). Principal components (PCs) analysis was used to quantify patterns of population structure, and the first two PCs were used as covariate to avoid spurious marker-trait associations that could arise from population structure. Co-localization of ROH islands with trait-associated SNPs, and overlap with previously reported QTLs (reviewed by De Franceschi and Dondini, 2019), was also investigated.

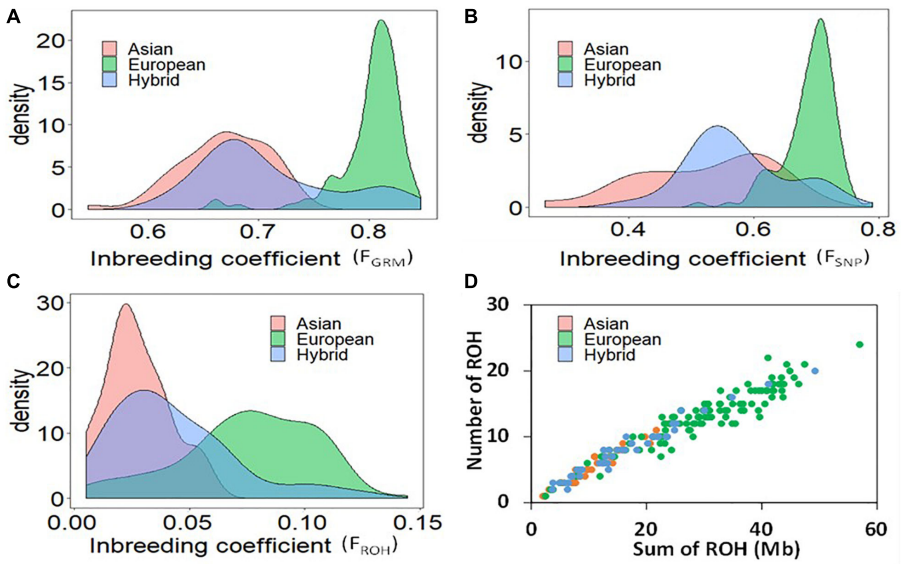

FIGURE 2 | Distribution of the measures of genomic inbreeding coefficients based on genomic relationship matrix **(A)**, Method of moments **(B)**, and runs of homozygosity **(C)**; and the relationship between the total genomic length covered by ROH and the total number of ROH per individual **(D)** in Asian, European and hybrid pear population.

RESULTS

ROH in Different Genetic Groups

After quality controls (i.e., missing data frequency < 20%, minor allele frequency > 0.05), about 8500 SNPs distributed across the genome (**Supplementary Figure S1**) were retained for further analysis. The first principal component (PC1) grouped the Asian and European accessions in two non-overlapping clusters (**Supplementary Figure S2**). The hybrids resided in between the two main clusters, but many hybrid accessions grouped closely with either Asian or European species. **Figures 2A–C** displays the distributions of F_{GRM}, F_{SNP}, and F_{ROH}, respectively, with means of 0.52, 0.67, and 0.03 in Asian accessions; 0.69, 0.80, and 0.08 in European accessions; and 0.57, 0.71, 0.04 in the hybrid population. The accessions with smaller F_{ROH} were considered as the least inbred, whereas accessions with larger F_{ROH} were considered as the most inbred accessions. Correlations between the three measures of genomic inbreeding were large, with correlations between F_{SNP} and F_{ROH} of 0.64, F_{SNP} and F_{GRM} of 0.94, and F_{ROH} and F_{GRM} of 0.74.

The relationship between the total genomic length (Mb) covered by ROH and the total number of ROH per accession demonstrates separation between the European and Asian pear species (**Figure 2D**). The average number of ROH per accession was 6, 13 and 8 for Asian, European and hybrid accessions, respectively. The average total genomic length of ROH was 11, 30 and 17 Mb for Asian, European and hybrids accessions, respectively. One accession of European pear ("Nellie") presented total ROH length of about 60 Mb. The average length of ROH in Asian, European and hybrid accessions was 1.96, 2.29, and 2.08 Mb, respectively. The occurrence of ROH on different linkage groups was generally similar between the two species, but a much higher frequency was observed in Asian accessions on LG4 (12 vs. 4.8%) and LG17 (13.2 vs. 8.7%), while European accessions displayed a higher proportion of ROH on LG13 (6.4 vs. 2.8%) and LG15 (11.4 vs. 4.5%) (**Supplementary Figure S3**).

Classification of ROH by length showed that the majority of ROHs were shorter than 4 Mb in all three pear genetic groups (**Figure 3**). For ROHs shorter than 2 Mb, Asian and European accessions had the highest (ca. 50%) and the lowest (ca. 40%) proportions, respectively. ROH segments 4–6 Mb long were more frequent in European (5%) accessions compared with the Asians (1%). None of the ROH in Asian and hybrid accessions were longer than 6 Mb, but the European accessions showed few ROH close to 10 Mb length.

Trait-F_{ROH} Associations

Trait values were regressed on F_{ROH} to estimate the effect of inbreeding/selection on each of the 32 traits considered in this study, with 23 reaching experiment-wise significance threshold

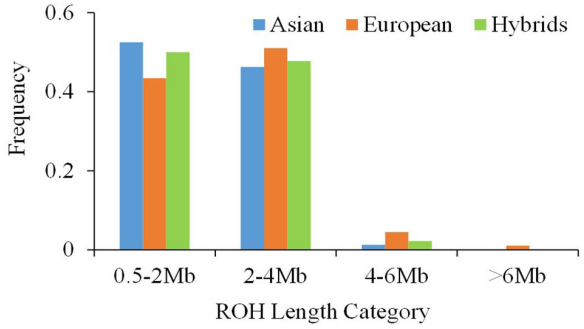

FIGURE 3 | The distribution of the single runs of homozygosity (ROH) length classes within the Asian, European and hybrid pear (*Pyrus* spp.) populations.

(0.01/32 = p < 3.1e-04). Effect size, in phenotypic standard deviation units (σ_p), corresponding to F_{ROH} = 0.15 (equivalent to the maximum value observed in this study) are shown in **Figure 4**. An increase of 0.15 in F_{ROH} was associated with $0.70\sigma_p$ and $1.8\sigma_p$ increase in fruit weight and fruit firmness, respectively. Non-volatile compounds (Brix and TA), which partly influence sensory flavor intensity, increased by about $1.0\sigma_p$ at F_{ROH} = 0.15. Volatile compounds, alcohols and non-ethyl esters (esters not derived from esterification with ethanol), increased with increasing F_{ROH}. Skin bitterness and ethyl esters decreased significantly with increases in F_{ROH} (**Figure 4**).

Islands of ROH

In the significant ROH islands described here using combined samples from all three groups, each SNP showed a percentage of occurrence > 15% (**Figure 5**). This approach resulted in the identification of 20 ROH islands, with a maximum of two ROH islands on some linkage groups (e.g., LGs 1, 4, 8, 10, 12, and 17) and no significant island on LGs 2, 9, and 11. The smallest and the longest significant ROH island were observed on LG5 (0.770–0.772 Mb) and LG15 (12.718–17.131 Mb), respectively (**Supplementary Table S1**). Within the ROH island on LG15, a homozygous haplotype (GCGAAT) comprising six SNPs spanning over a 71 bp region (14,017,541–14,017,612 bp) was shared by 48, 91, and 90% accessions of Asian, European and hybrid populations, respectively. The occurrence of ROH islands was also investigated in each genetic group separately, which revealed some key differences between these groups (**Supplementary Figure S4**).

A search for co-localization of ROH islands with previously published QTLs revealed that 18 out of 20 ROH islands overlapped with QTLs for various traits (**Figure 6**). In addition to numerous QTLs, candidate genes *Vnk* (which confer resistance to *Venturia nashicola*) and *PpAIV3* (controlling conversion of sucrose to hexose in mature fruit) resided within ROH islands on LG1. A 3.2 Mb long ROH island at the bottom of LG17 overlapped with the self-incompatibility gene (*SI-locus*). Using conventional GWA, with a genome-wide significance threshold of p < 0.05, a total of 294 SNPs were found to be significantly associated with at least one of the 33 traits considered in this study. Thirty-seven out of 294 significant SNPs resided within the significant ROH islands (**Figure 6**; **Supplementary Table S2**). A SNP associated with skin russet resided within an ROH island in the upper region of LG8, which has previously been reported a harboring a QTL for this trait.

DISCUSSION

The intense selection in commercial species has necessitated strategies to characterize and monitor inbreeding and maintain genetic diversity in long-term breeding and conservation programs (De Cara et al., 2013; Bosse et al., 2015). The absence of pedigree information on wild and semi-wild accessions makes it difficult to estimate observed levels of inbreeding in the current germplasm resources of fruit crops. Results from studies on livestock species (Toro and Varona, 2010; Bjelland et al., 2013; Peripolli et al., 2017) have shown that

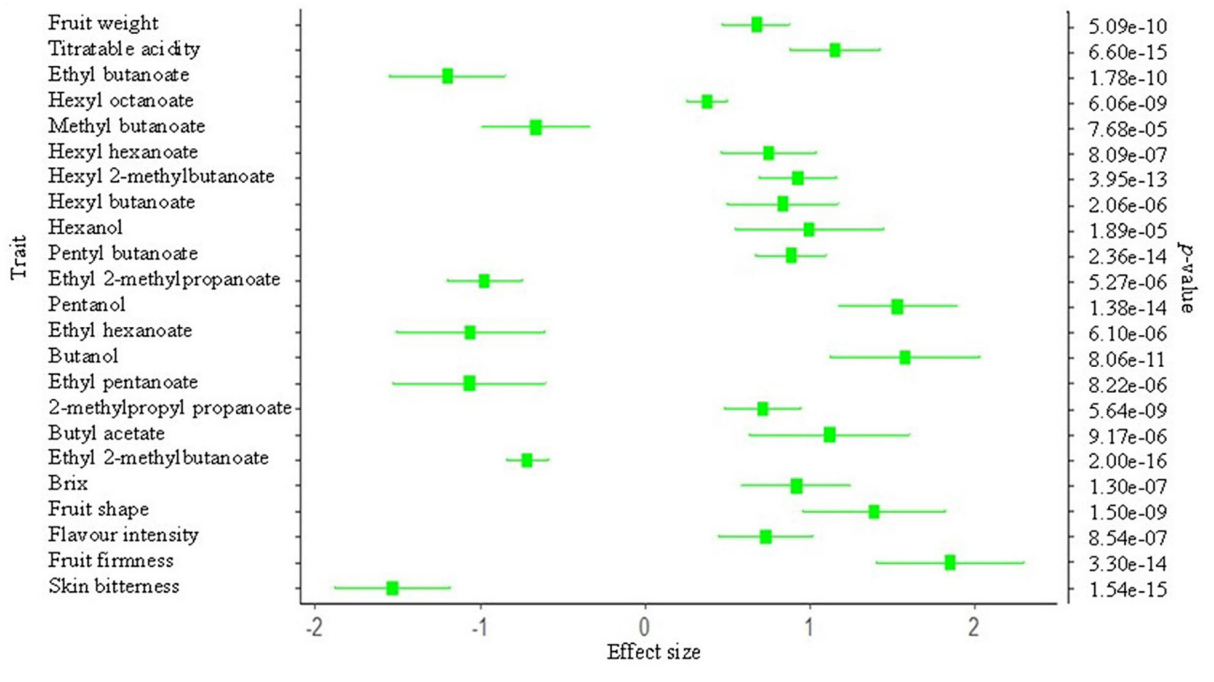

FIGURE 4 | Effect of inbreeding (F_{ROH})/selection on various pear (*Pyrus* spp.) phenotypes. The x-axis represents the effect size (in phenotypic standard deviation unit) estimate of F_{ROH} = 0.15. The secondary *y*-axis shows the probability of significance of the effect size for traits that reached Bonferroni-corrected significance of probability value (=0.01/32 traits).

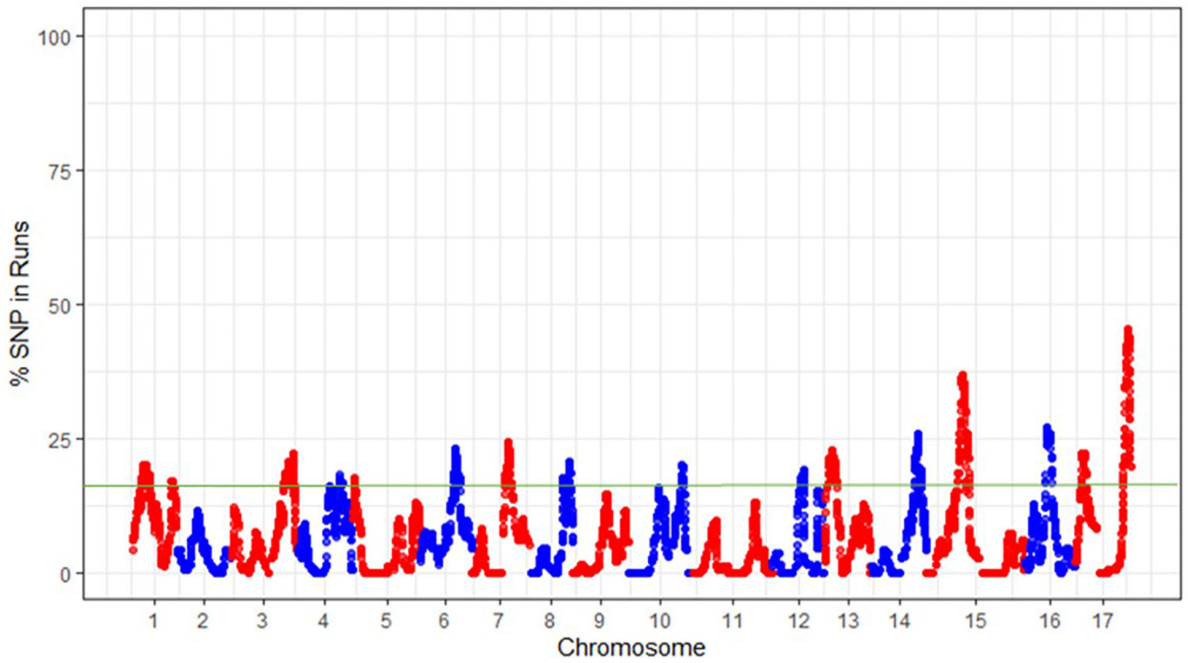

FIGURE 5 | The frequency (%) of single nucleotide polymorphisms (SNPs) occurrence into runs of homozygosity (ROH) islands across the pear (*Pyrus* spp.) genome. The green horizontal line indicates the adopted threshold (15%), which defines the ROH islands.

using genomic inbreeding estimates (F_{ROH}) improves mating decisions and gene conservation efforts. The average inbreeding coefficient (F_{ROH}) in the inter-specific hybrid population was similar to the Asian accessions, but lower compared with the European accessions. The first generation (F_1) hybrid population is expected to display lower F_{ROH} than the parental species. Most of the hybrids used in this study were second-generation accessions selected from the F_2 or BC1 crosses. The clustering patterns from PCA analysis also showed greater resemblance of many hybrid accessions with either the Asian or European species. Therefore, the artificial selection and the mating scheme in the inter-specific breeding program would have led to the observed higher level of F_{ROH} in the studied hybrid accessions. Use of genome sequence information has been advocated for monitoring and utilization of gene-banks of plant species (Henry, 2012; Mascher et al., 2019). Results from our study suggested that F_{ROH} derived from using a high-density genotyping platform would provide a novel tool for managing diversity of plant genetic resources.

Assessment of Runs of Homozygosity

In our study, we used a window size of 48 SNPs to identify ROH in *Pyrus* spp. Studies on dairy cattle have shown that for a window size of 20–50 SNPs, F_{ROH} was more accurate than F_{GRM} derived from the observed genomic relationship matrices (Kim et al., 2013; Forutan et al., 2018). The correlation between F_{ROH} and F_{GRM} was high (0.74), but the substantially higher value of F_{GRM} compared to F_{ROH} (**Figure 2**) is likely due to the fact that the base allele frequencies were not known and F_{GRM} cannot distinguish between alleles that are identity-by-descent and identity-by-state. Very similar correlations were observed between F_{ROH} and F_{GRM} in studies on dairy cattle (Bjelland et al., 2013; Forutan et al., 2018). In our study, the average F_{ROH} in European accessions (0.08) was higher compared with that of the Asian (0.03) accessions. There are no reports of F_{ROH} in self-incompatible plant species, but the much higher F_{ROH} (often $>$ 0.15) for livestock species reflects stronger artificial selection (review by Peripolli et al., 2017) compared with plant species investigated in this study.

Different mating systems, selection directions and population development histories will form unique ROH distribution patterns in the genomes of geographically isolated populations (Bosse et al., 2012). Our results showed that the number and total length of ROH in Asian accessions were shorter than in European accessions, and longer ROH segments ($>$2 Mb) were more frequent in European accessions; suggesting relatively lower nucleotide diversity in European pear. Using whole genome re-sequencing, Wu et al. (2018) observed that Asian pears had a higher nucleotide diversity than European pears, which supports ROH patterns observed in this study. Compared with the Asian group, the European accessions clearly displayed higher individual sums of ROH per accession (**Figure 2**). Taken together, these results suggest that systematic breeding of European pears would have started earlier than Asian pears. The pattern of ROH in the European pears could also be a result of population bottlenecks due to glaciation in Europe as compared to Asia. Similar inferences were drawn when ROH patterns were compared between Asian and European livestock populations (Groenen et al., 2012; Peripolli et al., 2017).

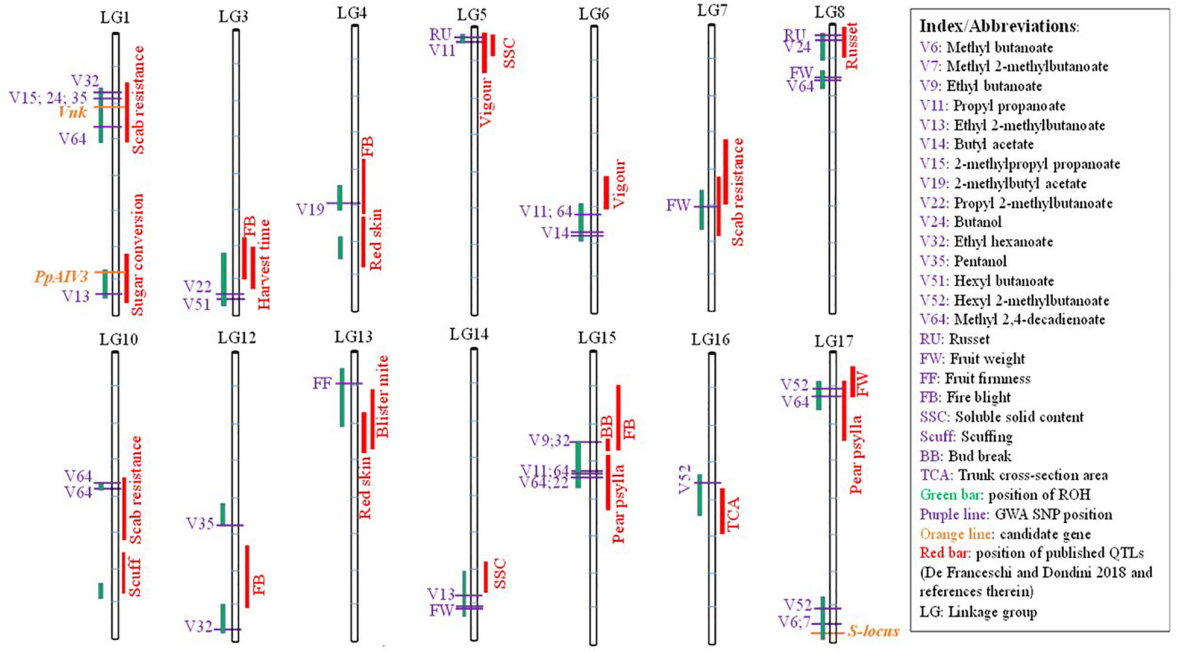

FIGURE 6 | Co-localization of runs of homozygosity islands (green bars), earlier published QTLs for various traits (red bars), marker-trait associations (purple horizontal lines) and candidates genes (orange horizontal lines).

Further investigation of ROH islands showed that generally different haplotypes were observed in accessions of the two species. The longest (71 bp) haplotype shared by both the Asian and European species resided on LG15. The small size of the shared ROH haplotype would indicate these two species shared ancestry many thousands of generations ago. Assuming an average genetic map size 1,350 cM and a genome size of 527 Mb (Li et al., 2017; Xue et al., 2018), one Mb physical distance approximately equates to 2.5 cM genetic distance in *Pyrus*. So, an ROH of length 0.0001775 cM would have arisen from a common ancestor occurring approximately 282,000 generations ago (g=100/(2×0.0001775); Fisher, 1954). Assuming a generation interval of 10 years in *Pyrus*, this would suggest that Asian and European species would have shared common ancestors at least 2.8 million years ago, which is supported by a study on the domestication history of pears (Wu et al., 2018).

Association Between ROH and Phenotypes

Genomic regions that are selection targets tend to generate ROH islands around the selected locus compared to the rest of the genome. Based on the observed similarity of nucleotide diversity between wild and cultivated pears, Wu et al. (2018) suggested relatively weak selection during pear domestication—which is supported by our observations of relatively low F_{ROH} compared with commercial livestock species (Peripolli et al., 2017). The ROH patterns in Asian and European populations showed some major differences on some linkage groups (**Supplementary Figures S3, S4**), which are supported by previous analysis of selective sweeps showing that genomic regions were differentially selected between Asian and European pears for traits such as cell wall degradation, fruit size, sugar biosynthesis, stone cells, acid and volatile compounds (Wu et al., 2018). ROH size and frequency were also reported to vary between Asian and European livestock populations (Bosse et al., 2012; Groenen et al., 2012; Peripolli et al., 2017).

Significant effects of increase in homozygosity were observed on various fruit phenotypes in this study. Increasing F_{ROH} significantly increased various traits such as fruit firmness, Brix, fruit weight, sensory flavor intensity, and TA. We also found strong evidence ($P < 1 \times 10^{-6}$) of negative selection for traits, including skin bitterness and ethyl esters, with an increase in F_{ROH}. Flavor volatiles have not directly been the target of artificial selection in pear breeding (Brewer and Palmer, 2010), however, these phenotypes are indirectly influenced by selection for other traits such as sensory flavor. Interestingly, ethyl esters were adversely correlated with sensory flavor intensity, suggesting that these particular esters are not important contributors to flavor intensity. Non-ethyl esters and the alcohols (hexanol, pentanol, and butanol) showed signatures of positive selection, which could largely be due to their favorable association with the breeding target traits (e.g., soluble solids and sensory flavor).

ROH Islands and Candidate Genes

The genomic regions with high occurrence of ROH have been shown to contain important genes associated with phenotypes in humans (Ceballos et al., 2018; Clark et al., 2019) and livestock species (Purfield et al., 2012; Kim et al., 2013; Biscarini et al., 2014). Overlap of ROH islands with marker-trait associations identified in this study, as well as with the previously published

QTLs for pear fruit/tree phenotypes (De Franceschi and Dondini, 2019; Kumar et al., 2019), adds to the robustness of ROH mapping as a complementary strategy for GWA studies in outbred fruit crops.

An ROH island at the bottom of LG17 harbors the self-incompatibility (SI) gene family which includes S-RNase and S-locus F-Box Brothers (SFBB) genes (Yamamoto et al., 2002; De Franceschi et al., 2011). It has been suggested that in addition to artificial selection, ROH islands could also be an indication of a lower recombination rate in those regions (Peripolli et al., 2017), which makes perfect sense in the case of the *Pyrus* SI-locus. Recombination suppression in the SI-locus region is essential because the pistil and pollen genes must inherit as one single unit in order to maintain the functionality of the SI system (Roalson and McCubbin, 2003; Claessen et al., 2019). Wu et al. (2013) reported highly repetitive sequences in the SI-locus region of pear, and hypothesized that suppression of recombination in the SI-locus region may be related to the presence of many repetitive sequences. Recombination between the pistil-*S* and pollen-*S* determinant genes would result in non-functional *S*-haplotypes and loss of self-incompatibility. The recombination suppressed region in *Pyrus* is predicted to be much larger compared to some other fruit species (Matsumoto and Tao, 2016).

Selection would result in selective sweeps, which refers to the genomic regions that have reduced nucleotide diversity compared with randomly evolving regions. Sabeti et al. (2002) developed the extended haplotype homozygosity (EHH) method to identify selective sweep regions in the human genome, and this tool has also been used to detect population-specific signatures of selection in livestock populations (Qanbari et al., 2010; Bomba et al., 2015). Short ROH regions were shown to overlap with EHH regions, suggesting complimentary nature of these two approaches to identify genomic regions under selection (Zhang et al., 2015). However, ROH patterns provide a guide to the population history (demography) and selection, which makes it a powerful tool for management of plant genetic resources, as well as for trait architecture studies in self-incompatible plants.

A key strength of ROH mapping is that long homozygous segments of genomes can be reliably identified using relatively low marker densities. However, the reduced-representation low-coverage genotyping platform used in our study could have missed many shorter ROH, which would result in an underestimation of F_{ROH}. Studies on humans (Ceballos et al., 2018) and livestock species (Purfield et al., 2012) have shown that high-density genotyping would be desirable, especially for mapping of shorter ROH. Further studies using the recently developed 200K SNP array (Li et al., 2019) or whole genome sequence data should help map shorter ROH for more accurate estimate of F_{ROH} and enable us to detect many more signals of natural and/or artificial selection in *Pyrus*.

CONCLUSION

In summary, this first application of the ROH approach in self-incompatible fruit crop species enabled us to compare genomic inbreeding coefficients between *Pyrus* species differing in domestication and breeding histories. For outbred fruit crops, genomic measure of inbreeding (F_{ROH}) would serve as a novel tool for breeding and management of gene-banks lacking reliable pedigree information. Association between F_{ROH} and phenotypes provides a simple mechanism to evaluate the direction of phenotypic change because of increased inbreeding levels. Co-localization of ROH islands and GWA signals agreed with results from studies in humans and livestock populations, which suggested that ROH mapping offers a complementary strategy to understand the genetic architecture of complex traits. Distribution of ROH islands in different species or populations of fruit crops can effectively be used to evaluate signatures of differential selection.

DATA AVAILABILITY STATEMENT

The genotyping-by-sequencing data presented in the study are deposited in the https://zenodo.org/ repository, digital object identifier https://zenodo.org/record/4302655, https://zenodo.org/record/4308154, and https://zenodo.org/record/4304904.

AUTHOR CONTRIBUTIONS

SK conceived the study, conducted ROH and GWA analyses, and wrote the first draft. SK and LB designed the study. CW and CK led leaf collection, DNA extraction, and GBS library preparation. CD conducted bioinformatics analysis. JW conducted QTL and candidate gene searches. DR, MH, and LB led the phenotyping. All authors helped to edit the manuscript, and read and approved the manuscript.

FUNDING

This research was partly funded by the New Zealand Ministry of Business, Innovation and Employment (MBIE). Funding from the National Science Foundation of China (31820103012) is also thankfully acknowledged.

ACKNOWLEDGMENTS

Constructive feedback from Paul Datson and Luis Gea is gratefully acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.590846/full#supplementary-material

Supplementary Figure S1 | Distribution of about 8500 single nucleotide polymorphism (SNP) markers on *Pyrus bretschneideri* "DangshanSuli"v1. 0 genome. The *x*-axis indicates linkage group number and *y*-axis length in base-pairs.

Supplementary Figure S2 | Population structure of Asian, European and inter-specific hybrid pear accessions using principal components analysis (PCA).

Supplementary Figure S3 | Distribution of the runs of homozygosity (ROH) across different linkage groups in Asian, European and Hybrid pear (*Pyrus* spp.) accessions.

Supplementary Figure S4 | The frequency (%) of single nucleotide polymorphisms (SNPs) occurrence into runs of homozygosity (ROH) islands within the Asian, European and hybrid pear (*Pyrus* spp.) population. The blue and red color represent chromosomes.

Supplementary Table S1 | The list of genomic regions of extended homozygosity (runs of homozygosity (ROH) islands). The overlap between genomic locations of ROH islands and previously published quantitative trait loci (QTLs) are approximated based on a recent review published by De Franceschi and Dondini (2019).

Supplementary Table S2 | Genomic positions of trait-associated single nucleotide polymorphisms (SNPs) which resided within the runs of homozygosity (ROH) islands on different linkage groups (LG). The probability of significance and the minor allele frequency (MAF) at each SNP loci is also shown.

REFERENCES

Angeloni, F., Ouborg, N. J., and Leimu, R. (2011). Meta-analysis on the association of population size and life history with inbreeding depression in plants. *Biol. Conserv.* 144, 35–43. doi: 10.1016/j.biocon.2010.08.016

Beynon, S. E., Slavov, G. T., Farré, M., Sunduimijid, B., Waddams, K., Davies, B., et al. (2015). Population structure and history of the Welsh sheep breeds determined by whole genome genotyping. *BMC Genet.* 16:65. doi: 10.1186/s12863-015-0216-x

Biscarini, F., Biffani, S., Nicolazzi, E. L., Morandi, N., and Stella, A. (2014). "Applying runs of homozygosity to the detection of associations between genotype and phenotype in farm animals," in *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production*, Vancouver, BC.

Bjelland, D. W., Weigel, K. A., Vukasinovic, N., and Nkrumah, J. D. (2013). Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *J. Dairy Sci.* 96, 4697–4706. doi: 10.3168/jds.2012-6435

Bomba, L., Nicolazzi, E. L., Milanesi, M., Negrini, R., Mancini, G., Biscarini, F., et al. (2015). Relative extended haplotype homozygosity signals across breeds reveal dairy and beef specific signatures of selection. *Genet. Sel. Evol.* 47:25. doi: 10.1186/s12711-015-0113-9

Bosse, M., Megens, H.-J., Madsen, O., Paudel, Y., Frantz, L. A., Schook, L. B., et al. (2012). Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet.* 8:e1003100. doi: 10.1371/journal.pgen.1003100

Bosse, M., Megens, H.-J., Madsen, O., Crooijmans, R. P. M. A., Ryder, O. A., Austerlitz, F., et al. (2015). Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. *Genome Res.* 25, 1–12. doi: 10.1101/gr.187039.114

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

Brewer, L. R., and Palmer, J. W. (2010). Global pear breeding programmes: goals, trends and progress for new cultivars and new rootstocks. *Acta Hortic.* 909, 105–119. doi: 10.17660/actahortic.2011.909.10

Broman, K. W., and Weber, J. L. (1999). Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am. J. Hum. Genet.* 65, 1493–1500. doi: 10.1086/302661

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., and Wilson, J. F. (2018). Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet.* 19:220. doi: 10.1038/nrg.2017.109

Charlesworth, D. (2003). Effects of inbreeding on the genetic diversity of populations. *Philos. Trans. R. Soc. London Ser. B: Biol. Sci.* 358, 1051–1070. doi: 10.1098/rstb.2003.1296

Claessen, H., Keulemans, W., Van de Poel, B., and De Storme, N. (2019). Finding a compatible partner: self-incompatibility in European pear (*Pyrus communis*); molecular control, genetic determination, and impact on fertilization and fruit set. *Front. Plant Sci.* 10:407. doi: 10.3389/fpls.2019.00407

Clark, D. W., Okada, Y., Moore, K. H., Mason, D., Pirastu, N., Gandin, I., et al. (2019). Associations of autozygosity with a broad range of human phenotypes. *Nat. Commun.* 10, 1–7. doi: 10.1002/9783527678679.dg01212

Curik, I., Ferencakovic, M., and Solkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest. Sci.* 166, 26–34. doi: 10.1016/j.livsci.2014.05.034

De Cara, M. A. R., Villanueva, B., Toro, M. A., and Fernandez, J. (2013). Using genomic tools to maintain diversity and fitness in conservation programmes. *Mol. Ecol.* 22, 6091–6099. doi: 10.1111/mec.12560

De Franceschi, P., Pierantoni, L., Dondini, L., Grandi, M., Sanzol, J., and Sansavini, S. (2011). Cloning and mapping multiple S-locus F-box genes in European pear (*Pyrus communis* L.). *Tree Genet. Genomes* 7, 231–240. doi: 10.1007/s11295-010-0327-5

De Franceschi, P., and Dondini, L. (2019). "Molecular mapping of major genes and QTLs in pear," in *The Pear Genome*, ed. S. S. Korban (Cham: Springer), 113–131. doi: 10.1007/978-3-030-11048-2_6

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:e5. doi: 10.1371/journal.pone.0019379

Fisher, R. A. (1954). A fuller theory of "junctions" in inbreeding. *Heredity* 8, 187–197. doi: 10.1038/hdy.1954.17

Forutan, M., Mahyari, S. A., Baes, C., Melzer, N., Schenkel, F. S., and Sargolzaei, M. (2018). Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. *BMC Genomics* 19:98. doi: 10.1186/s12864-018-4453-z

Gibson, J., Newton, E. M., and Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* 15, 789–795. doi: 10.1093/hmg/ddi493

Grilz-Seger, G., Neuditschko, M., Ricard, A., Velie, B., Lindgren, G., Mesarič, M., et al. (2019). Genome-wide homozygosity patterns and evidence for selection in a set of European and near eastern horse breeds. *Genes* 10:491. doi: 10.3390/genes10070491

Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–398. doi: 10.1038/nature11622

Henry, R. J. (2012). Next-generation sequencing for understanding and accelerating crop domestication. *Brief. Funct. Genomics* 11, 51–56. doi: 10.1093/bfgp/elr032

Huson, H. J., Kim, E. S., Godfrey, R. W., Olson, T. A., McClure, M. C., Chase, C. C., et al. (2014). Genome-wide association study and ancestral origins of the slick-hair coat in tropically adapted cattle. *Front. Genet.* 5:101. doi: 10.3389/fgene.2014.00101

Kim, E.-S., Cole, J. B., Huson, H., Wiggans, G. R., Van Tassell, C. P., Crooker, B. A., et al. (2013). Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. *PLoS One* 8:e80813. doi: 10.1371/journal.pone.0080813

Kirin, M., McQuillan, R., Franklin, C., Campbell, H., McKeigue, P. M., and Wilson, J. F. (2010). Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5:e13996. doi: 10.1371/journal.pone.0013996

Kumar, S., Kirk, C., Deng, C., Wiedow, C., Knaebel, M., and Brewer, L. (2017). Genotyping-by-sequencing of pear (*Pyrus* spp.) accessions unravels novel patterns of genetic diversity and selection footprints. *Hortic. Res.* 4:17015. doi: 10.1038/hortres.2017.15

Kumar, S., Kirk, C., Deng, C. H., Shirtliff, A., Wiedow, C., Qin, M., et al. (2019). Marker-trait associations and genomic predictions of interspecific pear (*Pyrus*) fruit characteristics. *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-45618-w

Lander, E. S., and Botstein, D. (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236, 1567–1570. doi: 10.1126/science.2884728

Lencz, T., Lambert, C., DeRosse, P., Burdick, K. E., Morgan, T. V., Kane, J. M., et al. (2007). Runs of homozygosity reveal highly penetrant recessive loci in

schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19942–19947. doi: 10.1073/pnas.0710021104

Li, L., Deng, C. H., Knabel, M., Chagne, D., Kumar, S., Sun, J., et al. (2017). Integrated high-density consensus genetic map of Pyrus and anchoring of the 'Bartlett' v1.0 (*Pyrus communis*) genome. *DNA Res.* 24, 289–301. doi: 10.1093/dnares/dsw063

Li, X., Singh, J., Qin, M., Li, S., Zhang, X., Zhang, M., et al. (2019). Development of an integrated 200K SNP genotyping array and application for genetic mapping, genome assembly improvement and genome wide association studies in pear (*Pyrus*). *Plant Biotechnol. J.* 17, 1582–1594. doi: 10.1111/pbi.13085

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444

Mascher, M., Schreiber, M., Scholz, U., Graner, A., Reif, J. C., and Stein, N. (2019). Gene-bank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nat. Genet.* 51, 1076–1081. doi: 10.1038/s41588-019-0443-6

Matsumoto, D., and Tao, R. (2016). Recognition of a wide-range of S-RNases by S locus F-box like 2, a general-inhibitor candidate in the *Prunus*-specific S-RNase-based self-incompatibility system. *Plant Mol. Biol.* 91, 459–469. doi: 10.1007/s11103-016-0479-2

Metzger, J., Karwath, M., Tonda, R., Beltran, S., Agueda, L., Gut, M., et al. (2015). Runs of homozygostiy reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics* 16:764. doi: 10.1186/s12864-015-1977-3

Moore, J. N., and Janick, J. (1975). *Advances in Fruit Breeding*. West Lafayette, IN: Purdue University Press.

Peripolli, E., Munari, D. P., Silva, M. V., Lima, A. L., Irgang, R., and Baldi, F. (2017). Runs of homozygosity: current knowledge and applications in livestock. *Anim. Genet.* 48, 255–271. doi: 10.1111/age.12526

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Purfield, D. C., Berry, D., McParland, S., and Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genetics* 13:70. doi: 10.1186/1471-2156-13-70

Qanbari, S., Pimentel, E. C. G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A. R., et al. (2010). A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim. Genet.* 41, 377–389. doi: 10.1111/j.1365-2052.2009.02016.x

Roalson, E. H., and McCubbin, A. G. (2003). S-RNases and sexual incompatibility: structure, functions, and evolutionary perspectives. *Mol. Phylogenet. Evol.* 29, 490–506. doi: 10.1016/s1055-7903(03)00195-7

Rowan, D. D., Hunt, M. B., Dimouro, A., Alspach, P. A., Weskett, R., Volz, R. K., et al. (2009). Profiling fruit volatiles in the progeny of a 'Royal Gala' x 'Granny Smith' apple (*Malus x domestica*) cross. *J. Agric. Food Chem.* 57, 7953–7961. doi: 10.1021/jf901678v

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837. doi: 10.1038/nature01140

Smýkal, P., Nelson, M. N., Berger, J. D., and Von Wettberg, E. J. (2018). The impact of genetic changes during crop domestication. *Agronomy* 8:119. doi: 10.3390/agronomy8070119

Toro, M. A., and Varona, L. (2010). A note on mate allocation for dominance handling in genomic selection. *Genet. Sel. Evol.* 42:33. doi: 10.1186/1297-9686-42-33

Van de Wouw, M., Kik, C., van Hintum, T., van Treuren, R., and Visser, B. (2010). Genetic erosion in crops: concept, research results and challenges. *Plant Genet. Resour.* 8, 1–5. doi: 10.1017/s1479262109990062

VanRaden, P. M., Olson, K. M., Wiggans, G. R., Cole, J. B., and Tooker, M. E. (2011). Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. *J. Dairy Sci.* 94, 5673–5682. doi: 10.3168/jds.2011-4500

Wu, J., Wang, Z. W., Shi, Z. B., Zhang, S., Ming, R., Zhu, S. L., et al. (2013). The genome of the pear (Pyrus bretschneideri Rehd.). *Genome Res.* 23, 396–408. doi: 10.1101/gr.144311.112

Wu, J., Wang, Y., Xu, J., Korban, S. S., Fei, Z., Tao, S., et al. (2018). Diversification and independent domestication of Asian and European pears. *Genome Biol.* 19:77. doi: 10.1186/s13059-018-1452-y

Xue, H., Wang, S., Yao, J. L., Deng, C. H., Wang, L., Su, Y., et al. (2018). Chromosome level high-density integrated genetic maps improve the *Pyrus bretschneideri* 'DangshanSuli'v1. 0 genome. *BMC Genomics* 19:833. doi: 10.1186/s12864-018-5224-6

Yamamoto, T., Kimura, T., Shoda, M., Imai, T., Saito, T., Sawamura, Y., et al. (2002). Genetic linkage maps constructed by using an interspecific cross between Japanese and European pears. *Theor. Appl. Genet.* 106, 9–18. doi: 10.1007/s00122-002-0966-5

Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M. S., and Sahana, G. (2015). Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics* 16:542. doi: 10.1186/s12864-015-1715-x

Conflict of Interest: 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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 08 January 2021
doi: 10.3389/fpls.2020.617664

Resistance to '*Candidatus* Liberibacter asiaticus,' the Huanglongbing Associated Bacterium, in Sexually and/or Graft-Compatible *Citrus* Relatives

*Mônica N. Alves[1], Silvio A. Lopes[1], Laudecir L. Raiol-Junior[2], Nelson A. Wulff[1], Eduardo A. Girardi[2], Patrick Ollitrault[3] and Leandro Peña[1,4]**

[1] Fundo de Defesa da Citricultura, Araraquara, Brazil, [2] Empresa Brasileira de Pesquisa Agropecuária, Cruz das Almas, Brazil, [3] Centre de Coopération Internationale en Recherche Agronomique pour le Développement, BIOS Department, UPR amélioration génétique des espèces à multiplication végétative, Montpellier, France, [4] Instituto de Biologia Molecular y Celular de Plantas – Consejo Superior de Investigaciones Científicas, Universidad Politécnica de Valencia, Valencia, Spain

OPEN ACCESS

Edited by:
Agata Gadaleta,
University of Bari Aldo Moro, Italy

Reviewed by:
Ed Stover,
Horticultural Research Laboratory (USDA-ARS), United States
Muqing Zhang,
University of Florida, United States

***Correspondence:**
Leandro Peña
lpenya@fundecitrus.com.br;
lpenya@ibmcp.upv.es

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

Received: 16 October 2020
Accepted: 10 December 2020
Published: 08 January 2021

Citation:
Alves MN, Lopes SA, Raiol-Junior LL, Wulff NA, Girardi EA, Ollitrault P and Peña L (2021) Resistance to 'Candidatus Liberibacter asiaticus,' the Huanglongbing Associated Bacterium, in Sexually and/or Graft-Compatible Citrus Relatives. Front. Plant Sci. 11:617664. doi: 10.3389/fpls.2020.617664

Huanglongbing (HLB) is the most destructive, yet incurable disease of citrus. Finding sources of genetic resistance to HLB-associated '*Candidatus* Liberibacter asiaticus' (Las) becomes strategic to warrant crop sustainability, but no resistant *Citrus* genotypes exist. Some *Citrus* relatives of the family Rutaceae, subfamily Aurantioideae, were described as full-resistant to Las, but they are phylogenetically far, thus incompatible with *Citrus*. Partial resistance was indicated for certain cross-compatible types. Moreover, other genotypes from subtribe Citrinae, sexually incompatible but graft-compatible with *Citrus*, may provide new rootstocks able to restrict bacterial titer in the canopy. Use of seedlings from monoembryonic species and inconsistencies in previous reports likely due to Las recalcitrance encouraged us to evaluate more accurately these *Citrus* relatives. We tested for Las resistance a diverse collection of graft-compatible Citrinae species using an aggressive and consistent challenge-inoculation and evaluation procedure. Most Citrinae species examined were either susceptible or partially resistant to Las. However, *Eremocitrus glauca* and Papua/New Guinea *Microcitrus* species as well as their hybrids and those with *Citrus* arose here for the first time as full-resistant, opening the way for using these underutilized genotypes as Las resistance sources in breeding programs or attempting using them directly as possible new Las-resistant *Citrus* rootstocks or interstocks.

Keywords: HLB, Greening, Rutaceae, *Citrus* breeding, Aurantioideae, *Microcitrus*, *Eremocitrus*

INTRODUCTION

Huanglongbing (HLB) is the most destructive disease of citrus worldwide. Its occurrence is associated with the infection of trees with one of the following Gram-negative intracellular α-proteobacteria, '*Candidatus* Liberibacter asiaticus' (Las), '*Ca.* L. americanus' (Lam) or '*Ca.* L. africanus' (Laf), which colonize the phloem of their host plants (Bové, 2006). Las and Lam are

naturally transmitted by the Asian citrus psyllid *Diaphorina citri* Kuwayama (Sternorrhyncha: Liviidae) (Capoor et al., 1967; Yamamoto et al., 2006), while Laf is transmitted by the African citrus psyllid *Trioza erytreae* Del Guercio (Sternorrhyncha: Triozidae) (McClean and Oberholzer, 1965). In Brazil, Las and Lam have been detected (Coletta-Filho et al., 2004; Teixeira et al., 2005), Las being the most common HLB-associated bacterium, currently present in over 99.9% of all 'Liberibacter'-positive field samples analyzed at Fundecitrus (Bassanezi et al., 2020). Las tolerates higher temperatures, reaches higher titers in *Citrus* and is more efficiently transmitted than Lam (Lopes et al., 2009, 2013), all this making the former associated with the most severe damages caused by HLB in Asia and the Americas (Gottwald et al., 2007). HLB-infected, severely affected trees produce small and irregularly shaped fruits with a thick peel that remains green. Leaves shows a blotchy mottle, yellowing and may become thicker and with enlarged veins. The canopy show a premature defoliation and dieback of twigs (Bové, 2006). In areas where no control of the insect vector is done, the severity of symptoms and disease progression in the orchards increase rapidly (Gottwald et al., 1989, 1991).

The control of HLB is based on planting of healthy trees from insect-proof nurseries, monitoring of *D. citri* and tree flushing, application of insecticides to reduce the population of the insect vector when reaching an unacceptable threshold and rapid eradication of symptomatic trees. In Brazil, since the disease was first reported in 2004, these practices have been adopted in the main citrus producing region, comprising the São Paulo State and Southwest of Minas Gerais State (Bassanezi et al., 2020). However, the cost of implementing these strategies is high and in spite of them, the incidence of the disease remains around 20% annually in the region (Fundecitrus, 2020). These palliative and preventive measures are employed for HLB management because there is neither cure nor resistant cultivars within *Citrus*, which may be used to control the disease durably.

It has been reported that some *Citrus* relatives belonging to the family Rutaceae, subfamily Aurantioideae may be partially, transiently or totally resistant to Las. In the case of some *Poncirus trifoliata* (L.) Raf. accessions the bacterium reaches inconsistent infections with low titers and uneven distribution (Folimonova et al., 2009). For *Murraya paniculata* (L.) Jack or *Swinglea glutinosa* (Blanco) Merr. (Cifuentes-Arenas et al., 2019) infection is just transient and after a few months plants become Las-free. In *Clausena excavata* Burm. F., *Glycosmis pentaphylla* (Retz.) Corr. and other Aurantioideae *Citrus* relatives the bacterium is undetectable and thus unable to replicate in the genotype under test (Ramadugu et al., 2016).

Consistent resistance to Las characterized by lack of detectable bacterial replication in germplasm sexually compatible with *Citrus* is of major interest because it may be used in sexual breeding programs aimed to generate hybrid and backcross citrus-like populations, which could provide new rootstocks or scion varieties resistant to HLB. Moreover, segregating progenies could be useful to identify genetic loci involved in the resistance trait. However, claims for resistance to Las in close relatives to *Citrus* are discrepant depending on the use of different genetic backgrounds, challenge inoculation systems, environmental conditions, plant ages, number of replicates, field vs. greenhouse tests, seedlings vs. mature plants and grafted vs. rooted stocks (Hung et al., 2000; Folimonova et al., 2009; Shokrollah et al., 2009; Feng et al., 2015; Ramadugu et al., 2016; Miles et al., 2017). For example, *Poncirus trifoliata* has been reported as resistant (Albrecht and Bowman, 2012), partially resistant (Folimonova et al., 2009) or showing different levels of resistance, recovery or delayed infection depending on the accession (Ramadugu et al., 2016). The monoembryonic, Australian native *Eremocitrus glauca* (Lindl.) Swingle as well as *Microcitrus australasica* (F. Muell.) Swingle and other *Microcitrus* species have been considered partially resistant, with transient replication or variable responses among seedlings, respectively (Ramadugu et al., 2016), representing the first results opening the way for their use in breeding programs for Las resistance, as they are cross-compatible with *Citrus* (Swingle and Reece, 1967). However, the use of seedlings for Las challenge inoculation in monoembryonic species implies that segregating individuals were actually evaluated, likely explaining their variable responses to Las (Ramadugu et al., 2016). Moreover, Las infection through *D. citri* in the field, being the natural challenge inoculation system, does not allow distinguishing resistance to the bacterium from resistance to the psyllid vector or to both vector and bacterium. Furthermore, the influence of other abiotic and biotic factors present in the field may affect seedling performance even in absence of clear Las infection (Miles et al., 2017).

Resistant *Citrus* relatives graft-compatible with *Citrus* may provide interstocks or new rootstocks (other than the widely used *P. trifoliata* accessions and their hybrids with sweet orange and grapefruit) able to restrict bacterial titer and distribution in the scion and thus potentially reduce disease damages. *Eremocitrus glauca* and several *Microcitrus* species as *M. australasica* and the 'Sydney' hybrid have performed well as interstocks. Moreover, *E. glauca*, *Microcitrus* types and their hybrids warrant consideration as citrus rootstocks (Bitters et al., 1964, 1977). Additionally, the most promising Aurantioideae genera for use as citrus rootstocks were *Atalantia* (some of its species previously classified in the *Severinia* genus), *Naringi*, *Citropsis*, *Limonia* (previously known as *Feronia*) and *Swinglea* (Bitters et al., 1964, 1969, 1977). Responses to Las infection have been tested for several species from these genera (Koizumi et al., 1996; Hung et al., 2000; Feng et al., 2015; Ramadugu et al., 2016; Cifuentes-Arenas et al., 2019). However, the use of seedlings from monoembryonic species (*Citropsis*), inconsistencies in results of different reports (*Limonia*), and low number of plant replications in some cases (*Naringi* and *Atalantia),* make it advisable to perform a more accurate evaluation of these *Citrus* relatives for resistance to Las. In addition to the identification of promising germplasm to be used as parents in sexual breeding programs or as rootstocks/interstocks, a better knowledge of the distribution of the response to Las in the gene pool within Citrinae considering phylogenetic relations, would provide valuable orientation to decipher the determinants of resistance/susceptibility.

To gain more insight into Las resistance irrespective of the insect vector within close *Citrus* relatives, we have selected cross- and/or graft-compatible Aurantioideae species of the Citrinae

subtribe as well as intergeneric hybrids and inoculated them with Las. A challenge and evaluation system allowing unequivocal demonstration of either resistance or susceptibility was used. For this, we have selected mature buds from each genotype of interest, and grafted them onto 'Rangpur' lime (*Citrus* × *limonia* Osbeck), a well-known, Las-susceptible rootstock. Challenge inoculation was performed by grafting Las-infected budwood pieces both onto the rootstock and onto the scion, so infection may come to the scion under test directly from the infected grafted budwood or from the bacterial flow moving up from the susceptible 'Rangpur' lime rootstock. Evaluation for Las multiplication was done regularly by qPCR over 24 months after inoculation (MAI). By using such method, we identify here which Citrinae species are the most indicated to be used as sources of full-resistance to Las either in breeding programs or directly as potential *Citrus* rootstocks/interstocks. The distribution of susceptibility/resistance responses to Las within the Citrinae germplasm is discussed according to chloroplast-based phylogeny.

MATERIALS AND METHODS

Plant Genotypes

The cross-compatible *Citrus* relatives selected were four *Poncirus trifoliata* (L.) Raf. accessions, 'Pomeroy,' 'Rubidoux,' 'Barnes,' and 'Benecke,' commonly used as rootstocks and in breeding programs with *Citrus* to generate new hybrid rootstocks (Phillips and Castle, 1977; Soares Filho et al., 2003; Bowman et al., 2016) and belonging to two different genetic groups, 'Rubidoux' and 'Barnes' to group 3 and 'Pomeroy' and 'Benecke' to group 4, according to Fang et al. (1997); the Australian limes, including (1) the pure species: *Microcitrus australasica* (F. Muell.) Swingle, *M. australasica* 'Sanguinea,' *M. australasica* 'True Sanguinea,' *M. inodora* (F.M. Bail) Swingle, *M. warburgiana* (F.M. Bailey) Tanaka, *M. papuana* Winters, *M. australis* (A. Cunn. ex Mudie) Swingle and *Eremocitrus glauca* (Lindl.) Swingle, (2) three hybrids among them: *M. virgata* (*M. australis* × *M. australasica*), known as the 'Sydney' hybrid, a *Microcitrus* sp. × *E. glauca* and an *E. glauca* × *Microcitrus* sp. hybrids, and (3) two Australian lime hybrids with *Citrus*: an *Eremocitrus glauca* × *Citrus* × *sinensis* hybrid (eremorange) and the 'Faustrimedin' hybrid [*M. australasica* × 'Calamondin' (*Fortunella* sp. × *C. reticulata* Blanco); *C.* × *oliveri*] (**Table 1** and **Figure 1**). Among those Citrinae genotypes cross-incompatible but graft-compatible with *Citrus*, we selected those recommended as suitable rootstocks and/or interstocks by Bitters et al. (1964, 1969, 1977), including *Atalantia citroides* Pierre ex Guillaumin, *A. ceylanica* (Arn.) Oliv., *Limonia acidissima* L., *Citropsis gilletiana* Swingle & M. Kellerm, and *Naringi crenulata* (Roxb.) Nicolson. *Limonia acidissima* is classified in the Balsamocitrinae subtribe by Swingle and Reece (1967). However, it appears to be closely related with *Atalantia* species according to chloroplastic phylogenetic studies and clearly included in the Citrinae clade (Bayer et al., 2009). As Las susceptible controls, we used two *Citrus* genotypes, the Brazilian sweet orange varieties 'Pera' and 'Tobias' [*C.* × *sinensis* (L.) Osbeck] (Donadio et al., 1995). We also added the 'Mountain' citron (*C. halimii* B.C. Stone), which was the less Las-susceptible type within *Citrus* according to Ramadugu et al. (2016) (**Table 1** and **Figure 1**).

Plant Material, Grafting and Las Challenge Inoculation

In a preliminary experiment, the 25 accessions from *Citrus* and *Citrus* relatives shown in **Table 1**, belonging to the family Rutaceae, subfamily Aurantioideae, subtribe Citrinae, were selected from the virus/viroid-free Fundecitrus germplasm collection and propagated by grafting onto 'Rangpur' lime (*Citrus* × *limonia* Osbeck) nucellar rootstocks to study graft-compatibility. The Las-susceptible, graft-compatible sweet orange varieties 'Pera' and 'Tobias' [*C.* × *sinensis* (L.) Osbeck]

TABLE 1 | Citrinae genotypes/accessions tested for resistance to '*Candidatus* Liberibacter asiaticus.'

Genotype/Accession[a]	Common name[a]
Citrus × *sinensis* (L.) Osbeck 'Pera' **(P[c])**	'Pera' sweet orange
C. × *sinensis* 'Tobias' **(P)**	'Tobias' sweet orange
Citrus halimii B.C. Stone **(M[d])**	'Mountain' citron
Poncirus trifoliata (L.) Raf. 'Pomeroy' **(P)**	'Pomeroy' trifoliate orange
P. trifoliata 'Benecke' **(P)**	'Benecke' trifoliate orange
P. trifoliata 'Barnes' **(P)**	'Barnes' trifoliate orange
P. trifoliata 'Rubidoux' **(P)**	'Rubidoux' trifoliate orange
Microcitrus australasica (F. Muell.) Swingle **(M)**	Australian finger lime
M. australasica 'Sanguinea' **(M)**	'Sanguinea' Australian finger lime
M. australasica 'True Sanguinea' **(M)**	'True Sanguinea' Australian finger lime
M. australasica × (*Fortunella* sp. × *Citrus reticulata*) 'Calamondin'; *C.* × *oliveri* **(PM[e])**	'Faustrimedin' hybrid
M. inodora (F.M. Bail) Swingle **(M)**	Australian large-leaf wild lime
M. warburgiana (F.M. Bailey) Tanaka **(M)**	New Guinean wild lime
M. papuana Winters **(M)**	Brown river finger lime
M. australis (A. Cunn. Ex Mudie) Swingle **(M)**	Australian round lime
M. virgata (M. australis × *M. australasica)* **(PM)**	'Sydney' hybrid
Microcitrus sp. × *E. glauca* hybrid **(PM)**	Australian lime hybrid BGC 695[b]
E. glauca (Lindl.) Swingle **(M)**	Australian desert lime
E. glauca × *C.* × *sinensis* hybrid **(PP)**	Eremorange
E. glauca × *Microcitrus* sp. hybrid **(PM)**	Australian desert lime hybrid BGC 682[b]
Atalantia citroides Pierre ex Guillaumin **(PP[f])**	Cochin China atalantia
A. ceylanica (Arn.) Oliv. **(PP)**	Ceylon atalantia
Limonia acidissima L. **(P)**	Indian wood apple or elephant apple
Citropsis gilletiana Swingle & M. Kellerm. **(M)**	Gillet's cherry orange
Naringi crenulata (Roxb.) Nicolson **(PM)**	Hesperethusa

[a] *The nomenclature used follows Swingle and Reece (1967) and Bayer et al. (2009).*
[b] *Original accession number at the Citrus Germplasm Bank (BGC) of Embrapa Cassava & Fruits in Cruz das Almas, Bahia.*
[c] *P, polyembryonic;* [d] *M, monoembryonic;* [e] *PM, possibly monoembryonic;* [f] *PP, possibly polyembryonic; according to Swingle and Reece (1967) and Bitters (1986).*

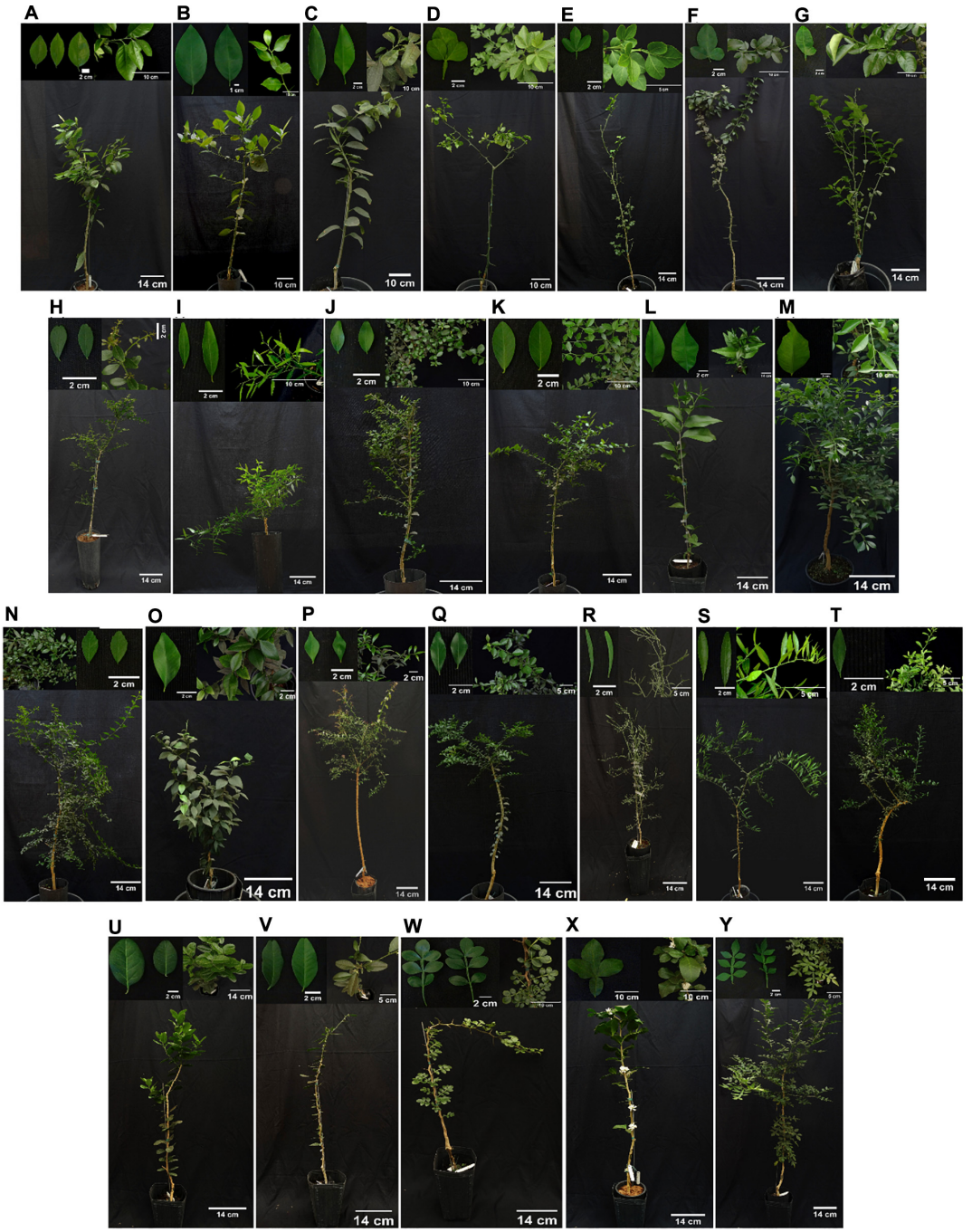

FIGURE 1 | Representative photographs of each of the selected Citrinae genotypes/accessions [cross-compatible **(A–T)**; and cross-incompatible, putative graft-compatible **(U–Y)** with *Citrus* about 1 year after grafting onto 'Rangpur' lime rootstock]. **(A)** *Citrus* × *sinensis* 'Pera'; **(B)** *C.* × *sinensis* 'Tobias'; **(C)** *C. halimii;* **(D)** *Poncirus trifoliata* 'Pomeroy'; **(E)** *P. trifoliata* 'Benecke'; **(F)** *P. trifoliata* 'Barnes'; **(G)** *P. trifoliata* 'Rubidoux'; **(H)** *Microcitrus australasica;* **(I)** *M. australasica* 'Sanguinea'; **(J)** *M. australasica* 'True Sanguinea'; **(K)** 'Faustrimedin' hybrid; **(L)** *M. inodora;* **(M)** *M. warburgiana;* **(N)** *M. papuana;* **(O)** *M. australis;* **(P)** *M. virgata* hybrid; **(Q)** *Microcitrus* sp. × *E. glauca* hybrid; **(R)** *Eremocitrus glauca;* **(S)** *E. glauca* × *C.* × *sinensis* hybrid; **(T)** *E. glauca* × *Microcitrus* sp. hybrid; **(U)** *Atalantia citroides*; **(V)** *A. ceylanica*; **(W)** *Limonia acidissima*; **(X)** *Citropsis gilletiana*; **(Y)** *Naringi crenulata*.

(Donadio et al., 1995) were also propagated on 'Rangpur' lime as controls.

Twenty to thirty plants per accession from Citrinae genotypes that were graft-compatible with 'Rangpur' lime were propagated on this same rootstock using buds from a single donor mother plant per genotype, and kept at a greenhouse in Araraquara, São Paulo State, Brazil (**Figure 1**). True-to-typeness was assessed by analyzing all Citrinae species morphologically

(Swingle and Reece, 1967) and in the case of Australian lime hybrids also by using SSR molecular markers (Fanciullino et al., 2005; Froelicher et al., 2008; Luro et al., 2008; Ollitrault et al., 2008) (**Supplementary Table 1**). Plants were grown in 4 L polyethylene bags filled with coir, irrigated and fertilized twice a week, and sprayed monthly with insecticides and miticides. Las challenge-inoculation experiments were conducted in a greenhouse in which the mean daily air temperature varied between 18.5°C to 34.4°C and illumination was natural.

The original source of inoculum was Las-positive sweet orange budwood from a farm located in São Paulo (Brazil), propagated and kept at Fundecitrus since 2006 (Lopes et al., 2009). Las-free and Las-positive *D. citri* populations are continuously reared at Fundecitrus, as described in Parra et al. (2016). Las-free psyllids are reared on healthy *Murraya paniculata* (L.) Jack seedlings while Las-positive psyllids are reared on Las-positive 'Valencia' sweet orange plants grafted on 'Rangpur' lime. For this work, greenhouse-grown 'Rangpur' lime seedlings were inoculated by exposing them to Las-infected *D. citri* insects. Budwood from these seedlings was then used to inoculate 'Valencia' sweet orange nursery plants grafted on 'Rangpur' lime, which were used as source of inoculum for the 25 Citrinae accessions. Las-infection in 'Rangpur' lime seedlings and 'Valencia' sweet orange plants was confirmed by qPCR (Li et al., 2006). Seven to sixteen plants per genotype were selected based on regular and homogeneous growth within each accession to be graft-inoculated with Las using two budwood pieces ca. 3 to 5 cm long per plant.

In all plants to be tested, one Las-infected (qPCR-positive) budwood piece was grafted on the 'Rangpur' lime rootstock, and another one on the scion variety, both at 5 cm below and above the scion-rootstock junction, respectively. At the same time, 4 to 14 uniform plants of each genotype were grafted with Las-negative budwood from healthy greenhouse-grown 'Valencia' sweet orange plants grafted on 'Rangpur' lime, which were used as negative controls. In the case of 'Pera' sweet orange, 41 plants were Las graft-inoculated as described above and 33 plants were grafted with healthy budwood pieces. Three months after graft challenge inoculation, plants were pruned at 0.5–1.5 m from the rootstock to promote new shoot growth and thus Las translocation from the infected budwood pieces and rootstock to the scion in the Las-challenge inoculated plants (Johnson et al., 2014; Raiol-Junior et al., 2021). Control plants were pruned likewise. Only those plants showing Las infection in the 'Rangpur' lime rootstock at 12 months after graft-inoculation (MAI) were considered as successfully Las-inoculated. The exact number of plants used per genotype for resistance evaluation as well as that of healthy controls are detailed in **Supplementary Table 2**.

Sample Collection and Evaluation by qPCR

A random but representative sampling of 16–20 leaf pieces from actively growing shoots per plant scion were collected at 4, 6, 8, 10, and 12 MAI with Las-infected or Las-free (control) budwood. Those accessions that remained with few or no Las-positive scions at 12 MAI, and continued to grow well, were re-evaluated at 24 MAI, including *Microcitrus australasica* 'True Sanguinea,' *M. warburgiana*, *M. papuana*, *M. australis*, *Microcitrus* sp. × *E. glauca* hybrid, *E. glauca* × *C.* × *sinensis* hybrid and *E. glauca* × *Microcitrus* sp. hybrid as well as *C.* × *sinensis* 'Tobias' and 'Pera' controls.

To be sure that Las challenge-inoculated plants were actually infected, the fibrous root tissue from the susceptible 'Rangpur' lime rootstock was evaluated at 12 MAI as well as its bark at 12 and 24 MAI. Only plants with Las-positive roots at 12 MAI had their leaves evaluated. Las graft-infection success was calculated as the percentage of plants with Las-positive rootstocks per the total number of Las-graft-inoculated composite plants (**Supplementary Table 2**).

To assess movement of Las from the rootstock to the scion through the vascular system, bark samples were collected at 5 and 30 cm above the bud union and at the scion canopy (21–152 cm) from each plant that remained without Las-multiplication in leaf samples including *M. warburgiana*, *M. papuana*, *M. australis*, *Microcitrus* sp. × *E. glauca* hybrid, *Eremocitrus glauca*, *E. glauca* × *C.* × *sinensis* hybrid and *E. glauca* × *Microcitrus* sp. hybrid as well as *C.* × *sinensis* 'Tobias' and 'Pera' as Las positive controls.

Samples were all subjected to DNA extraction. Total DNA was extracted from 0.5 g of leaf midribs, 0.3 g of fibrous roots or 0.3 g of bark tissue and processed using a cetyltrimethylammonium bromide (CTAB) extraction buffer (Murray and Thompson, 1980) as described by Teixeira et al. (2005). DNA quality was checked with the NanoDrop (Thermo Scientific) (Desjardins and Conklin, 2010). Real-time polymerase chain reactions (qPCR) for detection of the 16S rDNA from Las were ran using 1 μL of total DNA (100 ng/μL), TaqMan® PCR Master Mix (1x) (Invitrogen, Carlsbad, CA, United States) and HLB as Las-specific primer/probe (0.5 μM/0.2 μM) in a StepOnePlus thermocycler (Applied Biosystems) as described by Li et al. (2006). A positive and a negative sample were included as quality controls during DNA extractions and qPCR assays. As an internal control, the mitochondrial gene cytochrome oxidase (COX) was used (Li et al., 2006). As plant housekeeping controls, primers from the Actin and Rubisco small subunit genes were designed based on homologous sequences of six and nine different plant species, respectively, that were aligned using DNA MAN®, and were also used (**Supplementary Tables 3, 4**).

Leaf, bark and root samples were considered Las-positive when their qPCR cycle threshold (Ct) was lower than 34.0. Las quantification was done based on the linear relation among Ct and the 16S rRNA log, according to $y = -0.2998\text{Ct} + 11.042$, $R^2 = 0.9981$ (Lopes et al., 2013). Target gene concentrations below 0.9 16S rRNA log, which corresponded to a Ct value of 34.0, produced variable results (Lopes et al., 2013), so this was used as the lower detection limit for Las-positive samples. However, Ct values between 34.0 and 36.0 were considered as suspicious to be positive, so only when all samples from other remaining time points showed Ct values over 36.0, the plant was considered as Las-negative. For root samples, a standard curve was generated using the amplified PCR product of the gene 16S rRNA from Las. Eight- to ten-fold serial dilutions were prepared in triplicate and mixed in total DNA at 100 ng/μl for healthy 'Rangpur' lime roots (**Supplementary Figure 1**). The linear relation among Ct

and the log 16S rRNA was $y = 1.073Ct + 25.098$, $R^2 = 0.9929$. Ct values higher than 34.0 were also inconsistent for roots. When Ct values were 34.0 or close to 34.0, bark pieces from the rootstock were further analyzed at 12 and 24 MAI by qPCR to confirm Las-infection of the 'Rangpur' lime.

Data Analysis

Data were analyzed with the statistical software RStudio (RStudio Team, 2015). The data were assessed for homoscedasticity (Levene, 1960), and normality (Shapiro and Wilk, 1965). All multiple comparisons were first subjected to analysis of variance (ANOVA). Contrast analysis was performed to compare means between groups (Rosenthal and Rosnow, 1985) and when significant differences were found, the means were compared by the *t*-test ($p < 0.05$). Analysis was done using the Las titer average at 12 MAI in log10 of amplicon copies per gram of plant tissue estimated based on a standard curve described by Lopes et al. (2013).

Phylogenetic Tree Construction

To establish the phylogenetic tree, we used the sequences of eight chloroplastic regions (atpB-coding region, rbcL-atpB spacer, rps16 spacer, trnL-F region, rps4-trnT spacer, matK-5′trnK spacer, psbM-trnDGUC spacer and trnG intron) published by Bayer et al. (2009) selecting those accessions for which informations on HLB resistance were available from our study and that from Ramadugu et al. (2016). We added a few accessions within the True Citrus group to cover more species as well as some representatives of the Triphasiinae subtribe to maintain the global structure of the Aurantioideae phylogenetic tree. According to the Swingle and Reece (1967) classification, the Clauseneae tribe was represented by three subtribes: Clauseninae (six accessions), Merrilliinae (one accession) and Micromelinae (one accession). The Citreae tribe was represented by three subtribes: Triphasilinae (nine accessions), Balsamocitrinae (five accessions) and Citrinae (37 accessions). For the 59 selected accessions (**Supplementary Table 5**), the sequences were obtained from the National Center of Biotechnology Information (NCBI). For each genome fragment, the sequences were aligned to the *C.* × *sinensis* reference chloroplast genome sequence (GenBank: NC008334) using BioEdit software (Hall, 1999) and curated manually in InDel areas. The resulting alignment was used to establish a Neighbor Joining (NJ) tree according to Kimura (1980) genetic distances and considering deletions as missing data, using DarWin 6 software (Perrier and Jacquemoud-Collet, 2006). One thousand bootstraps were performed to test the robustness of each branching.

RESULTS

Graft-Compatibility Onto 'Rangpur' Lime Rootstock

Most genotypes used were graft-compatible on 'Rangpur' lime (**Figure 2A**). Incompatibility reactions were observed in two species, *L. acidissima* and *C. halimii*, but only about 1 to 2 years after propagating them and once they had been already Las-inoculated. For *L. acidissima*, progressive overgrowth of the scion resulted in five of the 10 Las challenge-inoculated plants dying before the twelfth month of evaluation (**Figure 2B**). Even considering that two plants were positive ($Ct \leq 34.0$) and another one was considered suspicious ($Ct > 34$ and ≤ 36.0) at 10 MAI, the insufficient number of replications alive at 12 MAI, led us to disregard this species in our analysis. At the end of the experiment, all grafts of these accessions were affected by overgrowth of the scion. For *C. halimii*, incompatibility reaction took almost 1 year to start appearing, but in this case, it affected to 100% of the grafts of 'Valencia' sweet orange budwood pieces on the scion but not on the 'Rangpur' lime rootstock, neither on the bud union between *C. halimii* and 'Rangpur' lime. It was characterized by a profuse exudation of gum at the graft union in all plants (**Figure 2C**). Therefore, only after being propagated, inoculated, and analyzed by qPCR for Las infection during months, we realized that these plants had a problem which was seriously affecting their growth and general aspect and which finally killed all of them between 11 and 15 MAI (**Supplementary Table 6**). The 'Rangpur' lime rootstocks that survived resulted Las-positive at 12 MAI. Moreover, all their *L. acidissima* and *C. halimii* scions were also Las-positive at 12 MAI (results not shown; **Supplementary Table 6**). A possible role of Las in these incompatibility reactions could be discarded because uninfected budwood controls suffered the same incompatibility problems with identical frequencies for both genotypes (results not shown).

Response of Graft-Compatible Citrinae Species to Challenge-Inoculation of Las

All graft-compatible accessions were propagated on 'Rangpur' lime rootstock and challenge-inoculated by grafting two budwood pieces, one on the rootstock and another one on the scion, from Las-infected 'Valencia' sweet orange trees. Uninfected, healthy 'Valencia' trees were used as a source of budwood pieces to be grafted in the corresponding controls for each genotype. Las infection in budwood pieces was confirmed by testing a small piece from each one by qPCR. Using this challenge inoculation system, Las infection in the scion may come from the infected budwood pieces grafted on it and/or from the 'Rangpur' lime stock once infected. In any case, infection of the rootstock would ensure a continuous flow of bacteria moving up to the scion. Root infection was evaluated just at 12 MAI, because root sampling damaged the composite plants and at that time-point scion Las-infection outcomes were already obtained. To confirm consistent rootstock infection, 'Rangpur' lime bark was also evaluated at 12 MAI. However, Las infection in the rootstock was successful in 63 to 100% of the graft-inoculated plants, irrespective of the phylogenetic relation of scion types with *Citrus* (**Supplementary Table 2**). As all scions grafted on challenge inoculated but Las-negative 'Rangpur' lime rootstocks resulted to be Las-negative, for further analyses of Las-resistance we only considered those plants with Las-positive rootstocks. This result revealed largely inefficient Las-infection of the scions through graft-inoculation of infected budwood and how difficult was to get 100% infection of a well-known

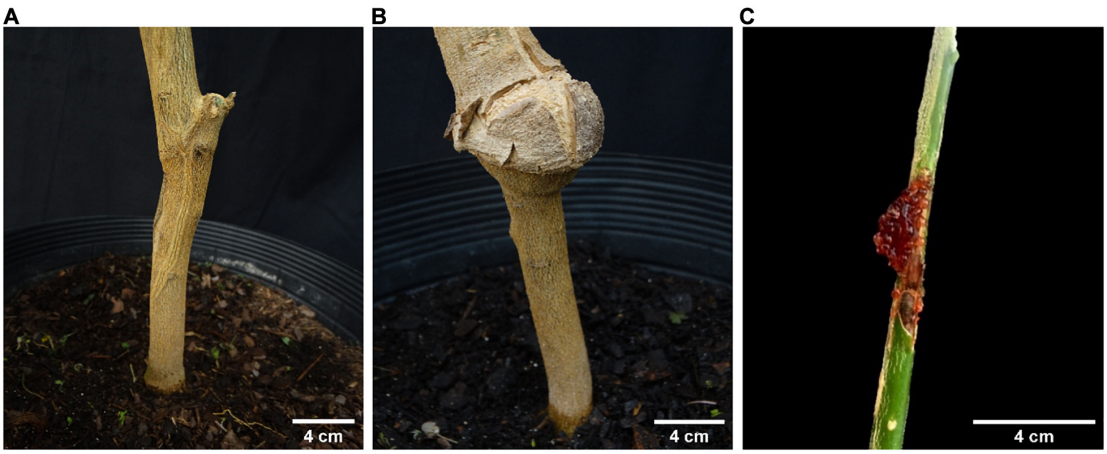

FIGURE 2 | (A) An *Eremocitrus glauca* × *Citrus* × *sinensis* scion propagated on 'Rangpur' lime as an example of good rootstock/scion compatibility between two genotypes. **(B)** *Limonia acidissima* scion propagated on 'Rangpur' lime showing overgrowth incompatibility at the graft union. **(C)** Profuse gum exudation in the graft union between 'Valencia' sweet orange budwood and *Citrus halimii* scion.

Las host as 'Rangpur' lime even under controlled experimental conditions and forced challenge-inoculations, and therefore how important is to use proper controls of infection to avoid getting false negatives when testing genotypes for resistance to this bacterium.

Based on Las infection results, we classified the Citrinae genotypes used as scions in three categories (**Table 2**):

Category 1. Susceptible

This comprised all genotypes with 100% of the clonal propagations resulting Las-positive at 12 MAI, when their evaluation finalized. It included nine genotypes: the two sweet orange controls 'Pera' and 'Tobias,' *C. halimii,* the four accessions of *P. trifoliata* ('Pomeroy,' 'Benecke,' 'Barnes,' and 'Rubidoux'), and the two *Atalantia* species (*A. citroides* and *A. ceylanica*). Ct values in the scion canopy leaves and 'Rangpur' lime roots and bark for each plant and evaluation date are detailed in **Supplementary Table 6**. HLB-like symptoms on the infected scions at 10–12 MAI were masked by nutritional deficiencies probably due to the severe root loss caused by the bacterium (**Figure 3**), so sensitivity to Las infection in the scion was not evaluated. Because of the severe root damage, one *A. citroides* composite plant died about 10 MAI. As mentioned above, *C. halimii* scions were seriously affected by an incompatibility problem with 'Valencia' budwood pieces, irrespective of being infected or not with Las, which finally killed four infected composite plants at 11–12 MAI and the remaining infected ones before 15 MAI. *C. halimii* non-inoculated control composite plants, which had been grafted with Las-free 'Valencia' budwood pieces, died between 11 and 14 months after grafting.

The contrast analysis showed not significant differences in bacterial titers when comparing all Category 1 accessions together ($p < 0.0884$), and *Citrus* with *P. trifoliata* accessions ($p < 0.12$). However, they were significant when comparing *P. trifoliata* versus *Atalantia* ($p < 0.000183$) and also when comparing *Citrus* versus *Atalantia* accessions ($p < 1e{-}04$).

Category 2. Partially Resistant

Eight Citrinae accessions were considered as partially resistant, because they showed Las infection in just part of the clonally propagated scions. In Las-positive propagations, infection was usually delayed and bacterial titers were generally lower than those found in sweet orange controls ($p < 0.0377$). The three accessions of Australian finger lime (*Microcitrus australasica, M. australasica* 'Sanguinea' and *M. australasica* 'True Sanguinea'), 'Faustrimedin,' *M. inodora, M. virgata*, the cherry orange *Citropsis gilletiana* and *Naringi crenulata* were included in this group (**Table 2** and **Supplementary Table 7**). Such variable responses may be considered as a genetic resistance trait, which was partially overcome in some clonal propagations by aggressive Las challenge inoculation. However, no correlation was found between higher bacterial titer in the rootstock and resistance breaking in the scion (**Supplementary Table 8**).

The three *M. australasica* accessions showed similar response patterns, with six out of 11 scions being Las-positive at 8 MAI, and the other five by 12 MAI. However, Las infection was inconsistent because some scions that were infected at 8 MAI resulted qPCR-negative at 12 MAI. Moreover, five scions resulted Las-positive, with Ct > 32.0, just at 1–2 out of five time points evaluated (**Supplementary Table 7**). As Las titers were generally low, with Ct > 30.0 at 10–12 MAI, inconsistent detection may be derived from uneven distribution of a few bacterium cells infecting those scions. The 'Sanguinea' accession showed four out of nine Las positive scions but the 'True Sanguinea' showed just one (Ct > 30.0) out of 11 scions (**Supplementary Table 7**). Because of its apparent higher resistance compared to the other accessions, we decided to maintain and re-evaluate the 'True Sanguinea' accession again at 24 MAI. At this time, three out of 10 were Las-positive (Ct > 32.0). The remaining one died possibly due to the severe damages caused by the bacterium to the 'Rangpur' lime rootstock (**Table 3** and **Supplementary Table 10**).

The trigeneric hybrid 'Faustrimedin,' with half of its genome coming from *M. australasica*, showed only three scions out of 10

being Las-positive, but just at one time point and with Ct > 32.0 in the three cases. When re-evaluated at 24 MAI, four scions resulted to be Las-positive, having been three of them Las-negative until 12 MAI (**Tables 2**, **3** and **Supplementary Tables 7, 10**). *M. virgata,* another hybrid with half of *M. australasica* genome, showed three out 10 scions being Las-positive at 12 MAI. Only five of them could be re-evaluated at 24 MAI, being Las-negative, and the other five died due to Las-induced damages to the rootstock (**Tables 2**, **3** and **Supplementary Tables 7, 10**). *M. inodora* also was partially resistant, as six out of 13 scions resulted Las-positive at least at one time point, but as in the case of the *M. australasica* accessions and hybrids mentioned above, infections were inconsistent though bacterial titers were not so low as compared to those of the other genotypes. Las-induced damages in the rootstock killed two plants at 8 and 12 MAI and other two before 24 MAI, with three out of 10 scions resulting Las-positive. All accessions that were re-evaluated at 24 MAI had the bark tissue from the 'Rangpur' lime rootstocks also analyzed by qPCR and all were confirmed as Las-infected (**Table 3** and **Supplementary Table 10**).

Regarding the cross-incompatible *Citrus* relatives within this category, *Citropsis gilletiana* had three plants out of 13 that were Las-positive at 12 MAI, and other three with Ct ranging between 34.8 and 35.7 at least at one time point. Likewise, five out of nine *Naringi crenulata* scions were Las-positive but only at 12 MAI, and the bacterial titers were usually low, with Ct > 30.0 (**Table 2** and **Supplementary Table 7**). As in Category 1, HLB-like symptoms resembling mineral deficiencies appeared in infected scions likely due to the severe root loss caused by the bacterium, which precluded evaluating sensitivity to Las infection in the scion.

Contrast analysis among Category 2 accessions showed that differences observed in bacterial titers were not significant within the group ($p < 0.08$), but they were significant when comparing Category 1 and 2 ($p < 0.001$). Moreover, there were significant differences in bacterial titers when specifically comparing 'Pera'

TABLE 2 | '*Candidatus* Liberibacter asiaticus' infection in the Citrinae genotypes evaluated as determined through detection of the 16S DNA by qPCR at 12 months after infection.

Category	Accession	Freq.[a]	Scion		Rootstock			
			Leaves		Root		Bark	
			Ct avg[b] ± SEM[c]	Log avg[d] ± SEM	Ct avg ± SEM	Log avg ± SEM	Ct avg ± SEM	Log avg ± SEM
1	*Citrus* × *sinensis* 'Pera'	41/41	25.7 ± 0.6	4.7 ± 0.2	30.0 ± 0.6	3.4 ± 0.2	27.4 ± 0.4	4.2 ± 0.1
	C. × *sinensis* 'Tobias'	09/09	22.3 ± 0.4	5.7 ± 0.1	29.9 ± 1.1	3.4 ± 0.3	29.6 ± 0.9	3.5 ± 0.3
	Citrus × *halimii*	05/05	30.0 ± 1.1	3.4 ± 0.3	31.4 ± 0.9	3.0 ± 0.3	30.6 ± 0.3	3.2 ± 0.1
	Poncirus trifoliata 'Pomeroy'	06/06	27.9 ± 1.4	4.0 ± 0.4	30.6 ± 1.5	3.2 ± 0.5	31.0 ± 0.4	3.1 ± 0.1
	P. trifoliata 'Benecke'	06/06	31.2 ± 1.0	3.0 ± 0.3	29.7 ± 1.4	3.5 ± 0.4	29.5 ± 0.8	3.5 ± 0.2
	P. trifoliata 'Barnes'	08/08	32.0 ± 1.1	2.8 ± 0.3	31.0 ± 0.9	3.1 ± 0.3	29.6 ± 1.0	3.5 ± 0.3
	P. trifoliata 'Rubidoux'	06/06	26.2 ± 1.4	4.5 ± 0.4	30.6 ± 1.1	3.2 ± 0.3	29.4 ± 0.9	3.6 ± 0.3
	Atalantia citroides	10/10	29.9 ± 0.7	3.4 ± 0.2	26.0 ± 0.9	4.6 ± 0.3	26.2 ± 0.6	4.5 ± 0.2
	A. ceylanica	12/12	29.4 ± 1.2	3.6 ± 0.4	25.8 ± 0.4	4.7 ± 0.1	26.0 ± 0.7	4.6 ± 0.2
2	*Microcitrus australasica*	06/11*	31.6 ± 0.7	2.9 ± 0.2	27.4 ± 0.6	4.1 ± 0.2	30.8 ± 0.5	3.2 ± 0.2
	M. australasica 'Sanguinea'	04/09*	26.0 ± 0.9	4.6 ± 0.3	29.0 ± 1.4	3.7 ± 0.4	28.7 ± 0.6	3.6 ± 0.1
	M. australasica 'True Sanguinea'	01/11*	33.8 ± 0.0	2.3 ± 0.0	26.8 ± 0.0	4.4 ± 0.0	29.3 ± 0.0	3.6 ± 0.0
	Faustrimedin hybrid; *C.* × *oliveri*	02/10*	33.1 ± 0.9	2.5 ± 0.3	33.5 ± 0.6	2.4 ± 0.2	30.8 ± 0.2	3.2 ± 0.1
	Microcitrus inodora	04/12*	25.4 ± 2.9	4.8 ± 0.9	31.6 ± 1.4	2.9 ± 0.4	28.0 ± 1.0	4.0 ± 0.3
	Microcitrus virgata hybrid	03/10*	30.0 ± 1.7	3.4 ± 0.5	30.1 ± 1.9	3.4 ± 0.6	31.9 ± 1.1	2.8 ± 0.4
	Citropsis gilletiana	05/13*	32.2 ± 1.9	2.7 ± 0.6	24.8 ± 0.3	5.0 ± 0.1	27.6 ± 0.5	4.1 ± 0.4
	Naringi crenulata	05/09*	32.1 ± 0.7	2.7 ± 0.2	26.8 ± 0.3	4.4 ± 0.1	28.2 ± 0.7	3.8 ± 0.2
3	*Microcitrus warburgiana*	00/09	nd[e]	nd	31.9 ± 0.6	2.8 ± 0.2	30.1 ± 0.6	3.4 ± 0.2
	Microcitrus papuana	00/08	nd	nd	29.1 ± 0.8	3.7 ± 0.2	30.0 ± 0.5	3.4 ± 0.1
	Microcitrus australis	00/10	nd	nd	32.5 ± 0.5	2.7 ± 0.2	30.7 ± 0.7	3.2 ± 0.2
	Microcitrus × *Eremocitrus* hybrid	00/07	nd	nd	31.8 ± 0.7	2.8 ± 0.2	29.4 ± 0.7	3.6 ± 0.2
	Eremocitrus glauca	00/07	nd	nd	25.3 ± 0.3	4.8 ± 0.1	29.8 ± 0.8	3.5 ± 0.2
	E. glauca × *C.* × *sinensis* hybrid	00/12	nd	nd	30.7 ± 0;4	3.2 ± 0.1	28.9 ± 0.5	3.7 ± 0.1
	Eremocitrus × *Microcitrus* hybrid	00/08	nd	nd	33.1 ± 0.3	2.5 ± 0.1	31.1 ± 0.3	3.1 ± 0.1

[a]*Freq., number of Las-positive scions (Ct ≤ 34.0)/total number of plants evaluated (with Las-positive rootstocks, Ct ≤ 34.0).*
[b]*Ct avg, cycle threshold average determined through detection of the 16S DNA by qPCR.*
[c]*SEM, standard error of the mean.*
[d]*Log, Las titer average in log10 of amplicon copies per gram of plant tissue estimated based on a standard curve as described by Lopes et al. (2013).*
[e]*nd, non-detected.*
Rootstock roots and bark avg ± SEM data from accessions in the category 2 were obtained exclusively from plants with positive (Ct ≤ 34.0) or suspicious to be positive (34.0 < Ct ≤ 36.0) leaves. Details on values of each plant are in the* ***Supplementary Tables 6, 7, 9**.*

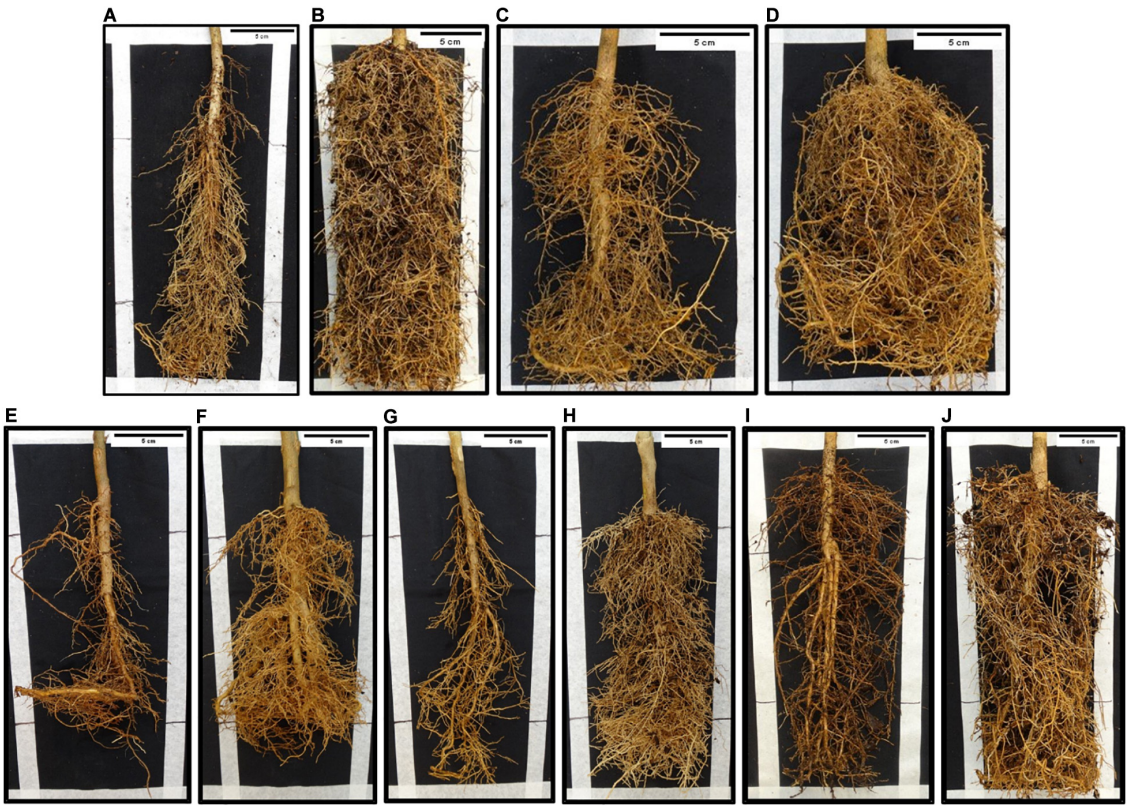

FIGURE 3 | HLB-like symptoms in roots from 'Rangpur' lime rootstocks graft-inoculated either with Las-infected or healthy control budwood in composite plants with infected vs. healthy control scions from three susceptible, the partially resistant *Microcitrus australasica* and the full-resistant *E. glauca* × *C.* × *sinensis* genotypes at 12 months after challenge-inoculation. **(A)** *Citrus* × *sinensis* 'Pera', Las-infected; **(B)** *C.* × *sinensis* 'Pera', control; **(C)** *C.* × *sinensis* 'Tobias', Las-infected; **(D)** *C.* × *sinensis* 'Tobias', control; **(E)** *Atalantia citroides*, Las-infected; **(F)** *Atalantia citroides*, control; **(G)** *Microcitrus australasica*, Las-infected; **(H)** *Microcitrus australasica*, control; **(I)** *E. glauca* × *C.* × *sinensis* hybrid, Las-inoculated; **(J)** *E. glauca* × *C.* × *sinensis* hybrid, control.

and 'Tobias' sweet orange controls versus Category 2 ($p < 0.0377$) and when comparing *P. trifoliata* versus Category 2 accessions ($p < 0.001$).

Category 3. Resistant

The seven accessions included into this group were *Microcitrus warburgiana*, *M. papuana*, *M. australis*, a *Microcitrus* sp. × *E. glauca* hybrid, *Eremocitrus glauca,* an eremorange hybrid (*E. glauca* × *Citrus* × *sinensis*), and an *E. glauca* × *Microcitrus* sp. hybrid. A variable number of plants from these accessions showed Las-infected 'Rangpur' lime rootstocks, confirming the partial success of Las inoculation (**Supplementary Table 1**), but none of the leaf samples from scions grafted on Las-positive rootstocks resulted positive for Las at 12 MAI (**Table 2** and **Supplementary Table 9**). Because of this, they were classified as resistant to Las. Although *E. glauca* was graft-compatible with 'Rangpur' lime, its growth and development was generally poor, being this an intrinsic characteristic of the accession and not related to bud union problems (Bitters et al., 1964, 1969). The other six accessions were further evaluated at 24 MAI and the full resistance was confirmed for all of them (**Table 3** and **Supplementary Table 10**). The eremorange, the *Microcitrus* sp. × *E. glauca* hybrid and the *E. glauca* × *Microcitrus* sp. hybrid evaluated in this work for the first time, showed the best graft-compatibility with 'Rangpur' lime as scion growth was vigorous. Considering their genetic backgrounds, all the seven accessions are probably sexually compatible with *Citrus*.

To confirm that these accessions were truly resistant, namely that there was vascular connection at the grafts and there was bacterial movement from rootstock to scions through the vascular system, bark tissue from each scion was evaluated by qPCR at 5 and 30 cm above the grafting propagation line and at the canopy (21–152 cm, depending on the accession) at 24 MAI to assess the presence of the bacterium. Las was detected in the scion bark close to the bud union (5 cm) in most plants but as sampling was performed farther from the grafting line, Las was found in very few plants at low concentration (30 cm) or was not detected (canopy) (**Table 4** and **Supplementary Table 11**).

Distribution of the Response to Las Categories According to Phylogeny

A phylogenic tree was established using available data for eight genic and intergenic chloroplastic sequences (Bayer et al., 2009). Sequence alignment for the 59 accessions was performed with BioEdit software and manually curated for InDel regions. We

identified 6430 positions with single nucleotide polymorphism and used them to establish the NJ tree (**Figure 4**). The True Citrus clade grouping the Oceanian genera *Microcitrus*, *Eremocitrus*, *Oxanthera* and *Clymenia* and the Asian genera *Citrus*, *Poncirus* and *Fortunella* was very well defined (bootstrap value = 100%). It was linked with lower support to a cluster joining three *Atalantia* species and *L. acidissima* and then a third cluster including two *Citropsis* species, *N. crenulata* and *Pleiospermium latialatum*. These three clusters constitute a clade quite well defined (bootstrap value = 67%) and differentiated from the other species of the Balsamocitrinae and Triphasiinae subtribes and the Clauseneae tribe (**Figure 4**).

Resistance to Las (category 1 here and categories 6 to 8 for Ramadugu et al., 2016) appeared to be concentrated in the clade joining *Atalantia* species with the True Citrus (**Figure 4**). Within this clade, there was a very strong phenotypic differentiation between the Oceanian species tested that were all classified as full resistant or partially resistant while all

TABLE 3 | *'Candidatus* Liberibacter asiaticus' infection in the Citrinae genotypes re-evaluated at 24 months after inoculation, as determined through detection of the 16S rDNA by qPCR.

Category	Accession	Freq.[a]	Scion		Rootstock	
			Leaves		Bark	
			Ct avg[b] ± SEM[c]	Log avg[d] ± SEM	Ct avg ± SEM	Log avg ± SEM
1	*Citrus* × *sinensis* 'Pera'	15/15	25.5 ± 0.7	4.7 ± 0.2	27.9 ± 0.5	4.0 ± 0.1
	C. × *sinensis* 'Tobias'	09/09	23.8 ± 1.4	5.2 ± 0.4	29.1 ± 0.5	3.7 ± 0.2
2	*M. australasica* 'True Sanguinea'	03/10*	33.0 ± 0.1	2.5 ± 0.0	30.9 ± 0.7	3.1 ± 0.2
	Faustrimedin hybrid; *C.* × *oliveri*	04/10*	30.3 ± 1.1	3.3 ± 0.3	31.2 ± 1.1	3.0 ± 0.3
	Microcitrus inodora	03/07*	25.3 ± 1.7	4.8 ± 0.5	27.4 ± 1.3	4.2 ± 0.4
	Microcitrus virgata hybrid	00/09*	nd[e]	nd	29.0 ± 0.4	3.7 ± 0.1
3	*Microcitrus warburgiana*	00/06	nd	nd	30.1 ± 0.0	3.4 ± 0.0
	Microcitrus papuana	00/04	nd	nd	31.2 ± 0.8	3.0 ± 0.2
	Microcitrus australis	00/08	nd	nd	30.3 ± 0.1	3.3 ± 0.0
	Microcitrus × *Eremocitrus* hybrid	00/07	nd	nd	31.0 ± 0.9	3.1 ± 0.3
	E. glauca × *C.* × *sinensis* hybrid	00/11	nd	nd	28.5 ± 0.4	3.8 ± 0.1
	Eremocitrus × *Microcitrus* hybrid	00/08	nd	nd	33.4 ± 0.1	2.4 ± 0.0

[a]*Freq., number of Las-positive scions (Ct ≤ 34.0)/total number of plants evaluated (with Las-positive rootstocks, Ct ≤ 34.0).*
[b]*Ct avg, cycle threshold average determined through detection of the 16S DNA by qPCR.*
[c]*SEM, standard error of the mean.*
[d]*Log, Las titer average in log10 of amplicon copies per gram of plant tissue estimated based on a standard curve as described by Lopes et al. (2013).*
[e]*nd, non-detected.*
Rootstock bark avg ± SEM data from accessions in the category 2 were obtained exclusively from plants with positive (Ct ≤ 34.0) or suspicious to be positive (34.0 < Ct ≤ 36.0) leaves. Details on values of each plant are in the* ***Supplementary Table 10**.*

TABLE 4 | '*Candidatus* Liberibacter asiaticus' detection in scions from each plant of Category 3 accessions plus sweet orange controls. Bark samples were taken at different distances from the rootstock (5 cm, 30 cm and at the canopy) at 24 months after inoculation.

Accession	Freq.[a]	05 cm		Freq.	30 cm		Freq.	Canopy (21–152 cm)	
		Ct avg[b] ± SEM[c]	Log avg[d] ± SEM		Ct avg ± SEM	Log avg ± SEM		Ct avg ± SEM	Log avg ± SEM
Citrus × *sinensis* 'Pera'	41/41	25.5 ± 0.4	4.8 ± 0.1	41/41	25.4 ± 0.3	4.8 ± 0.1	41/41	25.0 ± 0.3	4.8 ± 0.1
Citrus × *sinensis* 'Tobias'	09/09	26.7 ± 1.1	4.4 ± 0.33	09/09	25.2 ± 0.9	4.8 ± 0.3	09/09	23.8 ± 0.6	5.3 ± 0.2
Microcitrus warbugiana	04/06	31.7 ± 0.6	2.9 ± 0.2	00/06	nd[e]	nd	00/06	nd	nd
Microcitrus papuana	02/04	32.5 ± 0.4	2.6 ± 0.1	00/04	nd	nd	00/04	nd	nd
Microcitrus australis	07/08	32.8 ± 0.6	2.6 ± 0.2	00/08	nd	nd	00/08	nd	nd
Microcitrus × *Eremocitrus* hybrid	06/07	30.8 ± 0.5	3.2 ± 0.1	03/07	27.8 ± 1.0	4.1 ± 0.3	00/07	nd	nd
Eremocitrus glauca	06/07	31.6 ± 0.8	2.9 ± 0.3	00/07	nd	nd	00/07	nd	nd
E. glauca × *C.* × *sinensis* hybrid	11/11	31.2 ± 0.4	3.0 ± 0.1	04/11	30.6 ± 0.4	3.2 ± 0.1	00/11	nd	nd
Eremocitrus × *Microcitrus* hybrid	08/08	32.1 ± 0.3	2.8 ± 0.1	00/08	nd	nd	00/08	nd	nd

[a]*Freq., number of Las-positive bark scion samples (Ct ≤ 34.0)/total number of plants evaluated (with Las-positive rootstocks, Ct ≤ 34.0).*
[b]*Ct avg, cycle threshold average determined through detection of the 16S DNA by qPCR.*
[c]*SEM, standard error of the mean.*
[d]*Log, Las titer average in log10 of amplicon copies per gram of plant tissue estimated based on a standard curve as described by Lopes et al. (2013).*
[e]*nd, non-detected.*

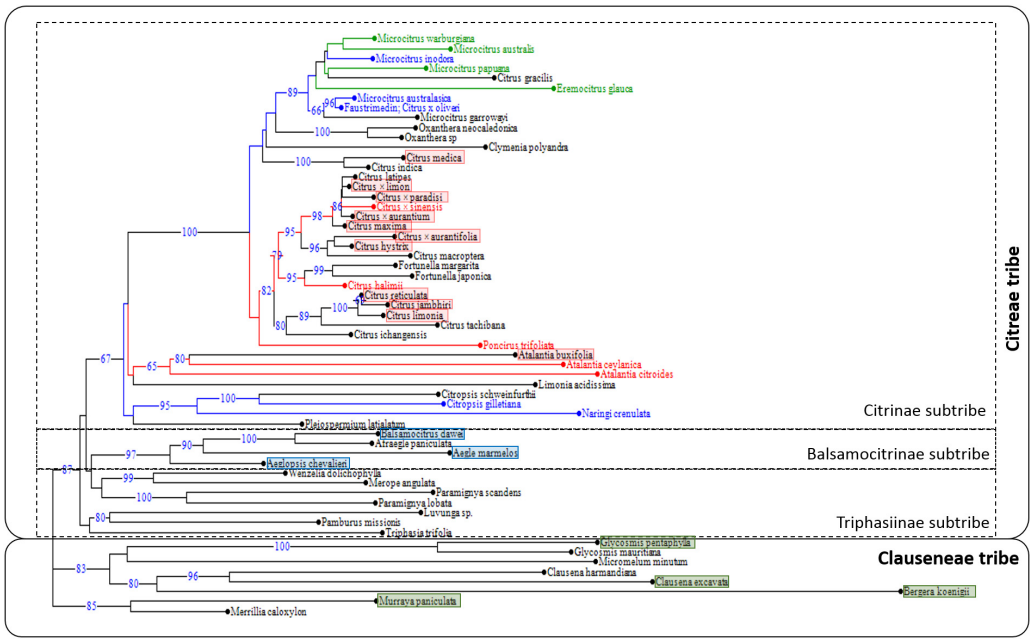

FIGURE 4 | Distribution on the phylogenetic tree of the Aurantioideae sub-family (NJ tree based on eight chloroplast genome fragments) based on accession responses to Las challenge inoculation according to this study (red text: susceptible, blue text: partially resistant, green text: full resistant) and Ramadugu et al. (2016) (red rectangle: susceptible –categories 6 to 8; blue rectangle: tolerant –categories 3, 4; green rectangle: resistant –categories 1, 2).

Asian citrus species analyzed here or previously assessed by Ramadugu et al. (2016) were found to be susceptible. In the Oceanian clade, all resistant species were grouped in a same sub-clade, differentiated (bootstrap value = 66) from another one grouping the partially resistant *M. australasica*, *C.* × *oliveri* and the non-tested *M. garrawayi*. Intriguingly, within the Citrinae subtribe, *C. gilletiana* and *N. crenulata* were partially resistant to Las (**Figure 4**). During their field evaluation under natural Las inoculation through *D. citri*, Ramadugu et al. (2016) found that all analyzed species of the Balsamocitrinae and Triphasiinae subtribes as well as those of Clauseneae tribe were resistant or tolerant to HLB (categories 1–2 and 3–4 of their study, respectively).

DISCUSSION

Searching for resistance to HLB within *Citrus* and its relatives of the family Rutaceae, subfamily Aurantioideae, has been an active area of research due to the severe damages caused by the disease on tree performance, production and fruit quality. Citriculture costs in HLB-affected regions have been increased due to the implementation of treatments to keep citrus groves economically productive but curative methods capable of overcoming losses are lacking (Bassanezi et al., 2020).

Therefore, there is a need for confident and reliable sources of resistance to either the Las bacterium, the insect vector *D. citri* or both, which could be used for introgression into the *Citrus* germplasm (i) to generate new *Citrus*-like cultivars that may be useful as rootstocks or scions, (ii) to map and identify the genes involved in the resistance trait for direct modification of well-known elite *Citrus* cultivars using modern biotechnology tools, especially those which cannot be improved by sexual breeding due to their high heterozygosity, or (iii) to be used promptly as new rootstocks or interstocks to potentially alleviate HLB-induced damages.

However, the characteristics of Las infection in *Citrus* genotypes, cultivars and relatives have generally rendered confounding results, mainly because symptoms appear many months after infection which delays disease development (Hung et al., 2001; Folimonova et al., 2009; Cifuentes-Arenas et al., 2019), they can be mistaken with those derived from nutritional deficiencies, at least at the beginning of infection (da Graça, 1991; Bové, 2006), the uneven distribution of the bacterium within the infected trees (Tatineni et al., 2008; Li et al., 2009; Raiol-Junior et al., 2021), its active multiplication quite restricted to new flushes and developing roots (Hilf and Luo, 2018; Raiol-Junior et al., 2021), the environmental influences on bacterial multiplication and plant colonization (Lopes et al., 2009; Gasparoto et al., 2012) and largely unknown plant host-pathogen molecular interactions which certainly may affect their outcome. To further complicate this interplay, the psyllid vector *D. citri* shows preference for specific colors and volatile compounds emitted by the host plants (Wenninger et al., 2009; Patt and Sétamou, 2010; Hall et al., 2011), and once settled it prefers to feed and reproduce on young shoots rather than mature leaves (Hall et al., 2016; Cifuentes-Arenas et al., 2018). Moreover, psyllids exhibit a clear preference for some hosts within Rutaceae, subfamily Aurantioideae (*Bergera*, *Murraya*) over others (*Citrus*), while some Aurantioideae species

were reported as intermediate [such as *Glycosmis pentaphylla* (Retz.) DC., *Clausena harmandiana* (Pierre) Guillaumin and *Zanthoxylum ailanthoides* (L.)] and other *Citrus* relatives as highly resistant to the psyllid, such as certain *Poncirus trifoliata* accessions (Westbrook et al., 2011; Richardson and Hall, 2013; Hall et al., 2015; Felisberto et al., 2019). Furthermore, the use of seedlings, especially in the case of Aurantioideae relatives, introduces another factor of variation, not only due to the different morphological, development and physiological features of juvenile vs. mature plants, but more importantly because many *Citrus* relatives are monoembryonic, so propagation through seeds does not provide clonal plants but segregating progenies genetically different to the mother genotype. Taking all this into consideration and aiming to evaluate resistance to Las within *Citrus* relatives undoubtably, we decided to center our study on the bacterium-host interaction in a way avoiding any interference of the above mentioned factors.

The Citrinae genotypes of interest were first propagated clonally onto the 'Rangpur' lime rootstock and then inoculated by grafting well-controlled Las-infected wood onto the selected plants. However, this was not trivial. In addition to difficulties on grafting success for clonal propagation due to the genetic distances among some of the Citrinae species used and *Citrus*, problems in transmitting Las to them also occurred. Because of the irregular distribution of Las in plant tissues, use of symptomatic and qPCR-positive segments from donor plants are recommended for challenge inoculation. However, even with this, variation in Las transmission efficiencies averaging 70 to 90% within *Citrus*, *Fortunella* and *Poncirus* germplasm, have been found (Folimonova et al., 2009). In this work, we decided to use the Las-susceptible 'Rangpur' lime as rootstock because this allowed ascertaining which composite plants were actually infected by Las. Although graft-inoculation was performed both in the scion and in the rootstock, only those plants with infected root system were considered for further resistance evaluation, as in no case we detected a Las-positive scion on a 'Rangpur' lime rootstock free from Las. Moreover, this procedure precluded erroneous categorization of false-negative plants as resistant. Furthermore, the upward movement of Las from the rootstock to the scion 5 cm above the graft unions was confirmed in those accessions showing full-resistance to Las.

In more than 30 *Citrus, Poncirus* and *Fortunella* hosts tested, Folimonova et al. (2009) determined that Las was unevenly distributed, with higher titers of the bacterium found in symptomatic tissue. This erratic spread within a plant led to classification of *P. trifoliata* initially as resistant, but after a subsequent test it resulted to be susceptible, thus being results on infection of this genotype inconsistent. In our experiments, the four *P. trifoliata* accessions evaluated were categorized as susceptible as the bacterium readily multiplied in all propagations tested. The delay of about 2–4 months to reach 100% infection in *P. trifoliata* genotypes as compared to the two sweet orange cultivars used as controls, may be attributed to the deciduous nature of the former, which made them to flush much less frequently than most citrus types. However, Ramadugu et al. (2016) also tested two *P. trifoliata* accessions for resistance to HLB, in this case by psyllid-mediated natural infection under field conditions, and classified 'Simmons' as resistant and 'Littleleaf' as showing delayed infection. Notably, in other studies these two accessions were among the most resistant *Poncirus* ones to oviposition by *D. citri* (Westbrook et al., 2011; Richardson and Hall, 2013; Hall et al., 2015), so the field-resistance attributed to these two accessions may be explained by a lack of preference by the vector, especially when exposed mixed with other more preferred hosts, and perhaps not to bacterial resistance. In other works on *D. citri* biology, the four *Poncirus* accessions used here were considered partial resistant to *D. citri* as the female insects laid significantly fewer eggs than in sweet orange controls (Hall et al., 2015, 2017, 2019). Therefore, *P. trifoliata* has interest as a possible source of genetic resistance to *D. citri* rather than to Las. Nevertheless, other *P. trifoliata* accessions should be tested confidently for Las-resistance, especially those being widely used in rootstock breeding programs.

Ramadugu et al. (2016) showed in their field experiment that *Microcitrus* and *Eremocitrus* genera may be useful sources of resistance to HLB, though they could not clarify whether they were resistant to *D. citri* or to Las. Moreover, they observed variation in seedlings disease response within *Microcitrus* and transient infection in *E. glauca*, likely due to segregation as they used zygotic seedlings. We confirmed here using clonal plants that Las-resistance is widely spread within the germplasm of both genera. Remarkably, our results on response to Las challenge-inoculation separated the species from New Guinea, *E. glauca* and *M. australis* in the group of resistant genotypes while *M. inodora* and *M. australasica* were included in the category of partially resistant types. Response to Las of *M. australasica* accessions and hybrids reinforced generally its categorization as partially resistant, with the only exception of their hybrids with *E. glauca*, which were full-resistant. Interestingly, all *E. glauca* hybrids used in this study were fully resistant to Las, suggesting that the determinants of resistance in *E. glauca* may have dominant inheritance, which is particularly engaging for introgression breeding schemes. Therefore, *E. glauca* and its hybrids as well as *M. australis* may be the most indicated ones among the Australian limes as parents in breeding efforts for generating *Citrus*-like cultivars resistant to Las. Conversely, those *Microcitrus* species and hybrids presenting partial resistance to Las would be less indicated as they would probably confer incomplete resistance to their progenies.

Although sexual-compatibility with *Citrus* is restricted to some True Citrus fruit trees (Swingle and Reece, 1967; Bayer et al., 2009), graft-compatibility is widely reported, at least to most Citrinae genera (sensu Bayer et al., 2009, after Swingle and Reece, 1967). There are also reports of graft-compatibility of *Citrus* on *Clausena* (Bitters et al., 1964) and on *Murraya paniculata* (Swingle and Reece, 1967), which are farther relatives to *Citrus* and have been reported as resistant to Las (Ramadugu et al., 2016; Cifuentes-Arenas et al., 2019), but being suitable hosts for *D. citri* (Felisberto et al., 2019). However, *Citrus* grafts on *Clausena* and *Murraya* could be kept alive under greenhouse conditions but none of them seemed recommendable in a field situation due to poor bud unions and progressive incompatibility problems (Bitters et al., 1969). Other promising, HLB-resistant "Remote Citroid fruit trees" (Swingle and Reece, 1967), such as *Glycosmis*,

are graft-uncongenial with *Citrus* (Bitters et al., 1964). Based on Bitters et al. (1964, 1969, 1977), our studies of graft-compatibility with *Citrus* and response to Las challenge inoculation were centered on Citrinae. Most species used were graft-compatible on 'Rangpur' lime, with the exception of *Limonia acidissima. L. acidissima* has been described as compatible with *Citrus* both as a scion and as a rootstock (Bitters et al., 1964; Swingle and Reece, 1967; Yoshida, 1996; Siebert et al., 2015), but its compatibility with 'Rangpur' lime was not previously tested. Our experience with *L. acidissima* exemplifies that close phylogenetic relations could be used as an approach to foresee graft-compatibility, but it does not always predict successful bud unions (Bitters et al., 1977). *C. halimii* showed another type of unexpected incompatibility, not derived from the scion-rootstock union with 'Rangpur' lime, but from the use of 'Valencia' sweet orange budwood to challenge-inoculate Las onto *C. halimii* scions. At the beginning, we associated the weak growth of propagations to the high sensitivity of *C. halimii* to Las infection, as indicated by Folimonova et al. (2009), but gum exudation started to appear around 11 MAI at sweet orange-*C. halimii* graft unions, irrespective of whether the grafted budwood was Las-infected or not. We included this *Citrus* type in our studies because it had displayed some resistance to HLB in the field experiments performed by Ramadugu et al. (2016), but it showed to be clearly susceptible in our challenges, as also indicated by Folimonova et al. (2009). Regarding graft-union problems, further experiments should be performed to attempt to reveal the causes of the incompatibilities. From the cross-incompatible, graft-compatible *Citrus* relatives tested for resistance to Las, the two *Atalantia* species used were susceptible, as already suggested for *Atalantia citroides* by Feng et al. (2015) using zygotic seedlings, while *Citropsis gilletiana* and *Naringi crenulata* were considered as partially resistant. According to Bitters et al. (1964), both genotypes may be excellent rootstocks for *Citrus*, but clearly, those cross-compatible *Citrus* relatives included in the full-resistant category offered the best alternatives to be tested as *Citrus* rootstocks and interstocks, particularly those showing excellent rootstock-scion compatibility and vigor in our studies, which were the *E. glauca* hybrids with *Microcitrus* and sweet orange.

Looking at the distribution of the different types of responses to Las challenge inoculation with a phylogenetic view, we found that susceptible accessions (category 1 here and categories 6 to 8 for Ramadugu et al., 2016) were concentrated in a clade joining True Citrus and *Atalantia* species plus *L. acidissima*. The sister position of *L. acidissima* and *Atalantia* species as well as the monophyly of True Citrus plus *Atalantia* species was previously described from chloroplast phylogeny by Pfeil and Crisp (2008) as well as Bayer et al. (2009), which conduced them to include *L. acidissima* in the Citrinae subtribe. The sister position of *Atalantia* clade with the True Citrus clade was also validated at nuclear level (Nagano et al., 2018). For the subsequent node of the phylogenetic tree, with a clade including *Citropsis* species, *N. crenulata* and *Pleiospermium latialatum*, the two species evaluated in our study were found partially resistant. Even lacking data of the response of Triphasiinae subtribe to Las challenge inoculation, it seemed that the determinants of susceptibility to Las, understood to mean efficient bacterial multiplication and colonization of the whole tree, appeared in an ancestor of the clade joining *Atalantia* and True Citrus species. All *Atalantia* species have Indian and south East Asia origin while *Citropsis* species are from Africa. It may be hypothesized that the Las susceptibility determinants arose in a common ancestor of *Atalantia* and True Citrus species, in India/South East Asia after the separation of *Citropsis* from Asian genera of the Citrinae subtribe. Pfeil and Crisp (2008) estimated that this separation occurred 12.9 My ago (with a quite large probability interval from 7.0 to 20.7 My).

The distribution of the categories of Las response to challenge inoculation in the True Citrus clade was also highly contrasting, with all Oceanian species considered as resistant or partially resistant while all Asian *Citrus* species were susceptible to Las (Folimonova et al., 2009; Ramadugu et al., 2016; this work), including *C. medica*, which is sister species of Oceanian species according to chloroplast phylogeny (Pfeil and Crisp, 2008; Bayer et al., 2009; Carbonell-Caballero et al., 2015; this work), but allied in a robust clade with *C. maxima* and *C. micrantha* for nuclear phylogeny, as concluded from nuclear gene sequencing (Garcia-Lor et al., 2013), restriction site-associated DNA sequencing (Nagano et al., 2018) and whole-genome sequencing (Wu et al., 2018). Based on chloroplastic phylogeny, Bayer et al. (2009) included *Microcitrus* and *Eremocitrus* within the group of True Citrus fruit trees (sensu Swingle and Reece, 1967), and proposed to include them in the genus *Citrus* following Mabberley (2004). The monophylly of the True Citrus species and the cross-compatibility between Oceanian and Asian citrus species are strong arguments sustaining this suggestion (Ollitrault et al., 2020). However, nuclear phylogeny clearly identified two sister clades within the True Citrus, one for the Asian species and the other for the Oceanian ones, with either full or partial resistance only present in the Oceanian clade (and observed for all Oceanian species tested). Under the hypothesis that Las susceptibility determinants were present but not fixed in the ancestral population of the True Citrus plus *Atalantia* clade, the differentiation between Australian and Asian sub-clades may result from a founder effect in the two geographic regions or/and genetic drift.

According to Swingle and Reece (1967), *Microcitrus* and *Eremocitrus* evolved from a common ancestor probably resembling *M. warburgiana*, which together with *M. papuana* are native to New Guinea. From such an ancestral form, one line of evolution produced the Australian round lime (*M. australis*), another line culminated in *M. inodora*, and a third line of evolution led to the Australian finger limes, *M. australasica*. On the other hand, *E. glauca* rapidly evolved from the common ancestor with marked xerophytic adaptations to the Australian deserts. The chloroplast phylogeny, with the monophylly of all the previously mentioned species, confirm their common origin but, as previously described by Bayer et al. (2009), we observed two main clades for the Australian and New Guinean species. The first one included all full resistant species (*M. warbugiana*, *M. australis, M. papuana* and *E. glauca*) plus the partially resistant *M. inodora* while the species and hybrids tested for the second

clade were all partially resistant (*M. australasica*, its hybrids with *M. australis, M. virgata* and with 'Calamondin,' *C.* × *oliveri*). Considering the historical lack of interactions between HLB-associated pathogens/vectors and the citrus germplasm native to Australia/New Guinea, as they are still lacking in Australia and were detected for the first time with limited spread in Papua/New Guinea only in 2003 (EPPO Global Database, 2020), it may be speculated that resistance to Las in Australian limes is actually due to lack of functional susceptibility genes, likely derived from commonly inactive genetic loci among them.

In conclusion, these results demonstrate that there is consistent, complete and unequivocal resistance to Las, that is absence of bacterial multiplication, in a small group of *Citrus* relatives including *E. glauca* as well as its hybrids with *Citrus* and *Microcitrus* tested here, and also in some *Microcitrus* species, which may be used directly to be assessed as possible *Citrus* rootstocks/interstocks, to breed them with *Citrus* types to generate new *Citrus*-like cultivars and to map specific loci involved in the Las resistance (or lack of susceptibility) phenotype/s. Further studies on the interaction of the Las-resistant vs. susceptible genotypes with *D. citri* and with Las at molecular level would also help in understanding host-pathogen-vector interactions, identify effectors and metabolites prone to genetic modulation in *Citrus* and therefore get full profit of Citrinae genetic resources to produce new citrus cultivars resistant to this ravaging bacterium.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

LP and MA conceptualized and designed the work. MA and LR-J collected the data. MA, SL, LR-J, NW, EG, PO, and LP contributed to data analysis and interpretation, drafting the article and critical revision of the article. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by Fundecitrus, grant no. 817526 from the European Union H2020 Innovation Action Program and project PID2019-104569RB-I00 from the AEI-Spain.

ACKNOWLEDGMENTS

We thank André Luís Sanches and Miguel Antonio Mendes for their excellent help with grafting for propagation, Las challenge inoculations and plant maintenance, Lourdes Carmona and Elsa Pons for helping with SSR marker analysis of True Citrus hybrids, and Fernanda Benedito and Priscila Alves for assistance with DNA extraction and qPCR analysis. We also thank to Embrapa Cassava & Fruits for providing some genetic resources.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.617664/full#supplementary-material

REFERENCES

Albrecht, U., and Bowman, K. D. (2012). Tolerance of trifoliate citrus rootstock hybrids to *Candidatus Liberibacter asiaticus*. *Sci. Hortic.* 147, 71–80. doi: 10.1016/j.scienta.2012.08.036

Bassanezi, R. B., Lopes, S. A., Miranda, M. P., Wulff, N. A., Volpe, H. X. L., and Ayres, A. J. (2020). Overview of citrus huanglongbing spread and management strategies in Brazil. *Trop. Plant Pathol.* 45, 251–264. doi: 10.1007/s40858-020-00343-y

Bayer, R. J., Mabberley, D. J., Morton, C., Miller, C. H., Sharma, I. K., Pfeil, B. E., et al. (2009). A molecular phylogeny of the orange subfamily (Rutaceae: Aurantioideae) using nine cpDNA sequences. *Am. J. Bot.* 96, 668–685. doi: 10.3732/ajb.0800341

Bitters, W. P. (1986). *Citrus Rootstocks: Their Characters and Reactions. UC Riverside Science Library*. Available at: https://citrusvariety.ucr.edu/links/documents/Bitters.pdf (accessed May 13, 2020).

Bitters, W. P., Brusca, J. A., and Cole, D. A. (1964). The search for new citrus rootstocks. *Citrograph* 49, 443–448.

Bitters, W. P., Cole, D. A., and Brusca, J. A. (1969). The citrus relatives as citrus rootstocks. *Proc. First Int. Citrus Symposium* 1, 411–415.

Bitters, W. P., Cole, D. A., and McCarty, C. D. (1977). Citrus relatives are not irrelevant as dwarfing stocks or interstocks for citrus. *Proc. of Int. Soc. of Citricult.* 2, 561–567.

Bové, J. M. (2006). Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. *J. Plant Pathol.* 88, 7–37. doi: 10.4454/jpp.v88i1.828

Bowman, K. D., Faulkner, L., and Kesinger, M. (2016). New citrus rootstocks released by USDA 2001-2010: field performance and nursery characteristics. *HortScience* 51, 1208–1214. doi: 10.21273/HORTSCI10970-16

Capoor, S., Rao, D., and Viswanath, S. (1967). Diaphorina citri Kuway., a vector of the greening disease of citrus in India. *Indian J. Agric. Sci.* 37, 572–579.

Carbonell-Caballero, J., Alonso, R., Ibanez, V., Terol, J., Talon, M., and Dopazo, J. (2015). A phylogenetic analysis of 35 chloroplast genomes elucidates the relationships between wild and domestic species within the genus Citrus. *Mol. Biol. Evol.* 32, 2015–2035. doi: 10.1093/molbev/msv082

Cifuentes-Arenas, J. C., Beattie, G. A. C., Peña, L., and Lopes, S. A. (2019). *Murraya paniculata* and *Swinglea glutinosa* as short-term transient hosts of '*Candidatus Liberibacter asiaticus*' and implications for the spread of huanglongbing. *Phytopatology* 109, 2064–2073. doi: 10.1094/PHYTO-06-19-0216-R

Cifuentes-Arenas, J. C., de Goes, A., de Miranda, M. P., Beattie, G. A. C., and Lopes, S. A. (2018). Citrus flush shoot ontogeny modulates biotic potential of *Diaphorina citri*. *PLoS One* 13:e0190563. doi: 10.1371/journal.pone.0190563

Coletta-Filho, H. D., Targon, M. L. P. N., Takita, M. A., De Negri, J. D., Pompeu, J. R., Carvalho, S. A., et al. (2004). First report of the causal agent of huanglongbing (*Candidatus Liberibacter asiaticus*) in Brazil. *Plant Dis.* 88:1382. doi: 10.1094/PDIS.2004.88.12.1382C

da Graça, J. V. (1991). Citrus greening disease. *Annu. Rev. Phytopathol.* 29, 109–136. doi: 10.1146/annurev.py.29.090191.000545

Desjardins, P., and Conklin, D. (2010). NanoDrop microvolume quantitation of nucleic Acids. *J. Vis. Exp.* 45:e2565. doi: 10.3791/2565

Donadio, L. C., Figueiredo, J. O., and Pio, R. M. (1995). Variedades cítricas brasileiras. *Jaboticabal: FUNEP* 1995:228.

EPPO Global Database, (2020). *Papua New Guinea: Organisms present*. Available at: https://gd.eppo.int/country/PG/organisms (accessed July 19, 2020).

Fanciullino, A.-L., Gancel, A.-L., Froelicher, Y., Luro, F., Ollitrault, P., and Brillouet, J.-M. (2005). Effects of nucleo-cytoplasmic interactions on leaf volatile compounds from citrus somatic fiploid hybrids. *J. Agric. Food Chem.* 53, 4517–4523. doi: 10.1021/jf0502855

Fang, D. Q., Roose, M. L., Krueger, R. R., and Federici, C. T. (1997). Fingerprinting trifoliate orange germ plasm accessions with isozymes, RFLPs, and inter-simple sequence repeat markers. *Theor. Appl. Genet.* 95, 211–219. doi: 10.1007/s001220050550

Felisberto, P. A. C., Girardi, E. A., Peña, L., Felisberto, G., Beattie, G. A. C., and Lopes, S. A. (2019). Unsuitability of indigenous South American rutaceae as potential hosts of *Diaphorina citri*. *Pest Manag. Sci.* 75, 1911–1920. doi: 10.1002/ps.5304

Feng, Y. C., Tsai, C. H., Vung, S., Hung, T. H., and Su, H. J. (2015). Cochin *China atalantia* (*Atalantia citroides*) as a new alternative host of the bacterium causing citrus huanglongbing. *Plant Path.* 44, 71–80. doi: 10.1007/s13313-014-03138

Folimonova, S. Y., Robertson, C. J., Gamsey, S. M., and Dawson, W. O. (2009). Examination of the responses of different genotypes of citrus to huanglongbing (citrus greening) under different conditions. *Phytopathology* 23, 1346–1354. doi: 10.1094/PHYTO-99-12-1346

Froelicher, R., Dambier, D., Bassene, J. B., Costantino, G., Lotfy, S., Didout, C., et al. (2008). Characterization of microsatellite markers in mandarin orange (*Citrus reticulata Blanco*). *Mol. Ecol. Resour.* 8, 119–122. doi: 10.1111/j.1471-8286.2007.01893.x

Fundecitrus, (2020). *Doenças: Greening/HLB*. Available at: https://www.fundecitrus.com.br/levantamentos/greening (accessed March 21, 2020).

Garcia-Lor, A., Curk, F., Snoussi-Trifa, H., Morillon, R., Ancillo, G., Luro, F., et al. (2013). A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the 'true citrus fruit trees' group (Citrinae, Rutaceae) and the origin of cultivated species. *Ann. Bot-London* 111, 1–19. doi: 10.1093/aob/mcs227

Gasparoto, M. C. G., Coletta-Filho, H. D., Bassanezi, R. B., Lopes, S. A., Lourenço, S. A., and Amorin, L. (2012). Influence of temperature on infection and establishment of '*Candidatus Liberibacter americanus*' and '*Candidatus Liberibacter asiaticus*' in citrus plants. *Plant Pathol.* 61, 658–664. doi: 10.1111/j.1365-3059.2011.02569.x

Gottwald, T. R., Aubert, B., and Huang, K. L. (1991). "Spatial pattern analysis of citrus greening in shantou, China," in *Proceedings of the 11th Conference of the International Organization of Citrus Virologists*, eds R. H. Brlansky, R. F. Lee, and L. W. Timmer, (California: Riverside), 421–427.

Gottwald, T. R., Aubert, B., and Xue-Yaun, Z. (1989). Preliminary analysis of citrus greening (Huanglungbin) epidemics in the people's Republic of China and French Reunion Island. *Phytopathology* 79, 687–693. doi: 10.1094/Phyto-79-687

Gottwald, T. R., da Graça, J. V., and Bassanezi, R. B. (2007). Citrus huanglongbing: the pathogen and its impact. *Plant Health Prog.* 8, 1–36. doi: 10.1094/PHP-2007-0906-01-RV

Hall, D. G., Albrecht, U., and Bowman, K. D. (2016). Transmission rates of '*Ca. Liberibacter asiaticus*' by asian citrus psyllid are enhanced by the presence and developmental stage of citrus flush. *J. Econ. Entomol.* 109, 558–563. doi: 10.1093/jee/tow009

Hall, D. G., George, J., and Lapointe, S. L. (2015). Further investigations on colonization of *Poncirus trifoliata* by the Asian citrus psyllid. *Crop Prot.* 72, 112–118. doi: 10.1016/j.cropro.2015.03.010

Hall, D. G., Hentz, M. G., and Stover, E. (2017). Field survey of Asian citrus psyllid (Hemiptera: Liviidae) infestations associated with six cultivars of *Poncirus trifoliata*. *Fla. Entomol.* 100, 667–668. doi: 10.1653/024.100.0328

Hall, D. G., Ramadugu, C., Hentz, M. G., Gmitter, F. G., and Stover, E. (2019). Survey of *Poncirus trifoliata* hybrids for resistance to colonization by Asian citrus psyllid. *Fla. Entomol.* 102, 635–637. doi: 10.1653/024.102.0339

Hall, D. G., Wenninger, E. J., and Hentz, M. G. (2011). Temperature studies with the Asian citrus psyllid, *Diaphorina citri*: cold hardiness and temperature thresholds for oviposition. *J. Insect Sci.* 11:83. doi: insectscience.org/11.83

Hall, T. A. (1999). BioEdit: a User-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41, 95–98. doi: 10.14601/Phytopathol_Mediterr-14998u1.29

Hilf, M. E., and Luo, W. (2018). Dynamics of '*Candidatus Liberibacter asiaticus*' colonization of new growth of Citrus. *Phytopathology* 108, 1165–1171. doi: 10.1094/PHYTO-12-17-0408-R

Hung, T. H., Wu, M. L., and Su, H. J. (2000). Identification of alternative hosts of the fastidious bacterium causing citrus greening disease. *J. Phytopathol.* 148, 321–326. doi: 0931-1785/2000/4806-4321

Hung, T. H., Wu, M. L., and Su, H. J. (2001). Identification of the Chinese box orange (*Severinia buxifolia*) as an alternative host of the bacterium causing citrus huanglongbing. *Eur. J. Plant Pathol.* 107, 183–189. doi: 10.1023/A:1011283906502

Johnson, E. G., Wu, J., Bright, D. B., and Graham, J. H. (2014). Association of '*Candidatus Liberibacter asiaticus*' root infection, but not phloem plugging with root loss on huanglongbing-affected trees prior to appearance of foliar symptoms. *Plant Pathol.* 63, 290–298. doi: 10.1111/ppa.12109

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. doi: 10.1007/BF01731581

Koizumi, M., Prommintara, M., and Ohtsu, Y. (1996). "Wood apple, *Limonia acidissima* L.: a new host for the huanglongbing (greening) vector, *Diaphorina citri*," in *Thirteenth IOCV Conference, 13, 271-275. Abstract Retrieved from Abstracts in International Organization of Citrus Virologists Conference Proceedings database*. Accession No. 2313-5123. (IOCV)

Levene, H. (1960). "Robust tests for equality of variance," in *Contributions to Probability and Statistics*, ed. I. Olkin, (Palo Alto, CA: Stanford University Press), 278–292.

Li, W., Levy, L., and Hartung, J. S. (2009). Quantitative distribution of "*Candidatus Liberibacter asiaticus*" in citrus plants with citrus huanglongbing. *Phytopathology* 99, 139–144. doi: 10.1094/PHYTO-99-2-0139

Li, W., Li, W., Hartung, J. S., and Levy, L. (2006). Quantitative real-time PCR for detection and identification of *Candidatus Liberibacter* species associated with citrus huanglongbing. *J. Microbiol. Methods* 66, 104–115. doi: 10.1016/j.mimet.2005.10.018

Lopes, S. A., Frare, G. F., Bertolni, E., Cambra, M., Fernandes, N. G., Ayres, A. J., et al. (2009). Liberibacters associated with citrus huanglongbing in Brazil: '*Candidatus Liberibacter asiaticus*' is heat tolerant, '*Ca. L. americanus*' is heat sensitive. *Plant Dis.* 3, 257–262. doi: 10.1094/PDIS-93-3-0257

Lopes, S. A., Luiz, F. Q. B. Q., Martins, E. C., Fassini, C. G., Barbosa, J. C., and Beattie, G. A. C. (2013). *Candidatus Liberibacter asiaticus* titers in citrus and acquisition rates by *Diaphorina citri* are decreased by higher temperature. *Plant Dis.* 97, 1563–1570. doi: 10.1094/PDIS-11-12-1031-RE

Luro, F. L., Costantino, G., Terol, J., Argout, X., Allario, T., Wincker, P., et al. (2008). Transferability of the EST-SSRs developed on nules clementine (Citrus clementina Hort ex Tan) to other citrus species and their effectiveness for genetic mapping. *BMC Genom.* 9:287. doi: 10.1186/1471-2164-9-287

Mabberley, D. J. (2004). Citrus (Rutaceae): a review of recent advances in etymology, systematics and medical applications. *Blumea* 49, 481–498. doi: 10.3767/000651904X484432

McClean, A. P. D., and Oberholzer, P. C. J. (1965). Citrus psylla, a vector of the greening disease of sweet orange. *S. Afr. J. Agric Sci.* 8, 297–298.

Miles, G. P., Stover, E., Ramadugu, C., Keremane, M. L., and Lee, R. F. (2017). Apparent tolerance to huanglongbing in citrus and citrus-related germplasm. *HortScience* 52, 31–39. doi: 10.21273/HORTSCI11374-16

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4325. doi: 10.1093/nar/8.19.4321

Nagano, Y., Mimura, T., Kotoda, N., Matsumoto, R., Nagano, A. J., Honjo, M. N., et al. (2018). Phylogenetic relationships of *Aurantioideae* (Rutaceae) based on RAD-Seq. *Tree Genet. Genomes* 14, 1–11. doi: 10.1007/s11295-017-1223-z

Ollitrault, P., Curk, F., and Krueger, R. (2020). ""Citrus taxonomy," in *The Genus Citrus*, eds M. Talon, M. Caruso, and F. Gmitter, (Elsevier), 57–81.

Ollitrault, P., Lotfy, S., Costantino, G., Federici, C. T., Mu, L., Chen, C., et al. (2008). "International effort toward a SSR-based linkage map for *C. clementina*," in *Proceedings of the XIth International Citrus Congress*, (Wuhan: Chine), 26–30.

Parra, J. R. P., Alves, G. R., Diniz, A. J. F., and Vieira, J. M. (2016). *Tamarixia radiata* (Hymenoptera: Eulophidae) x *Diaphorina citri* (Hemiptera: Liviidae):

mass rearing and potential se of the parasitoid in Brazil. *J. Integr. Pest Manag.* 7:5. doi: 10.1093/jipm/pmw003

Patt, J. M., and Sétamou, M. (2010). Responses of the Asian citrus psyllid to volatiles emitted by the flushing shoots of its rutaceous host plants. *Environ. Entomol.* 39, 618–624. doi: 10.1603/EN09216

Perrier, X., and Jacquemoud-Collet, J. P. (2006). *DARwin Software*. Available at: http://darwin.cirad.fr/ (accessed September 01, 2020).

Pfeil, B. E., and Crisp, M. D. (2008). The age and biogeography of *Citrus* and the orange subfamily (Rutaceae: Aurantioideae). *Am. J. Bot.* 95, 1621–1631. doi: 10.3732/ajb.0800214

Phillips, R. L., and Castle, W. S. (1977). Evaluation of twelve rootstocks for dwarfing citrus. *J. Amer. Soc. Hort. Sci.* 102, 526–528.

Raiol-Junior, L. L., Cifuentes-Arenas, J. C., Carvalho, E. V., Girardi, E. A., and Lopes, S. A. (2021). Evidence that '*Candidatus Liberibacter asiaticus*' moves predominantly towards new tissue growth in citrus plants. *Plant Dis.* doi: 10.1094/PDIS-01-20-0158-RE Online ahead of print.

Ramadugu, C., Keremane, M. L., Halbert, S. E., Duan, Y. P., Roose, M. L., Stover, E., et al. (2016). Long-term field evaluation reveals huanglongbing resistance in Citrus relatives. *Plant Dis.* 9, 1858–1869. doi: 10.1094/PDIS-03-16-0271-RE

Richardson, M. L., and Hall, D. G. (2013). Resistance of *Poncirus* and *Citrus* × *Poncirus germplasm* to the asian citrus psyllid. *Crop Sci.* 53, 183–188. doi: 10.2135/cropsci2012.02.0091

Rosenthal, R., and Rosnow, R. (1985). *Contrast Analysis: Focused Comparisons in the Analysis of Variance*. Cambridge: Cambridge University Press.

RStudio Team (2015). *RStudio: Integrated Development Environment for R*. Available online at: http://www.rstudio.com/

Shapiro, S. S., and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611.

Shokrollah, H., Abdullah, T. L., Sijam, K., Abdullah, S. N. A., and Abdullah, N. A. P. (2009). Differential reaction of citrus species in Malaysia to huanglongbing (HLB) disease using grafting method. *Am. J. Agri. Biol. Sci.* 4, 32–38. doi: 10.3844/ajabssp.2009.32.38

Siebert, T. J., Kahn, T. L., and Krueger, R. R. (2015). Observations of graft compatibility between *Citrus spp*. and related *Aurantioideae taxa*. *Acta Hortic.* 1065, 173–179. doi: 10.17660/ActaHortic.2015.1065.17

Soares Filho, W. d. S., Vilarinhos, A. D., Alves, A. A. C., Cunha Sobrinho, A. P., Oliveira, A. A. R., Souza, A. S., et al. (2003). *Citrus Breeding Program at Embrapa Cassava & Fruits: Development of Hybrids*. Cruz das Almas: Embrapa Mandioca e Fruticultura.

Swingle, W. T., and Reece, C. (1967). "The botany of Citrus and its wild relatives," in *The Citrus Industry*, eds W. Reuther, H. J. Webber, and L. D. Batchelor, (Berkeley and Los Angeles: Univ. Calif. Press), 191–430.

Tatineni, S., Sagaram, U. S., Gowda, S., Robertson, C. J., Dawson, W. O., Iwanami, T., et al. (2008). In planta distribution of '*Candidatus Liberibacter asiaticus*' as revealed by polymerase chain reaction (PCR) and real-time PCR. *Phytopathology* 98, 592–599. doi: 10.1094/PHYTO-98-5-0592

Teixeira, D. C., Danet, J. L., Eveillard, S., Martins, E. C., Junior, W. C. J., Yamamoto, P. T., et al. (2005). Citrus huanglongbing in São Paulo State, Brazil: PCR detection of the '*Candidatus*' *Liberibacter* species associated with the disease. *Mol. Cell. Probes* 19, 173–179. doi: 10.1016/j.mcp.2004.11.002

Wenninger, E. J., Stelinski, L. L., and Hall, D. G. (2009). Roles of olfactory cues, visual cues, and mating status in orientation of *Diaphorina citri* Kuwayama (Hemiptera: Psyllidae) to four different host plants. *Environ. Entomol.* 38, 225–234. doi: 10.1603/022.038.0128

Westbrook, C. J., Hall, D. G., Stover, E., Duan, Y. P., and Lee, R. F. (2011). Colonization of *Citrus* and *Citrus*-related germplasm by *Diaphorina citri* (Hemiptera: Psyllidae). *Hortscience* 46, 997–1005. doi: 10.21273/HORTSCI.46.7.997

Wu, G. A., Terol, J. F., Ibáñez, V., López-García, A., Pérez-Román, E., Borredá, C., et al. (2018). Genomics of the origin and evolution of *Citrus*. *Nature* 554, 311–136. doi: 10.1038/nature25447

Yamamoto, P. T., Felippe, M. R., Garbim, L. F., Coelho, J. H. C., Ximenes, N. L., and Martins, E. C. (2006). "*Diaphorina citri* (Kuwayama) (Hemiptera: Psyllidae): vector of the bacterium *Candidatus Liberibacter americanus*," in *Proceedings of the Huanglongbing - Greening International Workshop*, (São Paulo).

Yoshida, T. (1996). Graft compatibility of *Citrus* with plants in the *Aurantioideae* and their susceptibility to citrus tristeza virus. *Plant Dis* 80, 414–517. doi: 10.1094/PD-80-0414

Conflict of Interest: 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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 08 January 2021
doi: 10.3389/fpls.2020.607213

Genome-Wide Association Study Reveals Candidate Genes Involved in Fruit Trait Variation in Persian Walnut (*Juglans regia* L.)

*Anthony Bernard[1,2], Julie Crabier[1], Armel S. L. Donkpegan[1], Annarita Marrano[3], Fabrice Lheureux[2] and Elisabeth Dirlewanger[1]**

[1] Univ. Bordeaux, INRAE, Biologie du Fruit et Pathologie, UMR 1332, Villenave d'Ornon, France, [2] CTIFL, Centre Opérationnel de Lanxade, Prigonrieux, France, [3] Department of Plant Sciences, University of California, Davis, Davis, CA, United States

Elucidating the genetic determinants of fruit quality traits in walnut is essential to breed new cultivars meeting the producers and consumers' needs. We conducted a genome-wide association study (GWAS) using multi-locus models in a panel of 170 accessions of *Juglans regia* from the INRAE walnut germplasm collection, previously genotyped using the Axiom^{TM} *J. regia* 700K SNP array. We phenotyped the panel for 25 fruit traits related to morphometrics, shape, volume, weight, ease of cracking, and nutritional composition. We found more than 60 marker-trait associations (MTAs), including a highly significant SNP associated with nut face diameter, nut volume and kernel volume on chromosome 14, and 5 additional associations were detected for walnut weight. We proposed several candidate genes involved in nut characteristics, such as a gene coding for a beta-galactosidase linked to several size-related traits and known to be involved in fruit development in other species. We also confirmed associations on chromosomes 5 and 11 with nut suture strength, recently reported by the University of California, Davis. Our results enhance knowledge of the genetic control of important agronomic traits related to fruit quality in walnut, and pave the way for the development of molecular markers for future assisted selection.

Keywords: texture analyzer, gas chromatography, HPLC, X-ray CT, fruit traits, GWAS, *Juglans regia* L., walnut

OPEN ACCESS

Edited by:
Agata Gadaleta,
University of Bari Aldo Moro, Italy

Reviewed by:
Aziz Ebrahimi,
Purdue University, United States
Shouvik Das,
Indian Agricultural Research Institute (ICAR), India
Ksenija Gasic,
Clemson University, United States

****Correspondence:***
Elisabeth Dirlewanger
elisabeth.dirlewanger@inrae.fr

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

Received: *16 September 2020*
Accepted: *10 December 2020*
Published: *08 January 2021*

Citation:
Bernard A, Crabier J, Donkpegan ASL, Marrano A, Lheureux F and Dirlewanger E (2021) Genome-Wide Association Study Reveals Candidate Genes Involved in Fruit Trait Variation in Persian Walnut (Juglans regia L.). Front. Plant Sci. 11:607213. doi: 10.3389/fpls.2020.607213

INTRODUCTION

Persian walnut (*Juglans regia* L.) is one of the oldest food sources known (Aradhya et al., 2006) and is subject to myths and legends since ancient times (Rondeau, 1997). It is a tree species with 2n = 2x = 32 chromosomes (Woodworth, 1930), growing in temperate region (McGranahan and Leslie, 2009). Walnut is an important tree crop in France after apple, with a worldwide in-shell walnut production led by China, California, and Iran[1]. If the ability of adapting to changing climatic conditions is among breeding priorities, larger fruit size, larger filling ratio, and easiness of cracking remain the main goals of most walnut breeding programs (Bernard et al., 2018b; Vahdati et al., 2019).

[1] www.fao.org/faostat/

Effective management of genetic and phenotypic diversity within germplasm repositories is of valuable assistance to breeders. Typically, germplasm collections are first evaluated using morphological descriptors (Bernard et al., 2018b). Measurements of nut-related traits (e.g., shell thickness, nut length, nut diameter, etc.) in walnut were broadly conducted in Iran and Turkey using a caliper or a micrometer (Eskandari et al., 2005; Ghasemi et al., 2012; Ahandani et al., 2014; Khadivi-Khub et al., 2015; Mahmoodi et al., 2019). Similar studies were conducted in Europe (Zeneli et al., 2005; Poggetti et al., 2017), but the task is labor-intensive and not accurate. Careful consideration of the phenotyping method is of great importance since the heritability of a given trait may depend on the accuracy of the data (Furbank and Tester, 2011; Burghardt et al., 2017). Phenotyping techniques are evolving along with genotyping technologies, and X-ray computed tomography (CT) is one of the imaging techniques applied in food sciences to evaluate internal quality (Kotwaliwale et al., 2014), used notably for studying nut species such as almond, hazelnut, pecan (Harrison et al., 1993; Kim and Schatzki, 2001; Khosa and Pasero, 2014), and recently walnut (Bernard et al., 2020a).

Determination of genetic architecture of such traits has been the focus of several recent studies. A new era in walnut genetics started with the release of a high-density Axiom^{TM} *J. regia* 700K SNP genotyping array (Marrano et al., 2019a). By using this genomic tool, several association mappings studies have been possible. For instance, Arab et al. (2019) applied this SNP array to dissect the genetic architecture of nut-related traits, such as nut length, nut weight, shell thickness, shell texture, and kernel percentage. In addition, Marrano et al. (2019b) carried out a gene-mapping study to decipher the genetic control of yield, phenology, and kernel pellicle color. In addition, Sideli et al. (2020) investigated the genetic control of shell suture strength phenotyped using a texture analyzer, identifying many candidate genes. By combining acquisition of accurate phenotypic data using innovative techniques with the unique walnut germplasm collection from INRAE of Bordeaux, we studied the genetic control of 25 traits related to nut quality, and we propose several candidate genes involved in those traits. These achievements are the starting point for the selection and development of walnut cultivars with desirable and improved fruit quality features.

MATERIALS AND METHODS

Plant Material and Phenotypic Data Acquisition

The INRAE walnut germplasm collection is publicly available and the GWAS panel is made of 170 unique *J. regia* accessions previously used to study phenological traits and lateral bearing genetic architecture (Bernard et al., 2020b). All the accessions are located in the Fruit Experimental Unit of the INRAE-Bordeaux research center, at Toulenne located 50 km south-west from Bordeaux, France. The accessions were classified into three groups according to the breeding level: Selection, Landrace and Modern varieties. All the accessions were collected from 1988 and 2000 from a collecting work in 23 countries.

Walnuts were phenotyped for 25 traits, important for producers or consumers but also for walnut industry. These traits can be classified into six groups as follows (**Supplementary Table 1**):

- Nut morphometric traits (nut length, nut face diameter, nut profile diameter, and nut surface area). These traits are particularly important for cultivar identification.
- Nut shape and texture traits (nut shape VA3D, nut Feret shape VA3D, nut sphericity, and shell rugosity), decisive for the attractiveness to the consumers.
- Volume traits (nut, shell, kernel, and empty space volumes, and kernel filling ratio), that determine the breaking yield.
- Weight traits (nut weight, x3 extreme groups, kernel weight, and breaking yield).
- Shell cracking related traits (nut face strength, nut suture strength, and shell thickness), crucial for walnut industry.
- Nutritional components (saturated fatty acids, polyunsaturated fatty acids, monounsaturated fatty acids, tocopherols, and vitamin E activity).

In-shell walnut sampling (n = 100 nuts/accession) was performed during harvest seasons in September 2017, 2018, 2019 for each accession. Walnuts were dried for 2 days at 25°C using a food dryer. They were stored in a cold room set to 2°C. We phenotyped 25 traits.

Nut morphometric traits were evaluated on nuts collected in 2017, nut length, nut face diameter and nut profile diameter were measured using an electronic caliper (accuracy 10^{-2} mm) on the sample of 100 nuts/accession. On nuts collected in 2018, all traits were measured using an X-ray CT method on a subsample of 50 nuts randomly selected from the 100 nuts original sample collected. X-ray CT scans and analyses were performed as described by Bernard et al. (2020a). Nut shape and texture traits (nut shape VA3D, nut Feret shape VA3D, nut sphericity, and shell rugosity) were measured using the X-ray CT method on nuts collected in 2018. Volume traits were measured using the X-ray CT method on nuts collected in 2018. Weight traits were measured on nuts collected in 2017, 2018, and 2019 using an electronic scale (accuracy 10^{-1} mm). A subsample of 50 nuts from the 100 nuts sampled every year was randomly selected to obtain the weight of 50 kernels and the breaking yield. The French walnut industry considers seven size groups depending on the diameter: $<$ 28 mm, 28–30 mm, 30–32 mm, 32–34 mm, 34–36 mm, 36–38 mm, $>$ 38 mm. The trait "x3 extreme groups" is the ratio of the number of walnuts in the three extreme groups (34–36 mm, 36–38 mm, and $>$ 38 mm) to the total number of walnuts in all groups. For shell cracking related traits a subsample of 50 randomly selected nuts was halved to determine the force needed to crack the nut (initial rupture) on the face (25 nuts) and on the suture (25 nuts) for 2017, 2018, and 2019 samples using a texture analyzer (TA-PLUS model, TA1 Texture Analyzer series, from Lloyd Materials Testing^{TM}, Ametek®). The following parameters were retained: compression speed of 75 mm/min and detection of the initial break when the force drops abruptly by 80%. For the 2018 harvest, the shell thickness was obtained using the X-ray CT method.

Nutritional components were measured on nuts collected in 2018 using the following methods: fatty acids were extracted by 3/2 (v/v) hexane/isopropanol mixture, miscella washed with 0.8% potassium chloride, vacuum evaporated with a Buchi® rotary evaporator. Fatty acid methyl esters were analyzed using gas chromatography according to NF EN ISO 12966-2 and 12966-4 standards, whereas tocopherols/tocotrienols were processed using high-performance liquid chromatography according to the NF EN ISO 9936 standard. Quantifications were performed at ITERG laboratory (Canéjan, France).

Data Analysis, SNP Genotyping, Population Structure, and Kinship Analyzes

Data management and visualization were performed using "R" software with the package "tidyverse" (Wickham, 2017). The Pearson correlation matrix was performed using the package "corrplot" (Wei and Simko, 2017). The Principal Component Analysis (PCA) was performed using the package "FactoMineR" (Lê et al., 2008). The means of genotypic effects were obtained for each accession by adjusting for year effect using the Best Linear Unbiased Predictions (BLUPs), considering the following mixed linear model:

$$P_{ik} = \mu + Y_i + g_k + e_{ik}$$

where P_{ik} refers to the observed phenotype of the kth accession in the ith year; μ is the mean value of the trait; Y_i is the fixed effect of the ith year, g_k is the random effect of the k genotype; and e_{ik} is the residuals of the model. The BLUPs were performed using the package "lme4" (Bates et al., 2015). The broad-sense heritability of the traits phenotyped during at least 2 years was estimated using the variance components obtained by the previous mixed linear model:

$$H^2 = \sigma^2_G/\sigma^2_P$$

where σ^2_G is the genotypic effect variance and σ^2_P is the phenotypic variance.

The accessions were genotyped using the Axiom™ *J. regia* 700K SNP array containing 609,658 SNPs uniformly distributed over the 16 *J. regia* chromosomes (Marrano et al., 2019a). These SNPs were then filtered through several criteria described previously (Bernard et al., 2020b). Finally, 364,275 robust SNPs were retained for the GWAS.

The population structure was investigated using the "fastSTRUCTURE" software and the most likely K was determined using the ΔK method (Bernard et al., 2020b). The identity-by-descent (IBD) proportions between all pairwise comparisons are already described in Bernard et al. (2020b).

Genome-Wide Association Analysis, LD Blocks, and Search of Annotations

GWAS was carried out using the R package "GAPIT" (Lipka et al., 2012). Two multi-locus models were tested using the BLUPs as phenotypic data: the Multi-Locus Mixed Model (MLMM) (Segura et al., 2012) and the Fixed and random model Circulating Probability Unification method (FarmCPU) (Liu et al., 2016), as already described (Bernard et al., 2020b). Familiar relatedness was accounted for using a kinship matrix estimated with the VanRaden algorithm implemented in GAPIT. In order to correct for population structure, the best number of principal components to include in our models was selected using the "model.selection" function implemented in GAPIT according to the Bayesian Information Criterion (BIC). Significant marker-trait associations (MTAs) were determined using both 1 and 5% Bonferroni correction, and previous knowledge based on literature for cracking related traits. The percentage explained variance R^2 was corrected for genome-wide background.

Each physical position of the identified MTAs was investigated to explore the extension of the surrounding linkage disequilibrium (LD) blocks using "solid spine of LD" method implemented in HaploView v4.2 software (Barrett et al., 2005). We searched the defined LD blocks for candidate genes using the walnut nuclear gene annotation and mapped into the new chromosome-scale reference "Chandler" genome v2.0 (Marrano et al., 2020)[2].

RESULTS

Phenotypic Data

A large variability was observed for all traits within the INRAE walnut germplasm collection (**Table 1**). Considering the range of the nut length in 2017, the values vary from ∼26 to 53 mm according to the accessions. We observed higher variation for nut volume in 2018 (from ∼10,400 to 43,000 mm^3), nut weight in 2017 (from ∼522 to 2,278 g), and for nut suture strength in 2019 (from ∼101 to 777 N). On the contrary, there was low variation for nut sphericity (from 0.84 to 0.93) and the proportion of saturated fatty acids (from 8.60 to 11.29%). The distribution of the 25 studied traits is shown in **Supplementary Figure 1**.

The Pearson correlation matrix (**Figure 1**) indicated strong positive correlations between all nut morphometrics related traits (group a) and volume related traits (group c). For instance, nut length was positively correlated with nut face diameter (0.62), nut volume (0.74), and kernel volume (0.64). Nut weight was also positively correlated with kernel weight (0.81). Nut suture and nut face strengths were positively correlated (0.73) but both were negatively correlated with breaking yield (−0.53 and −0.64 respectively). Nut sphericity was negatively correlated with shell rugosity and nut shape "VA3D" (−1 for both). Kernel filling ratio was negatively correlated with nut volume (−0.57). Finally, breaking yield, which is the ratio between the weight of the kernel and the nut, was not correlated with kernel filling ratio, which is the same ratio but based on volumes.

The broad-sense heritability values estimated for the nine traits measured in at least two years showed differences (**Table 1**). Nut morphometrics related traits had high H^2-values, ranging from 0.79 to 0.91, as those related to the weight (from 0.77 to 0.86). However, nut face strength and suture strength had smaller

[2]https://hardwoodgenomics.org/bio_data/2539069?tripal_pane=group_downloads

TABLE 1 | Descriptive statistics and broad-sense heritability values of the 25 traits studied in the walnut GWAS analysis.

Trait	Unit[a]	Year	Mean ± SD[b]	Range	H^2
Nut morphometrics related traits					
Nut length	mm	2017	37.53 ± 4.20	25.99–52.69	0.91
		2018	38.39 ± 3.91	28.57–51.43	
Nut face diameter	mm	2017	31.54 ± 2.99	23.01–40.55	0.79
		2018	32.27 ± 2.55	25.99–40.75	
Nut profile diameter	mm	2017	32.25 ± 3.41	24.10–42.87	0.83
		2018	33.29 ± 2.88	27.06–42.84	
Nut surface area	mm^2	2018	4,019.54 ± 701.70	2,622.59–7,093.53	–
Nut shape and texture related traits					
Nut shape VA3D	–	2018	1.47 ± 0.10	1.24–1.69	–
Nut feret shape 3D	–	2018	1.25 ± 0.08	1.12–1.48	–
Nut sphericity	–	2018	0.88 ± 0.02	0.84–0.93	–
Shell rugosity	–	2018	1.14 ± 0.03	1.08–1.19	–
Volume related traits					
Nut volume	mm^3	2018	19,400.02 ± 4,889.76	10,382.05–42,813.08	–
Shell volume	mm^3	2018	4,076.78 ± 877.94	2,390.66–9,051.88	–
Kernel volume	mm^3	2018	5,723.89 ± 1,152.14	3,408.85–9,548.93	–
Empty space volume	mm^3	2018	9,599.35 ± 3,226.42	4,536.51–24,212.21	–
Kernel filling ratio	%	2018	30.00 ± 3.38	20.70–37.40	–
Weight related traits					
Nut weight	g	2017	1,109.64 ± 262.52	521.74–2,278.20	0.86
		2018	1,192.80 ± 255.37	624.40–2,251.40	
		2019	1,152.86 ± 263.43	539.98–2,288.86	
X3 extreme groups	%	2017	30.97 ± 35.88	0.00–100.00	0.82
		2018	46.51 ± 35.24	0.00–100.00	
		2019	33.65 ± 35.57	0.00–100.00	
Kernel weight	g	2017	249.27 ± 61.40	113.72–416.40	0.77
		2018	276.12 ± 56.51	138.50–440.70	
		2019	260.18 ± 56.84	146.73–428.03	
Breaking yield	%	2017	44.86 ± 5.23	30.03–65.85	0.80
		2018	45.45 ± 5.11	30.80–59.40	
		2019	46.03 ± 5.09	30.47–60.54	
Cracking related traits					
Nut face strength	N	2017	435.18 ± 106.69	205.72–763.34	0.71
		2018	424.24 ± 119.58	182.80–893.70	
		2019	409.53 ± 105.89	172.41–863.61	
Nut suture strength	N	2017	281.06 ± 102.16	87.49–614.37	0.72
		2018	250.31 ± 88.59	74.40–657.20	
		2019	308.62 ± 104.20	101.35–776.97	
Shell thickness	mm	2018	1.03 ± 0.11	0.73–1.49	–
Nutritional components					
Saturated fatty acids	%	2018	9.90 ± 0.57	8.60–11.29	–
Polyunsaturated fatty acids	%	2018	72.36 ± 2.70	63.91–78.13	–
Monounsaturated fatty acids	%	2018	17.60 ± 2.73	11.77–25.33	–
Tocopherols	mg/kg	2018	371.07 ± 110.19	40.00–646.00	–
Vitamin E activity	mg alpha-TE/100 g	2018	4.27 ± 1.62	0.20–8.30	–

[a]Units abbreviations: mm, millimeter; mm^2, square millimeter; mm^3, cubic millimeter; %, percentage; g, gram; N, Newton; alpha-TE, alpha-tocopherol equivalent. [b]SD, standard deviation.

H^2-values, 0.71 and 0.72, respectively, indicating a more complex genetic architecture.

BLUPs and Cofounding Factors

BLUPs were calculated for the nine traits with at least 2 years of phenotypic data using a linear mixed model that included the genotypic mean as a random effect and the year as a fixed effect. The density plots (**Figure 2**) showed a normal distribution for all traits, except for "x3 extreme groups" trait.

Using a BIC, we defined that the best number of principal components to include in the models for accounting for structure was zero, as previously found for phenological traits

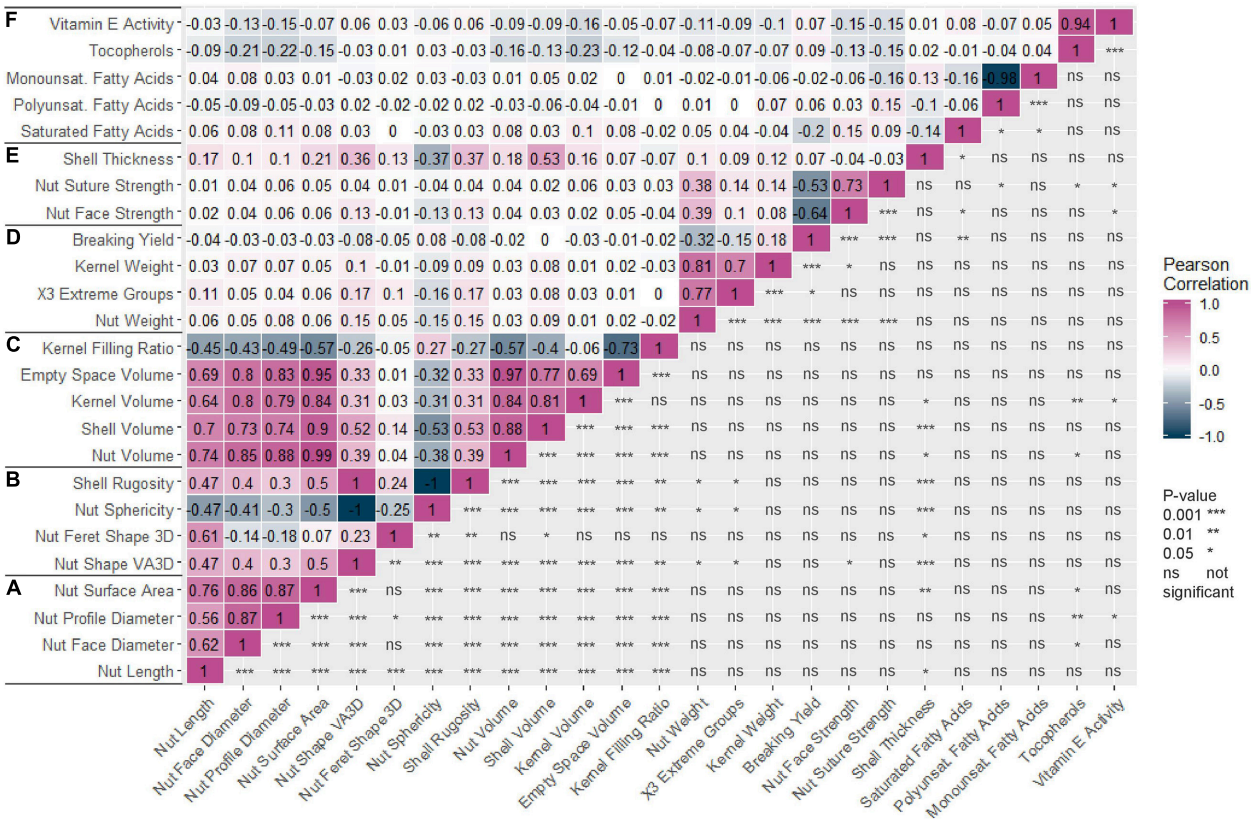

FIGURE 1 | Correlation matrix of the 25 traits studied in the walnut GWAS analysis. **(A)** Nut morphometric traits. **(B)** Nut shape and texture traits. **(C)** Volume traits. **(D)** Weight traits. **(E)** Cracking traits. **(F)** Nutritional components. Pearson correlation coefficient was used and *p*-value is indicated as follows: 0.001 (***), 0.01 (**), 0.05 (*), not significant (ns).

(Bernard et al., 2020b). We also determined the best K as $K = 2$ with a group consisting of the accessions coming from Western Europe and America and the second with the accessions from Eastern Europe and Asia (Bernard et al., 2020b). As a result, we decided to perform a Principal Component Analysis (PCA) using the 25 traits and structure-based individual clustering (**Figure 3**). The first two dimensions of the PCA explained 44.6% of the total variance (**Figure 3A**). Along the first dimension, individuals grouped according to the morphometrics, volume and weight-related traits, whereas the dimension 2 separated the walnut accessions based on shape, texture, and cracking related traits (**Figure 3B**). We could also notice that all groups (Western Europe and America "W," Eastern Europe and Asia "E," and the admixed "A") contained accessions with large or small-sized nuts (**Figure 3C**), since the 95% confidence ellipses intersected and blended together. This means that the structure of our germplasm did not influence the traits, supporting the best number of PCs to include of zero.

MTAs for Fruit Traits

In Manhattan plots, the $-\log_{10}$ of the *p*-value were equal to 7.56, giving 37 MTAs using 1% Bonferroni threshold automatically implemented in GAPIT. By using 5% Bonferroni threshold, the $-\log_{10}$ of the *p*-value were equal to 6.86, giving 58 significant MTAs in total (**Figures 4–6**). We found on average 3.2 MTAs per trait. For instance, we found MTAs on chromosomes (Chr) 5, 8, 11, and 15 for nut length (**Figure 4**), on Chr6, 7, 9, 14 for nut volume (**Figure 5**) and on Chr1, 3, 6, 8, and 12 for nut weight (**Figure 6**). Most of the significant MTAs were obtained using the FarmCPU model, whereas the MLMM approach identified two SNPs associated with both tocopherols content and vitamin E activity (**Supplementary Figure 3**). However, the Manhattan plots indicated strong similarities between both MLMM and FarmCPU models for all traits, even if the signals obtained with the MLMM were below the significant thresholds. We found no significant associations for 6 traits: nut shape VA3D, nut sphericity, nut rugosity, shell volume, saturated and polyunsaturated fatty acids content (**Supplementary Figure 3**).

For some of the correlated traits, we detected MTAs which were very close or even at the same physical position, that we called loci of interest (LOI) (**Table 2**). For instance, MTA AX-171125096 (Chr4, 26,041,265 bp) was associated with both nut face diameter ($R^2 = 19.80$) and nut profile diameter ($R^2 = 16.50$), even though the single traits exhibited additional associations signals on other chromosomes (e.g., on Chr1 for the nut face diameter and on Chr7 for the nut profile diameter). This locus was referred as LOI2 in **Table 2**. Moreover, MTA AX-170690867 (Chr14, 1,248,953 bp) was associated with nut face diameter,

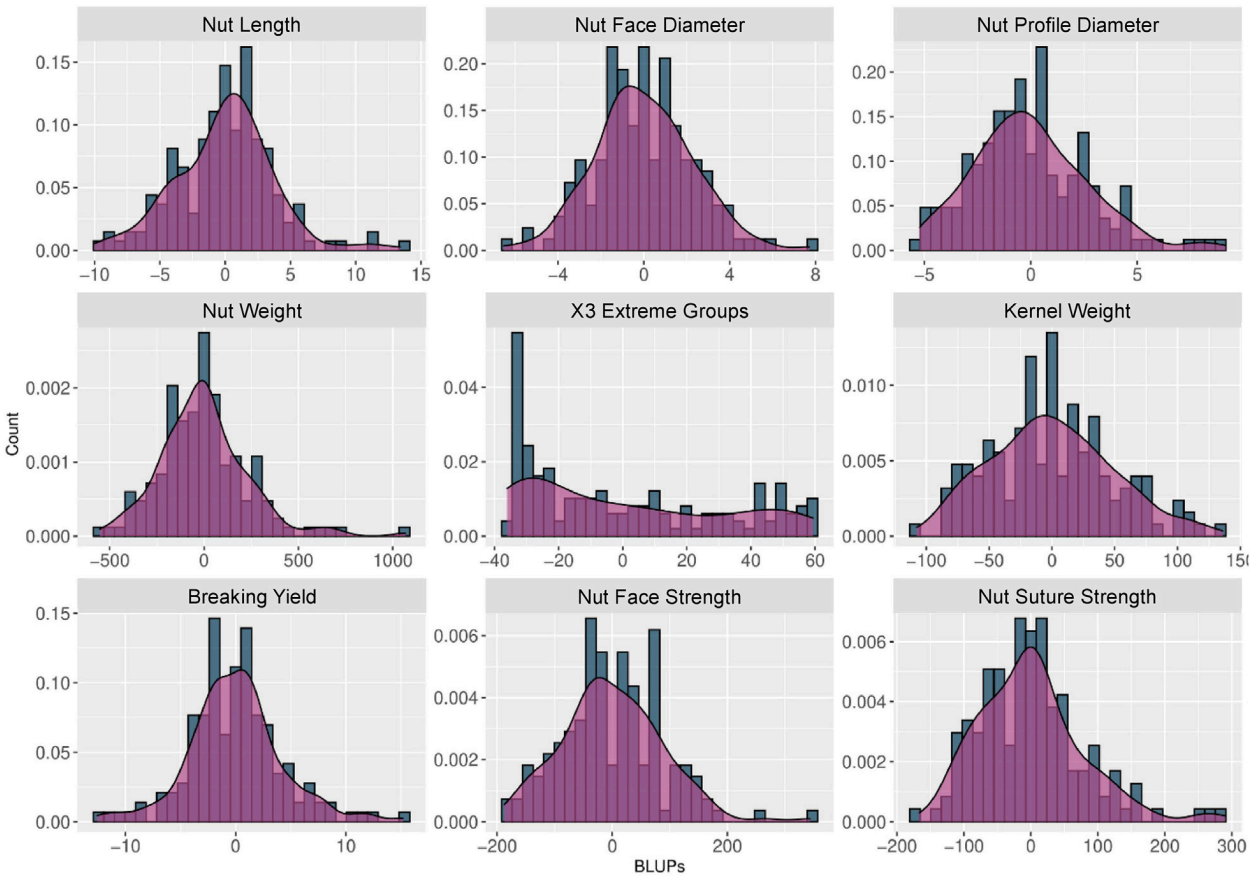

FIGURE 2 | Histograms and density plots of the BLUPs used for the walnut GWAS analysis. The BLUPs were calculated for the 9 traits observed during at least 2 years using a linear mixed model, considering the genotypic mean as a random effect and the year as a fixed effect.

nut volume, and kernel volume, and was only 119 bp apart from MTA AX-171170293 associated with "x3 extreme groups" trait. This locus included these two MTAs, was referred to LOI3 (**Table 2**). Similarly, the MTA AX-170984383 on Chr9 for nut surface area was also found for nut volume and empty space volume (referred as LOI4).

For the cracking-related traits, no signals passed the significant thresholds. However, we observed interesting peaks on Chr2 and 11 for nut face strength, on Chr5 and 11 for nut suture strength, and on Chr2 for shell thickness, which was very close to the one found with the nut face strength (**Table 2**). Since significant associations on Chr5 and 11 for the suture strength have been previously reported, we decided to explore and discuss further these signals.

LD Blocks and Candidate Genes

Using the "solid spine of LD" method, we found differences in LD level around the MTAs. The LOI3, including the MTA AX-170690867 (Chr14, 1,248,953 bp) associated with nut face diameter, nut volume, and kernel volume, and the MTA AX-171170293 (Chr14, 1,248,839 bp) linked with "x3 extreme groups," belong to a very short LD block of 814 bp (**Table 3**). On the contrary, the LD block around the MTA AX-171514816 on Chr7, associated with nut volume, extends for 139,342 bp. We also found several loci in complete linkage equilibrium, such as the MTA AX-171001086 (Chr6, 30,535,497 bp), associated with the nut surface area, which did not belong to any block.

We identified a total of 75 different candidate genes within all the LD blocks (**Table 3**), with 13 candidate genes including some of the MTAs. Regarding nut morphometrics, shape, volume, and weight related traits, a gene coding for a *beta-galactosidase* was found to be associated with nut face diameter, nut volume, kernel volume, and the "x3 extreme groups." We found a *receptor-like cytosolic serine/threonine-protein kinase RBK1* encoding gene for nut length, and an *L-Ala-D/L-amino acid epimerase* encoding gene associated with nut surface area and nut volume. We also found a *probable E3 ubiquitin-protein ligase ARI1* encoding gene linked with the nut Feret shape 3D, and a *protein TIFY 4B* encoding gene linked with empty space volume. Then, a *chaperonin 60 subunit alpha 2, chloroplastic* and a *BEL1-like homeodomain protein 4* encoding genes were found to be associated, respectively, with nut weight and kernel weight. In addition, for the cracking related traits, we determined a gene coding for a *protein TPX2-like* being involved in nut face strength, and

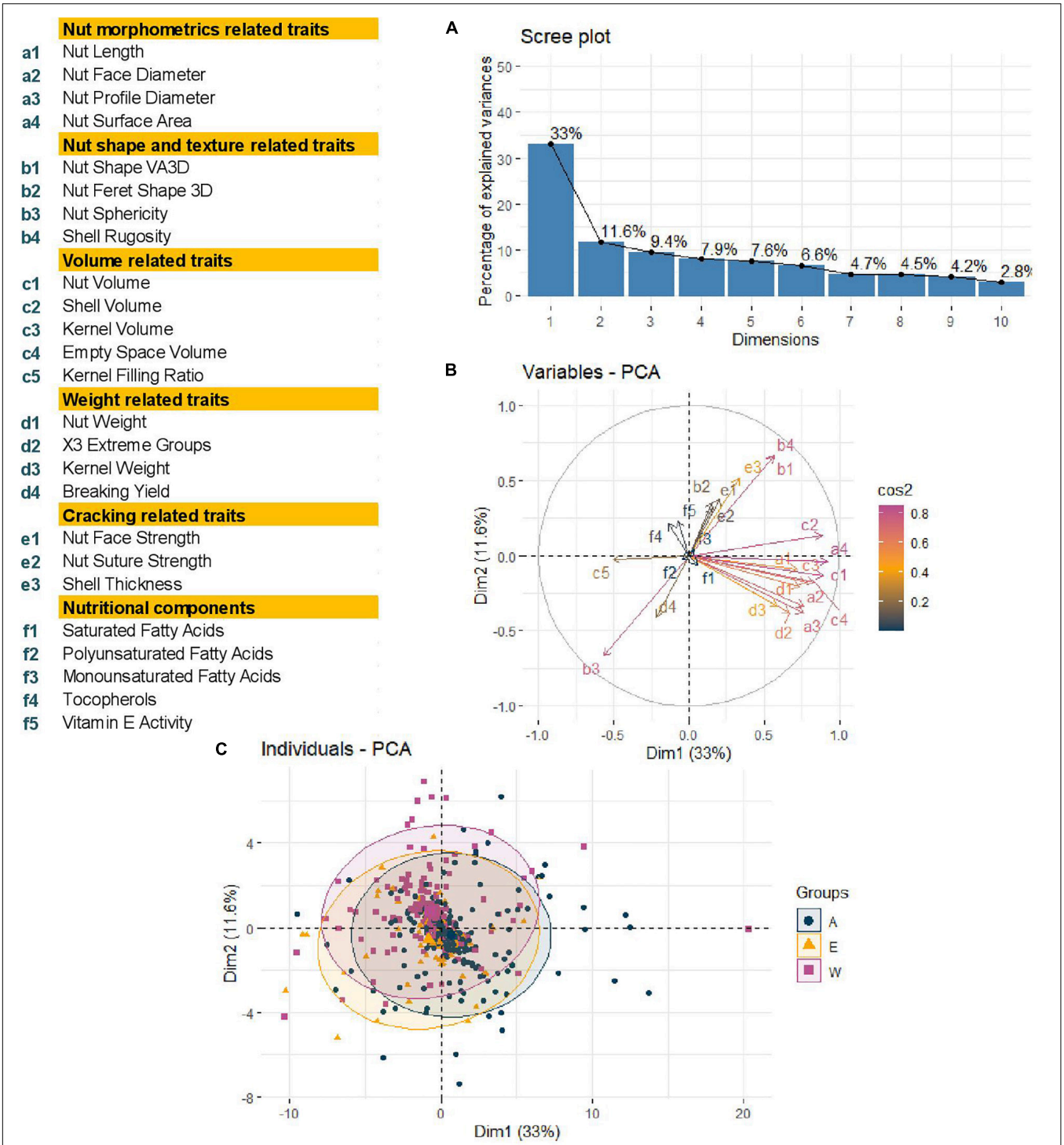

FIGURE 3 | Principal Component Analysis performed using phenotypic raw dataset of the walnut GWAS analysis. **(A)** Scree plot indicating the percentage of explained variances of the first 10 dimensions. **(B)** Variables circular plot indicating the cos^2 of the 25 traits for the two first two dimensions. **(C)** Individual scatterplot indicating the accessions colored by results of structure analysis previously performed. Groups: A, admixed, E, Eastern Europe and Asia, W, Western Europe and America.

a *TVP38/TMEM64 family membrane protein* encoding gene linked with the shell thickness. Finally, we found a candidate gene linked with the monounsaturated fatty acids content such as a *calmodulin* encoding gene. Interestingly, the SNP associated with this trait fell within an exonic sequence of this *calmodulin* encoding gene.

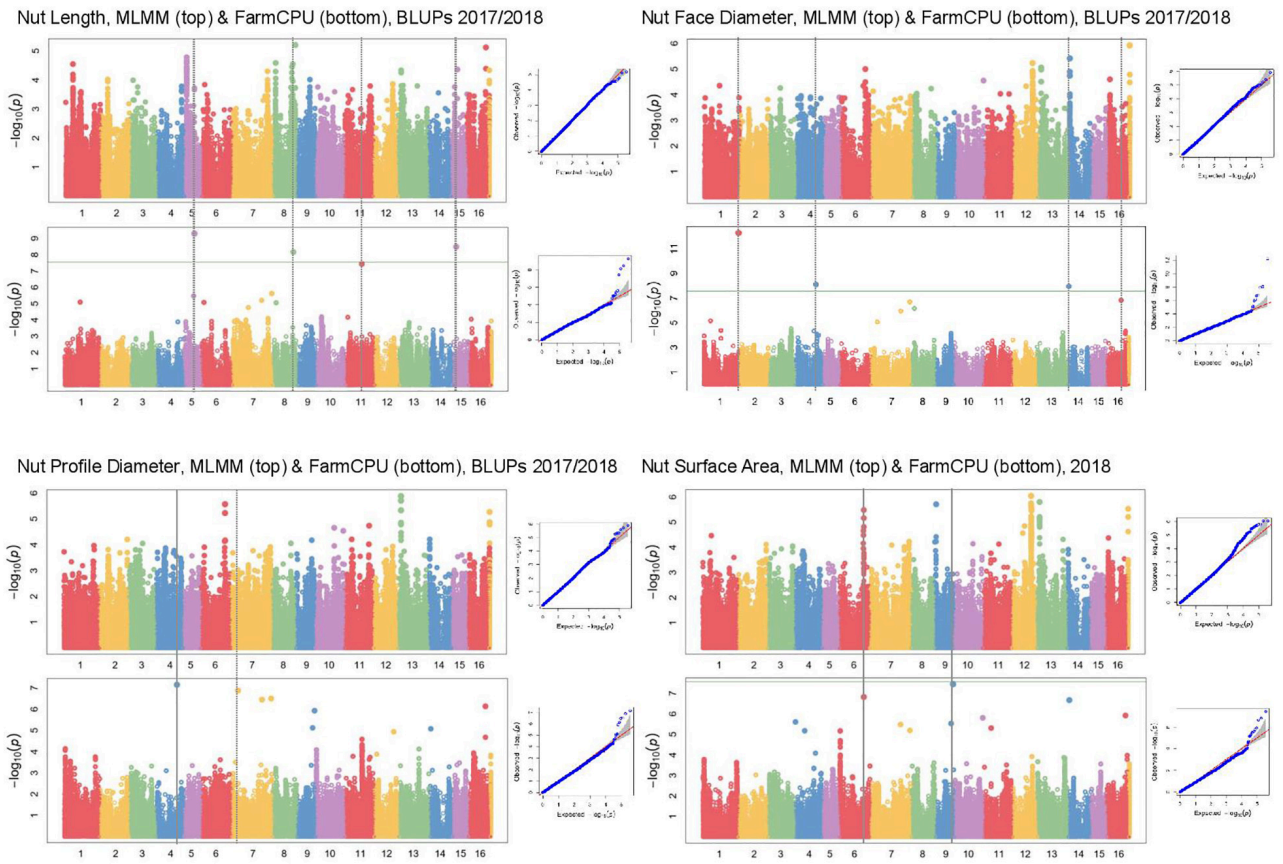

FIGURE 4 | Manhattan plots of nut morphometrics related traits in the walnut GWAS analysis. For each trait, MLMM Manhattan plot and Q-Q plot are shown at the top, and FarmCPU Manhattan plot and Q-Q plot are shown at the bottom. Horizontal green line corresponds to 1% Bonferroni threshold automatically implemented in GAPIT.

DISCUSSION

The INRAE Walnut Germplasm Collection Has a High Degree of Fruit Phenotypic Variability

The walnut genetic resources of INRAE form an *ex situ* collection near Bordeaux, France, and gather numerous accessions of different geographical origin (Bernard et al., 2018a). By comparison with other walnut panels used for GWAS, we observed a larger degree of fruit phenotypic variability. For instance, in the panel used in Iran, the range for nut weight was between 7.71 and 20.11 g (Arab et al., 2019), whereas in our panel, combining all the years studied, the range was between 5.22 and 22.51 g. Similarly, in the work on suture strength reported by the University of California, Davis, the use of a texture analyzer to phenotype suture strength showed a variation between 6.94 to 63.14 kg-force (i.e., between 68.06 to 619.19 N) among the 556 accessions studied (mainly from 39 biparental progenies) corresponding to a variation of 551.13 N (Sideli et al., 2020). In our walnut germplasm, with only 170 accessions and combining all the years, the variation extends from 74.40 to 776.97 N. With the high phenotypic variability and the diversity of the INRAE walnut panel used for GWAS analysis, knowing that most of these traits are highly quantitative, we may expect to find major loci involved in their variation.

MTAs of Correlated Traits Are Not Impacted by Different Phenotyping Methods

In our study, we attempted to improve the accuracy of phenotyping procedures for several traits across the following phenotyping years. This is the case for three of the nut morphometrics related traits: in 2017, we used an electronic caliper to measure nut length, nut face diameter, and nut profile diameter, but we decided to develop an X-ray CT method for the nuts phenotyped in 2018. We had some doubts concerning the impact of changing our phenotyping method on the GWAS results. However, the resulting BLUPs were reliable since we found a significant association on Chr14 for nut face diameter, which was also detected for nut volume and kernel volume, only measured in 2018 and for which no BLUPs were calculated. In addition, the three traits were significantly positively correlated.

Nut face diameter, nut volume, and kernel volume were all correlated traits and the MTA AX-170690867 on Chr14 in the

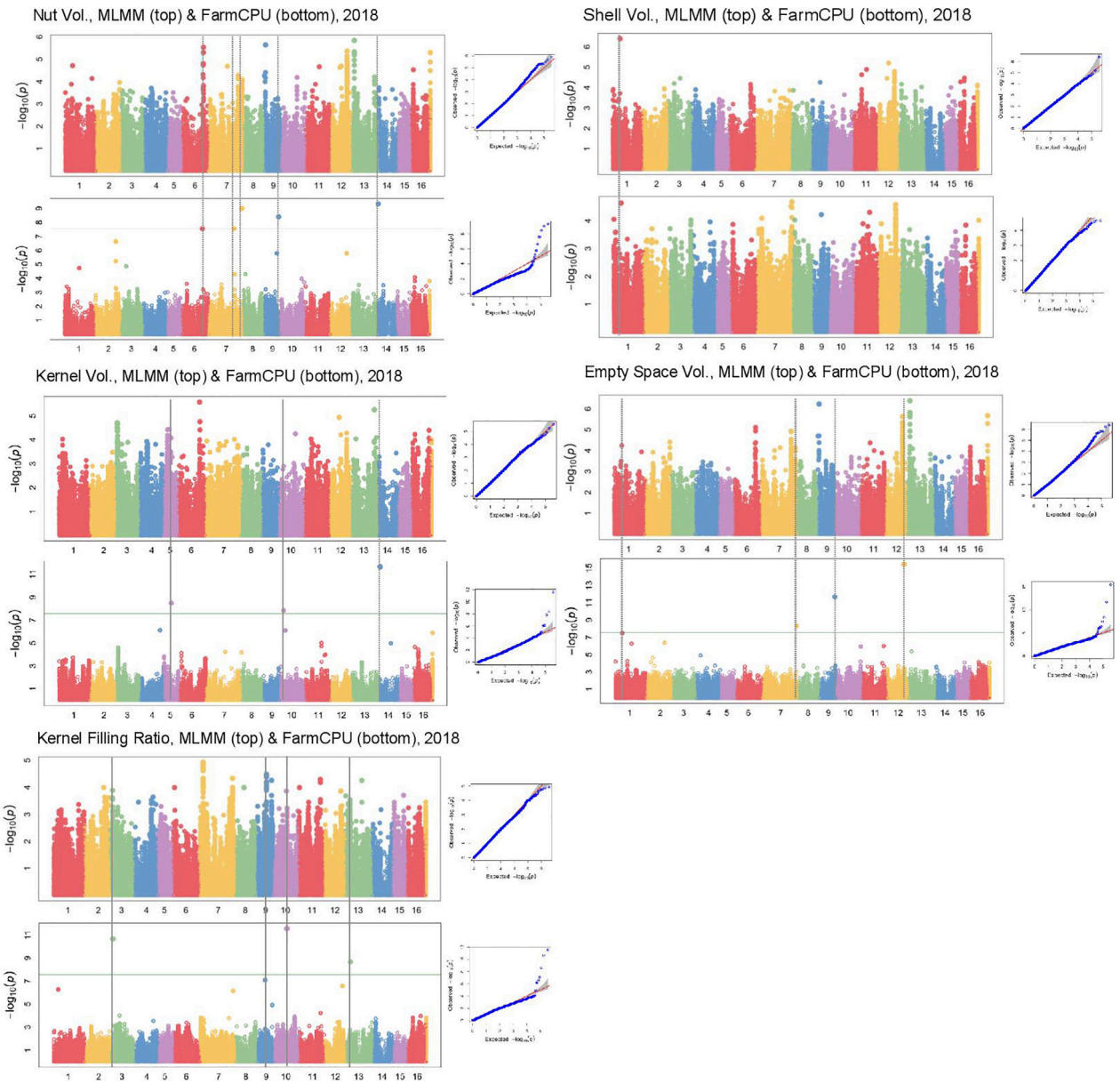

FIGURE 5 | Manhattan plots of volume related traits in the walnut GWAS analysis. For each trait, MLMM Manhattan plot and Q-Q plot are shown at the top, and FarmCPU Manhattan plot and Q-Q plot are shown at the bottom. Horizontal green line corresponds to 1% Bonferroni threshold automatically implemented in GAPIT.

LOI3 is associated with all of them (minor allele G: frequency 0.30, and major allele T: frequency 0.70), with major allele having a positive effect (+0.74 BLUPs unit, +2,101.08 mm^3 and + 521.01 mm^3, respectively). The trait "x3 extreme groups" was associated with the MTA AX-171170293 on Chr14 also in the LOI3 (minor allele A: frequency 0.47, and major allele G: frequency 0.53), with major allele having a negative effect (−9.83 BLUPs unit). The LOI3 including the two MTAs (AX-170690867 and AX-171170293) and two additional SNPs, was a haplotype block with five haplotypes, of which the one of interest had a frequency of only 0.23. The two MTAs were in moderate LD (r^2 = 0.37) and belonged to the sequence of a *beta-galactosidase* encoding gene. This finding suggests that almost equivalent traits might be considered as a simple "repetition" of the phenotyping, leading preferentially to the detection of the major locus.

Although nut size is a complex trait, the association on Chr14 controlled three different dimensions of nut (nut face diameter, nut volume, and kernel volume), suggesting its crucial role in determining nut size. The explained phenotypic variance (R^2) at this locus was high for each of the three traits (i.e., for nut face diameter: 26.36 vs. 19.80, 5.88, and 3.22). However, by phenotyping correlated traits, we were able to detect minor and more specific loci involved in nut size. For instance, if nut volume and kernel volume are sharing the major association on

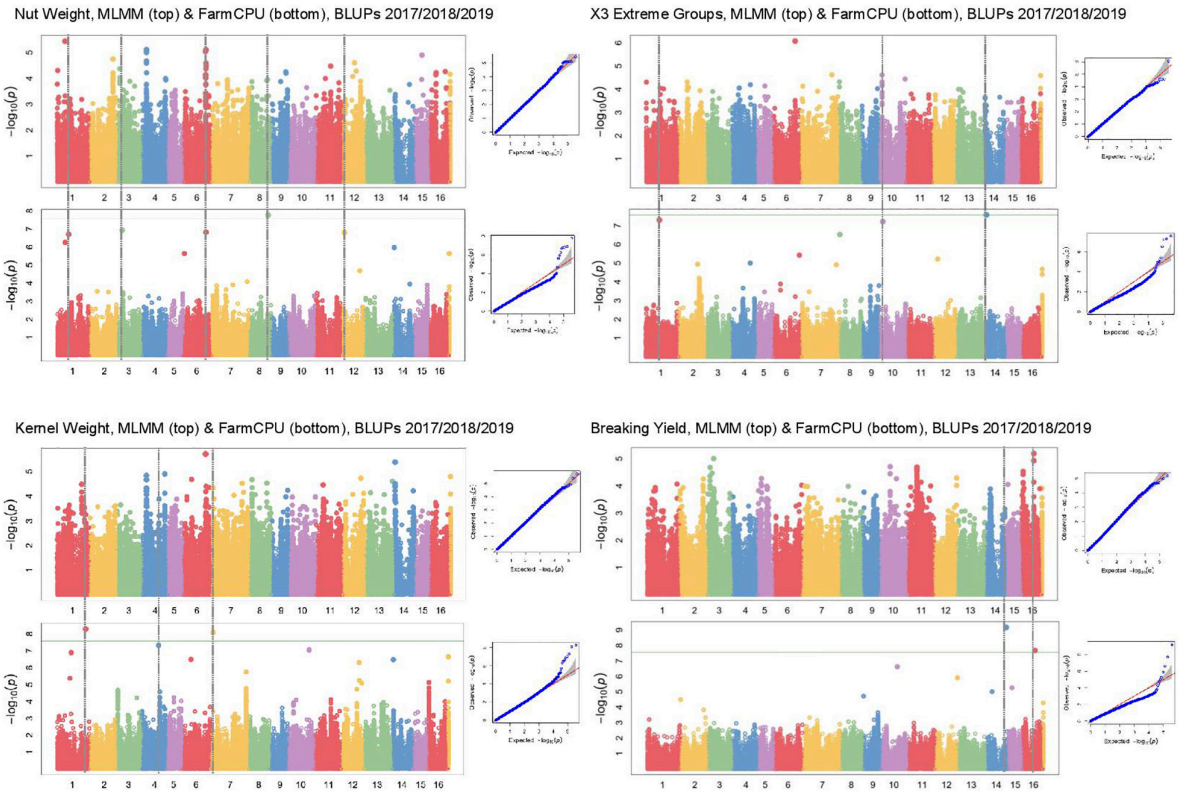

FIGURE 6 | Manhattan plots of weight related traits in the walnut GWAS analysis. For each trait, MLMM Manhattan plot and Q-Q plot are shown at the top, and FarmCPU Manhattan plot and Q-Q plot are shown at the bottom. Horizontal green line corresponds to 1% Bonferroni threshold automatically implemented in GAPIT.

Chr14, we found other specific minor associations. These findings have implications for selection: if it is possible to select superior genotypes for bigger nut size in general (MTA on Chr14), it remain possible for instance, to select for bigger kernel volume using only minor loci.

Comparison of MTAs Detected in Other Germplasm Collections for Similar Traits

For cracking related traits, using different phenotyping methods and FarmCPU model, the MTA AX-170748528 (Chr5, 13,023,760 bp) was found to be associated with the suture strength (Sideli et al., 2020). This MTA was only 762 bp apart from the SNP just below the significant threshold obtained for the suture strength initial rupture. These two SNPs are located in the same region close to a gene coding for a *lamin-like protein*. Similarly, three MTAs were identified on the Chr11 using FarmCPU model, depending on the phenotype recorded with texture analyzer (Sideli et al., 2020). The genomic region spans from 10,995,387 bp to 13,714,234 bp. We found one suggestive association signal within this interval (at 12,909,083 bp). The walnut germplasm collections of INRAE and University of California, Davis, share numerous accessions, such as "Chandler," "Robert Livermore," "Vina," and diverse representing accessions from China, Japan, and France. The two germplasm collections are also related, since many parents of common California varieties are French (Marrano et al., 2019a). The shared genetic background can explain why we found the same major loci. However, our GWAS panel was smaller compared to Sideli et al. (2020), and this could be the reason we did not obtain significant marker association with this trait.

On the contrary, we did not find any association in common with the association study performed on nut traits by Arab et al. (2019). This study was based on a panel of 95 genotypes representing local populations across four different Iranian areas from valleys to mountains. The panel represented a large part of Iranian walnut populations' genetic diversity that is not represented in our panel, explaining why we found different results. This means that our markers can be used at University of California, Davis, for the selection of superior genotypes, but they will be less efficient at the University of Tehran, and vice versa. This finding proves the difficulty to develop markers for selection on a global scale (Mohammadi et al., 2020), and that it will be important to create a large and more diverse collection of walnut accessions for future work, representing the genetic diversity across the world.

Contribution to Better Understanding of the Genetic Control of Fruit Quality

Nut size results from a complex signaling pathway involving cell division and expansion. Serine/threonine kinases are

TABLE 2 | List of the MTAs identified for the 25 traits studied in the walnut GWAS analysis with their position, significance and effect.

MTAs	Chr[a]	Physical position	Significance/Model[b]	i^{2c}	Alleles/Effect[d]	LOI[e] for several traits
Nut morphometrics related traits						
Nut length						
AX-170748449	5	13,134,800	5.02E-10/FarmCPU**	14.45	T,C/2.87	LOI1
AX-170808025	15	4,444,194	3.23E-09/FarmCPU**	15.40	T,G/1.47	
AX-171077810	8	25,778,592	6.88E-09/FarmCPU**	< 0.10	A,G/1.23	
AX-171191763	11	20,831,595	3.66E-08/FarmCPU*	14.50	C,T/2.44	
Nut face diameter						
AX-171110506	1	44,683,085	5.17E-13/FarmCPU**	5.88	A,G/0.90	
AX-171125096	4	26,041,265	7.98E-09/FarmCPU**	19.80	T,C/-1.41	LOI2
AX-170690867	14	1,248,953	1.11E-08/FarmCPU**	26.36	G,T/0.74	LOI3
AX-170799076	16	19,169,530	1.47E-07/FarmCPU*	3.22	T,G/0.67	
nut profile diameter						
AX-171125096	4	26,041,265	7.10E-08/FarmCPU*	16.50	T,C/-1.48	LOI2
AX-171059229	7	7,823,815	1.33E-07/FarmCPU*	< 0.10	T,C/0.72	
Nut surface area						
AX-170984383	9	22,804,956	3.58E-08/FarmCPU*	16.51	A,C/250.39	LOI4
AX-171001086	6	30,535,497	1.50E-07/FarmCPU*	16.11	T,G/-327.60	LOI5
Nut shape and texture related traits						
Nut shape VA3D	No association	–	–	–	–	
Nut feret shape 3D						
AX-170668694	13	4,932,826	2.20E-11/FarmCPU**	13.16	C,T/-0.07	
Nut sphericity	No association	–	–	–	–	
Nut rugosity	No association	–	–	–	–	
Volume related traits						
Nut volume						
AX-170690867	14	1,248,953	4.66E-10/FarmCPU**	17.33	G,T/2,101.08	LOI3
AX-171514816	7	50,526,399	1.00E-09/FarmCPU**	16.39	A,T/2,018.71	LOI6
AX-170984383	9	22,804,956	3.95E-09/FarmCPU**	16.38	A,C/1,783.58	LOI4
AX-170810656	7	38,570,409	2.76E-08/FarmCPU*	<0.10	T,C/-1,574.13	
AX-171001086	6	30,535,497	2.76E-08/FarmCPU*	17.09	T,G/-2,228.97	LOI5
AX-170719786	2	31,237,408	2.35E-07/FarmCPU*	<0.10	G,A/-1,760.19	
Shell volume	No association	-	-	-	-	
Kernel volume						
AX-170690867	14	1,248,953	2.13E-12/FarmCPU**	24.97	G,T/521.01	LOI3
AX-170640054	5	10,793,840	3.24E-09/FarmCPU**	21.02	G,A/-586.65	
AX-170940799	10	6,432,676	1.41E-08/FarmCPU**	19.54	C,A/-594.42	
Empty space volume						
AX-170593629	12	24,730,815	3.45E-16/FarmCPU**	15.03	A,C/-1,717.71	
AX-170984383	9	22,804,956	2.01E-12/FarmCPU**	17.01	A,C/1,462.60	LOI4
AX-171514726	7	50,189,503	4.52E-09/FarmCPU**	17.06	C,G/1,324.30	LOI6
AX-171017658	1	10,234,077	3.27E-08/FarmCPU**	16.07	C,A/-2,014.48	
Kernel filling ratio						
AX-171048523	10	19,688,911	2.75E-12/FarmCPU**	14.98	C,T/0.02	
AX-171583500	3	3,603,319	2.12E-11/FarmCPU**	20.44	T,A/0.01	
AX-171540343	13	6,539,102	2.16E-09/FarmCPU**	0.13	A,G/-0.01	
AX-171117250	9	12,230,952	8.33E-08/FarmCPU*	7.17	T,C/0.01	
Weight related traits						
Nut weight						
AX-171175345	8	26,618,252	1.72E-08/FarmCPU**	0.59	C,T/-176.41	
AX-170573680	3	7,355,283	1.20E-07/FarmCPU*	2.10	G,T/-64.13	
AX-171083810	6	31,463,066	1.54E-07/FarmCPU*	4.23	T,G/-104.12	
AX-170973887	12	3,304,218	1.62E-07/FarmCPU*	17.13	A,G/67.16	
AX-170605550	1	15,926,539	2.02E-07/FarmCPU*	3.47	A,G/50.24	

(Continued)

TABLE 2 | Continued

MTAs	Chr[a]	Physical position	Significance/Model[b]	i^{2c}	Alleles/Effect[d]	LOI[e] for several traits
X3 extreme groups						
AX-171170293	14	1,248,839	2.68E-08/FarmCPU**	21.7	A,G/-9.83	LOI3
AX-170834489	1	19,207,870	5.05E-08/FarmCPU*	7.02	T,C/8.82	
AX-170723157	10	4,802,938	7.00E-08/FarmCPU*	5.01	T,C/-12.56	
Kernel weight						
AX-171207844	1	39,963,556	5.17E-09/FarmCPU**	11.50	G,T/15.27	
AX-171547969	7	3,283,684	8.25E-09/FarmCPU**	4.66	T,A/-11.89	LOI7
AX-170806411	4	23,487,069	4.85E-08/FarmCPU*	10.54	C,A/28.31	
Breaking yield						
AX-171005810	14	26,821,528	6.80E-10/FarmCPU**	0.72	C,A/2.55	LOI8
AX-170746651	16	17,458,649	2.11E-08/FarmCPU**	27.23	A,G/-2,09	
Cracking related traits						
Nut face strength						
AX-170722428	11	13,511,848	9.10E-06/FarmCPU	16.13	G,A/-53.53	LOI9
AX-170865366	2	32,096,896	2.49E-06/MLMM	11.24	G,A/NA	LOI10
Nut suture strength						
AX-170748526	5	13,024,522	2.82E-07/MLMM	12.26	C,A/NA	LOI1
AX-170908256	11	12,909,083	6,72E-05/FarmCPU	14.75	G,A/-49.90	LOI9
Shell thickness						
AX-170865411	2	32,128,087	1.46E-07/MLMM	13.69	A,C/NA	LOI10
Nutritional components						
Saturated fatty acids	No association	–	–	–	–	
Polyunsaturated fatty acids	No association	–	–	–	–	
Monounsaturated fatty acids						
AX-170701328	14	26,730,713	2.73E-12/FarmCPU**	17.73	C,T/−1.43	LOI8
AX-171494826	9	21,100,618	3.81E-10/FarmCPU**	13.58	A,T/−2.19	
AX-170973656	12	3,155,205	1.42E-08/FarmCPU**	13.35	T,C/0.90	
AX-170996807	14	21,159,561	5.63E-08/FarmCPU*	< 0.10	A,G/0.86	
AX-171536319	7	47,712,322	7.00E-08/FarmCPU*	14.27	G,C/−1.72	
Tocopherols						
AX-171134134	10	34,010,861	7.21E-13/MLMM**	14.03	C,T/NA	LOI11
AX-170893393	9	15,280,292	6.22E-09/MLMM**	7.13	C,T/NA	LOI12
AX-170775823	13	23,062,305	5.98E-15/FarmCPU**	12.60	G,T/95.00	
AX-171016396	17[f]	–	8.10E-11/FarmCPU**	4.55	C,A/−40.48	
AX-170615496	8	29,149,403	1.02E-09/FarmCPU**	18.16	A,G/81.85	
AX-171592220	5	16,978,410	9.09E-09/FarmCPU**	< 0.10	C,G/−41.90	
AX-170909834	7	3,266,820	3.89E-08/FarmCPU*	11.70	T,C/73.88	LOI7
AX-171595297	14	7,508,539	4.40E-08/FarmCPU*	11.80	A,T/65.54	
Vitamin E activity						
AX-171134134	10	34,010,861	7.93E-13/MLMM**	13.78	C,T/NA	LOI11
AX-1708 93393	9	15,280,292	2.16E-08/MLMM**	5.32	C,T/NA	LOI12

***Association significant according to Bonferroni correction at 0.01 using the mentioned model. *Association significant according to Bonferroni correction at 0.05 using the mentioned model. No asterisk indicates that the association is not significant but of interest regarding Manhattan plot of the mentioned model and previous knowledge. [a]Chr, Chromosome. [b]The significance value indicated is the unadjusted p-value. [c]R^2, percentage explained variance corrected for genome-wide background. [d]The allelic effect is the difference in mean of measured trait between genotypes with one or other allele. The sign (±) is with respect to the major allele that is second-mentioned. [e]LOI, locus of interest associated to several traits referred to a unique MTA or to several MTAs with close physical position. [f]Chr 17, virtual chromosome corresponding to all unmapped SNPs.*

enzymes catalyzing the phosphorylation from ATP to an amino acid residue (Diallo and Prigent, 2011). Several studies suggested that receptor-like cytosolic kinases contribute to plant development, particularly in cell wall function (Muto et al., 2004; Steinwand and Kieber, 2010) and cell morphogenesis (Becraft, 2002). Similarly, beta-galactosidases are enzymes catalyzing hydrolyzation of galactosyl residues of hemicellulose and pectin from the cell wall (Smith and Gross, 2000; Yang et al., 2018), and L-Ala-D/L-amino acid epimerases are enzymes catalyzing the epimerization of various dipeptides (Lukk et al., 2012), involved in the peptidoglycan pathway, essential for cell wall integrity in bacteria, but also in land plants, although lacking in those peptidoglycans (Yang et al., 2013). The *beta-galactosidase*, *L-Ala-D/L-amino acid*

TABLE 3 | List of the MTAs for 25 walnut traits with the candidate genes located in the corresponding LD block.

MTAs	Chr[a]	Physical position[b]	LD block interval[b,c]	Gene ID v2.0 genome	Gene ID v1.0 genome	Gene interval[b]	Functional annotation[d]	LOI [e] forn several traits [e]
Nut morphometrics related traits								
Nut length								
AX-170748449	5	13,134,800	13,128,354–13,151,244	No gene	–	–	–	
AX-170808025	15	4,444,194	4,439,011–4,444,527	No gene	–	–	–	
AX-171077810	8	25,778,592	25,778,592–25,778,723	Jr08_18480, Jr08_18490, Jr08_18500	108979693	25,774,862–25,778,977	**Receptor-like cytosolic serine/threonine-protein kinase RBK1**	
AX-171191763	11	20,831,595	20,831,267–20,865,013	Jr11_12080	109017990	20,831,874–20,834,965	Death-associated protein kinase 1	
Nut face diameter								
AX-171110506	1	44,683,085	44,682,326–44,689,338	Jr01_34160	108982082	44,678,366–44,683,047	Transcriptional regulator STERILE APETALA	
AX-171125096	4	26,041,265	26,031,692–26,041,265	No gene	–	–	–	
AX-170690867	14	1,248,953	1,248,139–1,248,953	Jr14_01880	109012316	1,248,515–1,262,435	**Beta-galactosidase**	LOIA3
AX-170799076	16	19,169,530	19,167,864–19,170,797	No gene	–	–	–	
Nut profile diameter								
AX-171125096	4	26,041,265	26,031,692–26,041,265	No gene	–	–	–	
AX-171059229	7	7,823,815	7,819,426–7,829,301	No gene	–	–	–	
Nut surface area								
AX-170984383	9	22,804,956	22,785,041–22,811,656	Jr09_14670	108992222	22,786,325–22,787,008	Calcium-binding protein PBP1	LOI4
				Jr09_14680	108993282	22,793,835–22,797,249	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	
				Jr09_14690	108993306	22,807,448–22,811,241	NAD(P)H dehydrogenase (quinone) FQR1	
AX-171001086	6	30,535,497	No block	Jr06_14940	108998609	30,533,899–30,541,448	**L-Ala-D/L-amino acid epimerase**	LOI5
Nut shape and texture related traits								
Nut feret shape 3D								
AX-170668694	13	4,932,826	4,927,503–4,935,968	Jr13_06810, Jr13_06820	109010712	4,929,379–4,935,469	**Probable E3 ubiquitin-protein ligase ARI1 isoform X1**	
Volume related traits								
Nut volume								
AX-170690867	14	1,248,953	1,248,139–1,248,953	Jr14_01880	109012316	1,248,515–1,262,435	**Beta-galactosidase**	LOI3
AX-171514816	7	50,526,399	50,423,525–50,562,867	Jr07_36670	108986372	50,425,702–50,428,580	Receptor-like cytoplasmic kinase 176	
				Jr07_36680, Jr07_36690	108986373	50,429,399–50,432,932	Probable dual-specificity RNA methyltransferase RlmN	

(Continued)

TABLE 3 | Continued

MTAs	Chr[a]	Physical position[b]	LD block interval[b,c]	Gene ID v2.0 genome	Gene ID v1.0 genome	Gene interval[b]	Functional annotation[d]	LOI [e] forn several traits [e]
				Jr07_36700	108986346	50,434,181–50,439,305	Serine/arginine-rich SC35-like splicing factor SCL30	
				Jr07_36710	108986347	50,440,280–50,442,223	Elicitor-responsive protein 1	
				Jr07_36740	108986342	50,446,655–50,461,694	AP3-complex subunit beta-A	
				Jr07_36760, Jr07_36770	108986361	50,493,634–50,497,443	Protein INVOLVED IN *DE NOVO* 2	
				Jr07_36780	108986296	50,500,733–50,502,965	Expansin-A4	
				Jr07_36790	108986359	50,520,505–50,526,244	Telomere repeat-binding protein 4	
				Jr07_36800	108986374	50,540,228–50,542,459	Scarecrow-like protein 14	
				Jr07_36810, Jr07_36820	108986306	50,544,143–50,545,923	Ribosomal RNA large subunit methyltransferase E	
				Jr07_36830	108986354	50,547,343–50,550,514	Protein FIZZY-RELATED 3	
				Jr07_36840	108986353	50,551,233–50,557,760	Actin-related protein 5	
AX-170984383	9	22,804,956	22,785,041–22,811,656	Jr09_14670	108992222	22,786,325–22,787,008	Calcium-binding protein PBP1	LOI4
				Jr09_14680	108993282	22,793,835–22,797,249	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	
				Jr09_14690	108993306	22,807,448–22,811,241	NAD(P)H dehydrogenase (quinone) FQR1	
AX-170810656	7	38,570,409	38,568,209–38,618,466	Jr07_22410	108990569	38,582,842–38,586,425	Oxidation resistance protein 1	
				Jr07_22420, Jr07_22430	109019525	38,587,432–38,601,941	Pre-mRNA-processing factor 19	
				Jr07_22440	109019524	38,617,705–38,622,029	Zinc finger protein CO3	
AX-171001086	6	30,535,497	No block	Jr06_14940	108998609	30,533,899–30,541,448	**L-Ala-D/L-amino acid epimerase**	LOI5
AX-170719786	2	31,237,408	31,218,056–31,252,334	Jr02_18130	109014448	31,222,938–31,226,154	RNA-binding protein NOB1	
				Jr02_18140	109014447, 109004211	31,227,157–31,232,427	Probable LRR receptor-like serine/threonine-protein kinase At1g63430	
				Jr02_18150	109014446	31,244,247–31,248,754	Protein O-glucosyltransferase 2	
Kernel volume								
AX-170690867	14	1,248,953	1,248,139–1,248,953	Jr14_01880	109012316	1,248,515–1,262,435	**Beta-galactosidase**	LOI3
AX-170640054	5	10,793,840	10,792,541–10,794,446	No gene	–	–	–	
AX-170940799	10	6,432,676	6,429,804–6,441,312	Jr10_08990	108996735	6,431,741–6,432,166	Tubulin beta-1 chain	

(Continued)

TABLE 3 | Continued

MTAs	Chr[a]	Physical position[b]	LD block interval[b,c]	Gene ID v2.0 genome	Gene ID v1.0 genome	Gene interval[b]	Functional annotation[d]	LOI [e] forn several traits [e]
Empty space volume								
AX-170593629	12	24,730,815	24,688,102–24,738,026	Jr12_15420	109005084	24,689,734–24,690,792	Early light-induced protein 1, chloroplastic	
				Jr12_15430, Jr12_15440, Jr12_15450	109005081	24,691,848–24,699,404	Dymeclin	
				Jr12_15460	109005079	24,705,762–24,707,383	Heat stress transcription factor A-6b	
				Jr12_15470	109005080	24,709,136–24,712,428	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase 2	
				Jr12_15480	109005077	24,723,413–24,725,999	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase 1	
				Jr12_15490, Jr12_15500	109005076	24,728,030–24,733,224	**protein TIFY 4B**	
AX-170984383	9	22,804,956	22,785,041–22,811,656	Jr09_14670	108992222	22,786,325–22,787,008	Calcium-binding protein PBP1	LOI4
				Jr09_14680	108993282	22,793,835–22,797,249	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	
				Jr09_14690	108993306	22,807,448–22,811,241	NAD(P)H dehydrogenase (quinone) FQR1	
AX-171514726	7	50,189,503	50,141,099–50,199,957	Jr07_36190	108986377	50,141,279–50,143,398	Actin-depolymerizing factor 2	
				Jr07_36200	108986297	50,144,101–50,146,776	Pentatricopeptide repeat-containing protein At2g39620	
				Jr07_36210	108986366	50,148,093–50,152,954	Elongator complex protein 1	
				Jr07_36250, Jr07_36260, Jr07_36270	109015291	50,169,570–50,172,850	THO complex subunit 4B	
				Jr07_36280	108986386	50,171,698–50,172,477	THO complex subunit 4A	
				Jr07_36290, Jr07_36300, Jr07_36310	108986360	50,174,720–50,195,227	**Uncharacterized protein**	
AX-171017658	1	10,234,077	10,227,829–10,244,488	No gene	–	–	–	

(Continued)

TABLE 3 | Continued

MTAs	Chr[a]	Physical position[b]	LD block interval[b,c]	Gene ID v2.0 genome	Gene ID v1.0 genome	Gene interval[b]	Functional annotation[d]	LOI [e] forn several traits [e]
Kernel filling ratio								
AX-171048523	10	19,688,911	19,686,439–19,688,911	No gene	–	–	–	
AX-171583500	3	3,603,319	3,596,654–3,609,718	Jr03_04860	108980538	3,597,955–3,599,758	Alpha-1,3-arabinosyltransferase XAT3	
				Jr03_04870, Jr03_04880	109014312	3,606,572–3,611,958	Transcription initiation factor IIB-2	
AX-171540343	13	6,539,102	6,523,102–6,557,694	Jr13_09170	108993396	6,521,880–6,523,699	Probable aquaporin PIP2-8	
				Jr13_09180	108993404	6,525,283–6,528,372	Pre-mRNA-splicing factor SYF2	
				Jr13_09220	108993314	6,543,826–6,546,850	Blue copper protein	
				Jr13_09240	108993378	6,552,307–6,558,488	Probable ubiquitin-conjugating enzyme E2 23	
AX-171117250	9	12,230,952	12,208,938–12,255,984	Jr09_03370	108982558	12,251,655–12,255,887	Serine/threonine-protein kinase ATM	
Weight related traits								
Nut weight								
AX-171175345	8	26,618,252	26,611,559–26,618,252	Jr08_19090	108985800	26,608,414–26,613,583	Phosphoenolpyruvate carboxykinase (ATP)	
				Jr08_19100	108985801	26,615,257–26,618,938	**Uncharacterized protein**	
AX-170573680	3	7,355,283	7,338,541–7,356,913	Jr03_09310	109003589	7,339,392–7,342,382	14 kDa zinc-binding protein	
				Jr03_09320, Jr03_09330	109003588	7,343,193–7,344,784	Sulfiredoxin, chloroplastic/mitochondrial	
				Jr03_09340	109003587	7,345,868–7,351,659	Phosphoglycerate kinase 3, cytosolic	
				Jr03_09350, Jr03_09360, Jr03_09370	109003586	7,353,891–7,362,252	**Chaperonin 60 subunit alpha 2, chloroplastic**	
AX-171083810	6	31,463,066	31,364,479–31,464,703	Jr06_15230, Jr06_15240	108985243	31,384,208–31,426,380	myb-related protein 308	
AX-170973887	12	3,304,218	3,296,379–3,352,059	Jr12_02800	108998258	3,304,480–3,341,673	Peroxisome biogenesis protein 1	
				Jr12_02810	108998259	3,345,122–3,359,362	FHA domain-containing protein FHA2	
AX-170605550	1	15,926,539	15,898,312–15,966,124	Jr01_17520	109011118	15,897,845–15,907,090	ER lumen protein-retaining receptor	
				Jr01_17530	109006016	15,940,330–15,943,514	Protein XRI1	
				Jr01_17540	109011117	15,947,614–15,951,582	Protein SICKLE	
				Jr01_17550, Jr01_17560	109006017	15,954,073–15,961,294	Purple acid phosphatase 15	

(Continued)

TABLE 3 | Continued

MTAs	Chr[a]	Physical position[b]	LD block interval[b,c]	Gene ID v2.0 genome	Gene ID v1.0 genome	Gene interval[b]	Functional annotation[d]	LOI [e] forn several traits [e]
X3 extreme groups								
AX-170690867	14	1,248,953	1,248,139–1,248,953	Jr14_01880	109012316	1,248,515–1,262,435	**Beta-galactosidase**	LOI3
AX-170834489	1	19,207,870	19,202,146–19,218,695	No gene	–	–	–	
AX-170723157	10	4,802,938	4,802,821–4,802,938	No gene	–	–	–	
Kernel weight								
AX-171207844	1	39,963,556	39,934,883–39,963,556	Jr01_29160, Jr01_29170, Jr01_29180	109003484	39,947,452–39,948,585	Retrovirus-related Pol polyprotein from transposon RE1	
AX-171547969	7	3,283,684	3,279,632–3,290,086	Jr07_03110	108995343	3,280,892–3,285,736	**BEL1-like homeodomain protein 2**	
AX-170806411	4	23,487,069	23,484,729–23,487,069	No gene	–	–	–	
Breaking yield								
AX-171005810	14	26,821,528	26,821,528–26,851,219	No gene	–	–	–	
AX-170746651	16	17,458,649	No block	No gene	–	–	–	
Cracking related traits								
Nut face strength								
AX-170722428	11	13,511,848	13,503,419–13,528,275	Jr11_09630	109000368	13,524,918–13,540,197	Putative GPI-anchor transamidase	
AX-170865366	2	32,096,896	32,095,993–32,104,213	Jr02_19150, Jr02_19160, Jr02_19170	109010833	32,094,342–32,100,753	**Protein TPX2**	
Nut suture strength								
AX-170748526	5	13,024,522	13,017,020–13,109,943	Jr05_10970	108986888	13,045,730–13,046,379	Lamin-like protein	
				Jr05_10980	108986890	13,046,665–13,056,464	Protein argonaute 4A	
AX-170908256	11	12,909,083	12,884,235–12,916,294	No gene	–	–	–	
Shell Thickness								
AX-170865411	2	32,128,087	32,126,341–32,131,231	Jr02_19210	109010875	32,127,046–32,130,960	**TVP38/TMEM64 family membrane protein slr0305**	
Nutritional components								
Monounsaturated fatty acids								
AX-170701328	14	26,730,713	26,728,860–26,731,202	No gene	–	–	–	
AX-171494826	9	21,100,618	No block	No gene	–	–	–	

(Continued)

TABLE 3 | Continued

MTAs	Chr[a]	Physical position[b]	LD block interval[b,c]	Gene ID v2.0 genome	Gene ID v1.0 genome	Gene interval[b]	Functional annotation[d]	LOI [e] forn several traits [e]
AX-170973656	12	3,155,205	3,139,539–3,155,205	Jr12_02680	108998216	3,151,847–3,153,051	Putative clathrin assembly protein At4g40080	
AX-170996807	14	21,159,561	No block	No gene	–	–	–	
AX-171536319	7	47,712,322	No block	Jr07_32830	108992563	47,710,331–47,712,379	**Calmodulin**	
Tocopherols								
AX-171134134	10	34,010,861	34,010,496–34,015,586	Jr10_22690	–	34,011,505–34,043,955	Receptor-like protein 22	LOI11
				Jr10_22700	108983893	34,012,093–34,014,327	Receptor-like protein 7	
AX-170893393	9	15,280,292	15,271,076–15,281,282	Jr09_05900	108994350	15,271,151–15,280,248	E3 ubiquitin-protein ligase HOS1	LOI12
				Jr09_05910	–	15,279,302–15,280,386	**Uncharacterized protein**	
				Jr09_05920	108994314	15,280,768–15,284,023	2-oxoglutarate-Fe(II) type oxidoreductase hxnY	
AX-170775823	13	23,062,305	23,062,229–23,062,659	No gene	–	–	–	
AX-170615496	8	29,149,403	29,145,942–29,153,435	No gene	–	–	–	
AX-171592220	5	16,978,410	16,976,697–16,991,665	Jr05_12430, Jr05_12440, Jr05_12450	108983737	16,978,531–16,980,684	Glucan endo-1,3-beta-glucosidase 14	
AX-170909834	7	3,266,820	3,265,033–3,267,053	No gene	–	–	–	
AX-171595297	14	7,508,539	7,508,539–7,508,805	No gene	–	–	–	
Vitamin E activity								
AX-171134134	10	34,010,861	34,010,496–34,015,586	Jr10_22690	–	34,011,505–34,043,955	Receptor-like protein 22	LOI11
				Jr10_22700	108983893	34,012,093–34,014,327	Receptor-like protein 7	
AX-170893393	9	15,280,292	15,271,076–15,281,282	Jr09_05900	108994350	15,271,151–15,280,248	E3 ubiquitin-protein ligase HOS1	LOI12
				Jr09_05910	–	15,279,302–15,280,386	**Uncharacterized protein**	
				Jr09_05920	108994314	15,280,768–15,284,023	2-oxoglutarate-Fe(II) type oxidoreductase hxnY	

[a]Chr, Chromosome. [b]Physical position given in bp. [c]LD blocks are defined using the "solid spine of LD" method. [d]The candidate genes in bold overlap the physical position of the associated SNP. [e]LOI, locus of interest associated to several traits referred to a unique MTA or to several MTAs with close physical position.

epimerase-like, and *receptor-like cytosolic serine/threonine-protein kinase RBK1* are therefore interesting candidate genes involved in nut morphometrics.

Along with nut size, kernel weight and the easiness of cracking are important traits for walnut quality. We found a *BEL1-like homeodomain protein 4* encoding gene as involved in kernel weight. In *Arabidopsis thaliana*, BEL1 protein is a commonly found transcription factor required for ovule morphogenesis (Modrusan et al., 1994), whereas in potato, BEL1-type proteins enhance tuberization (Sharma et al., 2014). Therefore, BEL1 is likely involved in determining kernel weight in walnut. In addition, we found a *protein TPX2-like* encoding candidate gene for nut face strength. TPX2 is a protein phosphorylated during mitosis interphase acting as a spindle assembly factor for microtubules in the nucleus, and is therefore crucial for cell division.

Using the "solid spine of LD" for the definition of the LD blocks, we highlighted complementary candidate genes for several MTAs. Nut volume is the trait with the highest number of candidate genes found. In addition to the gene coding for a *beta-galactosidase* related to the most significant association on Chr14, we found an *expansin-A4* and a *scarecrow-like protein 14* encoding genes on Chr7, involved in loosening of cell wall during growth (Cosgrove, 2000) and flower bud transition (Quan et al., 2019). We also found a *calcium-binding protein PBP1* encoding gene on Chr9, supporting the central role of calcium during fruit growth when Ca^{2+} uptake increases (Song et al., 2018). As Sideli et al. (2020), we identified a gene encoding a *lamin-like protein* for nut suture strength. Lamins are proteins part of the lamina, providing a mechanical support to the nuclear envelope and also acting on nucleus size and shape (Ciska and Moreno Díaz de la Espina, 2014).

The selection of appropriate model and threshold levels is crucial in GWAS. Among the two models implemented, FarmCPU provided the highest number of significant associations. FarmCPU has a higher detection power than previous GWAS models since it controls for false positives while reducing false negatives; it uses a MLMM divided into fixed effect and random effect model iteratively (Liu et al., 2016). FarmCPU has not been widely used for complex traits in crops because of a lacking comparison with previous existing models. However, MLMM and FarmCPU models have been recently compared using well-known GWAS dataset of soybean and maize, and the authors confirm FarmCPU's ability to outperform previous GWAS models (Kaler et al., 2020). In addition, the authors claim that MLMM did not find any significant marker since it uses a too conservative multiple comparison adjustment method. For this reason, we decided to consider also signals below the 1% Bonferroni threshold to avoid false negatives.

First List of Candidate Genes Involved in Fatty Acids and Vitamin E Contents Variation in Kernel

For the first time in walnut, we discovered SNPs and candidate genes associated with monounsaturated fatty acids and tocopherols contents. Primary metabolites in plants such as lipids and vitamins are synthesized by numerous multi-enzymatic complexes, and we found that the significant SNP on Chr7 associated with monounsaturated fatty acids content falls within a gene coding for a *calmodulin*. In *de novo* fatty acids synthesis, the fatty acid synthase requires NAD(P)H as cofactor, and the dependent nicotinamide adenine dinucleotide (NAD) signaling is thought to be highly regulated by calmodulin (Tai et al., 2019). However, additional phenotyping data are necessary to confirm the MTAs and the candidate genes identified.

CONCLUSION

By combining a highly accurate characterization of crucial traits for walnut quality with a high diverse walnut germplasm collection, a dense SNP genotyping, and newly available multi-locus models, we identified numerous MTAs, particularly for traits related to nut size. We confirmed major loci involved in suture strength, and we proposed candidate genes for fruit quality, mainly linked to cell wall function and calcium signaling. After phenological traits, the INRAE walnut germplasm collection proved its suitability for GWAS. It will make easy to select genitors allowing future release of new walnut cultivars meeting the criteria required by both consumers and producers.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.org/10.15454/XPKII8, Portail Data INRAE.

AUTHOR CONTRIBUTIONS

AB performed all the genetic analyzes and wrote the manuscript. JC assisted AB in performing, analyzing, and interpreting the data of weight and cracking related traits. AD and AM provided their experience in GWAS to interpret the results. ED, FL, and AB conceived and coordinated the research. All authors revised the manuscript.

FUNDING

This work has been mainly funded by the Nouvelle-Aquitaine Region in the project "INNOV'noyer," coordinated by the CTIFL, and in partnership with the INRAE of Bordeaux. This work has been also funded by the "Cifre" convention number 2016/1558 of ANRT (Agence Nationale de la Recherche et de la Technologie). INRAE funded part of the Ph.D. grant of AB. Postdoctoral fellowship for AD was funded by the Nouvelle-Aquitaine Region project 2018-1R 20203 "CerGEn." We thank the GIS Fruits for the funding of the Master fellowship of JC. To finish, AB thanks the "Initiative d'Excellence" program of the University

of Bordeaux and the "Dufrenoy-Crédit Agricole d'Ile-de-France Mécénat" grant for contributing to the funding of his Ph.D.

ACKNOWLEDGMENTS

We thank the Fruit Tree Experimental Unit of the INRAE in Toulenne and the *Prunus/Juglans* Genetic Resources Center for the maintenance of the collection and for helping us to collect the fruits. We want to thank the GIS Fruits consortium for the funding of the Master's degree fellowship of JC. We acknowledge the GEVES for X-ray CT analyzes. The measurements were performed on a phenotyping platform, PHENOTIC, dedicated to instrumentation and imagery for seeds, seedlings and plants in Angers, France (https://www6.inrae.fr/phenotic/; https://doi.org/10.15454/U2BWFJ). We acknowledge the ITERG laboratory in Canéjan, France, for the fatty acids and tocopherols quantifications (http://iterg.com/index.php/fr/home/). Then, the CTIFL, holder of the project "INNOV'noyer," in partnership with the INRAE of Bordeaux, want to thank the "Région Nouvelle-Aquitaine," and "Cifre" convention of "ANRT" (Agence Nationale de la Recherche et de la Technologie). It is also important to note that the project is supported by the "Agri Sud-Ouest Innovation" competitiveness cluster.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.607213/full#supplementary-material

Supplementary Figure 1 | Histograms of the 25 studied traits for each year of evaluation for the walnut GWAS analysis.

Supplementary Figure 2 | Manhattan plots of traits related to fruit shape, cracking and nutritional components studied in the walnut GWAS analysis. For each trait, MLMM Manhattan plot and Q-Q plot are shown at the top, and FarmCPU Manhattan plot and Q-Q plot are shown at the bottom. Horizontal green line corresponds to 1% Bonferroni threshold automatically implemented in GAPIT.

Supplementary Table 1 | Description of the 25 studied traits with the methodologies used for the different years of evaluation, for the walnut GWAS analysis.

Supplementary Table 2 | Phenotypic raw dataset of the 25 traits studied in the walnut GWAS analysis.

REFERENCES

Ahandani, E. A., Ramandi, H. D., Sarmad, J., Samani, M. A., Yavari, A., Ahandani, R. A., et al. (2014). Evaluation of morphological diversity among some *Persian walnut* accessions (*Juglans regia* L.) in Guilan, northern Iran. *Int. J. Plant Biol. Res.* 2, 1015–1022.

Arab, M. M., Marrano, A., Abdollahi-Arpanahi, R., Leslie, C. A., Askari, H., Neale, D. B., et al. (2019). Genome-wide patterns of population structure and association mapping of nut-related traits in Persian walnut populations from Iran using the Axiom J. regia 700K SNP array. *Sci. Rep.* 9:6376. doi: 10.1038/s41598-019-42940-1

Aradhya, M. K., Potter, D., and Simon, C. J. (2006). "Cladistic biogeography of juglans (*juglandaceae*) based on chloroplast DNA intergenic spacer sequences," in *Darwins Harvest New Approaches Orig. Evol. Conserv. Crops*, eds T. J. Motley, N. Zerega, and H. Cross (New-York: Columbia University Press), 143–170. doi: 10.7312/motl13316-008

Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. doi: 10.1093/bioinformatics/bth457

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

Becraft, P. W. (2002). Receptor kinase signaling in plant development. *Annu. Rev. Cell Dev. Biol.* 18, 163–192. doi: 10.1146/annurev.cellbio.18.012502.083431

Bernard, A., Lheureux, F., and Dirlewanger, E. (2018b). Walnut: past and future of genetic improvement. *Tree Genet. Genomes* 14:1. doi: 10.1007/s11295-017-1214-0

Bernard, A., Barreneche, T., Lheureux, F., and Dirlewanger, E. (2018a). Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. *PLoS One* 13:e0208021. doi: 10.1371/journal.pone.0208021

Bernard, A., Hamdy, S., Le Corre, L., Dirlewanger, E., and Lheureux, F. (2020a). 3D characterization of walnut morphological traits using X-ray computed tomography. *Plant Methods.* 16:115. doi: 10.1186/s13007-020-00657-7

Bernard, A., Marrano, A., Donkpegan, A., Brown, P. J., Leslie, C. A., Neale, D. B., et al. (2020b). Association and linkage mapping to unravel genetic architecture of phenological traits and lateral bearing in *Persian Walnut* (*Juglans regia* L.). *BMC Genomics* 21:203. doi: 10.1186/s12864-020-6616-y

Burghardt, L. T., Young, N. D., and Tiffin, P. (2017). A Guide to Genome-Wide Association Mapping in plants. *Curr. Protoc. Plant Biol.* 2, 22–38. doi: 10.1002/cppb.20041

Ciska, M., and Moreno Díaz de la Espina, S. (2014). The intriguing plant nuclear lamina. *Front. Plant Sci.* 5:166. doi: 10.3389/fpls.2014.00166

Cosgrove, D. J. (2000). Loosening of plant cell walls by expansins. *Nature* 407, 321–326. doi: 10.1038/35030000

Diallo, A., and Prigent, C. (2011). The serine/threonine kinases that control cell cycle progression as therapeutic targets. *Bull. Cancer* 98, 1335–1345. doi: 10.1684/bdc.2011.1467

Eskandari, S., Hassani, D., and Abdi, A. (2005). Investigation on genetic diversity of *Persian walnut* and evaluation of promising genotypes. *Acta Hortic.* 705, 159–166. doi: 10.17660/ActaHortic.2005.705.18

Furbank, R. T., and Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644. doi: 10.1016/j.tplants.2011.09.005

Ghasemi, M., Arzani, K., and Hassani, D. (2012). Evaluation and identification of walnut (*Juglans regia* L.) genotypes in Markazi province of Iran. *Crop Breed. J.* 2, 119–124.

Harrison, R. D., Gardner, W. A., Tollner, W. E., and Kinard, D. J. (1993). X-ray computed tomography studies of the burrowing behaviour of fourth-instar pecan weevil (*Coleoptera: Curculionidae*). *J. Econ. Entomol.* 86, 1714–1719. doi: 10.1093/jee/86.6.1714

Kaler, A. S., Gilman, J. D., Beissinger, T., and Purcell, L. C. (2020). Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. *Front. Plant Sci.* 10:1794. doi: 10.3389/fpls.2019.01794

Khadivi-Khub, A., Ebrahimi, A., Sheibani, F., and Esmaeili, A. (2015). Phenological and pomological characterization of Persian walnut to select promising trees. *Euphytica* 205, 557–567. doi: 10.1007/s10681-015-1429-9

Khosa, I., and Pasero, E. (2014). "Feature extraction in X-ray images for hazelnuts classification," in *Proceedings of the. International Jointt Conference Neural. Networks*, (Beijing). doi: 10.1109/IJCNN.2014.6889661

Kim, S., and Schatzki, T. (2001). Detection of pinholes in almonds through X-ray imaging. *Trans. ASAE* 44, 997–1003. doi: 10.13031/2013.6232

Kotwaliwale, N., Singh, K., Kalne, A., Jha, S. N., Seth, N., Kar, A., et al. (2014). X-ray imaging methods for internal quality evaluation of agricultural produce. *J. Food Sci. Technol.* 51, 1–15. doi: 10.1007/s13197-011-0485-y

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: an r package for multivariate analysis. *J. Stat. Softw.* 25, 1–18. doi: 10.18637/jss.v025.i01

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12:e1005767. doi: 10.1371/journal.pgen.1005767

Lukk, T., Sakai, A., Kalyanaraman, C., Brown, S. D., Imker, H. J., Song, L., et al. (2012). Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily. *Proc. Natl. Acad. Sci. USA*. 109, 4122–4127. doi: 10.1073/pnas.1112081109

Mahmoodi, R., Dadpour, M. R., Hassani, D., Zeinalabedini, M., Vendramin, E., Micali, S., et al. (2019). Development of a core collection in Iranian walnut (*Juglans regia* L.) germplasm using the phenotypic diversity. *Sci. Hortic.* 249, 439–448. doi: 10.1016/j.scienta.2019.02.017

Marrano, A., Britton, M., Zaini, P. A., Zimin, A. V., Workman, R. E., Puiu, D., et al. (2020). High-quality chromosome-scale assembly of the walnut (*Juglans regia* L) reference genome. *GigaScience* 9:giaa050. doi: 10.1093/gigascience/giaa050

Marrano, A., Martínez-García, P. J., Bianco, L., Sideli, G. M., Di Pierro, E. A., Leslie, C. A., et al. (2019a). A new genomic tool for walnut (*Juglans regia* L.): development and validation of the high-density Axiom^{TM} J. regia 700K SNP genotyping array. *Plant Biotechnol. J.* 17, 1027–1036. doi: 10.1111/pbi.13034

Marrano, A., Sideli, G. M., Leslie, C. A., Cheng, H., and Neale, D. B. (2019b). Deciphering of the genetic control of phenology, yield and pellicle color in persian walnut (*Juglans regia* L.). *Front. Plant Sci.* 10:1140. doi: 10.3389/fpls.2019.01140

McGranahan, G., and Leslie, C. (2009). "Breeding Walnuts (Juglans regia)," in *Breeding Plantation Tree Crops: Temperate Species*, eds S. M. Jain and P. M. Priyadarshan (Berlin: Springer Science), 249–273. doi: 10.1007/978-0-387-71203-1_8

Modrusan, Z., Reiser, L., Feldmann, K. A., Fischer, R. L., and Haughn, G. W. (1994). Homeotic transformation of ovules into carpel-like structures in Arabidopsis. *Plant Cell* 6, 333–349. doi: 10.2307/3869754

Mohammadi, M., Xavier, A., Beckett, T., Beyer, S., Chen, L., Chikssa, H., et al. (2020). Identification, deployment, and transferability of quantitative trait loci from genome-wide association studies in plants. *Curr. Plant Biol.* 24:100145. doi: 10.1016/j.cpb.2020.100145

Muto, H., Yabe, N., Asami, T., Hasunuma, K., and Yamamoto, K. T. (2004). Overexpression of constitutive differential growth 1 gene, which encodes a RLCK VII-subfamily protein kinase, causes abnormal differential and elongation growth after organ differentiation in Arabidopsis. *Plant Physiol.* 136, 3124–3133. doi: 10.1104/pp.104.046805

Poggetti, L., Ermacora, P., Cipriani, G., Pavan, F., and Testolin, R. (2017). Morphological and carpological variability of walnut germplasm (*Juglans regia* L.) collected in North-Eastern Italy and selection of superior genotypes. *Sci. Hortic.* 225, 615–619. doi: 10.1016/j.scienta.2017.07.056

Quan, S., Niu, J., Zhou, L., Xu, H., Ma, L., and Qin, Y. (2019). Genome-wide identification, classification, expression and duplication analysis of GRAS family genes in *Juglans regia* L. *Sci. Rep.* 9:11643. doi: 10.1038/s41598-019-48287-x

Rondeau, A.-S. (1997). *Le Noyer. Eds*. Aries. Actes Sud.

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü, Long, Q., et al. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44, 825–830. doi: 10.1038/ng.2314

Sharma, P., Lin, T., Grandellis, C., Yu, M., and Hannapel, D. J. (2014). The BEL1-like family of transcription factors in potato. *J. Exp. Bot.* 65, 709–723. doi: 10.1093/jxb/ert432

Sideli, G. M., Marrano, A., Montanari, S., Leslie, C. A., Allen, B. J., Cheng, H., et al. (2020). Quantitative phenotyping of shell suture strength in walnut (*Juglans regia* L.) enhances precision for detection of QTL and genome-wide association mapping. *PLoS One* 15:e0231144. doi: 10.1371/journal.pone.0231144

Song, W.-P., Chen, W., Yi, J.-W., Wang, H.-C., and Huang, X.-M. (2018). Ca distribution pattern in litchi fruit and pedicel and impact of Ca channel inhibitor, La^{3+}. *Front. Plant Sci.* 8:2228. doi: 10.3389/fpls.2017.02228

Smith, D. L., and Gross, K. C. (2000). A family of at least seven β-galactosidase genes is expressed during tomato fruit development. *Plant Physiol.* 123, 1173–1183. doi: 10.1104/pp.123.3.1173

Steinwand, B. J., and Kieber, J. J. (2010). The role of receptor-like kinases in regulating cell wall function. *Plant Physiol.* 153, 479–484. doi: 10.1104/pp.110.155887

Tai, L., Li, B.-B., Nie, X.-M., Zhang, P.-P., Hu, C.-H., Zhang, L., et al. (2019). Calmodulin is the fundamental regulator of NADK-mediated NAD signaling in plants. *Front. Plant Sci.* 10:681. doi: 10.3389/fpls.2019.00681

Vahdati, K., Arab, M. M., Sarikhani, S., Sadat-Hosseini, M., Leslie, C. A., and Brown, P. J. (2019). "Advances in *Persian Walnut* (*Juglans regia* L.) Breeding Strategies," in *Advances in Plant Breeding Strategies: Nut and Beverage Crops*, eds J. Al-Khayri, S. M. Jain, and D. V. Johnson (Switzerland: Springer), 401–472. doi: 10.1007/978-3-030-23112-5_11

Wei, T., and Simko, V. R. (2017). *Package "corrplot": Visualization of a Correlation Matrix (Version 0.84)*. Available online at: https://github.com/taiyun/corrplot (accessed April 10, 2020).

Wickham, H. (2017). *Tidyverse: Easily Install and Load the 'Tidyverse'. R Package Version 1.2.1*. Available online at: https://CRAN.R-project.org/package=tidyverse (accessed April 10, 2020).

Woodworth, R. H. (1930). Meiosis of micro-sporogenesis within the Juglandaceae. *Am. J. Bot.* 17, 863–869. doi: 10.1002/j.1537-2197.1930.tb04927.x

Yang, H., Liu, J., Dang, M., Zhang, B., Li, H., Meng, R., et al. (2018). Analysis of β-Galactosidase during fruit development and ripening in two different texture types of apple cultivars. *Front. Plant Sci.* 9:539. doi: 10.3389/fpls.2018.00539

Yang, Z., Wang, Y., Zhou, Y., Gao, Q., Zhang, E., Zhu, L., et al. (2013). Evolution of land plant genes encoding L-Ala-D/L-Glu epimerases (AEEs) via horizontal gene transfer and positive selection. *BMC Plant Biol.* 13:34. doi: 10.1186/1471-2229-13-34

Zeneli, G., Kola, H., and Dida, M. (2005). Phenotypic variation in native walnut populations of Northern Albania. *Sci. Hortic.* 105, 91–100. doi: 10.1016/j.scienta.2004.11.003

Conflict of Interest: 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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 13 January 2021
doi: 10.3389/fpls.2020.612376

Pod Morphology, Primary and Secondary Metabolite Profiles in Non-grafted and Grafted Carob Germplasm Are Configured by Agro-Environmental Zone, Genotype, and Growing Season

Angelos C. Kyratzis[1], Chrystalla Antoniou[1], Lambros C. Papayiannis[1], Giulia Graziani[2], Youssef Rouphael[3] and Marios C. Kyriacou[1]**

[1] Agricultural Research Institute, Ministry of Agriculture, Rural Development and Environment, Nicosia, Cyprus, [2] Department of Pharmacy, University of Naples Federico II, Naples, Italy, [3] Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy

OPEN ACCESS

Edited by:
Maria Luisa Badenes, Instituto Valenciano de Investigaciones Agrarias, Spain

Reviewed by:
Shouvik Das, Indian Agricultural Research Institute (ICAR), India
Alicia Serrano Gómez, IFAPA Centro Alameda del Obispo, Spain

***Correspondence:**
Angelos C. Kyratzis
a.kyratzis@ari.gov.cy
Marios C. Kyriacou
m.kyriacou@ari.gov.cy

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

Received: *08 October 2020*
Accepted: *15 December 2020*
Published: *13 January 2021*

Citation:
Kyratzis AC, Antoniou C, Papayiannis LC, Graziani G, Rouphael Y and Kyriacou MC (2021) Pod Morphology, Primary and Secondary Metabolite Profiles in Non-grafted and Grafted Carob Germplasm Are Configured by Agro-Environmental Zone, Genotype, and Growing Season. Front. Plant Sci. 11:612376. doi: 10.3389/fpls.2020.612376

Carob is a predominantly rainfed tree crop of high nutritive value and a long history of adaptation to the edaphoclimatic stress conditions of the Mediterranean. However, declining attention to the carob tree in recent decades has aggravated genetic erosion. The extant *in situ* germplasm varies both in terms of pod morphology and composition, reflecting the genetic and physiological divide chiefly among grafted and non-grafted material, and possibly the impact of variable agro-environments. Accordingly, the present study aimed to establish a systematic categorization of the genetic and phenotypic diversity encountered across carob germplasm identified *in situ* throughout Cyprus, a historical center of production and genetic diversity for the species. Linking pod morphology, primary and secondary metabolite profiles with genotyped source material originating in different agro-environments and crop seasons would provide a framework for interpreting (a) the interaction of these factors in configuring carob pod physicochemical constitution, and (b) the relative stability of phenotypic traits against environmental and seasonal variation. Microsatellite analysis discriminated 36 genotypes out of the 124 trees located in nine traditional agro-environmental zones and revealed low genetic diversity within the grafted germplasm. Two landraces were identified: "Tillyria," which is widespread and predominant, and "Kountourka," which is mainly localized to the northeastern peninsula of Karpasia. Morphological traits, such as seeds-to-pod weight ratio, pod width and thickness were principally under genetic control. Contrarily, compositional traits, particularly total phenolic content—including condensed tannins, *in vitro* antioxidant capacity and to a lesser extent gallic acid, organic acids and minerals were under agro-environmental control. Agro-environmental zone also modulated principally fructose and glucose; sucrose was modulated equally by genotype and agro-environment, while total sugars were under genetic control. Statistically significant differences between seasons were detected for all traits except for the seeds-to-pod weight ratio, pod length and width. Hierarchical cluster analysis corroborates that Cyprus may be divided into two major agro-environmental zones

modulating the compositional properties of the carob pulp. The present study provides a comprehensive insight into the extant carob genetic resources of Cyprus and advances our understanding of how genetic, agro-environmental and seasonal factors interact in shaping carob pod morphology and composition.

Keywords: functional quality, genetic diversity, polyphenols, SSRs, sugars, tannins

INTRODUCTION

Carob (*Ceratonia siliqua* L.) is an evergreen, diploid species (2n = 48) that belongs to the *Fabaceae* family. The carob tree is an important component of the Mediterranean vegetation and it is predominately cultivated in dry and marginal areas, due to its low requirements in agronomic inputs compared to other fruit species. Its ability to thrive in the prevailing calcareous soils has amplified the socio-economic value of the carob in several Mediterranean mild climate dry lands (Batlle and Tous, 1997; Tous et al., 2013).

The carob tree occurred in the flora of the East Mediterranean basin long before the emergence of agriculture and it was naturally distributed westwards to the rest of Mediterranean countries (Zohary, 2002; Viruel et al., 2016). The domestication and cultivation of carob appeared relatively late in Hellenistic and Roman times as vegetative propagation was not effective for carob propagation and only after the discovery of scion grafting have selected phenotypes been cultivated. It is nonetheless worth noting that cultivation essentially comprised naturally and randomly distributed wild rootstocks grafted with select scion phenotypes rather than systematic orchards, which only came about in recent decades (Tous et al., 2013). Grafted carob trees differ from their wild ancestors mainly in phenotypic traits such as fleshiness, size and sweetness of the pod, local adaption and productivity (Batlle and Tous, 1997; Zohary, 2002; Tous et al., 2013).

For centuries, the carob fruit has been used as a regional culinary ingredient and as a livestock feed due to its high sugar content (Batlle and Tous, 1997). Nowadays, carob cultivation is expanding in response to growing demand for its compositional, functional, nutritional, and industrial value which makes it an economically important crop (Goulas et al., 2016; Stavrou et al., 2018). The mature carob pod is comprised of two parts: the pulp and the seeds in roughly 90/10 ratio w/w (Goulas et al., 2016). Carob seeds are exploited industrially for the production of carob bean gum (Locust Bean Gum—LBG), a widely used natural food thickening agent (Bouzouita et al., 2007). Recently, researchers have focused on carob pulp which is a low-cost by-product of the milling process. Carob pod is not only a rich source of sugars but also of bioactive molecules including dietary fibers, polyphenols and cyclitols and it has a low-fat content (Avallone et al., 1997). These bioactive compounds exhibit a wide range of biological properties with significant health-promoting effects, including the prevention of colon cancer and hepatocellular carcinoma, reduction of diarrheal symptoms, lowering of LDL cholesterol as well as antidiabetic effects (Zunft et al., 2003; Goulas et al., 2016; Theophilou et al., 2017). These findings have contributed to the valorization of carob pulp products such as carob powder, fiber, juice and molasses used by the food industry for developing a wide range of health-promoting or niche food products, including gluten-free ones (Yousif and Alghzawi, 2000; Nasar-Abbas et al., 2016).

The carob constitutes a genetic resource of long-standing adaptation to the edapho-climatic conditions of Cyprus, which has been for centuries one of the leading countries in carob production (Ticho, 1959; Orphanos, 1980). Carob pods produced from specific geographical areas of Cyprus are considered of premium quality and they are customarily exported intact to Egypt for human consumption (Gennadius, 1902; Personal communication with local stakeholders 2018, 2019). Based on morphological traits (Orphanos and Papaconstantinou, 1969) and other records (Gennadius, 1902; Ticho, 1959; Davies, 1970), the Cypriot carob germplasm can be categorized into four groups. The first group comprises wild trees producing short and thin pods of variable morphology and substandard quality, which farmers historically did not harvest. The second group contains non-grafted trees producing pods of reasonably good quality, identified by farmers as "Apostolika," a term alluding to their possible use as foodstuff by Christ's wandering apostles. Scattered trees of this group can be found within carob groves and farmers invariably collect their pods. The third group, which predominates on the island, encompasses exclusively grafted trees described in the literature as "Tillyria," reference to the prolific geographical area of the island's central northern coast (Gennadius, 1902). The name "Tillyria," however, is not ubiquitous among farmers, which instead tend to identify these trees under several local names. "Tillyria" produce pods of slightly variable morphology, and it remains unanswered if this variation is due to genetic variability or due to the variability of edapho-climatic conditions. The fourth group contains grafted trees producing shorter pods than "Tillyria," locally common to the Karpasia peninsula and referred to as "Kountourka." In addition, two scion phenotypes, morphologically proximate to "Tillyria," have been reported. These phenotypes are called "Mavroteratsia" and "Koumpota."

The preferential cultivation of irrigated cash crops (e.g., citrus) after the 1960s in Cyprus led to the depreciation of the carob crop and resulted in substantial reduction of the carob cultivated area (Davies, 1970). Moreover, wild fires, illegal logging and heavy rat infestation have further threatened the species' diversity (Orphanos, 1980). The preference of farmers for a specific select phenotype, in the case of Cyprus for "Tillyria," could pose an additional threat for the species' genetic diversity (Barracosa et al., 2007). However, the reviving interest in the carob warrants the investigation of its genetic diversity to establish conservation and breeding programs (Di Guardo et al., 2019). Further to the classical studies on genetic diversity based on morphological

traits (Barracosa et al., 2007), the use of molecular markers caries the advantage that they are more polymorphic and unaffected by the environment. Microsatellites (SSRs) have been successfully used to assess carob genetic diversity (La Malfa et al., 2014; Viruel et al., 2018; Di Guardo et al., 2019). Additionally, a prerequisite for revitalizing the carob industry is the assessment of the variation for morphological and compositional traits (Barracosa et al., 2007; Custodio et al., 2011; Benchikh et al., 2014). However, for rain fed crops, such as carob, grown under marginal environments, the agro-environmental effect owing to spatial variation in edapho-climatic conditions would be expected to predominate over the genotypic effect (Blum, 2010). Furthermore, annual variation in climatic conditions may have a significant effect on productivity (Orphanos, 1980), hence putatively also on compositional traits. However, very few studies have investigated the environmental effect on carob compositional traits (Avallone et al., 1997; El Bouzdoudi et al., 2016; Farag et al., 2019; Othmen et al., 2019) while information concerning the seasonal effect on compositional traits remains scarce (Nahla, 2014; Correia et al., 2018).

The current study combined genetic and phenotypic data collected from trees *in situ/on farm* to dissect the genotypic and the agro-environmental effects on morphological and compositional traits. Extensive sampling was performed from all the traditional environmental zones of carob cultivation, with emphasis on grafted material. Sampling was performed for two consecutive years to evaluate the seasonal effect on phenotypic traits. The genetic and phenotypic variation of Cypriot carob genetic resources is hereby presented. To the best of our knowledge, this is the first study aiming to investigate the genotypic, agro-environmental and seasonal effects on carob morphological and compositional traits.

MATERIALS AND METHODS

Sampling Strategy

Extensive survey of carob genetic resources was carried out during two seasons, 2018 and 2019. Trees were sampled from nine agro-environmental zones based on geographical and geological parameters (**Figure 1**). "Anogira," "Mountainous Larnaca," and "Mountainous Lemesos" were the inland zones of relatively high altitude. "Mountainous Paphos," "Mountainous Polis," and "Neo Chorio" were the zones of intermediate altitude, with the latter positioned closer to the sea than the two former zones. "Tillyria," "North zone," and "South zone" were the zones of low altitude where the majority of trees are grown relatively close to the sea. Trees in the "Tillyria" zone are grown on igneous formations with pillow lavas and diabase dikes while trees in the other zones are grown on calcareous formations and to lesser extent on red soils overlying a white soft highly calcareous layer. Average annual precipitation ranges between 400 and 600 mm across zones.

Vernacular knowledge on phenotypic diversity, cultivation practices, trade and processing were collected through interviews with farmers and stakeholders of the local industry. Farmers participated in joint field trips to locate trees of distinct phenotypic traits. Non-grafted trees were also sampled to assess the relatedness between grafted and non-grafted gene pools. Additionally, trees from the nursery plantations of the Department of Agriculture, Ministry of Agriculture, Rural Development and Environment (DOA) which provide budding wood for propagation purposes were sampled. Three accessions, representing the cultivated material in Sicily, were also included in the study. In total, 124 trees were sampled out of which 107 were grafted and 17 non-grafted. The trees were georeferenced and passport data were compiled (**Supplementary File 1**). Leaves were collected for DNA extraction. Around 30–40 carob pods were randomly harvested from each tree to assess morphological and compositional traits. Harvesting was performed during the major harvesting period (mid-August to mid-September) when pods were fully mature. The exact harvest date for each tree was recorded (**Supplementary File 1**). Sampling was repeated for two consecutive years with the exception of non-bearing trees due to severe pruning (trees from the DOA plantations), extensive damage from rats, or trees that were identified during the 2nd season of the survey.

Genetic Analysis

DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Venlo, Netherlands). DNA integrity was verified in agarose electrophoresis and DNA quality and quantity was determined by NanoDrop 1000 (Thermo Fisher Scientific, Wilmington, United States). Eighteen microsatellites were selected based on the available information from previous studies (La Malfa et al., 2014; Viruel et al., 2018). Amplification reactions were set up in a 12 μL volume of a mixture containing 50 ng of genomic DNA, 1x Type-it Multiplex PCR master mix (Type-it Microsatellite PCR kit, Qiagen, Venlo, Netherlands) and 0.2 μM of each primer (the forward primers were 5′-end labeled with FAM—5-and HEX carboxy-fluorescents). PCR amplification was performed in a PTC-200 thermocycler (Bio-Rad, Hercules, United States) under the following temperature profile: 5 min at 95°C, followed by 30 cycles, each one including 30 s at 95°C, 30 s at annealing temperature depending of the primer pair (**Supplementary File 2**), 30 s at 72°C and a final extension step for 30 min at 60°C. A negative control was included in each set of PCR amplification. A subset of accessions (10%) was amplified twice to check the reproducibility of the primers. Amplified PCR products were run on an ABI3130 genetic analyzer (Applied Biosystems, Foster City, CA, United States). Size standard GeneScan^{TM} 500LIZ (Applied Biosystems) was added to each sample to delineate allele sizes. Data were analyzed using GeneMapper Software version 4.1 (Applied Biosystems, Foster City, CA, United States).

Morphological Characterization and Processing of Plant Material

Carob pods were washed under tap water to remove debris, rinsed with deionized water and patted dry. The thickness of the pod (determined lengthwise in three sections: edge, groove, valley) and the length and width of the pod were recorded using a caliber. Then the pods were coarsely ground in a Vita Prep 3

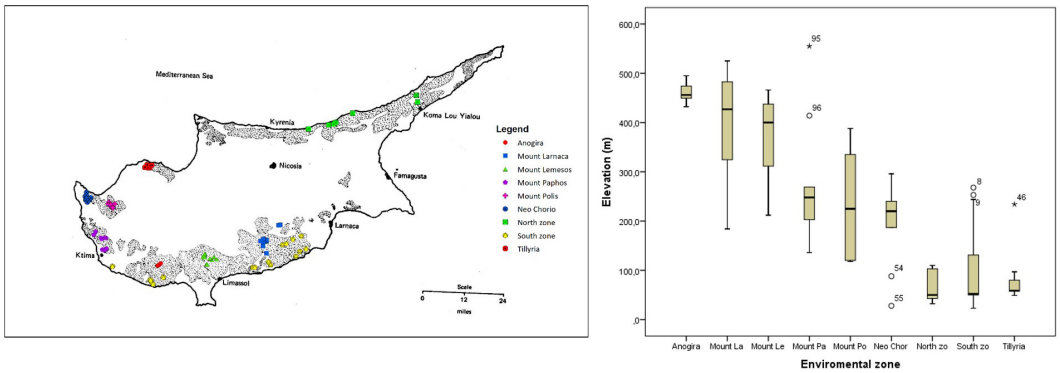

FIGURE 1 | Map of Cyprus showing the extent of sampling in the current study. Dashed pattern presents the traditional areas of carob cultivation (after Davies, 1970). Sampled trees in different agro-environmental zones are depicted in different color. Box plot shows the variation in altitude of sampled trees within the different agro-environmental zones.

blender (Vita-Mix Corp., Cleveland, United States) operated at low speed and deseeded. Pod and seed weight were measured using a Precisa XT120A analytical balance (Precisa Gravimetrics, Dietikon, Switzerland). The carob kibbles were lyophilized in a Christ, Alpha 1–4 lyophilizer (Osterode, Germany) to stable weight, ground to powder of 1–2 mm granulometry using a CT293 Cyclotech mill (Foss Analytical A/S, Hillerød, Denmark) and stored at −60°C.

The color lightness (L*) of the carob pod and seed were assessed using a Minolta CR-400 Chroma Meter (Minolta, Osaka, Japan) and that of the powder using a Minolta CR-410 both with a diffusion illumination 0° viewing angle geometry and the color space XYZ, Yxy, L*a*b*, Hunter, L*C*h, Munsel as the default.

Minerals, Organic Acids, and Protein Content

Analysis of the cations (K^+, Ca^{2+}, Mg^{2+}, Na^+, and NH_4^+), anions (NO^{-3}, PO_3^{-4}, SO_4^{-2}, and Cl^-) and organic acids (malic, citric and oxalic) on lyophilized carob powder was performed as previously described in detail by Rouphael et al. (2017). The monovalent and bivalent cations were separated by ion chromatography (ICS-3000, Dionex, Sunnyvale, CA, United States) and quantified through an electrical conductivity detector. Cations separation was performed using an IonPac CG12A (4 × 50 mm, Dionex, Corporation) guard column and IonPac CS12A (4 × 250 mm, Dionex, Corporation) analytical column, whereas for anions and organic acids an IonPac AG11-HC guard (4 × 50 mm) column and IonPac AS11-HC analytical column (4 × 250 mm) were used. Minerals and organic acids content results were expressed in g kg^{-1} dw. The total protein content was assessed by the Kjeldahl method, with nitrogen-to-protein conversion factor of 6.25 (Bremner, 1965). Analysis of cations, organic acids and protein content was performed only on samples harvested in 2018.

Total Phenols and Condensed Tannins

The total phenols content (TPC) of the carob powder was determined according to the method of Singleton et al. (1999), slightly modified as previously described by Kyriacou et al. (2016). Approximately 0.1 g of powder was extracted in solvent to a final volume of 25 mL. Extraction was performed in the dark at 4°C for 24 h. Two different extraction solvents were used: (a) methanol:H_2O:HCl (50:40:10); and (b) methanol:H_2O:acetate (80:19.5:0.5). In addition to the extraction of readily extractable phenolics, the former solution facilitated the extraction of acid-hydrolyzed condensed tannins. Absorbance was measured on a Jasco V-550 UV–vis spectrophotometer (Jasco Corp., Tokyo, Japan) and quantification was performed using linear calibration with external gallic acid standards over the range of 50-500 mg L^{-1}, yielding a regression coefficient $R^2 > 0.99$. TPC was expressed as gallic acid equivalents in g kg^{-1} pulp dw.

Condensed tannins (proanthocyanidins) were determined using a modification of the vanillin method described by Sun et al. (1998) and Sepperer et al. (2019). Extraction was performed as described above for phenolics using solvent (a) on 0.5 g of lyophilized powder to a final volume of 10 mL. One milliliter of extract was combined with 2.5 mL of 2% vanillin methanolic solution in 15 mL Falcon tubes placed on ice; then 2.5 mL of MeOH:HCl (90:10) were added, swirled and kept on ice for 5 min, followed by 15 min incubation at 30°C with gentle agitation (80 rpm). Absorbance was measured at 500 nm against calibration with seven catechin standards (0.025–0.5% w/v in methanol) replacing the sample. Methanol was used in the place of sample or standard as blank. The results were expressed in catechin equivalents as mg g^{-1} pulp dw.

Analysis of Polyphenols by UHPLC-Q-Orbitrap HRMS

Polyphenols were determined on methanolic extracts obtained using the extraction solvent (b) described above. An Ultimate 3000 UHPLC system (Thermo Fisher Scientific, Waltham, MA, United States) was employed equipped with a Kinetex 1.7 μm biphenyl (100 × 2.1 mm) column (Phenomenex, Torrance, CA, United States) maintained at 25°C. The injection volume was 2 μL and flow rate was 0.2 mL min^{-1}. The mobile phase consisted

of water (A) and methanol (B), both containing 0.1% formic acid. Gradient elution program was applied as follows: 0 min—5% B, 1.3 min—30% B, 9.3 min—100% B, 11.3 min—100% B, 13.3 min—5% B, 20 min—5% B. The UHPLC system was coupled to a Q Exactive Orbitrap liquid chromatography tandem-mass spectrometry (LC–MS/MS). A heated electrospray ionization source (HESI II, Thermo Fischer Scientific) operating in negative ion mode (ESI^-) was used. Ion source parameters were: sheath gas-flow rate 45 units, auxiliary gas-flow rate 10 units, spray voltage −3.5 kV, capillary temperature 275°C, S-lens (RF) level 50, auxiliary gas heater temperature 350°C.

Analysis was arranged setting two scan events (Full ion MS and All ion fragmentation, AIF) for all compounds of interest. Full MS data were acquired setting the following parameters: microscans, 1; AGC target, 1e6; maximum injection time, 200 ms; mass resolution, 35,000 FWHM at m/z 200 and m/z range, 80–1,200. AIF scan conditions were: microscans, 1; AGC target, 1e5; maximum injection time, 200 ms; mass resolution, 17,500 FWHM at m/z 200; HCD energy, 10, 20, and 45 and m/z range, 80–1,200. Calibration of the Q Exactive Orbitrap LC–MS/MS was checked daily, in both negative and positive modes, using the commercial calibration solutions provided by the manufacturer. Mass tolerance was kept at 5 ppm in both fullscan MS and AIF modes. Xcalibur software v. 3.1.66.10 (Xcalibur, Thermo Fisher Scientific) was used to perform data analysis and processing. Single phenolic compounds were quantified using calibration curves built with appropriate standards. As some standards were not available, quantitation for some compounds was carried out employing calibration curves of available standard belonging to the same chemical group and with similar response to the mass spectrometer. Analysis of polyphenols by UHPLC-Q-Orbitrap HRMS was performed only on samples harvested in 2018.

In vitro Antioxidant Activity

The *in vitro* Ascorbate Equivalent Antioxidant Activity (AEAC) of carob powder was assayed on pulp methanolic extracts according to the 1,1-diphenyl-2-picrylhydrazyl free-radical (DPPH) scavenging capacity assay of Brand-Williams et al. (1995) modified according to Kyriacou et al. (2020). Quantification was performed against ascorbate external standards (100–1,000 μM) based on the decrease in absorbance at 517 nm and expressed in ascorbate equivalents as g kg^{-1} pulp dw. The *in vitro* antioxidant activity of carob powder methanolic extracts was assayed also according to the ferric reducing antioxidant power (FRAP) assay of Benzie and Strain (1996). Quantification was performed at 593 nm against external standard curves of ascorbic acid at 100–1,000 μM. Results were expressed in ascorbate equivalents as g kg^{-1} pulp dw.

Water Soluble Carbohydrates

The analysis of water-soluble carbohydrates was performed on aqueous extracts clarified using the Carrez Clarification Kit (Sigma-Aldrich, St. Louis, MO, United States). Separation and quantification of glucose, fructose and sucrose was performed by liquid chromatography on an Agilent HPLC system (Agilent Technologies, Santa Clara, CA, United States) equipped with a 1,200 Series quaternary pump and a 1,260 Series Refractive Index detector operated by Chem-Station software as previously described by Kyriacou et al. (2016). Injection volume was 20 μL and separation was performed on a Waters 4.6 × 250 mm carbohydrate column (Waters, Milford, MA, United States) at 35°C using an acetonitrile:water (82:18) mobile phase at an isocratic flow rate of 1.5 mL min^{-1}. Quantification was performed against calibration curves of fructose, glucose and sucrose external standards (0.2–2.0 g 100^{-1} mL) with a coefficient of determination $R^2 > 0.999$ and expressed as g 100 g^{-1} pulp dw.

Statistical Analysis

Genetic analysis was conducted on allele fragments. Three out of the eighteen primers (Cesi 21 cttt7, C21, and C23) amplified more than two alleles per tree sample, therefore, in the analysis, it was assumed that these primers were multilocus. Number of alleles (Na), number of alleles with frequency less than five ($N \leq 0.05$), number of private alleles, effective number of alleles (Ne), observed (Ho) and expected heterozygosity (He), fixation index (Fis) and deviation for Hardy Weinberg equilibrium (HW sign) were calculated on the whole set of trees and separately for grafted and non-grafted trees using GenAlEx 6.4 (Peakall and Smouse, 2006). Polymorphic information content (PIC) and probability for null alleles F (null) were estimated with Cervus ver. 3.0.7 (Kalinowski et al., 2007). Analysis of Molecular Variance (AMOVA) was also performed to assess the within and between variance across grafted and non-grafted trees using GenAlEx. The significance of the resulting variance components and the inter-population genetic distances were tested using 999 random permutations. Simple matching dissimilarity index was used to calculate genetic distances between tree pairs and a weighted Neighbor-Joining dendrogram was constructed using the DARwin software, version 6 (Perrier, 2006). The robustness of the tree was tested using 10,000 bootstraps and bootstrapped values above 20% are presented. STRUCTURE software (ver. 2.3.4) was employed to investigate the genetic structure using the admixture model with 100,000 burn in followed by 100,000 iterations with 30 replicate runs (Pritchard et al., 2000). Ten clusters were tested ($k = 1$–10) and the validation of the most likely number of clusters K was performed with the Structure Harvester following the ΔK method proposed by Evanno et al. (2005). Trees were assigned to a cluster when the membership coefficient was ≥ 0.8 (Di Guardo et al., 2019).

Mean values, standard error of means, coefficients of variation, maximum and minimum values were calculated and frequency plots were constructed for all morphological and compositional traits. Two tailed *t*-tests were conducted to explore trait differences between seasons and Pearson correlations were estimated to investigate their relationships. Separate Analyses of Variance (ANOVA) were conducted to investigate the genetic and agro-environmental effects on the examined traits. The first ANOVA was executed using as factor treatments the clustering of the trees according to the structure analysis, while the second ANOVA was executed using as factor treatments the agro-environmental zones where

the trees are grown. The percentage of variance explained by the main effect, as percentage of the SS to the total SS, is presented. Mean values and standard error of the means were calculated for each structure cluster and agro-environmental zone. Box plots are presented in cases where the percentage of variance explained by the genotypic or agro-environmental effect was particularly high. Hierarchical cluster analysis was conducted to assess the genetic and agro-environmental effects on the overall phenotype. Squared Euclidean distances were calculated on standardized *Z*-values, with a mean of 0 and a standard deviation of 1. Clustering was performed using the "WARD" method. All analyses were carried out using SPSS (IBM, SPSS ver. 26).

RESULTS

Genetic Variation and Affinity Across Trees

Microsatellite analysis was used to genotype carob trees from nine agro-environmental zones. The 18 primer pairs used in the present study discriminated 36 genotypes. The primers distinguished all non-grafted trees. Moreover, 20 genotypes were identified among the 107 grafted trees. The most common genotype appeared with very high frequency (**Figure 2**). In total, 86 alleles were detected, out of which 49 appeared with low frequency and 15 where private. Average, Ho, He, PIC and Fis, where 0.547, 0.354, 0.295, and −0.319, respectively. For most loci, there was significant deviation from Hardy Weinberg equilibrium. The probability for null alleles was low (**Table 1**). Genetic diversity indices for grafted and non-grafted trees are shown in **Table 2**. Ho was higher for grafted trees and the number of alleles detected was equal, while the number of private alleles and He were higher for non-grafted trees. Contrary to grafted trees, Fis was close to zero and most primers did not show significant deviation from Hardy Weinberg equilibrium for non-grafted trees. Primers C21 and C23 depicted higher polymorphism and discriminating ability within both groups. AMOVA showed that 32% of variance accounted for variation between grafted and non-grafted trees, while the remaining 68% was due to variance within each group. There was significant divergence between grafted and non-grafted trees (*PhiPT* = 0.321, p = 0.01), and Nm value was 0.528.

A Bayesian approach was employed to examine population structure. The analysis grouped trees into two clusters (**Figure 3**). Ninety-eight trees had membership coefficient over 0.8 to cluster 1 (hereafter cluster G), 16 trees had membership coefficient over 0.8 to cluster 2 (hereafter cluster N) and 10 trees were admixtures (hereafter A). The vast majority of grafted trees grouped into cluster (G), along with two non-grafted trees. Contrary, the majority of non-grafted trees grouped into cluster (N) along with the three Italian varieties, the three trees identified by farmers as "Kountourka" and one grafted tree from the DOA plantations with the id name "Tala." Admixtures (A) were four grafted and six non-grafted trees (**Figure 2**). Clustering of trees according to the Neighbor-Joining dendrogram was consistent with the results of the structure analysis (**Figure 4**). Grafted trees were discriminated from non-grafted trees, with one exception, a tree identified by farmers as "Apostoliki—id code 3_NG_Kal_Ap." The vast majority of the grafted trees, identified by farmers and stakeholders with the names "Imeri," "Tillyrias," "Koumpota," "Empa," "Lefkaritiki," "Mavroteratsia," "Vaklaes," "Sarakina," "Eftakoili," "Koutsoulia," and "Koumparkes" were genetically identical or similar. Most of these trees exhibited similar pod morphological traits, except from "Koutsoulia" and "Koumparkes." The latter entries produced very short pods. Despite genetic similarity, it is noteworthy that differences on morphological and compositional traits were identified in orchards where "Lefkaritiki" and "Mavroteratsia" were grown side by side (**Supplementary File 3**). The grafted trees that were nonetheless genetically differentiated from the core of the grafted trees produced pods of variable and atypical morphology (e.g., very long "17_Gr_Ax_Talas", intermediate "6_Gr_Tox" and very short "22_Gr_Ape_Koutsoulia." Entries identified as "Kountourka" constitute a divergent genepool of grafted material. Genetic variation was also present within "Kountourka." These entries produced pods with distinct morphological traits compared to the other grafted material (**Supplementary File 4**).

Variation for Morphological and Compositional Traits

Descriptive statistics of the morphological and compositional traits are shown in **Tables 3**, **4** and their frequency plots are presented in **Supplementary File 5**. Average pod weight, length and width over seasons were 17.35 (g), 15.24 (cm), and 23.25 (mm), respectively. Variation was higher for the ratio seeds/pod weight and for pod weight (**Table 3**). Average powder color lightness (L*) and total sugars content over seasons were 67.99 and 45.27 g 100 g^{-1} dw, respectively. Variation was very low for these traits. Sucrose and mallic acid were the predominant soluble carbohydrate and organic acid, respectively. Variation was lower for sucrose and fructose compared to glucose. Phosphorous (PO_4) and potassium were the major metals detected in carob pods. Average DPPH, FRAP and total phenolics extracted with MeOH-HCL over seasons were 9.04 g AE kg^{-1} dw, 45.76 g AE kg^{-1} dw and 29.85 g GAE kg^{-1} dw, respectively. Gallic acid was the main phenolic compound detected followed by the flavonoid kaempferol-7-O-glucoside. Variation was higher for total phenolic content, condensed tannins, antioxidant activity and individual polyphenols compared to the other compositional and morphological traits (**Table 4**).

Genotypic and Environmental Effect on Morphological and Compositional Traits

ANOVA showed that genotype had a stronger effect on morphological traits, particularly on the ratio seeds/pod weight, pod width and thickness (**Table 3**). Descriptive statistics for each structure cluster and agro-environmental zone are presented in **Supplementary File 6**. Pods of trees grouped into cluster

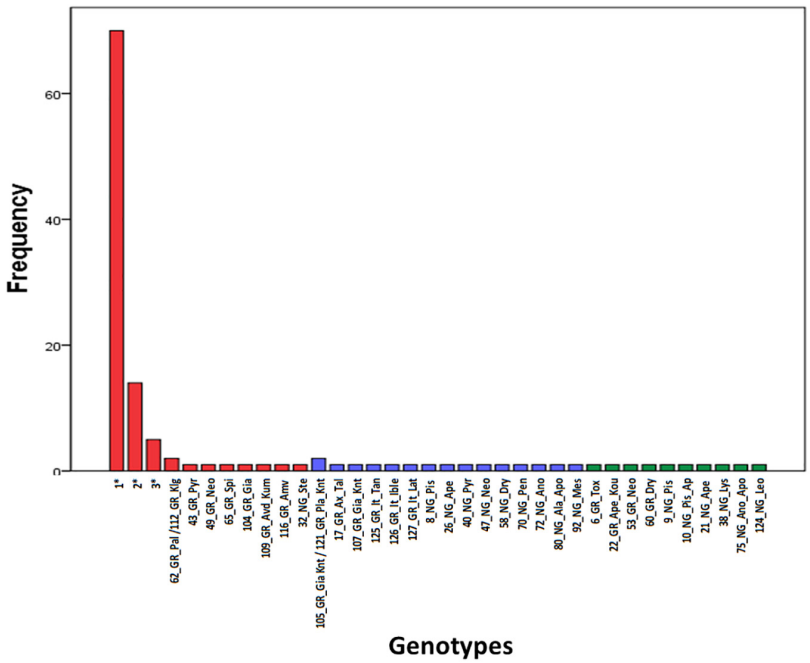

FIGURE 2 | Frequency of appearance of the 36 genotypes. Genotypes grouped into clusters (G) and (N) of the structure analysis are shown in red and blue color, respectively. Admixture (A) genotypes are shown in green color. The first component of the id code refers to tree number, the second to grafted (GR) or non-grafted trees (NG), the third to tree location and the fourth to the variety identification by farmers. Information concerning multiple tress assigned to genotypes 1, 2, and 3 can be found in **Supplementary File 1**.

TABLE 1 | Genetic diversity indices, deviation from Hardy Weinberg equilibrium (HW sign) and probability for null alleles F(null) among the trees sampled.

Locus	Na	Range	N ≤ 0.05	Private	Ne	Ho	He	PIC	Fis	HW sign	F(null)
Cesi 21 cttt7 L1	2	183–191	0	0	1.971	0.863	0.493	0.371	−0.751	***	−0.273
Cesi 21 cttt7 L2	2	280–288	0	0	1.994	0.927	0.498	0.374	−0.861	***	−0.301
Cesi 187 at15	6	139–158	4	1	2.330	0.952	0.571	0.478	−0.667	***	−0.267
Cesi 1187 at9	5	152–175	3	2	2.145	0.927	0.534	0.426	−0.737	***	−0.280
Cesi 98 gct6	3	161–171	1	0	2.149	0.895	0.535	0.429	−0.674	***	−0.264
Cesi 15 aaatag4	2	315–320	1	0	1.084	0.048	0.077	0.074	0.375	***	0.217
Cesi 976 ta5tg6	3	249–255	2	0	1.067	0.056	0.063	0.062	0.103	**	0.106
Cesi 74 ta7	2	277–279	1	0	1.093	0.073	0.085	0.081	0.144	ns	0.077
Cesi 17 tta7	2	170–173	0	0	1.990	0.831	0.497	0.374	−0.670	***	−0.251
C8	6	239–270	5	1	1.254	0.185	0.203	0.198	0.086	***	0.091
C10	3	211–227	1	0	2.092	0.879	0.522	0.412	−0.684	***	−0.264
C22	3	223–239	1	0	1.218	0.105	0.179	0.170	0.413	***	0.272
C23 L1	2	196–198	1	0	1.033	0.032	0.032	0.031	−0.016	ns	−0.002
C23 L2	12	209–262	9	4	3.166	0.893	0.684	0.632	−0.306	***	−0.167
C29	3	151–159	1	0	2.016	0.903	0.504	0.385	−0.792	***	−0.285
C31	3	240–248	2	1	1.025	0.024	0.024	0.024	−0.009	ns	−0.001
C4	5	191–209	2	0	2.425	0.976	0.588	0.501	−0.661	***	−0.267
C5	4	146–163	3	1	1.140	0.121	0.123	0.120	0.016	***	0.029
C21 L1	2	160–163	1	0	1.033	0.032	0.032	0.031	−0.016	ns	−0.002
C21 L2	12	174–231	9	4	3.166	0.893	0.684	0.632	−0.306	***	−0.167
C33	4	231–243	2	1	2.059	0.871	0.514	0.398	−0.694	***	−0.261
Mean	4.095		2.333	1.000	1.783	0.547	0.354	0.295	−0.319		−0.108
Total	86		49	15							

*Na, Number of alleles; N ≤ 0.05, number of alleles with frequency less than five; Private, number of private alleles; Ne, effective number of alleles; Ho, observed heterozygosity; He, expected heterozygosity; PIC, polymorphic information content; Fis, fixation index; ns, non-significant **P < 0.01, ***P < 0.001.*

TABLE 2 | Genetic diversity indices and deviation from Hardy Weinberg equilibrium (HW sign) among the grafted and the non-grafted trees.

	Grafted								Non-grafted							
Locus	**Na**	**N ≤ 0.05**	**Private**	**Ne**	**Ho**	**He**	**Fis**	**HW sign**	**Na**	**N ≤ 0.05**	**Private**	**Ne**	**Ho**	**He**	**Fis**	**HW sign**
Cesi 21 cttt7 L1	2	0	0	1.989	0.925	0.497	−0.861	***	2	0	0	1.710	0.471	0.415	−0.133	ns
Cesi 21 cttt7 L2	2	0	0	2.000	0.981	0.500	−0.963	***	2	0	0	1.710	0.588	0.415	−0.417	ns
Cesi 187 at15	5	3	1	2.196	0.981	0.545	−0.802	***	5	1	1	3.211	0.765	0.689	−0.111	ns
Cesi 1187 at9	4	2	1	2.056	0.972	0.514	−0.892	***	4	1	1	2.569	0.647	0.611	−0.059	ns
Cesi 98 gct6	3	1	0	2.110	0.953	0.526	−0.812	***	3	0	0	2.149	0.529	0.535	0.010	ns
Cesi 15 aaatag4	2	1	0	1.048	0.028	0.046	0.386	***	2	0	0	1.335	0.176	0.251	0.297	ns
Cesi 976 ta5tg6	3	2	1	1.038	0.037	0.037	−0.015	ns	3	0	0	1.273	0.176	0.215	0.177	ns
Cesi 74 ta7	2	1	0	1.028	0.028	0.028	−0.014	ns	2	0	0	1.562	0.353	0.360	0.019	ns
Cesi 17 tta7	2	0	0	2.000	0.935	0.500	−0.869	***	2	0	0	1.637	0.176	0.389	0.547	*
C8	6	5	2	1.132	0.103	0.117	0.120	***	5	0	0	2.546	0.706	0.607	−0.162	ns
C10	3	1	0	2.031	0.944	0.508	−0.859	***	3	0	0	2.181	0.471	0.542	0.131	ns
C22	3	2	0	1.088	0.019	0.081	0.770	***	3	0	0	2.379	0.647	0.580	−0.116	ns
C23 L1	2	1	1	1.009	0.009	0.009	−0.005	ns	2	0	0	1.192	0.176	0.161	−0.097	ns
C23 L2	10	7	3	2.936	0.944	0.659	−0.431	***	10	4	6	5.294	0.533	0.811	0.342	***
C29	3	1	1	2.015	0.953	0.504	−0.892	***	3	1	1	1.835	0.588	0.455	−0.293	ns
C31	1	0	0	1.000	0.000	0.000			3	1	1	1.197	0.176	0.164	−0.074	ns
C4	4	2	0	2.153	0.972	0.536	−0.815	***	5	0	0	4.313	1.000	0.768	−0.302	ns
C5	3	2	0	1.058	0.056	0.055	−0.022	ns	4	1	1	1.871	0.529	0.465	−0.138	ns
C21 L1	2	1	1	1.009	0.009	0.009	−0.005	ns	2	0	0	1.192	0.176	0.161	−0.097	ns
C21 L2	10	7	3	2.936	0.944	0.659	−0.431	***	10	4	6	5.294	0.533	0.811	0.342	***
C33	4	2	1	2.069	0.935	0.517	−0.809	***	2	0	0	1.993	0.471	0.498	0.056	ns
Mean	3.619	1.952	0.714	1.710	0.559	0.326	−0.411		3.667	0.619	0.809	2.307	0.471	0.472	−0.004	
Total	76	41	15						77	21	17					

*Na, Number of alleles; N ≤ 0.05, number of alleles with frequency less than five; Private, number of private alleles; Ne, effective number of alleles; Ho, observed heterozygosity; He, expected heterozygosity; Fis, fixation index; ns, non-significant *P < 0.05, ***P < 0.001.*

(G) had lower seeds/pod weight ratio (**Figure 5**), they were wider, thicker and heavier. On the other hand, the effect of the agro-environment was stronger on most of the compositional traits, particularly on total phenolics extracted with MeOH-HCL, FRAP, condensed tannins and powder color and to lesser extent on DPPH, total phenolics extracted with MeOH-acetate, malic, citric, oxalic, total organic acids, gallic acid, PO_4, Cl, Ca, Mg, and Na (**Table 4**). Variation in total phenolics extracted with MeOH-HCL, antioxidant activity, gallic acid and total organic acids concentrations between the different agro-environmental zones is shown in **Figure 6**. Most of the outliers are trees grouped in cluster N or they were admixtures (A). Concerning carbohydrates, the agro-environmental effect was stronger on reducing sugars fructose and glucose, the genotypic effect was stronger on total sugars, while genotypic and agro-environmental effects on sucrose were equal (**Table 4**). Carobs from trees grouped into cluster G had higher total sugars than those grouped into cluster N or they were admixtures (**Figure 5B**).

Seasonal Effect on Morphological and Compositional Traits

Overall, the seasonal effect was statistically significant for all traits except for the ratio seeds/pod weight, pod length and width (**Table 5**). Pods were heavier and thicker in 2019, while the concentrations of soluble carbohydrates, condensed tannins, total phenolics and antioxidant activity (DPPH and FRAP) were lower (**Tables 3**, **4**). Pearson correlations between seasons were strong and positive for most morphological traits and slightly lower for powder color, DPPH, FRAP, condensed tannins, total phenolics extracted with MeOH-HCL and total sugars. Among carbohydrates, sucrose had the strongest correlation, while fructose and glucose correlations were lower and similar. The correlations for valley thickness and average seed weight were non-significant.

Hierarchical Cluster Analysis on Phenotypic Traits

In 2018, hierarchical cluster analysis grouped trees in two major clusters of phenotypes (**Figure 7A**). The first cluster (Ia) contained the trees collected from the agro-environmental zones Neo Chorio, mountainous Paphos, north and south zones. The second cluster (Ib) contained exclusively trees collected from Anogira, mountainous Polis, Larnaca, Lemesos and Tillyria with one exception (10_NG_Pis_A_SZ_Ap). In 2019, cluster Ia1 contained trees collected almost exclusively from the agro-environmental zones Neo Chorio, mountainous Paphos, north and south zone while cluster Ib2 contained trees collected from mountainous Polis, Larnaca, Lemesos, and Tillyria (**Figure 7B**). Cluster Ib1 contained trees either collected from the latter agro-environmental zones or trees from the former

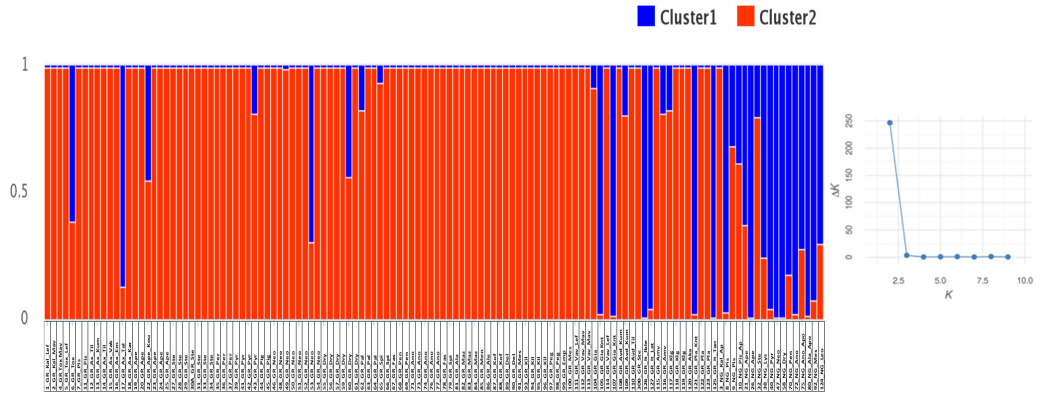

FIGURE 3 | Structure analysis of the 124 carob trees with optimum cluster $k = 2$. The first component of the id code refers to tree number, the second to grafted (GR) or non-grafted trees (NG), the third to tree location and the fourth to the variety identification by farmers (**Supplementary File 1** for code interpretation).

TABLE 3 | Means, standard errors (Std Error), Coefficients of Variation (CV), maximum and minimum values for morphological traits, percentage of sum of squares in ANOVA explained by the agro-environmental zones (Sign Env) and structure clusters (Sign Str).

Trait	N	Mean	Std Error	CV	Minimum	Maximum	Sign Env	Sign Str
Pod weight (g) (2018)	98	16.83	0.468	27.51	5.806	29.11	11.1	12.3**
Pod weight (g) (2019)	99	17.88	0.553	30.75	4.777	31.81	22.5**	14.4**
Ratio seeds/pod weight (2018)	98	8.758	0.283	31.94	3.975	16.72	10.6	55.4***
Ratio seeds/pod weight (2019)	100	8.532	0.288	33.75	4.508	17.08	9.4	48.0***
Length (cm) (2018)	98	15.12	0.237	15.50	7.540	20.41	12.7	0.6
Length (cm) (2019)	100	15.36	0.235	15.32	7.133	20.00	17.8*	0.3
Width (mm) (2018)	98	23.33	0.187	7.95	16.57	26.13	8.1	28.8***
Width (mm) (2019)	100	23.18	0.211	9.11	15.82	28.10	12.5	30.9***
Edge thickness (mm) (2018)	98	9.829	0.142	14.34	5.342	12.88	13.1	25.9***
Edge thickness (mm) (2019)	100	10.10	0.128	12.65	6.534	12.82	14.8	43.6***
Groove thickness (mm) (2018)	98	8.798	0.133	15.00	4.488	11.76	12.1	14.5**
Groove thickness (mm) (2019)	100	9.259	0.116	12.57	5.202	11.71	7.5	43.9***
Ratio edge/groove (2018)	98	1.138	0.011	9.77	0.826	1.443	22.8**	9.4*
Ratio edge/groove (2019)	100	1.103	0.008	7.23	0.899	1.330	30.7***	4.6
Valley thickness (mm) (2019)	98	5.735	0.071	12.23	2.577	8.097	9.1	6.6*
Valley thickness (mm) (2019)	100	8.849	0.093	10.51	6.969	12.03	7.6	4.0
Average seed weight (g) (2018)	98	0.202	0.004	20.93	0.126	0.492	11.1	9.8**
Average seed weight (g) (2019)	100	0.172	0.002	11.72	0.120	0.220	10.9	1.2
L* pod external (2018)	98	23.75	0.156	6.51	19.99	28.97	21.9**	7.7*
L* seed (2018)	98	34.21	0.311	9.00	27.25	39.78	23.5**	0.5

*$p < 0.05$, **$p < 0.01$, ***$p < 0.0001$.

agro-environmental zones that were grouped into cluster N or were admixtures.

DISCUSSION

SSRs Revealed Low Genetic Diversity Within Grafted Trees

Genetic diversity was assessed in a set of 107 grafted and 17 non-grafted trees distributed in nine agro-environmental zones of Cyprus, using 18 microsatellites. The SSRs discriminated 36 genotypes out of the 124 trees examined. They distinguished all non-grafted trees from grafted trees, of which one common genotype appeared in high frequency. La Malfa et al. (2014) reported that SSRs failed to distinguish some accessions presenting limited phenotypic variation. Likewise, SSRs in the current study did not discriminate neighboring trees showing slightly different morphological and compositional traits, as in the case of "Mavroteratsia" and "Lefkaritiki." Moreover, SSRs in some cases failed to distinguish accessions of grafted trees bearing pods of distinctive characters, e.g., "Koutsoulia" or "Koumbarkes" having very short pods (< 10 cm). Although morphological and compositional differentiation can also be attributed to other factors, such as rootstock-scion interaction (Emmanouilidou and Kyriacou, 2017), the discriminating ability of published SSRs for phenotypically similar grafted trees remains a concern. It can be

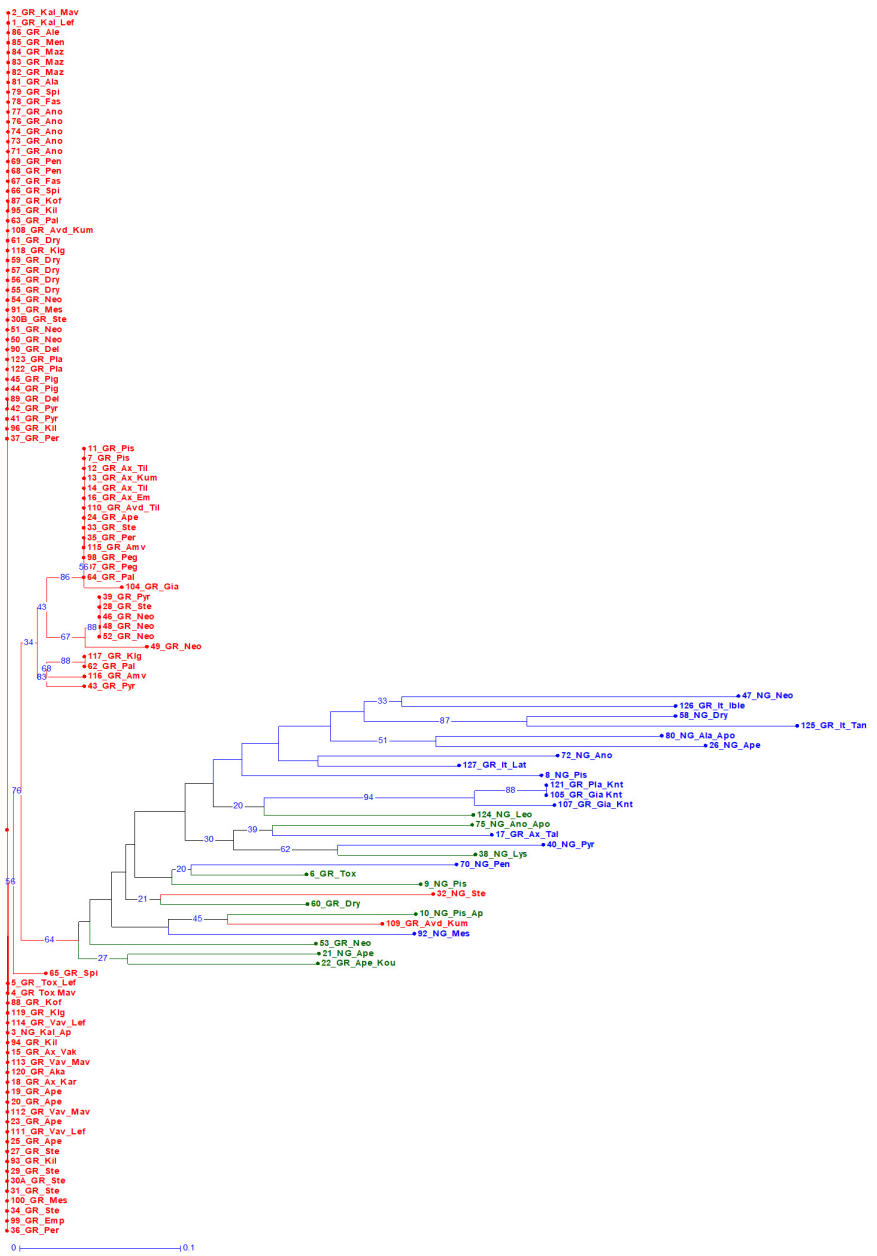

FIGURE 4 | Bootstrapped Neighbor Joining dendrogram of the 124 trees. Trees grouped into clusters (G) and (N) of the structure analysis are shown in red and blue color, respectively. Admixture (A) genotypes are in green color. The first component of the id code refers to tree number, the second to grafted (GR) or non-grafted trees (NG), the third to tree location, and the fourth to the variety identification by farmers (**Supplementary File 1** for code interpretation).

concluded that, SSRs can be effectively used for the assessment of carob genetic diversity relating to non-grafted genetic material (Viruel et al., 2018); however, other molecular techniques (e.g., next-generation genotyping) should also be employed to examine in more detail the genetic diversity of grafted material (La Malfa et al., 2014; Viruel et al., 2018; Di Guardo et al., 2019).

Grafted carob is a vegetatively (asexually) propagated tree with expectedly lower genetic diversity than an annual crop in similar environments, e.g., Cypriot durum wheat landraces (Kyratzis et al., 2019). Moreover, carob genetic diversity is also lower than that of olive (Anestiadou et al., 2017; Emmanouilidou et al., 2018) and pomegranate Cypriot genetic resources (Kyriacou et al., 2020). The lower genetic diversity compared to other tree species from Cyprus is consistent with the scenario of declining carob genetic resources purported by Viruel et al. (2018). Genetic variability within the cultivated genepool can be increased through hybridization between local or/and foreign material, and subsequent selection by farmers (Tous et al., 2013; Anestiadou et al., 2017). Farmers' selection criteria for carob were rather limited (Batlle and Tous, 1997; Tous et al., 2013), therefore

TABLE 4 | Means, standard errors (Std Error), Coefficients of Variation (CV), maximum and minimum values for compositional traits, percentage of sum of squares in ANOVA explained by the agro-environmental zones (Sign Env) and structure clusters (Sign Str).

Trait	N	Mean	Std. Error	CV	Minimum	Maximum	Sign Env	Sign Str
L* powder (2018)	98	67.26	0.583	8.59	49.10	76.78	67.0***	7.3*
L* powder (2019)	100	68.72	0.501	7.29	52.47	77.33	39.9***	16.2**
Sucrose (g 100 g^{-1} dw) (2018)	98	34.19	0.497	14.39	12.10	43.40	21.8**	20.1***
Sucrose (g 100 g^{-1} dw) (2019)	100	30.12	0.444	14.75	12.60	38.00	13.5	16.0**
Fructose (g 100 g^{-1} dw (2018)	98	9.983	0.132	13.13	6.600	14.00	29.9***	7.8*
Fructose (g 100 g^{-1} dw) (2019)	100	7.449	0.125	16.75	4.600	12.00	17.4*	3.5
Glucose (g 100 g^{-1} dw) (2018)	98	5.245	0.131	24.71	2.400	10.20	48.3***	9.5**
Glucose (g 100 g^{-1} dw) (2019)	100	3.547	0.109	30.71	1.500	9.300	22.6**	3.1
Total Sugars (g 100 g^{-1} dw) (2018)	98	49.42	0.430	8.62	27.80	57.00	9.8	29.7***
Total Sugars (g 100 g^{-1} dw) (2019)	100	41.12	0.401	9.75	30.50	49.20	10.2	20.4***
Malic acid (g kg^{-1} dw) (2018)	98	5.873	0.160	26.98	1.495	10.71	37.2***	1.6
Citric acid (g kg^{-1} dw) (2018)	98	1.598	0.047	29.15	0.351	2.897	36.5***	10.4**
Oxalic acid (g kg^{-1} dw) (2018)	98	0.304	0.005	15.89	0.206	0.442	29.0***	1.7
Total organic acid (g kg^{-1} dw) (2018)	98	7.774	0.180	22.89	2.568	13.45	41.1***	0.1
Protein% (w/w) (2018)	98	4.691	0.095	20.07	2.179	7.868	18.6*	1.2
PO_4 (g kg^{-1} dw) (2018)	98	1.213	0.036	29.65	0.568	2.611	16.1*	2.3
SO_4 (g kg^{-1} dw) (2018)	98	0.505	0.035	68.76	0.083	1.648	9.6	4.3
NH_4 (g kg^{-1} dw) (2018)	97	0.147	0.013	86.76	0.000	0.749	13.4	4.2
NO_3 (g kg^{-1} dw) (2018)	98	0.066	0.003	47.62	0.020	0.197	16.8	0.4
K (g kg^{-1} dw) (2018)	98	12.14	0.267	21.74	4.546	19.89	10	0.04
Cl (g kg^{-1} dw) (2018)	98	1.982	0.056	27.91	1.080	3.994	30.2***	5.0
Ca (g kg^{-1} dw) (2018)	98	1.640	0.066	39.82	0.231	4.453	28.9***	2.3
Mg (g kg^{-1} dw) (2018)	98	1.013	0.029	28.76	0.218	1.913	18.0*	1.3
Na (g kg^{-1} dw) (2018)	98	0.216	0.014	62.96	0.047	0.843	18.9*	5.2
DPPH (g AE kg^{-1} dw) (2018)	98	9.899	0.437	43.65	2.100	24.10	44.5***	11.6**
DPPH (g AE kg^{-1} dw) (2019)	100	8.181	0.629	76.85	0.250	39.04	32.6***	9.9**
FRAP (g AE kg^{-1} dw) (2018)	98	53.61	2.703	49.91	11.00	130.5	66.8***	3.6
FRAP (g AE kg^{-1} dw) (2019)	100	37.91	1.886	49.76	11.49	86.98	53.4***	0.5
Total Phenolics (MeOH-HCL) (g GAE kg^{-1} dw) (2018)	98	34.07	1.864	54.15	7.500	85.40	71.2***	2.8
Total Phenolics (MeOH-HCL) (g GAE kg^{-1} dw) (2019)	100	25.63	1.805	70.42	3.021	99.18	50.4***	1.7
Total Phenolics—(MeOH-acetate) (g GAE kg^{-1} dw) (2018)	98	11.08	0.429	38.26	3.700	29.10	39.8***	11.6**
Condensed tannins (mg Catechin g^{-1} dw) (2018)	98	47.89	2.871	59.38	5.000	108.0	68.9***	0.8
Condensed tannins (mg Catechin g^{-1} dw) (2019)	100	24.98	2.361	94.35	0.000	97.80	56.1***	0.03
Gallic acid ($\mu g\ g^{-1}$ dw) (2018)	98	959.6	56.64	58.43	52.89	4015	24.3**	5.9
Kaempferol-7-O-glucoside ($\mu g\ g^{-1}$ dw) (2018)	98	74.87	6.598	87.24	0.000	438.6	4.8	13.2**
Gallocatechin ($\mu g\ g^{-1}$ dw) (2018)	98	11.90	0.240	19.94	0.000	19.93	8.4	0.8
Naringenin diglucoside ($\mu g\ g^{-1}$ dw) (2018)	98	11.76	0.758	63.80	0.000	31.34	33.9***	0.6
Gallocatechin gallate ($\mu g\ g^{-1}$ dw) (2018)	98	10.23	0.943	91.34	0.000	41.30	24.3**	2.5
Catechin ($\mu g\ g^{-1}$ dw) (2018)	98	9.479	0.894	93.38	0.000	66.08	15.1	0.1
Isovitexin ($\mu g\ g^{-1}$ dw) (2018)	98	5.580	0.261	46.36	0.426	13.31	7.1	1.7
Luteolin-7-O-glucoside ($\mu g\ g^{-1}$ dw) (2018)	98	5.041	0.374	73.77	0.000	17.13	13.8	0.1
Epigallocatechin gallate ($\mu g\ g^{-1}$ dw) (2018)	98	3.213	0.089	27.51	0.000	6.228	5.8	0.7
Methyl gallate ($\mu g\ g^{-1}$ dw) (2018)	98	2.882	0.214	73.43	0.000	10.99	31.4***	7.2*
Epicatechin ($\mu g\ g^{-1}$ dw) (2018)	98	1.508	0.036	23.81	0.000	2.867	13.8	0.6
Vitexin ($\mu g\ g^{-1}$ dw) (2018)	98	0.955	0.076	79.19	0.436	5.045	8.6	0.3
Total ($\mu g\ g^{-1}$ dw) (2018)	98	1097	59.48	53.67	89.91	4139	22.1**	7.4*

*$p < 0.05$, **$p < 0.01$, ***$p < 0.0001$.

selection pressure for limited morphological and compositional traits could result in further declination of genetic diversity within the cultivated carob genepool compared to other tree species (e.g., pomegranate Kyriacou et al., 2020) presenting wider phenotypic variability, ease of propagation (e.g., by cuttings) and complexity of fruit quality traits. Furthermore, the carob population suffered a severe decline over the last 50 years mainly owing to the cultivation of irrigated cash crops, illicit logging and

land development, as evidenced in the substantial reduction of the carob cultivated area (Davies, 1970). For example, Tillyria was historically the most famous area for carob production in Cyprus (Gennadius, 1902), hence the variety "Tillyria" is synonymous with this area (Orphanos and Papaconstantinou, 1969); nowadays however, only few remaining scattered carob trees testify the area's past agricultural landscape.

The genetic diversity detected in the present study and the percentage of discriminated genotypes were lower than previous works on carob (La Malfa et al., 2014; Viruel et al., 2018; Di Guardo et al., 2019), which can be attributed to the fact that sampling was limited to Cyprus and emphasis was placed on grafted material. Italian varieties were genetically divergent from the core of the Cypriot grafted germplasm; however, they shared, to very large extent, polymorphism that exists within the non-grafted Cypriot material supporting the concept of multiple domestications of the carob tree from native populations throughout the Mediterranean basin (Viruel et al., 2018).

Non-grafted trees grouped at a different structure cluster than the core of the grafted trees or they were admixtures. Despite the smaller sample size, genetic diversity among non-grafted trees was higher, Fis was close to zero and most loci did not show significant deviation from Hardy Weinberg equilibrium indicating that contrary to grafted material wild populations were under random mating. Nevertheless, gene flow exists between grafted and non-grafted genetic material. Grafted trees producing pods of variable morphology were grouped as admixtures in structure analysis confirming that admixture is an important component of carob genetic diversity (Di Guardo et al., 2019). As carob is a cross pollinated species and grafted trees are not easily distinguished from non-grafted ones, farmers unconsciously select trees from segregating material once they discern desirable pod traits. Lower prices obtained in some geographical areas due to inferior quality (Gennadius, 1902) could also drive farmers to experiment with selection of local phenotypes, as was observed in the Karpasia peninsula (Ticho, 1959). This "new" genetic diversity appears with low frequency, less than 1% indicating a low selection pressure from farmers. Therefore, extensive sampling is needed to track genetic diversity within grafted carob genetic resources.

Genetic Analysis Suggests the Existence of Two Carob Landraces

Based on the definition of landrace by Villa et al. (2005), it can be concluded that two carob landraces exist in Cyprus grouped into separate genetic clusters, each composed of genetically similar or identical trees. The first landrace, which predominates on the island, encompasses the genetic material known under the local names "Mavroteratsia," "Lefkaritiki," "Tillyria," "Koumpota," "Imeri," "Kamateri," "Saradjina," and "Vakles." The second landrace, which is locally common to Karpasia peninsula, encompasses the genetic material known under the local name "Kountourka." These two landraces produced pods of distinct morphology and they can be easily distinguished also by the different tree morphology (Orphanos and Papaconstantinou, 1969). A slightly different compositional profile for these two landraces was also reported (Davies et al., 1971). Our results demonstrated that morphological and compositional differentiation of these two landraces are aligned with genetic differentiation. Contrary, Di Guardo et al. (2019) found that "Tillyria" and "Kountourka" are genetically similar varieties. This discrepancy, which is likely due to mislabeling of Cypriot accessions, confirms that passport data of *ex situ* collections should be treated with caution (Kyratzis et al., 2019). Mislabeling was also identified within the DoA plantations which provide budding wood for propagation purposes. Therefore, genetic identification of mother plants providing budding wood to nurseries is crucial to ensure access to certified material of the desirable genotype.

Morphological Traits Are Predominantly Under Genetic Control While Compositional Traits Are Mainly Under Agro-Environmental Control

Morphological traits, carbohydrate and protein content were comparable to those reported by Orphanos and Papaconstantinou (1969). In support of Tetik et al. (2011), variation in morphological traits was found larger for the ratio seed/pod weight and pod mass. Sucrose was the major sugar present while fructose and glucose were detected in lower concentrations, in agreement with Biner et al. (2007). The variation in total sugars was similar to the variation in soluble solids found in Turkish carobs (Tetik et al., 2011). Our results further confirm that carob is a rich source of polyphenols, including high concentrations of condensed tannins (Avallone et al., 1997; El Bouzdoudi et al., 2016). Gallic acid was the major phenolic compound and mallic acid the major organic acid

TABLE 5 | Pearson correlation coefficients (r), significance of correlations (Sig) and significance of t test for traits between seasons.

Trait	*r*	Sig.	*t*-test
Pod weight	0.809	***	**
Ratio seeds/pod weight	0.781	***	ns
Length	0.743	***	ns
Width	0.754	***	ns
Edge thickness	0.764	***	**
Groove thickness	0.646	***	***
Ratio edge/Groove	0.580	***	***
Valley thickness	−0.184	ns	***
Average seed weight	0.015	ns	***
L* powder	0.693	***	*
Sucrose	0.586	***	***
Fructose	0.389	***	***
Glucose	0.383	***	***
Total sugars	0.517	***	***
DPPH	0.685	***	**
FRAP	0.699	***	***
Total Phenolics (MeOH-HCL)	0.663	***	***
Condensed tannins	0.612	***	***

*$p < 0.05$, **$p < 0.01$, ***$p < 0.0001$.

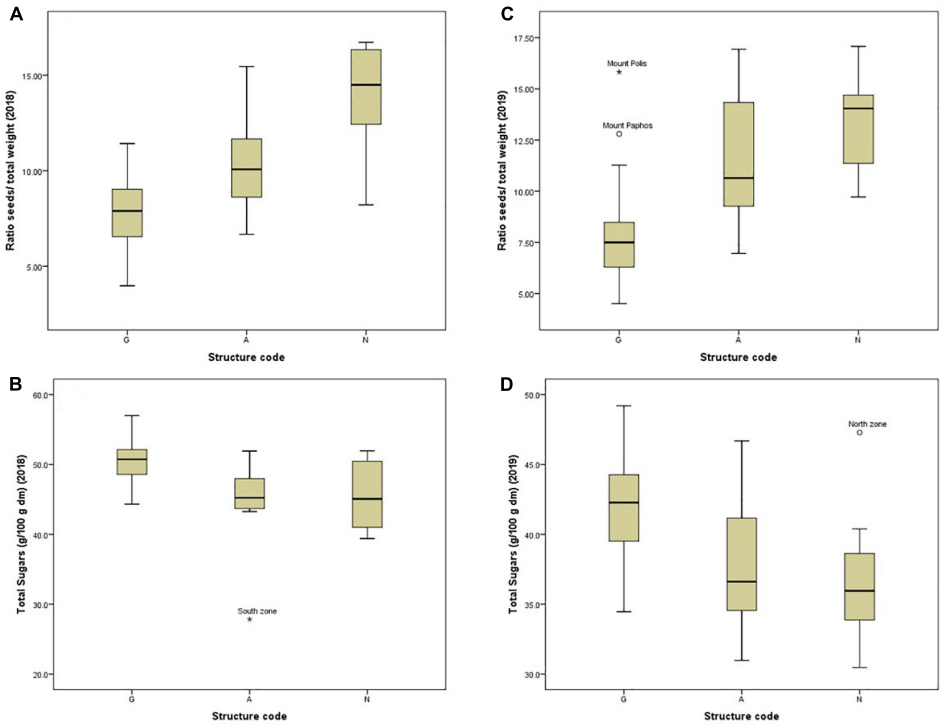

FIGURE 5 | Box plots showing the variation between and within structure clusters for the ratio seeds- to-pod weight **(A,B)** and for total sugars **(C,D)** during the first and the second season.

(Ayaz et al., 2007; Farag et al., 2019). Contrary to Portuguese carobs (Custodio et al., 2011), catechin appeared at very low concentration. Potassium was the major metal detected in the pulp accompanied by considerable concentration of calcium, phosphorus and magnesium (Oziyci et al., 2014; Goulas et al., 2016).

Morphological and compositional traits of carobs can be affected by genotype (Barracosa et al., 2007), gender (Custodio et al., 2011), climatic conditions (Tous et al., 2013), harvest stage (Benchikh et al., 2014; Farag et al., 2019), soil conditions, and season (Correia et al., 2018). In the present study, genotypic effect was high for pod morphological traits, particularly for the seeds/pod weight ratio, pod width and thickness, as well as total sugars. These traits were under strong selection pressure from farmers (Batlle and Tous, 1997; Tetik et al., 2011). During ripening, there is a slight reduction of glucose and fructose and a progressive accumulation of sucrose, with a significant effect of the environment on their relative concentrations (Othmen et al., 2019). Our results suggest that sucrose accumulation can be equally affected by the genotype and the agro-environment, contrary to the accumulation of monosaccharides that is mainly under agro-environmental control. Accordingly, Biner et al. (2007) found higher sucrose concentration on grafted material compared to wild trees, while differences in fructose and glucose were non-significant. Farag et al. (2019) reported that the discrimination of carob samples of different geographical origins is more effective when based on monosaccharaides, which can be attributed to the high agro-environmental effect on their accumulation. It may be inferred that the potential for sugar accumulation is to an extent genetically regulated, however the final concentration of total sugars deposited in the pod and especially the relative concentrations of reducing sugars to sucrose are influenced by the agro-environmental effect on the ripening process.

Agro-environment had a strong effect on total phenolic content, antioxidant activity, condensed tannins and color of the carob powder, and to lesser extent on polyphenolic composition. Farag et al. (2019) reported that tannins and flavonoids were the major compounds for discriminating carob samples of different origin. Previous studies reported significant differences between varieties or environments for condensed tannins, total phenolic content and antioxidant activity (Avallone et al., 1997; Custodio et al., 2011; Benchikh et al., 2014; Nahla, 2014; El Bouzdoudi et al., 2016; Othmen et al., 2019), nevertheless the reported variation between varieties or environments was larger in the present study, likely due to the extensive sampling performed in different agro-environments. Despite the high agro-environmental effect, substantial variation was also observed within each agro-environment (**Figure 6**) indicating the presence of genetic variation for these traits. However, as selection of promising genotypes is time consuming, describing the environments enhancing the accumulation of these secondary metabolites would be a faster approach to improve bioactive content than promoting the cultivation of improved genotypes rich on these compounds.

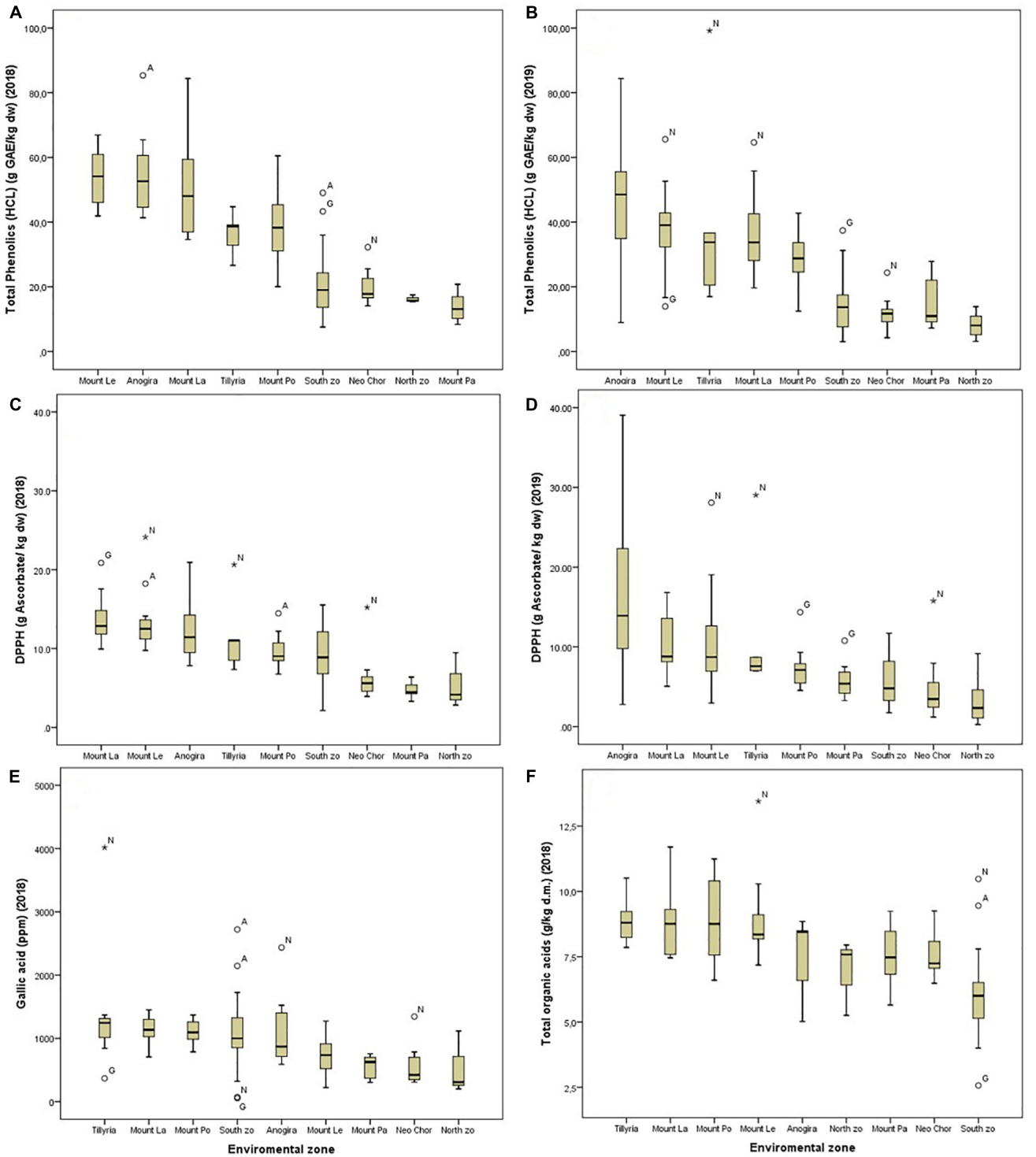

FIGURE 6 | Box plots showing the variation across different agro-environmental zones for total phenolics extracted with MeOH-HCL in 2018 **(A)** and in 2019 **(B)** for antioxidant activity (DPPH) in 2018 **(C)** and in 2019 **(D)** and for gallic acid concentration **(E)** and total organic acids **(F)** in 2018. Outlier codes refer to their structure cluster.

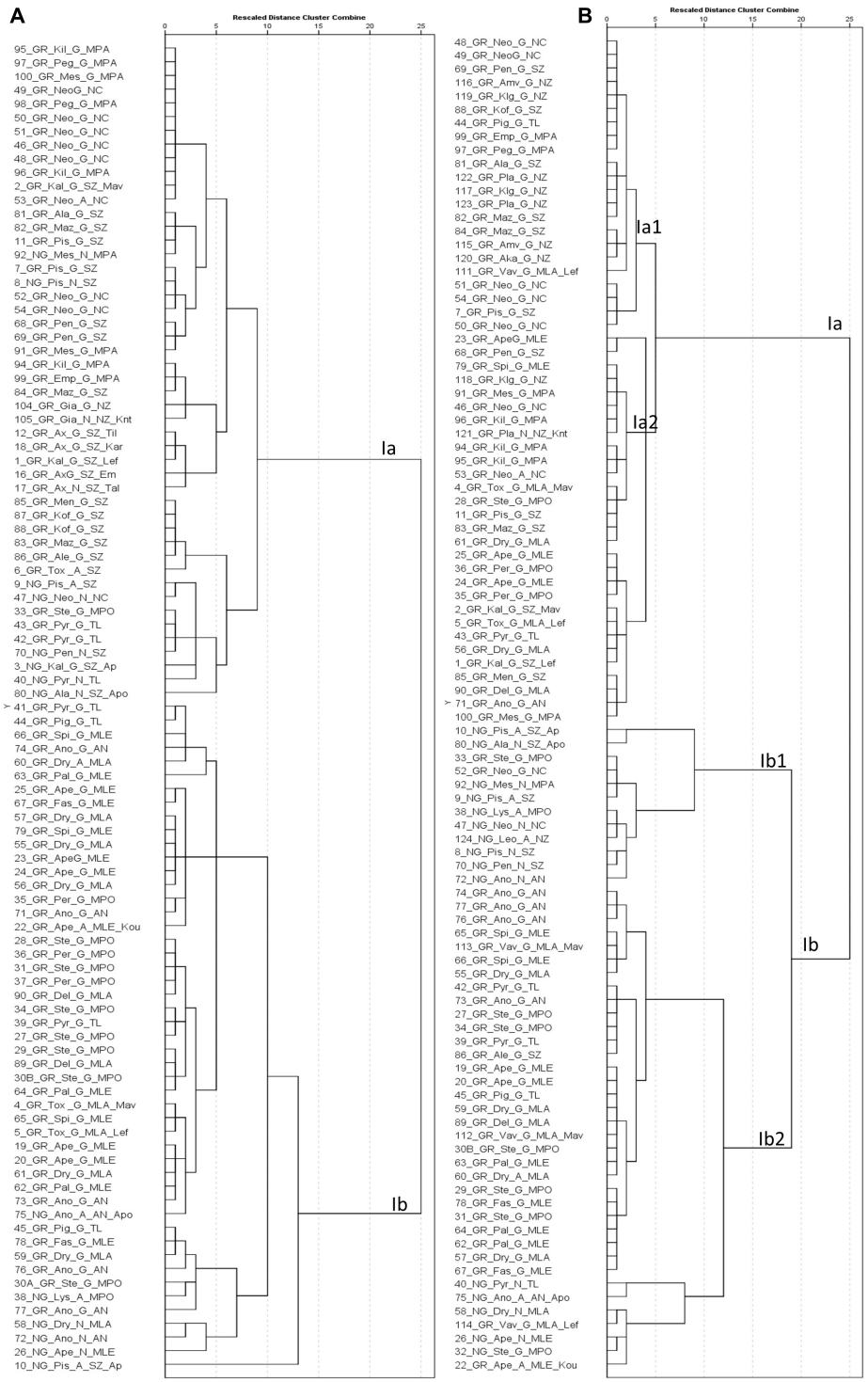

FIGURE 7 | Hierarchical cluster analysis based on phenotypic traits measured in 2018 **(A)** and 2019 **(B)**. The first component of the id code refers to tree number, the second to grafted (GR) or non-grafted trees (NG), the third to tree location, the fourth to the structure cluster, the fifth to the agro-environmental zone where the tree grows and the sixth to the variety identification by farmers (**Supplementary File 1** for code interpretation).

Morphological and Compositional Traits Vary Significantly Between Seasons

Carob pod size can be affected by environmental factors as well as level of pollination and fruit set (Batlle and Tous, 1997). In the present study, the effect of both the agro-environment and season on pod size were non-significant and strong correlations were observed between seasons indicating that pod morphology is rather stable, mainly driven by genetic factors. Average rainfall during the first season was slightly lower than normal (90%) while the second season was exceptionally rainy (158% of normal; Cyprus Meteorological, 2020 access 27/05/2020). As favorable environmental conditions increase the heritability of agronomic traits (Blum, 2010), it is likely that the environmental and seasonal effects on pod morphology diminished. Carob pods were heavier and thicker in 2019, contrary to the non-significant differences in pod length and width implying that carob trees modulate carob pod weight by adjusting pod thickness.

Seasonal effect was significant on antioxidant activity, total phenolic content and condensed tannins, which were lower during the 2nd season. The exceptionally high precipitation of the 2nd season likely reduced water stress resulting in lower phenolic content and antioxidant activity TE (Tavarini et al., 2011; Nasrabadi et al., 2019). The strong correlations between seasons for antioxidant activity, total phenolic content and condensed tannins suggest the presence of low season-by-tree interactions. Contrary, the lower correlations between seasons for carbohydrates indicates the presence of a strong season-by-tree interaction for these compounds, particularly for glucose and fructose and to lesser extent on sucrose. The lower concentration of carbohydrates in the season where precipitation was exceptionally high and average carob production was lower (Personal communication with stakeholders 2019) might be attributed to physiological factors related to alternate bearing (Von Haselberg, 1996), which triggered carob trees to invest carbohydrates in vegetative growth and development rather than storage in pods. The season-by-tree interactions for carbohydrates can be explained in the context that alternate bearing is not ubiquitous in the carob population (Rosenstock et al., 2010), thus in the same year both "on-season" and "off-season" trees are encountered.

Two Major Agro-Environmental Zones Shape the Compositional Profile of Carobs in Cyprus

Carob producing areas in Cyprus can be divided into two major agro-environmental zones based on the effect of the agro-environment on compositional traits and the overall performance depicted by hierarchical cluster analysis. The first major zone encompasses the agro-environmental zones mountainous Anogira, Larnaca, Polis, Lemesos and Tillyria where carobs are characterized by lighter-colored pulp (higher powder L*), higher antioxidant activity, higher concentration of total phenolics, condensed tannins, organic acids, calcium and sucrose. These compounds are highly significant for the characterization of carob as a functional food (Rasheed et al., 2019). The second major zone encompasses the agro-environmental zones south and north, mountainous Paphos and Neo Chorio where carobs are characterized by darker carob pulp, lower antioxidant activity, lower concentration of total phenolics, condensed tannins, organic acids and sucrose, and higher concentration of fructose and glucose. The vast majority of grafted trees in the two major zones were genetically similar or identical and were grouped in the same genetic cluster. Since the standardization of the crude pod material is important for the processing industry, the narrow genetic basis of grafted material in Cyprus implies that the local processing industry should categorize this material based on whether it was collected from grafted or non-grafted trees and according to the agro-environmental zone of origin.

Trees grouped into structure cluster (N) or those that were admixtures (A) behaved in many cases as outliers within particular environmental zones (**Figure 6**); this was further evident in their grouping in the hierarchical cluster analysis (**Figure 7**). Therefore, compositional traits are modulated by the agro-environment and the genotype. The present study also emphasizes that seasonal effect on compositional traits should also be considered. Despite previous efforts to categorize carobs according to their geographical origin (Farag et al., 2019; Kokkinofta et al., 2020), these works suffer from the limited description either of the genetic profile or the environmental conditions where trees were grown, moreover they lack replication in time. These limitations that are not easily manageable for an underutilized tree crop like carob, warrant international cooperation for the establishment of *ex situ* collections of the same divergent genetic material in different agro-environmental zones to allow accurate assessment of the genotypic, agro-environmental and seasonal effects on carob compositional traits and to facilitate association mapping studies.

CONCLUSION

Genotyping analysis of carob genetic resources, using microsatellites revealed low genetic diversity levels within Cypriot grafted germplasm. Two carob landraces were identified: "Tillyria" and "Kountourka" consisted of genetically similar or identical trees. "Tillyria" predominates on the island while "Kountourka" is locally common to the Karpasia peninsula. Genetically divergent genotypes from the abovementioned landraces were identified, which appeared though with very low frequency, indicating a low selection pressure from farmers. Morphological traits, particularly the seeds/pod weight ratio, pod width and thickness, and the total sugar content, were mainly under genetic control. On the other hand, agro-environmental conditions had a strong effect on compositional traits, particularly on total phenolics, antioxidant activity, condensed tannins and color of the carob powder. Concerning sugar profile, the agro-environmental effect was stronger on fructose and glucose, while sucrose was equally affected by genotype and agro-environment. Significant differences between seasons for morphological and compositional traits were observed revealing the existence of considerable seasonal effect. Correlations between seasons were stronger for morphological traits, intermediate for

powder color, antioxidant activity, total phenolic content and condensed tannins, and lower for carbohydrates suggesting the existence of a strong tree-by-season interaction on sugar content. Based on all the above findings, Cyprus can be divided to two major agro-environmental zones that modulate to a great extent the compositional properties of the carob pulp.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of the Agricultural Research Institute of the Cyprus Ministry of Agriculture, Rural Development and Environment. The participants provided written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AK, MK, and CA: conceptualization, methodology, and data curation. AK, MK, CA, LP, GG, and YR: formal analysis, writing—original draft preparation, and writing—review and editing. MK, LP, GG, and YR: resources. AK and MK: supervision and project administration. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was co-funded by the European Regional Development Fund and the Republic of Cyprus through the Cyprus Research and Innovation Foundation (Project: BlackGold INTEGRATED/0916/0019).

ACKNOWLEDGMENTS

The authors would like to thank the staff of the Vegetable Crops Sector, Mr. Demetris Kourtelarides, Ms. Loukia Hadjiyianni, and Ms. Maria Stefani, for their assistant in carrying out the experiments and data recording.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.612376/full#supplementary-material

REFERENCES

Anestiadou, K., Nikoloudakis, N., Hagidimitriou, M., and Katsiotis, A. (2017). Monumental olive trees of Cyprus contributed to the establishment of the contemporary olive germplasm. *PLoS One* 12:e0187697. doi: 10.1371/journal.pone.0187697

Avallone, R., Plessi, M., Baraldi, M., and Monzani, A. (1997). Determination of chemical composition of carob (*Ceratonia siliqua*): protein, fat, carbohydrates, and tannins. *J. Food Compost. Anal.* 10, 166–172. doi: 10.1006/jfca.1997.0528

Ayaz, F. A., Torun, H., Ayaz, S., Correia, P. J., Alaiz, M., Sanz, C., et al. (2007). Determination of chemical composition of Anatolian carob pod (*Ceratonia siliqua* L.): sugars, amino and organic acids, minerals and phenolic compounds. *J. Food Qual.* 30, 1040–1055. doi: 10.1111/j.1745-4557.2007.00176.x

Barracosa, P., Osório, J., and Cravador, A. (2007). Evaluation of fruit and seed diversity and characterization of carob (*Ceratonia siliqua* L.) cultivars in Algarve region. *Sci. Hortic.* 114, 250–257. doi: 10.1016/j.scienta.2007.06.024

Batlle, I., and Tous, J. (1997). *Carob Tree: Ceratonia siliqua L.-Promoting the Conservation and Use of Underutilized and Neglected Crops. Ceratonia siliqua*. Rome: Bioversity International, 92.

Benchikh, Y., Louaileche, H., George, B., and Merlin, A. (2014). Changes in bioactive phytochemical content and in vitro antioxidant activity of carob (*Ceratonia siliqua* L.) as influenced by fruit ripening. *Ind. Crops Prod.* 60, 298–303. doi: 10.1016/j.indcrop.2014.05.048

Benzie, I. F. F., and Strain, J. J. (1996). The ferric reducing ability of plasma (FRAP) as a measure of "antioxidant power": the FRAP assay. *Anal. Biochem.* 239, 70–76. doi: 10.1006/abio.1996.0292

Biner, B., Gubbuk, H., Karhan, M., Aksu, M., and Pekmezci, M. (2007). Sugar profiles of the pods of cultivated and wild types of carob bean (*Ceratonia siliqua* L.) in Turkey. *Food Chem.* 100, 1453–1455. doi: 10.1016/j.foodchem.2005.11.037

Blum, A. (2010). *Plant Breeding for Water-Limited Environments*. Berlin: Springer Science & Business Media.

Brand-Williams, W., Cuvelier, M.-E., and Berset, C. (1995). Use of a free radical method to evaluate antioxidant activity. *LWT Food Sci. Technol.* 28, 25–30. doi: 10.1016/s0023-6438(95)80008-5

Bouzouita, N., Khaldi, A., Zgoulli, S., Chebil, L., Chekki, R., Chaabouni, M. M., et al. (2007). The analysis of crude and purified locust bean gum: a comparison of samples from different carob tree populations in Tunisia. *Food Chem.* 101, 1508–1515. doi: 10.1016/j.foodchem.2006.03.056

Bremner, J. (1965). Total nitrogen. *Methods Soil Anal. 2 Chem. Microbiol. Prop.* 9, 1149–1178.

Correia, P. J., Saavedra, T., Gama, F., da Graça Miguel, M., de Varennes, A., and Pestana, M. (2018). Biologically active compounds available in *Ceratonia siliqua* L. grown in contrasting soils under Mediterranean climate. *Sci. Hortic.* 235, 228–234. doi: 10.1016/j.scienta.2018.03.010

Custodio, L., Fernandes, E., Escapa, A. L., Fajardo, A., Aligue, R., Albericio, F., et al. (2011). Antioxidant and cytotoxic activities of carob tree fruit pulps are strongly influenced by gender and cultivar. *J. Agric. Food Chem.* 59, 7005–7012. doi: 10.1021/jf200838f

Cyprus Meteorological (2020). *Average Rainfalls. Cyprus Meteorological Service*. Available online at: http://www.moa.gov.cy/moa/ms/ms.nsf/DMLindex_gr/DMLindex_gr?OpenDocument (accessed May 27, 2020).

Davies, W. N. L. (1970). The carob tree and its importance in the agricultural economy of Cyprus. *Econ. Bot.* 24, 460–470. doi: 10.1007/BF02860752

Davies, W. N. L., Orphanos, P. I., and Papaconstantinou, J. (1971). Chemical composition of developing carob pods. *J. Sci. Food Agric.* 22, 83–86. doi: 10.1002/jsfa.2740220210

Di Guardo, M., Scollo, F., Ninot, A., Rovira, M., Hermoso, J. F., Distefano, G., et al. (2019). Genetic structure analysis and selection of a core collection for carob tree germplasm conservation and management. *Tree Genet. Genomes* 15:41.

El Bouzdoudi, B., El Ansari, Z. N., Mangalagiu, I., Mantu, D., Badoc, A., and Lamarti, A. (2016). Determination of polyphenols content in carob pulp from wild and domesticated Moroccan trees. *Am. J. Plant Sci.* 07, 1937–1951. doi: 10.4236/ajps.2016.714177

Emmanouilidou, M. G., and Kyriacou, M. C. (2017). Rootstock-modulated yield performance, fruit maturation and phytochemical quality of 'Lane Late' and 'Delta' sweet orange. *Sci. Hortic.* 225, 112–121. doi: 10.1016/j.scienta.2017.06.056

Emmanouilidou, M. G., Kyriacou, M. C., and Trujillo, I. (2018). Characterization and identification of indigenous olive germplasm from Cyprus using morphological and simple sequence repeat markers. *Hortscience* 53, 1306–1313. doi: 10.21273/HORTSCI13192-18

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

Farag, M. A., El-Kersh, D. M., Ehrlich, A., Choucry, M. A., El-Seedi, H., Frolov, A., et al. (2019). Variation in *Ceratonia siliqua* pod metabolome in context of its different geographical origin, ripening stage and roasting process. *Food Chem.* 283, 675–687. doi: 10.1016/j.foodchem.2018.12.118

Gennadius, P. (1902). *The Carob-Tree*. Nicosia: The Goverment Printing Office.

Goulas, V., Stylos, E., Chatziathanasiadou, M. V., Mavromoustakos, T., and Tzakos, A. G. (2016). Functional components of carob fruit: linking the chemical and biological space. *Int. J. Mol. Sci.* 17:1875. doi: 10.3390/ijms17111875

Kalinowski, S. T., Taper, M. L., and Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16, 1099–1106. doi: 10.1111/j.1365-294X.2007.03089.x

Kokkinofta, R., Yiannopoulos, S., Stylianou, M. A., and Agapiou, A. (2020). Use of chemometrics for correlating carobs nutritional compositional values with geographic origin. *Metabolites* 10:62. doi: 10.3390/metabo10020062

Kyratzis, A. C., Nikoloudakis, N., and Katsiotis, A. (2019). Genetic variability in landraces populations and the risk to lose genetic variation. The example of landrace 'Kyperounda' and its implications for ex situ conservation. *PLoS One* 14:e0224255. doi: 10.1371/journal.pone.0224255

Kyriacou, M. C., Emmanouilidou, M. G., and Soteriou, G. A. (2016). Asynchronous ripening behavior of cactus pear (*Opuntia ficus-indica*) cultivars with respect to physicochemical and physiological attributes. *Food Chem.* 211, 598–607. doi: 10.1016/j.foodchem.2016.05.113

Kyriacou, M. C., Ioannidou, S., Nikoloudakis, N., Seraphides, N., Papayiannis, L. C., and Kyratzis, A. C. (2020). Physicochemical characterization and trait stability in a genetically diverse ex situ collection of pomegranate (*Punica granatum* L.) germplasm from Cyprus. *Sci. Hortic.* 263:109116. doi: 10.1016/j.scienta.2019.109116

La Malfa, S., Currò, S., Bugeja Douglas, A., Brugaletta, M., Caruso, M., and Gentile, A. (2014). Genetic diversity revealed by EST-SSR markers in carob tree (*Ceratonia siliqua* L.). *Biochem. Syst. Ecol.* 55, 205–211. doi: 10.1016/j.bse.2014.03.022

Nahla, A. A. (2014). Relationships among phenotypic, chemical and genetic characteristics of some selected and evaluated carob strains (*Ceratonia siliqua*). *Int. J. Plant Soil Sci.* 3, 1415–1427. doi: 10.9734/IJPSS/2014/10519

Nasar-Abbas, S. M., Huma, Z., Vu, T.-H., Khan, M. K., Esbenshade, H., and Jayasena, V. (2016). Carob kibble: a bioactive-rich food ingredient. *Compr. Rev. Food Sci. Food Saf.* 15, 63–72. doi: 10.1111/1541-4337.12177

Nasrabadi, M., Ramezanian, A., Eshghi, S., Kamgar-Haghighi, A. A., Vazifeshenas, M. R., and Valero, D. (2019). Biochemical changes and winter hardiness in pomegranate (*Punica granatum* L.) trees grown under deficit irrigation. *Sci. Hortic.* 251, 39–47. doi: 10.1016/j.scienta.2019.03.005

Orphanos, P. (1980). Practical aspects of carob cultivation in Cyprus. *Port. Acta Biol.* XVI, 221–228.

Orphanos, P., and Papaconstantinou, J. (1969). *Carob Varieties of Cyprus*. Nicosia: Cyprus Agricultural Research Institute.

Othmen, K. B., Elfalleh, W., Lachiheb, B., and Haddad, M. (2019). Evolution of phytochemical and antioxidant activity of Tunisian carob (*Ceratonia siliqua* L.) pods during maturation. *Eurobiotech J.* 3, 135–142. doi: 10.2478/ebtj-2019-0016

Oziyci, H. R., Tetik, N., Turhan, I., Yatmaz, E., Ucgun, K., Akgul, H., et al. (2014). Mineral composition of pods and seeds of wild and grafted carob (*Ceratonia siliqua* L.) fruits. *Sci. Hortic.* 167, 149–152. doi: 10.1016/j.scienta.2014.01.005

Peakall, R. O. D., 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: 10.1111/j.1471-8286.2005.01155.x

Perrier, X. (2006). *DARwin Software*. Available online at: http://darwin.cirad.fr/darwin

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–952.

Rasheed, D. M., El-Kersh, D. M., and Farag, M. A. (2019). "*Ceratonia siliqua* (carob-locust bean) outgoing and potential trends of phytochemical, economic and medicinal merits," in *Wild Fruits: Composition, Nutritional Value and Products*, ed. A. A. Mariod (Cham: Springer International Publishing), 481–498. doi: 10.1007/978-3-030-31885-7_36

Rosenstock, T. S., Rosa, U. A., Plant, R. E., and Brown, P. H. (2010). A reevaluation of alternate bearing in pistachio. *Sci. Hortic.* 124, 149–152. doi: 10.1016/j.scienta.2009.12.007

Rouphael, Y., Colla, G., Giordano, M., El-Nakhel, C., Kyriacou, M. C., and De Pascale, S. (2017). Foliar applications of a legume-derived protein hydrolysate elicit dose-dependent increases of growth, leaf mineral composition, yield and fruit quality in two greenhouse tomato cultivars. *Sci. Hortic.* 226, 353–360. doi: 10.1016/j.scienta.2017.09.007

Sepperer, T., Hernandez-Ramos, F., Labidi, J., Oostingh, G. J., Bogner, B., Petutschnigg, A., et al. (2019). Purification of industrial tannin extract through simple solid-liquid extractions. *Ind. Crops. Prod.* 139:111502. doi: 10.1016/j.indcrop.2019.111502

Singleton, V. L., Orthofer, R., and Lamuela-Raventós, R. M. (1999). Analysis of total phenols and other oxidation substrates and antioxidants by means of folin-ciocalteu reagent. *Methods Enzymol.* 299, 152–178. doi: 10.1016/s0076-6879(99)99017-1

Stavrou, I. J., Christou, A., and Kapnissi-Christodoulou, C. P. (2018). Polyphenols in carobs: a review on their composition, antioxidant capacity and cytotoxic effects, and health impact. *Food Chem.* 269, 355–374. doi: 10.1016/j.foodchem.2018.06.152

Sun, B., Ricardo-da-Silva, J M., and Spranger, I. (1998). Critical factors of vanillin assay for catechins and proanthocyanidins. *J. Agric. Food Chem.* 46, 4267–4274. doi: 10.1021/jf980366j

Tavarini, S., Gil, M. I., Tomas-Barberan, F. A., Buendia, B., Remorini, D., Massai, R., et al. (2011). Effects of water stress and rootstocks on fruit phenolic composition and physical/chemical quality in Suncrest peach. *Ann. Appl. Biol.* 158, 226–233. doi: 10.1111/j.1744-7348.2010.00457.x

Tetik, N., Turhan, I., Oziyci, H. R., Gubbuk, H., Karhan, M., and Ercisli, S. (2011). Physical and chemical characterization of *Ceratonia siliqua* L. germplasm in Turkey. *Sci. Hortic.* 129, 583–589. doi: 10.1016/j.scienta.2011.04.029

Theophilou, I. C., Neophytou, C. M., and Constantinou, A. I. (2017). Carob and its components in the management of gastrointestinal disorders. *J. Hepatol. Gastroenterol.* 1:005.

Ticho, R. J. (1959). *Report to the Government of Cyprus on Carob Production*. Rome: FAO.

Tous, J., Romero, A., and Batlle, I. (2013). The Carob tree: botany, horticulture, and genetic resources. *Hortic. Rev.* 41, 385–456. doi: 10.1002/9781118707418.ch08

Villa, T. C. C., Maxted, N., Scholten, M., and Ford-Lloyd, B. (2005). Defining and identifying crop landraces. *Plant Genet. Resour.* 3, 373–384. doi: 10.1079/PGR200591

Viruel, J., Haguenauer, A., Juin, M., Mirleau, F., Bouteiller, D., Boudagher-Kharrat, M., et al. (2018). Advances in genotyping microsatellite markers through sequencing and consequences of scoring methods for *Ceratonia siliqua* (Leguminosae). *Appl. Plant Sci.* 6:e01201. doi: 10.1002/aps3.1201

Viruel, J., Médail, F., Juin, M., Haguenauer, A., Feliner, G. N., Bou Dagher Kharrat, M., et al. (2016). "Mediterranean carob populations, native or naturalized? A continuing riddle," in *Proceedings of the OPTIMA MEETING ANR DYNAMIC - Deciphering Symbiotic Networks in Carob-Based Mediterranean Agro-Ecosystems*, Montpellier, 169–169.

Von Haselberg, C. (1996). "Factors influencing flower and fruit development in carob (*Ceratonia siliqua* L.)," in *Proceedings of the Third International Carob Symposium*, (Tavira: University of Lisbon).

Yousif, A. K., and Alghzawi, H. (2000). Processing and characterization of carob powder. *Food Chem.* 69, 283–287. doi: 10.1016/S0308-8146(99)00265-4

Zohary, D. (2002). Domestication of the carob (*Ceratonia siliqua* L.). *Isr. J. Plant Sci.* 50, 141–145. doi: 10.1560/BW6B-4M9P-U2UA-C6NN

Zunft, H., Lüder, W., Harde, A., Haber, B., Graubaum, H. J., Koebnick, C., et al. (2003). Carob pulp preparation rich in insoluble fibre lowers total and LDL cholesterol in hypercholesterolemic patients. *Eur. J. Nutr.* 42, 235–242. doi: 10.1007/s00394-003-0438-y

Conflict of Interest: 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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 23 February 2021
doi: 10.3389/fpls.2021.626811

Construction of a High-Density Genetic Map of *Acca sellowiana* (Berg.) Burret, an Outcrossing Species, Based on Two Connected Mapping Populations

*Marianella Quezada[1], Rodrigo Rampazo Amadeu[2], Beatriz Vignale[3], Danilo Cabrera[4], Clara Pritsch[1] and Antonio Augusto Franco Garcia[2]**

[1] Laboratorio de Biotecnología, Departamento de Biología Vegetal, Facultad de Agronomía, Universidad de la República, Montevideo, Uruguay, [2] Laboratório de Genética Estatística, Departamento de Genética, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, Brazil, [3] Mejoramiento Genético, Departamento de Producción Vegetal, Estación Experimental de la Facultad de Agronomía, Universidad de la República, Salto, Uruguay, [4] Programa de Investigación en Producción Fruticola, Instituto Nacional de Investigación Agropecuaria (INIA), Estación Experimental "Wilson Ferreira Aldunate", Canelones, Uruguay

OPEN ACCESS

Edited by:
Maria Luisa Badenes, Instituto Valenciano de Investigaciones Agrarias, Spain

Reviewed by:
Shouvik Das, Indian Agricultural Research Institute (ICAR), India
Jonathan Elias Maldonado, Pontificia Universidad Católica de Chile, Chile

***Correspondence:**
Antonio Augusto Franco Garcia
augusto.garcia@usp.br

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

Received: 06 November 2020
Accepted: 12 January 2021
Published: 23 February 2021

Citation:
Quezada M, Amadeu RR, Vignale B, Cabrera D, Pritsch C and Garcia AAF (2021) Construction of a High-Density Genetic Map of Acca sellowiana (Berg.) Burret, an Outcrossing Species, Based on Two Connected Mapping Populations. Front. Plant Sci. 12:626811. doi: 10.3389/fpls.2021.626811

Acca sellowiana, known as feijoa or pineapple guava, is a diploid, ($2n = 2x = 22$) outcrossing fruit tree species native to Uruguay and Brazil. The species stands out for its highly aromatic fruits, with nutraceutical and therapeutic value. Despite its promising agronomical value, genetic studies on this species are limited. Linkage genetic maps are valuable tools for genetic and genomic studies, and constitute essential tools in breeding programs to support the development of molecular breeding strategies. A high-density composite genetic linkage map of *A. sellowiana* was constructed using two genetically connected populations: H5 (TCO × BR, $N = 160$) and H6 (TCO × DP, $N = 184$). Genotyping by sequencing (GBS) approach was successfully applied for developing single nucleotide polymorphism (SNP) markers. A total of 4,921 SNP markers were identified using the reference genome of the closely related species *Eucalyptus grandis*, whereas other 4,656 SNPs were discovered using a *de novo* pipeline. The individual H5 and H6 maps comprised 1,236 and 1,302 markers distributed over the expected 11 linkage groups, respectively. These two maps spanned a map length of 1,593 and 1,572 cM, with an average inter-marker distance of 1.29 and 1.21 cM, respectively. A large proportion of markers were common to both maps and showed a high degree of collinearity. The composite map consisted of 1,897 SNPs markers with a total map length of 1,314 cM and an average inter-marker distance of 0.69. A novel approach for the construction of composite maps where the meiosis information of individuals of two connected populations is captured in a single estimator is described. A high-density, accurate composite map based on a consensus ordering of markers provides a valuable contribution for future genetic research and breeding efforts in *A. sellowiana*. A novel mapping approach based on an estimation of multipopulation recombination fraction described here may be applied in the construction of dense composite genetic maps for any other outcrossing diploid species.

Keywords: *Acca sellowiana*, feijoa, pineapple guava, genotyping by sequencing, composite genetic map, multiparent family, Myrtaceae

1. INTRODUCTION

Acca sellowiana (Berg.) Burret, commonly known as feijoa or pineapple guava, is an agronomically promising fruit tree species, native to Uruguay and southern Brazil. This diploid ($2n = 2x = 22$) outcrossing species presents a small genome (245 Mb) (da Costa et al., 2008) and the basic haploid number of $n = 11$, largely conserved in the Myrtaceae family (Grattapaglia et al., 2012). The species stands out in the novel market of health-promoting and functional food due to the nutraceutical value of its fruits, rich in vitamin C (28 mg/100 g), hydrocarbons, minerals, iodine (3 mg/100 g), and bioflavonoids (Weston, 2010). However, feijoa can still be considered a minor crop due to its seasonal presence on the market and a small cultivation area worldwide. The development of new cultivars with superior fruit quality traits and adapted for new environments is needed to accelerate the commercial exploitation of the species. For fruit tree species, like feijoa, the breeding process is slow and costly because of the long juvenile period, extensive phenotyping cost, and limited field space (Byrne, 2012). For this reason, the development of genetic and genomic resources, such as genetic maps, represents a suitable strategy to overcome these limitations and speed up the breeding progress.

As a minor crop, genomic resources in *A. sellowiana* are limited and not yet useful in breeding programs. Few studies document the molecular diversity of the species using low-throughput molecular markers (Dettori and Palombi, 2000; dos Santos et al., 2007; Pasquariello et al., 2015; Donazzolo et al., 2020; Saifert et al., 2020) and a limited number of molecular markers have been specifically designed for *A. sellowiana* (dos Santos et al., 2008; Klabunde et al., 2014). Amplified fragment length polymorphisms (AFLP), intersimple sequence repeat (ISSR), and a small number of simple sequence repeats (SSR) markers were used in the construction on the first genetic map of the species (Quezada et al., 2014). This map along with those developed for the species *Psidium guajava* (Padmakar et al., 2015) represent the only ones developed for fruit species within the Myrtaceae family. With the advance of new sequencing technologies, single nucleotide polymorphisms (SNPs) have become the most useful type of marker for genetic analysis. The next-generation sequencing (NGS) technology coupled with enzyme-based complexity reduction and DNA barcoding have been used to simultaneously discover and genotype a large number of SNPs in a single experiment (Baird et al., 2008; Elshire et al., 2011). This low cost strategy, implemented in protocols such as genotyping by sequencing (GBS) (Elshire et al., 2011), provides a fast, efficient, and cost-effective strategy to obtain a significant number of markers. It represent an invaluable asset for minor or underutilized crops, such as *A. sellowiana*, with relatively few genomic and genetic resources. Using SNPs from GBS, high-density linkage maps for many commercial fruit tree species (Barba et al., 2014; Gardner et al., 2014; Bielenberg et al., 2015) as well as for minor crops have been developed (Ward et al., 2013; Covarrubias-Pazaran et al., 2016).

High-density genetic maps are extremely valuable tools to investigate the composition and organization of genomes for comparative genetic mapping analysis, chromosome-based genome assembly, physical and genetic map integration, and candidate gene/QTL cloning (Bartholomé et al., 2014; Mathew et al., 2014; Fierst, 2015; Guajardo et al., 2015; Li et al., 2017; Jaganathan et al., 2020). Genetic maps are also useful tools in breeding programs, fundamentally to establish associations between molecular markers and agronomical traits, providing the basis for future strategies of QTL identification and marker-assisted selection (MAS) (Troggio et al., 2012). For outcrossing species, genetic maps are typically based on single full-sib mapping populations derived from crosses of two highly heterozygous parents. The pseudo-test cross strategy was first developed to construct genetic maps in these populations. In this strategy, markers heterozygous only in one parent, therefore segregating in a 1:1 ratio, are used to generate two separate individual linkage maps (Grattapaglia and Sederoff, 1994). Later, methods using information from all markers simultaneously (considering dominant and codominant markers, heterozygous either in one or both parents, with segregation ratios 1:1:1:1, 3:1, 1:2:1, and 1:1) were developed to construct integrated genetic linkage maps (Maliepaard et al., 1997; Wu et al., 2002a,b). The OneMap R (Margarido et al., 2007) package implements the maximum likelihood approach to simultaneously estimate linkage and linkage phases between markers; in addition, it provides a multipoint estimation for the recombination fraction using a hidden Markov model (HMM) (Wu et al., 2002b), which is a reliable procedure (Mollinari et al., 2009). This approach was successfully applied in the construction of integrated genetic maps for several outcrossing species (Palhares et al., 2012; Pereira et al., 2013; Quezada et al., 2014; Balsalobre et al., 2017).

The construction of accurate genetic maps faces two main difficulties. First, it is necessary to have a high number of polymorphic markers to obtain a comprehensive coverage of the genome. Second, large mapping populations need to be developed to accurately estimate genetic distances. Presently, developing a large number of markers is not a limiting factor, even for species with few genetic and genomic resources (Davey et al., 2011). However, most of the mapping populations of fruit trees that have been developed in breeding programs have a small to medium progeny size because of the high cost of maintaining these populations and the limited orchard space (Peace and Norelli, 2009). Capitalizing on existing breeding-mapping populations provides an opportunity to develop genetic maps including information of multiple populations, and to overcome the limitations of a reduced number of recombination events captured in single mapping populations. The integration of single-population genetic maps into a composite map improves the accuracy and resolution of maps by correcting the order and position of markers, increasing genome coverage, and filling out genomic regions lacking polymorphic markers in specific crosses (Khan et al., 2012; Pootakham et al., 2015). Moreover, composite genetic maps enable a more precise estimation of QTL effects and positions (Di Pierro et al., 2016) as well as facilitates the transfer of marker information and genetic predictions between populations (Kuhn et al., 2017). Many strategies have been proposed to construct consensus composite maps applied to outcrossing species (Khan et al., 2012;

Pootakham et al., 2015; Bodénès et al., 2016; Schlautman et al., 2017). In some cases, the recombination fraction is estimated combining all genotypic data sets from multiple populations (de Givry et al., 2005; Van Ooijen, 2006), whereas in others, map integration is achieved based on the marker order and distances from individual maps (Wu et al., 2011; Endelman and Plomion, 2014). The inability to solve inconsistencies or ordering conflicts between maps as well as inflated genetic distances can be considered the main limitations in both approaches (Khan et al., 2012).

The objective of this study was to develop a reference and composite linkage genetic map for *A. sellowiana*, integrating the genetic information from two connected F_1 populations. For this purpose, two mapping populations segregating for fruit quality traits with the same female parent were used. For the first time in the species, a high-throughput GBS approach was successfully applied to identify SNP markers distributed throughout the genome. A new statistical model was implemented into OneMap (Margarido et al., 2007) and used to build a composite map using information from multiparental outcrossing populations. The two saturated individual genetic maps and the composite one provide a framework for future genetic and genomic studies, and they will be useful for future studies of marker-trait association and QTL mapping, which in turn will speed up the breeding process in *A. sellowiana*.

2. MATERIALS AND METHODS

2.1. Mapping Populations

Two full-sib F_1 populations genetically connected by a common parental genotype and segregating for relevant fruit quality traits were used to construct an integrated composite map. These populations were generated by the crossing of non-inbred clones TCO × BR (H5 population) and TCO × DP (H6 population), and consist of 160 and 184 genotypes, respectively. TCO was used as female parent in both crosses. The three parental genotypes were chosen on the basis of their agronomic features: TCO, collected from the wild, has small fruits with tasty and smooth pulp, as well as a thin light-green skin; BR, collected from a commercial orchard, has large fruits with tasteless pulp and a rough, thick, dark-green skin; DP, collected from a commercial orchard has a high yield of medium size fruits, and a rough, medium-thick, dark-green skin. The mapping populations were developed in 2008 by the Native Fruits Breeding Program (Universidad de la República (UdelaR) and Instituto Nacional de Investigación Agropecuaria (INIA), Uruguay). H5 population was planted in the Experimental Station Salto of Facultad de Agronomía-UdelaR, Uruguay (31°19′ S, 57°41′ W), whereas H6 population was planted in the Experimental Station Salto Grande of INIA, Uruguay (31°25′ S, 57°37′ W).

2.2. DNA Extraction and Library Construction

Leaf material was collected from a total of 344 individuals, and maintained in paper bags with silica gel at −20° until DNA extraction. Genomic DNA was extracted using DNeasy Plant Kit (Qiagen, Germantown, MD, USA) according to the manufacturer's protocol. DNA quality was checked by electrophoresis on 0.8% agarose gel and quantified by NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). GBS libraries were prepared according to the original GBS protocol (Elshire et al., 2011) in the Institute for Genomic Diversity (Cornell University, Ithaca, NY, USA). For optimization of the GBS protocol, three different restriction enzymes were compared. Test libraries were prepared separately with the five-cutter *Ape*KI, and six-cutter *Eco*T22I and *Pst*I. The fragment size distribution for each test library was evaluated using an Experion system (Bio-Rad, Hercules, CA, USA). Both six-cutter enzymes produced the smallest fragment pool (adequate for higher SNP sequencing coverage), but the *Eco*T22I library contained a detectable amount of repetitive DNA. Thus, the library derived from *Pst*I was selected as it comprised the largest number of fragments minor to 500 bp, and presented a smooth profile for fragment size distribution indicative of very low content of repetitive DNA. DNA samples and blank negative controls were prepared in two 192-plex libraries and were sequenced (single-end reads of 100 bp in length) twice in a HiSeq2000 sequencer (Illumina R Inc., San Diego, CA, USA). To guarantee a satisfactory coverage and confidence in calling heterozygous genotypes, the three parental genotypes had 12 samples each, and all the samples reached a coverage equivalent to 96-plex.

2.3. SNP and Genotype Calling

The quality of the raw sequence (per base sequence quality, average read quality score, per base N content) was evaluated using the FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). For SNP calling, two strategies implemented in the Tassel GBS software were employed comprising the Discovery pipeline using a reference genome (Tassel v4.0) as well as the UNEAK network pipeline (Tassel v3.0) (Lu et al., 2013; Glaubitz et al., 2014).

Raw sequences were filtered, discarding reads lacking the barcode, not having the expected *Pst*I cut-site, or containing uncalled bases (i.e., Ns) within the first 64 bp subsequent to the barcode. The barcodes was removed in the raw sequences and the remainder of the sequences trimmed to 64 bases. Filtered reads were grouped in tags, and only those with a minimum coverage depth of 5 were retained (parameter -c 5 at the MergeTaxaTagCountPlugin step). For the Discovery pipeline, tags were aligned to the *E. grandis* reference genome (*E. grandis* v1.0, JGI, http://www.phytozome.net) using Bowtie v1.2.0 (Langmead and Salzberg, 2012) with default parameters. The *E. grandis* genome stands out as the reference genome for the Myrtaceae family (Grattapaglia et al., 2012).

In the UNEAK strategy, tags were aligned to each other and only a single base-pair mismatch was considered. Due to the presence of repeated or paralogous sequences in the genome, many tag networks are generated. These networks were filtered (error tolerance rate set at 0.03) and only reciprocal tags were maintained representing potential SNPs. Discovery and UNEAK data were exported as *vcf* files

(information-rich variant call format) where the read depth of the reference and alternative allele is saved (Danecek et al., 2011). SNP markers identified by the Discovery strategy were denoted as "Eg" followed by the chromosome number and the position (bp) on the *E. grandis* reference genome. SNPs identified using the UNEAK pipeline were named as "Un1" plus a number referring to the tags network. Redundant markers found in both Discovery and UNEAK pipelines were removed from UNEAK data set, and assigned both Discovery and UNEAK identification (e.g., **Eg8_53239405:*Un1_5092215*). Consequently, when total number of mapped markers is reported, redundant markers are counted once.

A quantitative genotype calling was performed for each population separately using SUPERMASSA software (Serang et al., 2012). For genotype calling, this software considers the expected allele distribution in an F_1 population, as well as the relative site coverage (read counts) of each allele (Garcia et al., 2013). In a previous step, markers were filtered out for quality; only biallelic markers with less than 25% of missing data were retained. The F_1 segregation model with diploid level was fitted in SUPERMASSA, where each of the 12 replicates of parental genotypes were analyzed separately, to provide additional constraints during estimation (Pereira et al., 2018). Following the recommendations of (Mollinari and Serang, 2015), SUPERMASSA naive posterior report threshold was set to zero and individual posterior probability values were estimated. The median of all individual posterior probabilities was used as a quality control, so only SNPs with posterior values higher than 0.8 were selected. The SUPERMASSA *vcf* file output was formatted for mapping purposes using VCFR (Knaus and Grünwald, 2017) and ONEMAP (Margarido et al., 2007) R packages (R Development Core Team, 2017), following the instructions given in the ONEMAP software tutorial (http ://augustogarcia.me/onemap/vignettes_highres/Outcrossing_ Populations.html#importing-data). Following the notation of (Wu et al., 2002a), markers were classified into four groups ("A-D") according to their respective cross type. For GBS-SNP markers in an F_1 population, only three marker cross types are informative for genetic map construction: "B3.7" ("ab" × "ab"), "D1.10" ("ab" × "aa"), and "D2.15" ("aa" × "ab"). The "D" group comprises markers segregating in a 1:1 ratio, i.e., heterozygous in only one parent, also called testcross markers. The "B3.7" markers are heterozygous and symmetric in both parents, with an expected 1:2:1 segregation ratio. In addition, for markers segregating on both populations, five new configuration types were defined, considering the simultaneous segregation pattern. If the common maternal genotype is heterozygous for a locus ("ab"), the paternal genotype could be heterozygous in both populations ("ab" × "ab"/"ab" × "ab", "B3.7-B3.7"), only heterozygous in H5 population ("ab" × "ab"/"ab" × "aa," "B3.7"-"D1.10"), only heterozygous in H6 population ("ab" × "aa"/"ab" × "ab," "D1.10"-"B3.7") or homozygous in both populations ("ab" × "aa"/"ab" × "aa," "D1.10"-"D1.10"). For markers homozygous in the female parent, the only possible configuration is that the paternal genotypes were heterozygous in both populations ("aa" × "ab"/"aa" × "ab," "D2.15"-"D2.15"), or otherwise there is no information to integrate the populations.

2.4. Construction of Maps for Individual Populations

For each population, a data set including the SNPs identified by Discovery and UNEAK pipelines was generated. For H5 population, 493 previously identified markers (ISSR, AFLP, and SSR) were also included (Quezada et al., 2014). Integrated genetic maps for H5 and H6 populations were constructed using ONEMAP (Margarido et al., 2007) R package (current version available at https://github.com/augusto-garcia/onemap). This software implement the method proposed by Wu et al. (2002a,b) for simultaneous multipoint estimation of the recombination fraction and linkage phases between markers. Using a multipoint approach, that uses information from multiple markers to estimate the recombination fractions, the most adjacent informative markers fill the lack of information in markers with missing data. Considering that both populations have a comparable number of individuals and genotyped markers, the same procedure was applied to construct both maps.

First, the segregation of each marker was evaluated for goodness-of-fit to the expected Mendelian segregation ratio by a chi-square test, followed by a Bonferroni correction for multiple testing considering 0.05 the overall significance level. Two-point linkage analysis was carried out between all pairs of markers using the RF_2PTS command. Linkage groups (LGs) were established with a logarithm of the odds (LOD) score of 7.5 (estimated value based on the number of two-point tests and multiple test correction) and a maximum recombination fraction of 0.35. Recombination frequency was converted to genetic map distances (cM) using the Kosambi function (Kosambi, 1944). To order groups, the ORDER_SEQ function was used; this selects a subset of informative markers (6–7 markers), wherein the remaining markers were introduced. A framework map was evaluated, and markers exhibiting a suspect position (ordering problems), producing gaps at the ends of the groups or inflating LGs size were removed. Once the framework map presents a reliable initial order, the TRY_SEQ algorithm (Lander and Green, 1987) was used to integrate previously removed markers and also markers that remained unmapped. Incorrect allocation of markers was visually inspected using a heatmap plot (graphical representation of the recombination fraction and LOD score between markers). Considering that the GBS-SNP markers could have genotyping errors, a probability of 0.05 (default = 0.01) of error was considered to construct the genetic maps. The RIPPLE algorithm was used to verify alternative local orders (Lander and Green, 1987). LGs were numbered according to the chromosome number of *E. grandis* based on the location of markers identified using this genome as reference. Finally, the linkage maps were drawn using the MAPCHART V2.3 software (Voorrips, 2002).

2.5. Construction of a Composite Integrated Genetic Map

A new approach to construct a composite genetic map using the information of the two connected populations was developed.

The proposed model assumes a homogeneous recombination fraction between populations, which is estimated using the segregation data for both populations simultaneously. Therefore, this method will have a higher statistical power, resulting in more accurate estimates of markers distances and order.

To build a composite genetic map, H5 and H6 individual integrated genetic maps were compared to identify common markers. These markers were identified by name, considering that SNP calling was performed on the whole data set. For these markers, the location in the same LG as well as the marker order was registered in both H5 and H6 maps. A new data set for each LG was created, containing information only of the shared markers presented in homologous LGs.

The two-point recombination fraction between common markers in each LG was estimated using the segregation data of the two populations simultaneously. The maximum likelihood estimator of the recombination fractions and LOD score formulas were adapted from Maliepaard et al. (1997), only for the marker configurations present in the data. To estimate recombination fractions in F_1 populations, estimates of linkage phase between markers is necessary. Considering two consecutive markers in a homologous chromosome, the alleles of a pair of loci can be present at coupling (C) or repulsion (R) phase configuration. Because in our populations the same female parental genotype is shared, the linkage phase of markers on this parent was equal in both crosses. Thus, considering simultaneously the crosses "TCO × BR" and "TCO × DP," eight linkage phase assignments were evaluated (*f* and *m* indicated female and male parent, respectively): (i) C*f*C*m*-C*f*C*m*, (ii) C*f*C*m*-C*f*R*m*, (iii) C*f*R*m*-C*f*C*m*, (iv) C*f*R*m*-C*f*R*m*, (v) R*f*C*m*-R*f*C*m*, (vi) R*f*C*m*-R*f*R*m*, (vii) R*f*R*m*-R*f*C*m*, and (viii) R*f*R*m*-R*f*R*m*. An illustration of the recombination fraction estimation between markers with different configuration type is provided in **Supplementary Data Sheet 1**.

A multipoint approach was implemented using an HMM (Lander and Green, 1987) with the expectation maximization (EM) algorithm (Dempster et al., 1977). To estimate the recombination fraction, and due to the meiosis in one population is independent to the meiosis of the another population, the expectation step was computed on each population separately and the maximization step was performed considering segregation information for both populations simultaneously. The new R function RF_2POPS implement this (**Supplementary Data Sheet 1**).

To implement the multipoint approach, the marker order of the individual H5 and H6 maps was tested. The marker order with the highest likelihood was selected as initial order of the composite map. Then, all the markers with ordering conflict between the individual maps were evaluated in all possible positions on the composite map and the position with highest likelihood was selected. A final local order verification was performed with RIPPLE algorithm implemented in the RIPPLE_2POPS command. The final marker order of the composite map was used to individual H5 and H6 maps to facilitate comparisons.

Finally, the markers present only in H5 or H6 individual maps were incorporated to the composite map. For these markers, we only had recombination information from one population, so it was not possible to estimate a multipopulation recombination fraction. To include these unique markers, genetic distances in cM were re-estimated by multiplying by a specific scaling factor. In the case of unique markers between consecutive common markers, the scaling factor was calculated by dividing the genetic distance in the composite map by the distance in the individual map for the same interval. For markers at the end of LGs, the total map length of the composite map divided by the total map length of individual map for each LG was used.

To perform the analysis, new functions of the ONEMAP package were developed to handle information for multiple populations. The functions developed in this work (RF_2POPS and RIPPLE_2POPS) alongside with a user-friendly tutorial (**Supplementary Data Sheet 2**) are available at https://github.com/augusto-garcia/onemap2pop. They will be integrated into ONEMAP package in a near future.

3. RESULTS

3.1. GBS Libraries

A total of 859,454,459 reads of 100 bp length were obtained from the complete sequencing of the two libraries. Read number and sequence quality were shown to be comparable for the two libraries and the two sequencing processes using FastQC. Read quality was similarly good in both libraries, with an average base quality score higher than 30 (99.9% base call accuracy). The total number of reads per progeny ranged from 60 thousand to 6.2 M. The 12 fold-repeated samples of parental genotypes resulted in a deeper coverage, with a 3.2 M average number of reads for the parental genotypes and 2.2 M per progeny. Two samples (H5_095 and H6_156) that presented a relatively low numbers of reads ($< 150{,}000$), representing less of the 10% of the mean reads per sample of the lane on which they were sequenced, were removed for subsequent analysis.

3.2. SNP Calling

As a first step, 3,051,475 high-quality tags (unique sequence from one or more high-quality reads) were identified. These GBS tags were aligned to the *E. grandis* reference genome in the Discovery pipeline. Of these, 643,863 (21.10%) aligned exactly once in the reference genome and 280,937 (9.21%) aligned in multiple positions, resulting in an overall alignment rate of 30.31% (924,800 tags). As a result of the Discovery approach, 43,377 putative SNPs were identified. **Figure 1** presents the SNP markers distribution across the *E. grandis* chromosomes. Using the UNEAK pipeline, 44,889 high-quality pairwise alignments were identified, of which 13,430 putative SNPs were obtained.

3.3. Genotype Calling

A total of 56,807 SNPs, comprising both Discovery and UNEAK SNP data sets, were available for genotype calling (**Table 1**). This number lowered to 53,051 (93.38%) after filtering out markers with no identified variation (missing alternative allele) or triallelic markers per genotype. After removing markers with more than 25% of missing data, a total of 27,595 (48.58%)

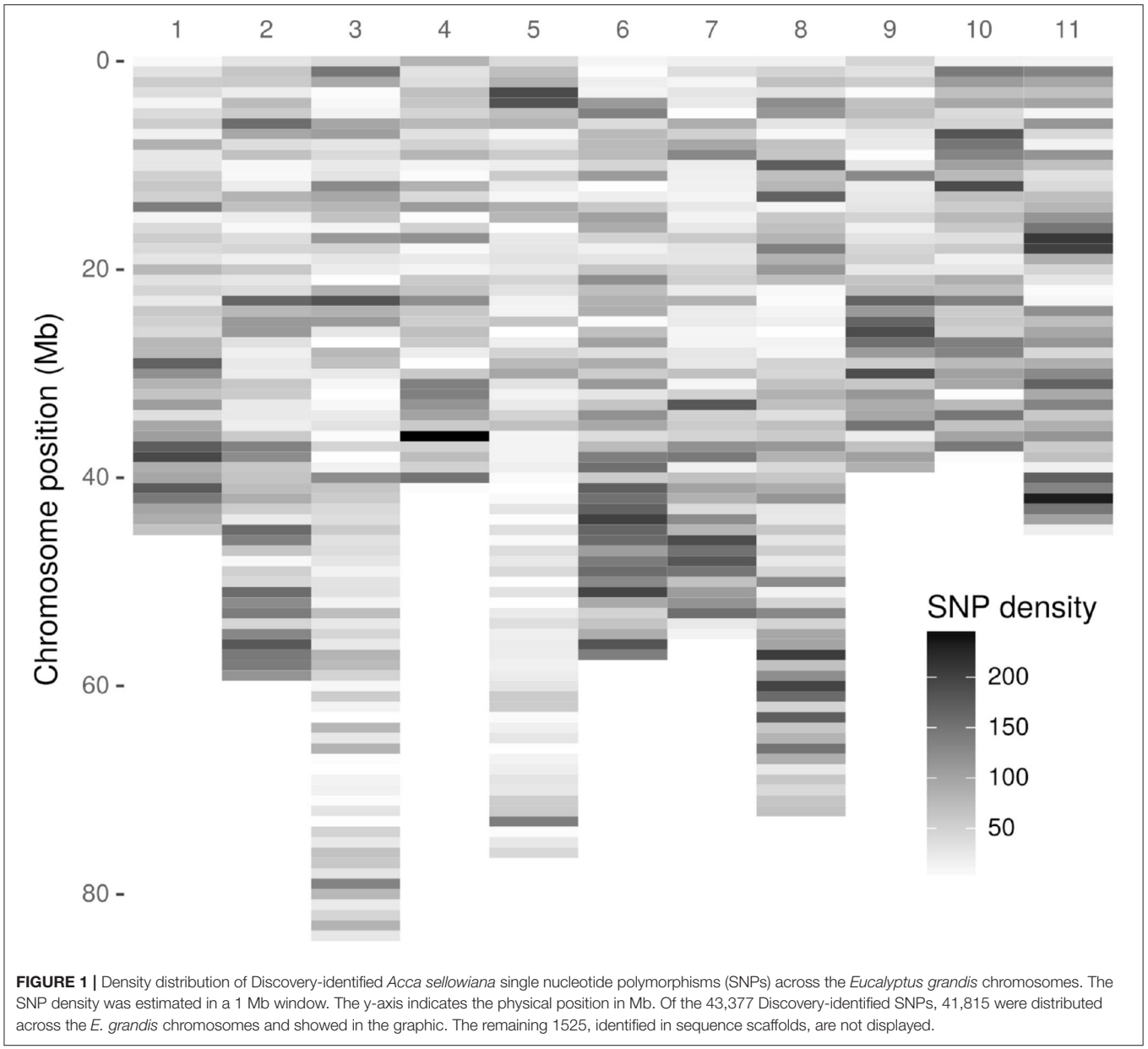

FIGURE 1 | Density distribution of Discovery-identified *Acca sellowiana* single nucleotide polymorphisms (SNPs) across the *Eucalyptus grandis* chromosomes. The SNP density was estimated in a 1 Mb window. The y-axis indicates the physical position in Mb. Of the 43,377 Discovery-identified SNPs, 41,815 were distributed across the *E. grandis* chromosomes and showed in the graphic. The remaining 1525, identified in sequence scaffolds, are not displayed.

and 25,670 (45.19%) SNPs were retained for H5 and H6 populations, respectively. For genotype calling, the filtered markers were subjected to quantitative genotyping following an F_1 model implemented in SUPERMASSA. Markers with the median of all individual posterior probabilities greater than 0.8 were selected. This criterion was chosen to ensure a high quality of genotypes as described by Garcia et al. (2013) and Mollinari and Serang (2015). Finally, a low number of overlapping markers between both Discovery and UNEAK SNP data sets was found for both H5 (174 SNPs) and H6 (88 markers) populations, and removed for the UNEAK data set. As a result, a set of 5350 (9.42%) and 4227 (7.44%) of high-quality SNPs was used for genetic mapping for H5 and H6 populations, respectively. Of the H5 population SNP set, 2875 (53.74%) markers were identified by the Discovery pipeline, so the remaining 2475 (42.26%) markers by the UNEAK pipeline. For the H6 population, 2046 (48.40%) markers and 2181 (51.60%) were identified using the Discovery and UNEAK approaches, respectively.

Considering the marker segregation patterns, SNP markers were classified into one of the three configuration marker types informative for genetic map construction: "B3.7" ("ab" × "ab"); "D1.10" ("ab" × "aa"); and "D2.15" ("aa" × "ab") (Wu et al., 2002a). Of the 5350 segregating markers identified in H5 population, 1520 (28.41%), 1752 (32.75%), and 2088 (39.03%) were classified as "B3.7," "D1.10," and "D2.15," respectively. For H6 population, a set of 4227 markers were identified, and of those: 993 (23.49%), 1520 (35.96%), and 1714 (40.55%) were

"B3.7," "D1," and "D2," respectively. Segregation distortion from the Mendelian expected ratios (1:2:1 for "B3.7" markers and 1:1 for "D1.10" and "D2.10" markers) was observed in 1263 (23.61%) markers of the H5 data set and 1112 (26.31%) markers of H6 data set.

Comparing H5 and H6 SNP data sets, 2427 SNP were common. Of these shared markers, 1204 (49.61%) were identified by Discovery pipeline and 1223 (50.93%) by the UNEAK pipeline (**Figure 2A**). The segregation pattern of the common markers were inspected based on a joint marker configuration type. Markers with the same configuration type in both populations were the most frequent. For instance, the joint configuration D1.10-D1.10 accounted for the 36.88%, whereas the configuration B3.7-B3.7 and D2.15-D2.15 accounted for 25.34 and 21.59%, respectively. The B3.7-D1.10 and D1.10-B3.7 represented 8.94 and 7.25% of the configuration patterns for all the common markers, respectively (**Figure 2B**).

TABLE 1 | Summary of single nucleotide polymorphism (SNP) marker number after filtering process for H5 and H6 populations.

	Population H5		Population H6	
Pipeline	**Discovery**	**UNEAK**	**Discovery**	**UNEAK**
Initial markers[a]	43,377	13,430	43,377	13,430
Triallelics[b]	39,650 (91.41%)	13,401 (99.78%)	39,650 (91.41%)	13,401 (99.78%)
25% missing data[c]	23,113 (53.28%)	4482 (33.37%)	21,395 (49.32%)	4275 (31.83%)
Probability genotype > 0.8[d]	2875 (6.63%)	2475 (18.43%)[e]	2046 (4.72%)	2181 (16.24%)[e]
Total	5350		4227	

[a] *Initial number of SNP markers identified simultaneously in both populations.*
[b] *Triallelics markers or with missing alternative allele were excluded.*
[c] *SNP markers with > 25% missing data were excluded.*
[d] *SNP markers with a median posterior probability > 0.8 from* SuperMASSA *software were selected.*
[e] *Redundant markers between Discovery and UNEAK sets were removed from UNEAK list.*

3.4. H5 and H6 Genetic Map

The 5,350 segregating GBS-SNPs of H5 population were combined with 493 (100 ISSR, 386 AFLP, and 7 SSR) makers previously reported by (Quezada et al., 2014) for the same population. Therefore, 5,843 markers were combined for linkage analysis. The resultant H5 genetic map, derived from the "TCO × BR" cross, comprised 1,236 markers encompassing 11 LGs, corresponding to the haploid chromosome number of *A. sellowiana* (**Table 2**, **Figure 3**). The total length of the map was 1593 cM, with LG4 being the smallest group (97.14 cM) and LG1 the largest (204.43 cM). The number of markers per group ranged from 68 (LG5) to 164 (LG10) with a mean of 112.36 markers per LG.

For the construction of the H6 genetic map, corresponding to the "TCO × DP" cross, 4227 GBS-SNPs were analyzed. A total of 1,302 markers were mapped in the H6 map, spanning the expected 11 LGs for the species (**Table 2**, **Figure 3**). The map covered a total distance of 1,567 cM, with the size of LGs ranging from 111.49 cM (LG3) to 230.28 cM (LG1). The average number of markers per LG was 118.36, with a minimum number of 74 markers in LG5 and a maximum of 172 markers in LG8.

Similar average distances between markers was achieved in both H5 and H6 genetic maps. The average marker densities were

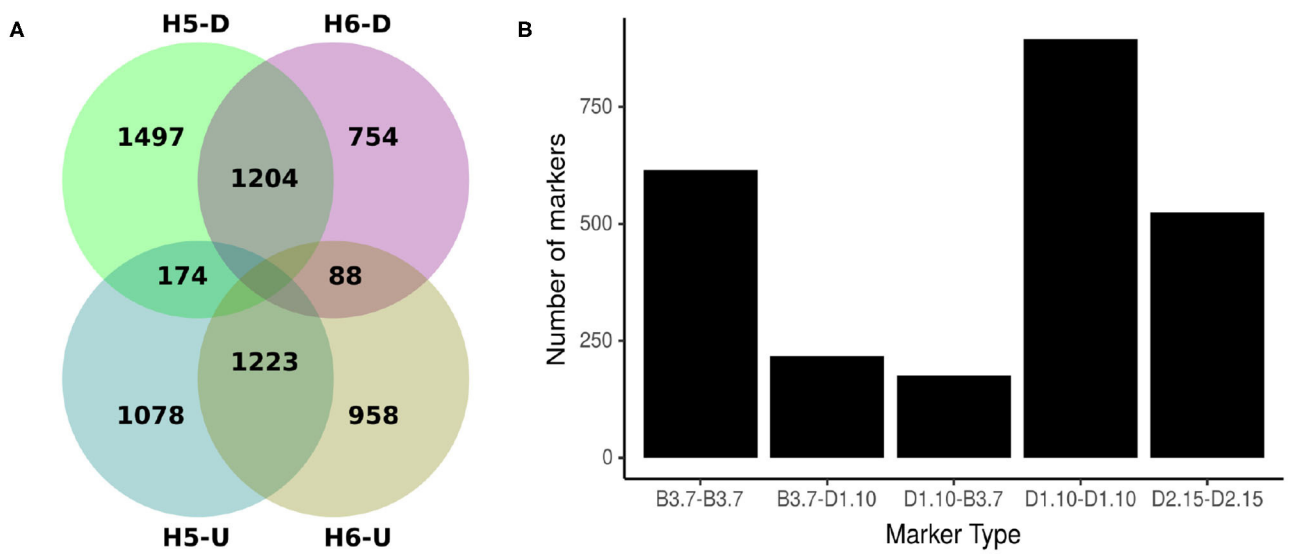

FIGURE 2 | Origin and types of marker configuration for GBS-SNP markers identified for H5 and H6 populations. **(A)** Venn diagram showing common and unique single nucleotide polymorphisms (SNPs) between H5 and H6 populations identified by Discovery (D) and UNEAK (U) pipelines. **(B)** Distribution of five joint types of marker configurations for common markers between H5 and H6 populations. The joint marker configuration types considered the marker types in H5 and H6 populations simultaneously.

TABLE 2 | Description of H5, H6, and composite linkage maps for *A. sellowiana*.

	H5 Map			H6 Map			Composite Map		
LG	**No. of Markers**	**Length (cM)**	**Marker density**	**No. of Markers**	**Length (cM)**	**Marker density**	**No. of Markers**	**Length (cM)**	**Marker density**
LG1	92	204.43	2.22	87	230.28	2.65	131	171.17	1.31
LG2	84	89.40	1.06	92	119.69	1.30	123	87.89	0.71
LG3	105	117.80	1.12	128	111.49	0.87	163	65.92	0.40
LG4	91	97.14	1.06	106	128.43	1.21	125	107.30	0.86
LG5	68	155.31	2.28	74	135.35	1.83	131	174.94	1.34
LG6	136	110.69	0.81	149	147.86	0.99	200	115.46	0.58
LG7	134	176.24	1.31	137	160.69	1.17	194	149.43	0.77
LG8	143	151.17	1.06	172	130.98	0.76	251	73.00	0.29
LG9	106	196.89	1.86	106	114.94	1.08	158	116.00	0.73
LG10	164	137.70	0.84	134	136.94	1.02	235	112.95	0.48
LG11	113	156.56	1.38	117	150.71	1.29	186	140.83	0.76
Total	1236	1593.35	1.29	1302	1567.38	1.20	1897	1314.89	0.69

1.29 and 1.20 cM for H5 and H6 map, respectively. The H5 map presented 15 gaps larger than 15 cM, whereas 10 were found in the H6 map. We also observed similarly low proportions of markers with segregation distortion in both H5 and H6 genetic maps. Thus, 3.80% (47 out of 1,256) and 3.69% (48 out of 1,302) of the mapped markers were distorted with a random distribution over the H5 and H6 map, respectively. In addition, both maps presented a similar distribution of the segregation patterns of the mapped markers. The D1 marker configuration type was the mostly mapped in both populations, with 711 (57.52%) and 705 (54,14%) markers in H5 and H6 genetic maps, respectively. H5 and H6 maps presented 423 (34.22%) and 522 (40.09%) D2 markers, respectively. The B3-type markers were the less frequently mapped, with 101 (8.17%) mapped markers in H5 maps and 75 (5.76%) in H6 map. Only one AFLP marker with C.8 configuration ("ao" × "ao") was mapped in H5 map.

Of the 1,236 mapped markers in H5 genetic map, only 23 markers (12 AFLP, 5 ISSR, and 1 SSR) were previously mapped in the former H5 map (Quezada et al., 2014). The LG9 and LG11 presented the largest number of markers previously mapped, with 8 and 7 markers mapped in H5 former map. The reaming markers were mapped in the LG5, LG1, and LG3 (**Supplementary Data Sheet 3**).

3.5. Composite Genetic Map

In order to construct a composite map, 641 common markers between H5 and H6 map were identified. All these shared SNPs were located within the same LG in both maps, with a largely consistent overall order between LGs. The recombination fractions between these markers was re-estimated using the information of both populations simultaneously and the best order was achieved based on the likelihood. After constructing the composite map, 595 and 661, H5 and H6 unique markers respectively, were incorporated. Consequently, the composite map included a total of 1,897 markers along 11 LGs with a total length of 1,314 cM (**Table 2**, **Figure 3**). The average group length was 119.53 cM, with LG2 the smallest (65.92 cM) and LG5 the largest ones (174.94 cM). The number of markers per group ranged from 123 (LG2) to 251 (LG8) with a mean of 172.45. The average distance between markers was 0.69 cM, with only three gaps larger than 15 cM.

The 641 common markers in the composite map presented all five possible joint marker configuration types defined considering the segregation pattern in both populations simultaneously. However, markers heterozygous in the maternal genotype and homozygous in both paternal genotypes ("ab" × "aa"/"ab" × "aa," "D1.10"-"D1.10") were largely the most frequent, with 449 (70.05%) of the 641 common mapped markers. The configuration type "D2.15"-"D2.15," homozygous in the female parent and heterozygous for the male parent in both populations, was the second most frequent marker type, with 70 (10.92%) mapped markers. Finally, the configuration types with heterozygous parents in one of both populations were the less frequent, with 62 (9.67%) markers defined as "B3.7"-"D1.10," 41 (6.39%) markers "D1.10"-"B3.7" and 19 (2.96%) described as "B3.7-B3.7."

A high agreement was observed between LG assignment of the Discovery-identified SNPs, mapped in the composite map and the physical positions of these markers among the *E. grandis* chromosomes. This same result was observed in the single population maps. Out of 1,897 markers mapped in the composite map, 474 (24.98%) SNPs corresponded to markers identified by the Discovery pipeline, and of these, 437 (92.19%) showed correspondence between the assignment to a LGs in *A. sellowiana* and *E. grandis* chromosomes (**Supplementary Data Sheet 4**). Other 32 markers were mapped in discordant reference chromosomes and the remaining five were assigned to *E. grandis* genome scaffolds. Despite the high level of synteny observed, the marker order (colinearity) in the *A. sellowiana* composite map was not consistent with the physical positions in the *E. grandis* genome.

4. DISCUSSION

Applied for the first time in *A. sellowiana*, the GBS approach demonstrated the advantage of coupled NGS with methods for reducing genome complexity to discover and genotype SNP markers. In our study, the rare-cutting enzyme *Pst*I was selected

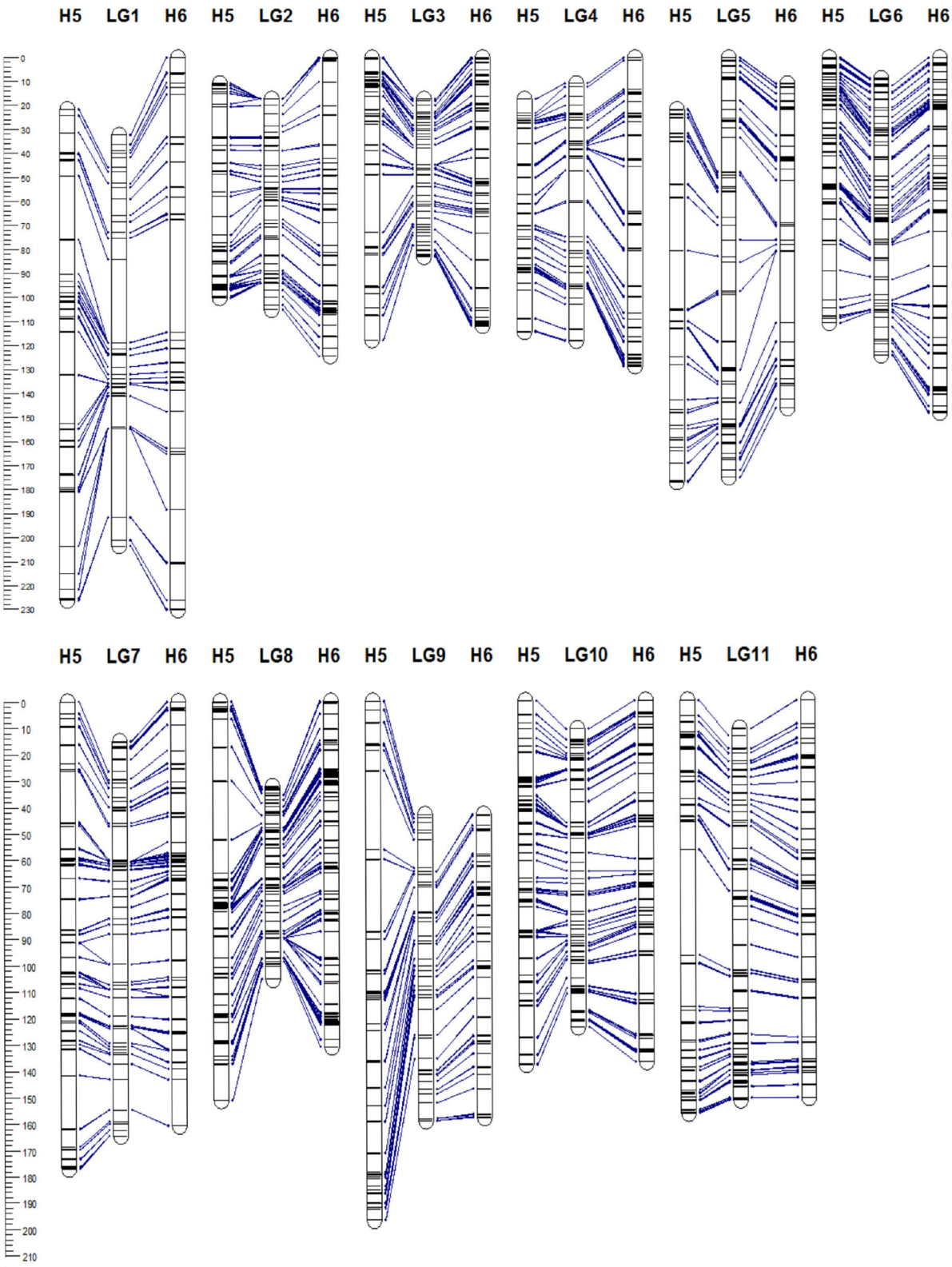

FIGURE 3 | Genetic linkage map of *Acca sellowiana*. The composite linkage map (in the middle) integrated the information of the single population H5 (on the left) and H6 (on the right) maps. Map distances (cM) are indicated at the left side of the figure. Single nucleotide polymorphisms (SNP) markers common between linkage groups (LGs) of the three genetic maps are connected with lines.

for targeting fewer sites with greater depth at each genomic site. As a result, although fewer SNPs were identified, a high site coverage was achieved, suitable for a full-sib mapping population of outcrossing species. It is relevant to note that this kind of populations present a high degree of linkage disequilibrium, so the number of markers required for covering the entire genome is relatively low. The total number of reads and average read per sample obtained in this study were comparable to the observed for other fruit tree species studies (Russell et al., 2014; Guajardo et al., 2015). However, these values are influenced not only by the restriction enzyme used but also by genome features of the species and the availability of a reference genome, among other factors (Gardner et al., 2014).

The original GBS protocol introduced the use of a sequenced reference genome to identify SNPs (Elshire et al., 2011). Considering that *A. sellowiana* lacks genome sequence information, the *E. grandis* genome was used as reference. As a result, an acceptable alignment value (30.1%) of *A. sellowiana* reads against the *E. grandis* genome was found. Using the same reference genome, the proportion of *E. urophylla* read alignment was twice larger (69.1%), as expected between more closely related species (Bartholomé et al., 2014). Differences in read alignment rates mainly depends on the quality of the genome assembly used as well as the phylogenetic relationship between the species. The occurrence of unaligned reads can be explained by poor-quality reads, poor-quality genome assembly, as well as the presence of divergent or species-specific sequences (Hyma et al., 2015). In this work, alignment distribution of *A. sellowiana* reads was homogeneous onto 11 *E. grandis* chromosomes (**Figure 1**). Chromosomes 8, 2, and 6 presented the highest number of identified SNPs, as was reported by Bartholomé et al. (2014), in the alignment of short reads obtained from the complete genome sequencing of *E. grandis* and *E. urophylla*. Although these chromosomes presented medium to large size, the number of identified markers had not a linear relationship with chromosome size. For instance, chromosome 5, which is one of the largest chromosome of *E. grandis*, presented a low *A. sellowiana* read alignment rates. This result can be explained by unique features of the *E. grandis* genome assembly (Bartholomé et al., 2014).

As a complementary strategy, the *de novo* based approach implemented in the UNEAK pipeline (Lu et al., 2013) was applied in this work. Using this approach, we identified ∼3-fold fewer markers than the number of markers identified using the *E. grandis* as reference genome. This notable difference was also observed when comparing the mostly widespread applied reference-free with reference-based GBS pipelines (Torkamaneh et al., 2016). The low UNEAK-identified SNP number may result of a rather stringent network filter, which allowed only one mismatch for pair-read, may not only reduce the SNP number but also errors in genotype calling, caused by paralogous or repetitive sequences. The two strategies implemented in our study allowed to successfully identify a high number of SNP markers for *A. sellowiana* and represented a robust strategy for other non-reference species.

The correct genotype calling from the NGS data presents unique challenges for highly heterozygous species. The low read coverage of GBS data results in a large proportion of missing data as well as heterozygous undercalling. The latter may be due to unequal allele sampling or to a high sequencing base error rate of NGS reads (Swarts et al., 2014). Both, missing values and heterozygous undercalling, hinder linkage analysis, with a substantial impact in marker ordering and phasing, increasing the total map length. To overcome this limitation, new approaches have been developed to improve genotype calling in outcrossing species as well as to impute missing genotypes (Swarts et al., 2014; Covarrubias-Pazaran et al., 2016; Kim et al., 2016; Gerard et al., 2018). Here, we used a genotype calling method that takes advantages of the relative abundance of each allele (read counts) and the Mendelian properties of the mapping populations (Serang et al., 2012). To call genotypes with a high confidence, this methodology simultaneously considers all the genetic information available, as the parental information, the site coverage of each allele and the expected frequencies of individual genotype for each locus (Garcia et al., 2013). Despite the fact that it was initially designed for genotype calling in polyploid species, this quantitative genotyping analysis has proven to be an efficient approach to overcome the main constraints presented by genotyping in highly heterozygous species.

As a result of both SNP and genotyping calling approaches, a large number of high reliable SNP markers, useful for mapping purposes was obtained. However, the final number of useful SNPs retained represented a small proportion of those initially identified. Of the 55,000 SNP discovered, only 9.42% (5350) and 7.44% (4227) of them from H5 and H6 populations, respectively, were retained. This dramatic reduction has also been reported for other fruit tree species, where although many SNPs were identified, robust genotype calling were generated for only a small proportion of them (Gardner et al., 2014). Despite this reduction, our results are in concordance with the number of GBS-SNPs expected in full-sib mapping populations, and are superior to those found in apple and grape, where only 6.0 and 4.2% of the markers were retained for linkage mapping purposes (Gardner et al., 2014; Hyma et al., 2015). Both, Discovery and UNEAK approaches, equally contributed to the final SNP set. However, these approaches significantly differed at the initial number of markers as well as the proportion of retained markers across the filtering process. Thus, only 6.63 and 4.72% of the markers identified in the Discovery approach were selected for H5 and H6 populations, respectively. This proportion increased to 18.43% for H5 and 16.24% for H6 populations, for the UNEAK approach. This result shows that combining both SNP calling approaches with a quantitative genotyping method allowed to obtain high-quality markers covering the entire genome.

The first saturated genetic maps with 1,256 and 1,302 markers for H5 and H6 populations, respectively, were obtained for *A. sellowiana*. Both maps established 11 LGs, matching the expected haploid chromosome number of the species ($n = 11$). The total map length and average distance between markers was similar between maps, as was the number of mapped markers. In addition, the number of markers on both maps was comparable to the high-density maps published for other non-model tree fruit species (Ward et al., 2013; Russell et al., 2014; Guajardo et al.,

2015). Nevertheless, the percentage of SNPs mapped from the initial data set was low (23.48% for H5 population and 30.80% for H6 population). This result could be explained by an uneven distribution of GBS-SNPs in the genome, resulting in large under-represented areas that could not be successfully covered in the genetic maps. Differences in the marker distribution could also explain the low number (23) of markers from the former H5 genetic map (Quezada et al., 2014) that could be mapped.

The comparison between single individual maps pointed out consistent pattern related to the total length and number of mapped markers among the LGs. For instance, both H5 and H6 LGs presented a slight variability in the number of mapped markers per LG, with the exception of the LG5 that presented a significantly lower number of markers. This result, also reported for other fruit tree species could be explained by a non-uniform distribution of SNP markers among chromosomes, occurrence of structural variations, as well by a local decrease of polymorphism in the same regions of the genome (Ward et al., 2013; Russell et al., 2014; Guajardo et al., 2015). Low proportion of markers with segregation distortion were observed in H5 (3.80%) and H6 (3.69%) maps, although distribution pattern for these markers was inconsistent between maps. Several reports have evaluated the impact of distorted segregation ratios in marker data on the construction of genetic linkage maps (Hackett and Broadfoot, 2003; Bodénès et al., 2016). However, the low number of distorted markers finally mapped is expected to have a minor impact in genetic distances, map length or marker order estimations, and therefore may not affect the construction of H5 or H6 genetic linkage maps.

The large number of common markers between H5 and H6 individual maps (51.86 and 49.23% of mapped markers in H5 and H6 map, respectively) was expected considering that both populations shared the same female parent. All the common markers had a complete correspondence with homologous LGs and showed a strong colinearity, confirming the robustness of the individual population maps. As commonly reported in mapping studies, some local inconsistencies in marker order were observed between these maps. The inconsistencies can be attributed to biological factors (such as chromosomal rearrangements, segregation distortion, and so on), sampling bias as well as to technical errors. Thus, structural variations have been reported in apple, where segmental duplications among different populations affect recombination frequencies, influencing marker order accuracy (Khan et al., 2012). In other study, local inconsistencies in marker order were attributed to a sampling bias for small mapping populations ($N = 50$) (Doligez et al., 2006). In our study, based on a comparative large mapping population ($N = 160$ and $N = 184$), the overall agreement between H5 and H6 marker order suggests that technical errors from genotyping methods and/or marker information content as the main causes of the local inconsistencies. The small inconsistencies observed can be a result of missing data, genotyping errors, or difference in content information among the markers and populations. Differences in recombination rates between the three parental genotypes were not considered, since they had little effect in incongruence of marker orders. Likewise, the effect of segregation distortion that can be considered another potential source of error in the ordering process was discarded because of the small proportion of distorted markers mapped.

Several mapping procedures for constructing composite genetic maps have been reported and applied to outcrossing species (Van Ooijen, 2006; Ronin et al., 2012; Endelman and Plomion, 2014). Despite differences in the statistical approaches, all of them are based on the combination of genetic distances estimation from single population maps. These approaches are negatively affected by differences in population sizes, marker information content, missing data, as well as the proportion of common markers. Consequently, marker order inconsistencies and map-distance inflation reduced the resolution and accuracy of the maps generated through these approaches (Doligez et al., 2006; De Keyser et al., 2010; Clark et al., 2014).

The innovative strategy to construct a composite map proposed in this work integrates the recombination information from the individuals of two genetically connected populations. The main advantage of our procedure is that the meiosis information of all individuals ($N = 344$) is captured in a single composite map, since the information data of common mapped markers was merged. Considering the complete correspondence between H5 and H6 LGs, the integration process was carried out for each LG independently. In a first step, only common markers were considered, with the objective of reducing missing information that hinders the marker ordering process (Hackett and Broadfoot, 2003). The resulting framework map provided the most accurate marker order that allowed the correction of the small order inconsistencies in the single population maps, resulting in robust high-density linkage maps. The improved estimation of marker positions can be explained by the larger number of recombination events evaluated in a composite approach, compared to single population approaches. Another possible explanation is that the multipoint maximum likelihood estimation was implemented using genetic information derived from two populations simultaneously, an efficient approach both for statistical and biological reasons. Although the two-point recombination fractions between common markers for H5 and H6 maps were estimated, the multipoint approach was preferred due to the higher accuracy of recombination fraction estimation (Mollinari et al., 2009). Finally, the incorporation of single population markers into the framework map enabled to saturate the composite map.

Using this strategy, a total of 1,897 markers were mapped in the high-density composite map, where 641 framework markers were common between H5 and H6 individual maps. This is by far the most saturated linkage map of *A. sellowiana* available to date. The composite genetic map resulted in a more comprehensive representation of the genome, including information of two mapping populations. This map included more markers than individuals ones, with a small total length (1,314 cM), reducing the average distance between markers from around 1.25 cM in the individual maps to 0.69 cM in the composite map. In addition, the composite map allowed the reduction of the number of gaps, to only three gaps larger than 15 cM. These results were in accordance with previous reports about the influence of population size in linkage maps, specifically in outcrossing

species. For instance, lower marker densities were observed when small population size were analyzed (Bartholomé et al., 2014). Besides, high correlations were reported between population size and number of mapped markers (Hyma et al., 2015). The progeny size has a direct effect on the number of detected recombinants events, since smaller populations have fewer recombinants than larger populations. For H5 ($N = 160$) and H6 ($N = 184$) genetic maps, the mapping resolution achieved was in accordance with the mapping population sizes. The construction of the composite map is adequate to capture the genetic information of a larger population, generating a map with higher precision in the order and distance between markers. Besides, the segregation analysis using connected mapping populations might increase the coverage of the genome, filling the gaps and increasing the mapping resolution. Our results are consistent with previous reports that showed that the number of recombination events evaluated, thus the population size is the current limiting factor to construct high-density linkage maps in outcrossing species (Bartholomé et al., 2014; Hyma et al., 2015).

The level of synteny between *A. sellowiana* ($n = 11$) and *E. grandis* ($n = 11$) was examined through the distribution of Discovery-identified SNP markers, which were assigned to a chromosomal position in the *E. grandis* genome. A high degree of synteny was found between the two species, considering that more than 90% of the Discovery-identified markers were syntenic to *E. grandis* chromosomes. Although this result suggested the conserved localization of markers, colinearity (order of the markers) within each LG was not observed (**Supplementary Data Sheet 4**). It is important to highlight our results considering the phylogenetic divergence and the genome size difference between *A. sellowiana* (245 Mb) (da Costa et al., 2008) and *E. grandis* (641 Mb). However, these results are consistent with the high level of synteny observed within the Myrtaceae family (Grattapaglia et al., 2012). In addition, the recent comparison of chloroplast genomes of myrtle species, including *A. sellowiana* and *E. grandis* also revealed a highly conserved genome content, gene order and genomic structure between these species (de Machado et al., 2017). The high synteny observed in this preliminary study provides the basis for using the *E. grandis* reference genome in genomic studies of minor crops in the Myrtaceae family.

Here, the composite genetic map provided a reference for *A. sellowiana*, a valuable tool for future genetic and genomic applications. The composite map especially improved the accuracy of order marker and genetic distances, facilitating the comparison between genetic maps. This map will be also a useful tool to guide the assembly of the *A. sellowiana* genome sequence, as a reference to anchor and orient the sequence scaffolds. With the development of composite genetic maps, the occurrence of genome structural variation or conserved synteny can be evaluated across divergent species. To date, the comparative mapping studies in the Myrtaceae family had only included dry fruit species of the tribe Eucalypteae (Grattapaglia et al., 2012). Therefore, the composite map developed in this work can be useful to extend the studies including more divergent species of the family. In addition, this composite map may allow to align the position of QTLs detected across variable genetic backgrounds, facilitating the transfer of genetic information from molecular markers and gene positions, and accelerating molecular breeding strategies such as marker-assisted selection (Diaz et al., 2011).

5. CONCLUSIONS

In this study, we constructed the first high-density genetic map for *A. sellowiana* using a genotyping by sequencing approach. The GBS protocol was an effective strategy for simultaneous SNP discovery and genotyping, identifying thousands of genome-wide polymorphic markers, in a species with limited genetic resources. We also developed a novel strategy for constructing a composite genetic map using the genetic information from two full-sib connected mapping populations. This approach provided a better estimation of recombination fraction, resulted in higher accuracy for marker distance and order as well as increased genome coverage. The composite map along with the H5 and H6 single population maps are the most comprehensive representations of the *A. sellowiana* genome and constitute key genetic resources for this minor species. These maps could be useful for future genetic studies, such as the detection of QTL of important agronomic traits, comparative genome analysis in the Myrtaceae family, genome assembly, and the acceleration of the breeding process in *A. sellowiana*.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA517479.

AUTHOR CONTRIBUTIONS

CP and AG conceived the study design. BV and DC provide the plant material. MQ performed the DNA extraction. MQ and AG designed the GBS experiments. MQ performed the GBS data analysis and the genetic map analysis. RA and MQ carried out the novel genetic map analysis, wrote computer code, and integrated the analysis of results. MQ wrote the manuscript draft. CP, RA, and AG edited and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by grants from the ANII (Agencia Nacional de Investigacion e Innovacion, Uruguay; FMV_3_2013_1_100571) and from CSIC-UdelaR-Uruguay (CSIC_I+D_375). MQ received doctoral fellowships from CAPES-Brazil (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and INIA-Uruguay (Instituto Nacional de Investigación Agropecuaria). RA received a master's fellowship from CNPq (Conselho Nacional

de Desenvolvimento Científico e Tecnológico). AG received a research fellowships from CNPq (Grant 310139/2018-0). The funders had no role in study design, data collection, analysis and interpretation of data, or writing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.626811/full#supplementary-material

REFERENCES

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e0003376. doi: 10.1371/journal.pone.0003376

Balsalobre, T. W. A., Pereira, G. S., Margarido, G. R. A., Gazaffi, R., Barreto, F. Z., Anoni, C. O., et al. (2017). GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. *BMC Genomics* 18:72. doi: 10.1186/s12864-016-3383-x

Barba, P., Cadle-Davidson, L., Harriman, J., Glaubitz, J. C., Brooks, S., Hyma, K., et al. (2014). Grapevine powdery mildew resistance and susceptibility loci identified on a high-resolution SNP map. *Theor. Appl. Genet.* 127, 73–84. doi: 10.1007/s00122-013-2202-x

Bartholomé, J., Mandrou, E., Mabiala, A., Jenkins, J., Nabihoudine, I., Klopp, C., et al. (2014). High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. *New Phytol.* 206, 1283–1296. doi: 10.1111/nph.13150

Bielenberg, D. G., Rauh, B., Fan, S., Gasic, K., Abbott, A. G., Reighard, G. L., et al. (2015). Genotyping by sequencing for SNP-based linkage map construction and QTL analysis of chilling requirement and bloom date in peach [*Prunus persica* (L.) Batsch]. *PLoS ONE* 10:e0139406. doi: 10.1371/journal.pone.0139406

Bodénès, C., Chancerel, E., Ehrenmann, F., Kremer, A., and Plomion, C. (2016). High-density linkage mapping and distribution of segregation distortion regions in the oak genome. *DNA Res.* 23, 115–124. doi: 10.1093/dnares/dsw001

Byrne, D. H. (2012). "Trends in fruit breeding," in *Fruit Breeding*, eds M. L. Badenes and D. H. Byrne (New York, NY: Springer-Verlag), 3–36.

Clark, M. D., Schmitz, C. A., Rosyara, U. R., Luby, J. J., and Bradeen, J. M. (2014). A consensus 'Honeycrisp' apple (*Malus* x *domestica*) genetic linkage map from three full-sib progeny populations. *Tree Genet. Genomes* 10, 627–639. doi: 10.1007/s11295-014-0709-1

Covarrubias-Pazaran, G., Diaz-Garcia, L., Schlautman, B., Deutsch, J., Salazar, W., Hernandez-Ochoa, M., et al. (2016). Exploiting genotyping by sequencing to characterize the genomic structure of the American cranberry through high-density linkage mapping. *BMC Genomics* 17:451. doi: 10.1186/s12864-016-2802-3

da Costa, I. R., Dornelas, M. C., and Forni-Martins, E. R. (2008). Nuclear genome size variation in fleshy-fruited Neotropical Myrtaceae. *Plant Syst. Evol.* 276, 209–217. doi: 10.1007/s00606-008-0088-x

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

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Publ. Gr.* 12, 499–510. doi: 10.1038/nrg3012

de Givry, S., Bouchez, M., Chabrier, P., Milan, D., and Schiex, T. (2005). CARTHAGENE: multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics* 21, 1703–1704. doi: 10.1093/bioinformatics/bti222

De Keyser, E., Shu, Q. Y., Van Bockstaele, E., and De Riek, J. (2010). Multipoint-likelihood maximization mapping on 4 segregating populations to achieve an integrated framework map for QTL analysis in pot azalea (*Rhododendron simsii* hybrids). *BMC Mol. Biol.* 11:1. doi: 10.1186/1471-2199-11-1

de Machado, L. O., do Nascimento Vieira, L., Stefenon, V. M., de Oliveira Pedrosa, F., de Souza, E. M., Guerra, M. P., et al. (2017). Phylogenomic relationship of feijoa (*Acca sellowiana* (O.Berg) Burret) with other Myrtaceae based on complete chloroplast genome sequences. *Genetica* 145, 163–174. doi: 10.1007/s10709-017-9954-1

Dempster, A., Laird, N., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc, Ser. B Methodol.* 39, 1–22. doi: 10.1111/j.2517-6161.1977.tb01600.x

Dettori, M. T., and Palombi, M. A. (2000). Identification of *Feijoa sellowiana* Berg accessions by RAPD markers. *Sci. Hortic.* 86, 279–290. doi: 10.1016/S0304-4238(00)00157-6

Di Pierro, E. A., Gianfranceschi, L., Di Guardo, M., Koehorst-van Putten, H. J., Kruisselbrink, J. W., Longhi, S., et al. (2016). A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. *Hortic. Res.* 3:16057. doi: 10.1038/hortres.2016.57

Diaz, A., Fergany, M., Formisano, G., Ziarsolo, P., Blanca, J., Fei, Z., et al. (2011). A consensus linkage map for molecular markers and Quantitative Trait Loci associated with economically important traits in melon (*Cucumis melo* L.). *BMC Plant Biol.* 11:111. doi: 10.1186/1471-2229-11-111

Doligez, A., Adam-Blondon, A., Cipriani, G., Di Gaspero, G., Laucou, V., Merdinoglu, D., et al. (2006). An integrated SSR map of grapevine based on five mapping populations. *Theor. Appl. Genet.* 113, 369–382. doi: 10.1007/s00122-006-0295-1

Donazzolo, J., Stefenon, V. M., Guerra, M. P., and Nodari, R. O. (2020). On farm management of *Acca sellowiana* (Myrtaceae) as a strategy for conservation of species genetic diversity. *Sci. Hortic.* 259:108826. doi: 10.1016/j.scienta.2019.108826

dos Santos, K. L., Santos, M. O., Laborda, P. R., Souza, A. P., Peroni, N., and Nodari, R. O. (2008). Isolation and characterization of microsatellite markers in *Acca sellowiana* (Berg) Burret. *Mol. Ecol. Resour.* 8, 998–1000. doi: 10.1111/j.1755-0998.2008.02134.x

dos Santos, K. L., Welter, L. J., Dantas, A. C. M., Guerra, M. P., Ducroquet, J. P. H. J., and Nodari, R. O. (2007). Transference of microsatellite markers from *Eucalyptus* spp to *Acca sellowiana* and the successful use of this technique in genetic characterization. *Genet. Mol. Biol.* 30, 73–79. doi: 10.1590/S1415-47572007000100014

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:e0019379. doi: 10.1371/journal.pone.0019379

Endelman, J. B., and Plomion, C. (2014). LPmerge: an R package for merging genetic maps by linear programming. *Bioinformatics* 30, 1623–1624. doi: 10.1093/bioinformatics/btu091

Fierst, J. L. (2015). Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. *Front. Genet.* 6:220. doi: 10.3389/fgene.2015.00220

Garcia, A. A. F., Mollinari, M., Marconi, T. G., Serang, O. R., Silva, R. R., Vieira, M. L. C., et al. (2013). SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Sci. Rep.* 3:3399. doi: 10.1038/srep03399

Gardner, K. M., Brown, P. J., Cooke, T. F., Cann, S., Costa, F., Bustamante, C. D., et al. (2014). Fast and cost-effective genetic mapping in apple using next-generation sequencing. *G3* 4, 1681–1687. doi: 10.1534/g3.114.011023

Gerard, D., Ferrão, L. F. V., Garcia, A. A. F., and Stephens, M. (2018). Genotyping polyploids from messy sequencing data. *Genetics* 210, 789–807. doi: 10.1534/genetics.118.301468

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:e0090346. doi: 10.1371/journal.pone.0090346

Grattapaglia, D., and Sederoff, R. (1994). Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137, 1121–1137. doi: 10.1093/genetics/137.4.1121

Grattapaglia, D., Vaillancourt, R. E., Shepherd, M., Thumma, B. R., Foley, W., Külheim, C., et al. (2012). Progress in Myrtaceae genetics and genomics: Eucalyptus as the pivotal genus. *Tree Genet. Genomes* 8, 463–508. doi: 10.1007/s11295-012-0491-x

Guajardo, V., Solís, S., Sagredo, B., Gainza, F., Muñoz, C., Gasic, K., et al. (2015). Construction of high density sweet cherry (*Prunus avium* L.) linkage maps using microsatellite markers and SNPs detected by Genotyping-by-Sequencing (GBS). *PLoS ONE* 10:e0127750. doi: 10.1371/journal.pone.0127750

Hackett, C. A., and Broadfoot, L. B. (2003). Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* 90, 33–38. doi: 10.1038/sj.hdy.6800173

Hyma, K. E., Barba, P., Wang, M., Londo, J. P., Acharya, C. B., Mitchell, S. E., et al. (2015). Heterozygous mapping strategy (HetMappS) for high resolution genotyping-by-sequencing markers: a case study in grapevine. *PLoS ONE* 10:e0134880. doi: 10.1371/journal.pone.0134880

Jaganathan, D., Bohra, A., Thudi, M., and Varshney, R. K. (2020). Fine mapping and gene cloning in the post-NGS era: advances and prospects. *Theor. Appl. Genet.* 133, 1791–1810. doi: 10.1007/s00122-020-03560-w

Khan, M. A., Han, Y., Zhao, Y. F., Troggio, M., and Korban, S. S. (2012). A multi-population consensus genetic map reveals inconsistent marker order among maps likely attributed to structural variations in the Apple genome. *PLoS ONE* 7:e47864. doi: 10.1371/journal.pone.0047864

Kim, C., Guo, H., Kong, W., Chandnani, R., Shuang, L. S., and Paterson, A. H. (2016). Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Sci.* 242, 14–22. doi: 10.1016/j.plantsci.2015.04.016

Klabunde, G. H. F., Olkoski, D., Vilperte, V., Zucchi, M. I., and Nodari, R. O. (2014). Characterization of 10 new nuclear microsatellite markers in *Acca sellowiana* (Myrtaceae). *Appl. Plant Sci.* 2:apps.1400020.doi: 10.3732/apps.1400020

Knaus, B. J., and Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* 17, 44–53. doi: 10.1111/1755-0998.12549

Kosambi, D. D. (1944). The estimation of map distances from recombination values. *Ann. Eugen.* 12, 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x

Kuhn, D. N., Bally, I. S. E., Dillon, N. L., Innes, D., Groh, A. M., Rahaman, J., et al. (2017). Genetic map of mango: a tool for mango breeding. *Front. Plant Sci.* 8:577. doi: 10.3389/fpls.2017.00577

Lander, E. S., and Green, P. (1987). Construction of multilocus genetic linkage maps in humans. *Proc. Natl. Acad. Sci. U.S.A.* 84, 2363–2367. doi: 10.1073/pnas.84.8.2363

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, L., Deng, C. H., Knäbel, M., Chagné, D., Kumar, S., Sun, J., et al. (2017). Integrated high-density consensus genetic map of *Pyrus* and anchoring of the 'Bartlett' v1.0 (*Pyrus communis*) genome. *DNA Res.* 24, 289–301. doi: 10.1093/dnares/dsw063

Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., et al. (2013). Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet.* 9:e1003215. doi: 10.1371/journal.pgen.1003215

Maliepaard, C., Jansen, J., and Van Ooijen, J. W. (1997). Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. *Genet. Res.* 70, 237–250. doi: 10.1017/S0016672397003005

Margarido, G. R. A., de Souza, A. P., and Garcia, A. A. F. (2007). Onemap: software for genetic mapping in outcrossing species. *Hereditas* 144, 78–79. doi: 10.1111/j.2007.0018-0661.02000.x

Mathew, L. S., Spannagl, M., Al-Malki, A., George, B., Torres, M. F., Al-Dous, E. K., et al. (2014). A first genetic map of date palm (*Phoenix dactylifera*) reveals long-range genome structure conservation in the palms. *BMC Genomics* 15:285. doi: 10.1186/1471-2164-15-285

Mollinari, M., Margarido, G. R. A., Vencovsky, R., and Garcia, A. A. F. (2009). Evaluation of algorithms used to order markers on genetic maps. *Heredity* 103, 494–502. doi: 10.1038/hdy.2009.96

Mollinari, M., and Serang, O. (2015). "Quantitative SNP genotyping of polyploids with MassARRAY and other platforms," in *Plant Genotyping*. Methods in Molecular Biology, ed J. Batley (New York, NY: Humana Press), 215–241.

Padmakar, B., Kanupriya, C., Latha, P. M., Prashant, K., Dinesh, M., Sailaja, D., et al. (2015). Development of SRAP and SSR marker-based genetic linkage maps of guava (*Psidium guajava* L.). *Sci. Hortic.* 192, 158–165. doi: 10.1016/j.scienta.2015.04.031

Palhares, A. C., Rodrigues-Morais, T. B., Van Sluys, M.-A., Domingues, D. S., Maccheroni, W., Jordão, H., et al. (2012). A novel linkage map of sugarcane with evidence for clustering of retrotransposon-based markers. *BMC Genet.* 13:51. doi: 10.1186/1471-2156-13-51

Pasquariello, M. S., Mastrobuoni, F., Di Patre, D., Zampella, L., Capuano, L. R., Scortichini, M., et al. (2015). Agronomic, nutraceutical and molecular variability of feijoa (*Acca sellowiana* (O. Berg) Burret) germplasm. *Sci. Hortic.* 191, 1–9. doi: 10.1016/j.scienta.2015.04.036

Peace, C., and Norelli, J. L. (2009). "Genomics approaches to crop improvement in the Rosaceae," in *Genet. Genomics Rosaceae*, eds K. M. Folta and S. E. Gardiner (New York, NY: Springer), 19-53.

Pereira, G. S., Garcia, A. A. F., and Margarido, G. R. A. (2018). A fully automated pipeline for quantitative genotype calling from next generation sequencing data in autopolyploids. *BMC Bioinformatics* 19:398. doi: 10.1186/s12859-018-2433-6

Pereira, G. S., Nunes, E. S., Laperuta, L. D. C., Braga, M. F., Penha, H. A., Diniz, A. L., et al. (2013). Molecular polymorphism and linkage analysis in sweet passion fruit, an outcrossing species. *Ann. Appl. Biol.* 162, 347–361. doi: 10.1111/aab.12028

Pootakham, W., Ruang-Areerate, P., Jomchai, N., Sonthirod, C., Sangsrakru, D., Yoocha, T., et al. (2015). Construction of a high-density integrated genetic linkage map of rubber tree (*Hevea brasiliensis*) using genotyping-by-sequencing (GBS). *Front. Plant Sci.* 6:367. doi: 10.3389/fpls.2015.00367

Quezada, M., Pastina, M. M., Ravest, G., Silva, P., Vignale, B., Cabrera, D., et al. (2014). A first genetic map of *Acca sellowiana* based on ISSR, AFLP and SSR markers. *Sci. Hortic.* 169, 138–146. doi: 10.1016/j.scienta.2014.02.009

R Development Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ronin, Y., Mester, D., Minkov, D., Belotserkovski, R., Jackson, B. N., Schnable, P. S., et al. (2012). Two-phase analysis in consensus genetic mapping. *G3* 2, 537–549. doi: 10.1534/g3.112.002428

Russell, J., Hackett, C., Hedley, P., Liu, H., Brennan, L. M., Bayer, M., et al. (2014). The use of genotyping by sequencing in blackcurrant (*Ribes nigrum*): developing high-resolution linkage maps in species without reference genome sequences. *Mol. Breed.* 33, 835–849. doi: 10.1007/s11032-013-9996-8

Saifert, L., Sánchez-Mora, F. D., Borsuk, L. J., Donazzolo, J., da Costa, N. C. F., Ribeiro, H. N., et al. (2020). Evaluation of the genetic diversity in the feijoa accessions maintained at Santa Catarina, Brazil. *Crop Sci.* 60, 345–356. doi: 10.1002/csc2.20088

Schlautman, B., Covarrubias-Pazaran, G., Diaz-Garcia, L., Iorizzo, M., Polashock, J., Grygleski, E., et al. (2017). Construction of a high-density American Cranberry (*Vaccinium macrocarpon* Ait.) composite map using genotyping-by-sequencing for multi-pedigree linkage mapping. *G3* 7, 1177–1189. doi: 10.1534/g3.116.037556

Serang, O., Mollinari, M., and Garcia, A. A. F. (2012). Efficient exact maximum a posteriori computation for Bayesian SNP genotyping in polyploids. *PLoS ONE* 7:e0030906. doi: 10.1371/journal.pone.0030906

Swarts, K., Li, H., Alberto Romero Navarro, J., An, D., Romay, M. C., and Hearne, S. (2014). Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *Plant Genome* 7, 1–12. doi: 10.3835/plantgenome2014.05.0023

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

Troggio, M., Gleave, A., Salvi, S., Chagné, D., Cestaro, A., Kumar, S., et al. (2012). Apple, from genome to breeding. *Tree Genet. Genomes* 8, 509–529. doi: 10.1007/s11295-012-0492-9

Van Ooijen, J. W. (2006). *JoinMap 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Wageningen: Kyazma BV.

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77

Ward, J. A., Bhangoo, J., Fernández-Fernández, F., Moore, P., Swanson, J., Viola, R., et al. (2013). Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. *BMC Genomics* 14:2. doi: 10.1186/1471-2164-14-2

Weston, R. J. (2010). Bioactive products from fruit of the feijoa (*Feijoa sellowiana*, Myrtaceae): a review. *Food Chem.* 121, 923–926. doi: 10.1016/j.foodchem.2010.01.047

Wu, R., Ma, C.-X., Painter, I., and Zeng, Z.-B. (2002a). Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. *Theor. Popul. Biol.* 61, 349–363. doi: 10.1006/tpbi.2002.1577

Wu, R., Ma, C.-X., Wu, S. S., and Zeng, Z.-B. (2002b). Linkage mapping of sex-specific differences. *Genet. Res.* 79, 85–96. doi: 10.1017/S0016672301005389

Wu, Y., Close, T. J., and Lonardi, S. (2011). Accurate construction of consensus genetic maps via integer linear programming. *IEEE ACM Trans. Comput. Biol. Bioinform.* 8, 381–394. doi: 10.1109/TCBB.2010.35

Conflict of Interest: 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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 25 February 2021
doi: 10.3389/fpls.2021.642019

Fine Mapping of the "black" Peel Color in Pomegranate (*Punica granatum* L.) Strongly Suggests That a Mutation in the *Anthocyanidin Reductase* (*ANR*) Gene Is Responsible for the Trait

*Taly Trainin[1], Rotem Harel-Beja[1], Irit Bar-Ya'akov[1], Zohar Ben-Simhon[1], Rami Yahalomi[1], Hamutal Borochov-Neori[2], Ron Ophir[3], Amir Sherman[3], Adi Doron-Faigenboim[4] and Doron Holland[1]**

[1] Unit of Fruit Tree Sciences, Institute of Plant Sciences, Agricultural Research Organization, Newe Ya'ar Research Center, Ramat Yishay, Israel, [2] Southern Arava Research and Development, Hevel Eilot, Israel, [3] Department of Fruit Tree Sciences, Institute of Plant Sciences, Agricultural Research Organization, Volcani Center, Rishon LeZion, Israel, [4] Department of Vegetable and Field Crops, Institute of Plant Sciences, Agricultural Research Organization, Volcani Center, Rishon LeZion, Israel

Anthocyanins are important dietary and health-promoting substances present in high quantities in the peel and arils of the pomegranate (*Punica granatum* L.) fruit. Yet, there is a high variation in the content of anthocyanin among different pomegranate varieties. The 'Black' pomegranate variety (P.G.127-28) found in Israel contains exceptionally high levels of anthocyanins in its fruit peel which can reach up to two orders of magnitude higher content as compared to that of other pomegranate varieties' peel anthocyanins. Biochemical analysis reveals that delphinidin is highly abundant in the peel of 'Black' variety. The pattern of anthocyanin accumulation in the fruit peel during fruit development of 'Black' variety differs from that of other pomegranates. High anthocyanin levels are maintained during all developmental stages. Moreover, the accumulation of anthocyanin in the fruit peel of 'Black' variety is not dependent on light. Genetic analysis of an F_2 population segregating for the "black" phenotype reveals that it is determined by a single recessive gene. Genetic mapping of the F_2 population using single nucleotide polymorphism (SNP) markers identified few markers tightly linked to the "black" phenotype. Recombination analysis of the F_2 population and F_3 populations narrowed the "black" trait to an area of 178.5 kb on the draft genome sequence of pomegranate *cv*. 'Dabenzi.' A putative *anthocyanidin reductase* (*ANR*) gene is located in this area. Only pomegranate varieties displaying the "black" trait

OPEN ACCESS

Edited by:
Inaki Hormaza, Institute of Subtropical and Mediterranean Horticulture La Mayora, Spain

Reviewed by:
Shangyin Cao, Zhengzhou Fruit Research Institute (CAAS), China
Dengfeng Hong, Huazhong Agricultural University, China

***Correspondence:**
Doron Holland
vhhollan@volcani.agri.gov.il

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

Received: *15 December 2020*
Accepted: *03 February 2021*
Published: *25 February 2021*

Citation:
Trainin T, Harel-Beja R, Bar-Ya'akov I, Ben-Simhon Z, Yahalomi R, Borochov-Neori H, Ophir R, Sherman A, Doron-Faigenboim A and Holland D (2021) Fine Mapping of the "black" Peel Color in Pomegranate (Punica granatum L.) Strongly Suggests That a Mutation in the Anthocyanidin Reductase (ANR) Gene Is Responsible for the Trait. Front. Plant Sci. 12:642019. doi: 10.3389/fpls.2021.642019

carry a base pair deletion toward the end of the gene, causing a frame shift resulting in a shorter protein. We propose that this mutation in the *ANR* gene is responsible for the different anthocyanin composition and high anthocyanin levels of the "black" trait in pomegranate.

Keywords: anthocyanins, pomegranate genetic variability, genetic mapping, fruit, *Punica granatum*, *anthocyanidin reductase*

INTRODUCTION

The pomegranate fruit is well known for its high content of health-promoting substances (Heber et al., 2006; Seeram et al., 2006). Among these are anthocyanins, which are produced in the peel and arils (the edible part of the pomegranate fruit) of the fruit (Gil et al., 1995; Hernández et al., 1999; Tzulker et al., 2007). However, there is high variation in the content and composition of anthocyanins among different pomegranate varieties (Bar-Ya'akov et al., 2019). The presence of anthocyanins in the fruit peel and arils contributes to their antioxidant activity and plays an important role in protecting the fruit from sunburns and in its attractiveness for pests (Winkel-Shirley, 2001; Koes et al., 2005). In addition, anthocyanins are the main metabolites that determine the color of the pomegranate fruit and high anthocyanin levels are a key factor in determining the economic value of the fruit. Anthocyanin level in the pomegranate fruit is developmentally regulated (Gil et al., 1995; Ben-Simhon et al., 2011; Qin et al., 2017; Yuan et al., 2017; Harel-Beja et al., 2019) and is dependent on environmental conditions such as abiotic stresses (Borochov-Neori et al., 2011; Bar-Ya'akov et al., 2019), salinity (Borochov-Neori et al., 2014), temperature and drought (Borochov-Neori et al., 2009; Holland and Bar-Ya'akov, 2014; Bar-Ya'akov et al., 2019). Numerous studies have shown that the biosynthesis of anthocyanins in maturing fruit such as apples or grapes is a light-dependent process (Mancinelli, 1985; Takos et al., 2006; Azuma et al., 2012; Jaakola, 2013).

The main anthocyanins produced in pomegranate consist of the mono- and di-glycosidic forms of cyanidin, delphinidin, and pelargonidin (Gil et al., 1995; Du et al., 2006; Seeram et al., 2006). Much effort was dedicated to understanding the control of anthocyanin in plants including pomegranate (Holton and Cornish, 1995; Ben-Simhon et al., 2011; Zhao et al., 2013; Yuan et al., 2017; Harel-Beja et al., 2019). Regulation of the expression of key enzymes in the anthocyanin biosynthetic pathway via a ternary complex of MYB-bHLH-WD40 transcription factors (MBW complex) was reported for various plants: grapes (Boss et al., 1996; Cutanda-Perez et al., 2009), apples (Ban et al., 2007), pomegranate (Ben-Simhon et al., 2011), and others (Petroni and Tonelli, 2011; Albert et al., 2014). Anthocyanin level can also be regulated by further metabolism or competition on parallel pathways that utilize the same substrate (Gao et al., 2020). The proanthocyanidins (condensed tannins) pathway splits from the anthocyanin pathway via enzymes including leucoanthocyanidin reductase (LAR) and ANR (**Figure 1**). LAR converts leucoanthocyanidin into catechin, while ANR acts downstream to leucoanthocyanidin dioxygenase (LDOX) and converts anthocyanidins (e.g., cyanidin. pelargonidin, delphinidin) into epicatechins (Xie et al., 2003, 2004; Bogs et al., 2005). By doing so ANR can compete with the enzyme UDP-glucose: flavonoid 3-*O*-glucosyltransferase (UFGT), which converts anthocyanidin into anthocyanin (Bogs et al., 2005), and to divert the metabolism away from production of anthocyanin toward production of epicatechin (**Figure 1**). Overexpression of *ANR* in flower petals of tobacco results in loss of color (Xie et al., 2003; Bogs et al., 2005; Han et al., 2012). In *Arabidopsis*, a mutation in the *ANR* gene results in a colored seed coat due to the presence of anthocyanins (Xie et al., 2003).

The genes encoding for anthocyanin biosynthesis were identified in many plant species and most of them show high homology in their encoded amino acid (aa) sequences. Most of the genes known today in pomegranate were identified by homology to known functional genes in other plant species (Ben-Simhon et al., 2011, 2015; Zhao et al., 2013; Qin et al., 2017; Yuan et al., 2017). The whole pomegranate genome sequence together with transcriptomic data provided a comprehensive view of the anthocyanin pathway in pomegranate and of the expression of the genes during fruit development (Ophir et al., 2014; Qin et al., 2017; Yuan et al., 2017; Harel-Beja et al., 2019). Despite this, the biological and biochemical function of anthocyanin genes in pomegranate was shown only for very few genes. These include the *LDOX* (Ben-Simhon et al., 2015) and the *WD40* homologous genes (Ben-Simhon et al., 2011). As in other plant species, it appears that anthocyanin biosynthesis is transcriptionally regulated (Ben-Simhon et al., 2011, 2015; Zhao et al., 2013; Rouholamin et al., 2015; Qin et al., 2017; Arlotta et al., 2020).

Pronounced color differences among various pomegranate cultivars were reported (Tzulker et al., 2007; Holland et al., 2009; Bar-Ya'akov et al., 2019), suggesting high genetic variability in the regulation of pomegranate fruit color. The variation found among pomegranate varieties is in both the amount and composition of anthocyanins (Gil et al., 1995; Tzulker et al., 2007; Holland et al., 2009; Zhao et al., 2013; Harel-Beja et al., 2019). It was found that the kinetics of color appearance during fruit development is different among cultivars and that this difference is also correlated with the expression of a *MYB* gene (Ben-Simhon et al., 2011; Khaksar et al., 2015) and other regulatory genes that belong to bHLH and MYB families (Harel-Beja et al., 2019). An interesting example is the "white" pomegranate variety, which does not produce anthocyanins in any of its tissues (Ben-Simhon et al., 2015). It was shown that this phenomenon

Abbreviations: HPLC, high performance liquid chromatography; PG, *Punica granatum*; ANR, anthocyanidin reductase; SNP, single nucleotide polymorphism; SSR, simple-sequence repeats; aa, amino acids; MQM, multiple QTL model; WGS, whole genome shotgun sequence; GO, gene ontology; LG2, linkage group 2.

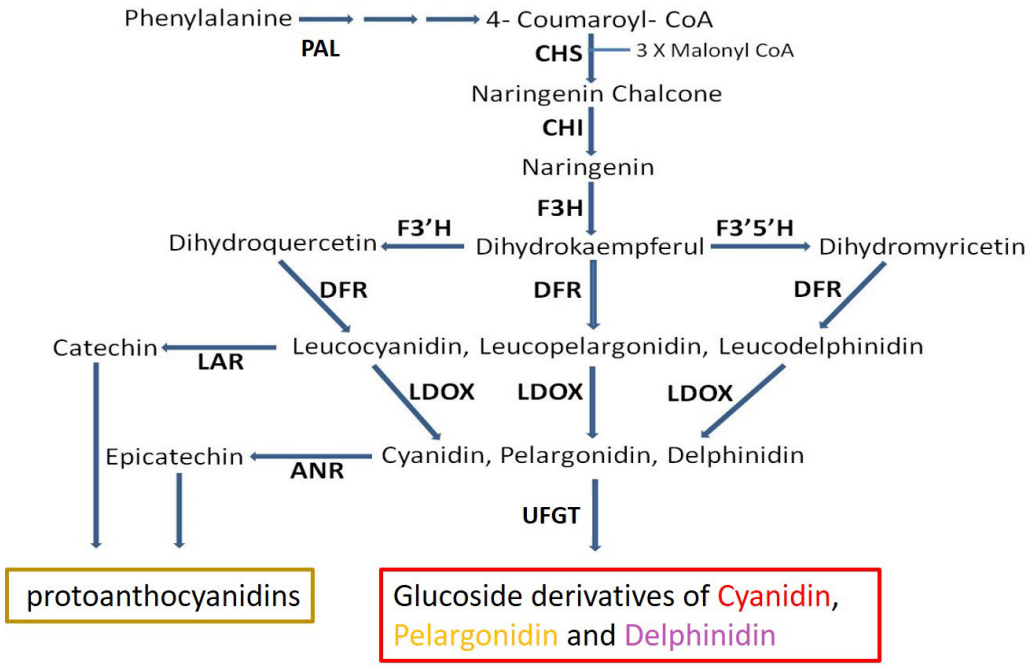

FIGURE 1 | Schematic representation of the flavonoid-biosynthesis of anthocyanins and proanthocyanidins. Enzyme name abbreviations are as follows: PAL, phenylalanine ammonia-lyase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; F3′5′H, flavonoid 3′,5′-hydroxylase; DFR, dihydroflavonol reductase; LDOX, leucoanthocyanidin oxidase; ANS, anthocyanidin synthase; LAR, leucoanthocyanidin reductase; ANR, anthocyanidin reductase; UFGT, UDP glucose:flavonoid 3-*O*-glucosyltransferase. Adapted from Ben-Simhon et al. (2011).

resulted from a mutation in the *LDOX* gene, which totally abolished its expression (Ben-Simhon et al., 2015). Despite the commercial and physiological importance of anthocyanin content and accumulation in pomegranate, the genetic functions responsible for most of the variability of color accumulation and production in pomegranate are not yet known.

"Black" pomegranates which are characterized by purple to black peel color are known from China in eastern Asia to western Asia (Iran, Afghanistan, Iraq, Israel) and differ from each other with respect to fruit characteristics such as taste, growth habit and dwarfism (Holland et al., 2009; Shams Ardekani et al., 2011; Özgüven et al., 2012; Zhu et al., 2015; Balli et al., 2020). In this manuscript, we focused our studies on the 'Black' pomegranate (P.G.127-28), which contains an exceptionally high amount of cyanidin and delphinidin mono- and di-glycosidic forms in its peel. Genetic studies have led to fine mapping of the region responsible for the "black" phenotype. These studies strongly suggest that a mutation in the *ANR* gene, situated in this region, is responsible for the "black" phenotype.

MATERIALS AND METHODS

Plant Material

The 'Black' pomegranate variety P.G.127-28 was used in this research. 'Black' is a late season ripening variety, with a dark purple ("black") peel color starting to accumulate in early stages of fruit development, pinkish-red arils, and a sweet taste. Other pomegranate varieties used in this study are described in **Supplementary Table 1**. All varieties are part of ARO's pomegranate germplasm collection, which is located at the Newe Ya'ar Research Center[1]. The Newe Ya'ar Research Center is located in the western Yizre'el Valley, lat.32°42′N, long.35°11′E, planted on clay grumusol (vertisol) soil at an elevation of about 100 m above sea level. Yizre'el Valley is characterized by a Mediterranean subtropical climate, with an average annual rainfall of about 580 mm concentrated from November through March. The mean diurnal minimum temperature in January is 6°C, and the mean diurnal maximum temperature in July is 33°C.

A segregating F_2 population, designated 'Nana × Black' (n = 204), was constructed from a cross between 'Black' (pollen donor) and 'Nana' seedling selection (P.G.232-243, *P. granatum* var. nana) (Harel Beja et al., 2015). Two F_3 segregating populations were generated from two F_2 plants as described in the "Results" Section.

Fruits were collected during the years 2008–2020 at various developmental stages, from flower to fully mature fruit (**Figure 2**) (Ben-Simhon et al., 2011). For RNA-seq analysis fruits were collected from a very early stage after setting and before stage 3. The number of fruits from each variety for each stage was different (ten from early stages to three from stages 8–12), due to the difference in size. The fruits were photographed. For further analysis, the thinnest possible colored peel skin was removed with a sharp knife and stored at −80°C.

[1]http://igb.agri.gov.il/main/index.pl?page=22

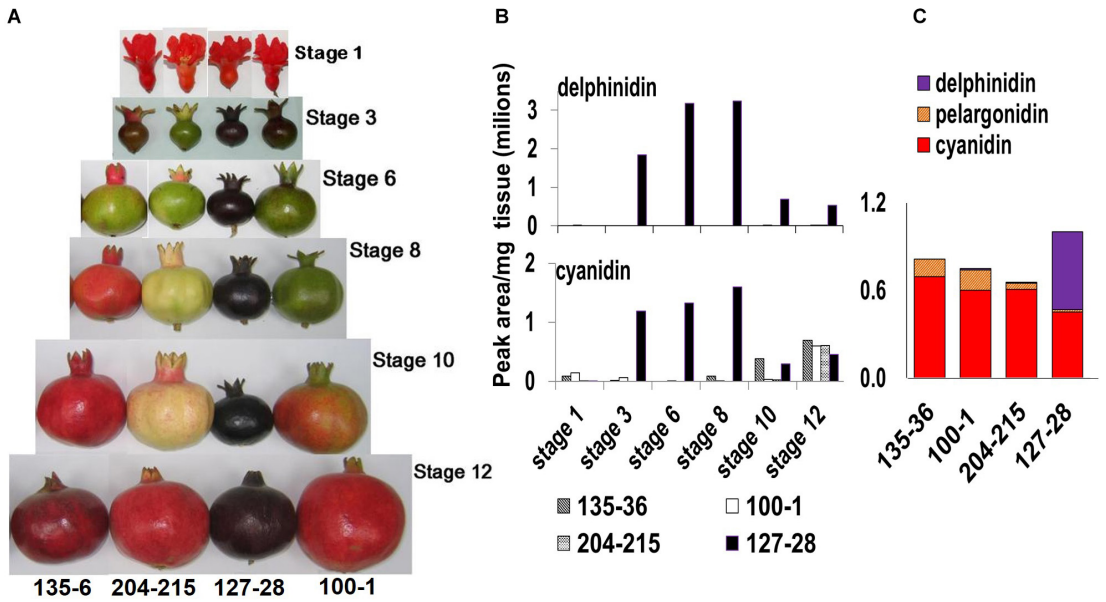

FIGURE 2 | Phenotypical and chemical characterization of the 'Black' pomegranate compared with various "red" pomegranate varieties. **(A)** Fruits of four pomegranate varieties at six different developmental stages (Ben-Simhon et al., 2011), from flower (stage 1) to fully mature fruit (stage 12). **(B)** Total level of delphinidin and its derivatives (purple color) and of cyanidin and its derivatives (red color) in the peel during the various developmental stages of the fruit in four pomegranate varieties. **(C)** Anthocyanin composition in the peel of ripe fruit- stage 12. The varieties shown are P.G.100-1, P.G.135-36, P.G.204-215, and P.G.127-28 ('Black').

For determining the timing of color accumulation in summer 2014, flowers of 'Black' and P.G.204-215 were hand pollinated and then photographed every 24 h (**Figure 3**).

The plant material used for the experiment investigating light influence on anthocyanin production, which was conducted in the field, is specified in **Supplementary Table 1**.

Phenotype Evaluation of Segregating Populations

Fruit peel color of the F_2 and F_3 populations was evaluated visually at ripening. Fruits were designated either "black" (black-purple peel color as in 'Black') or "red" (not black, including a largescale of pink- orange- red, pale to intense color). For quantitative analysis, three mature fruits from F_2 siblings (n = 82) were sampled and used for spectroscopic analysis of total anthocyanins.

DNA Extraction

DNA extraction was done according to Harel Beja et al. (2015) based on Porebski et al. (1997).

Reversed Phase HPLC Analysis of Anthocyanins

Sample preparation and analysis were carried out as previously described (Ben-Simhon et al., 2011). In short, methanolic extracts of pomegranate peel tissue were analyzed using a LaChrom Merck Hitachi HPLC system coupled with a diode array detector with 3D feature (Multiwavelength Detector, Jasco MD-2010 Plus), interface (Jasco LC-Net II/ADC) and scientific software (EZChrom Elite Client/Server version 3.1.6 build 3.1.6.2433). Extract aliquots were applied to a LiChrospher 100 RP-18 column with guard column (LichroCART cartridge, Merck Millipore), and eluted with a binary mobile phase consisting of phosphoric acid (0.1% v/v, pH 2.4) and acetonitrile. Anthocyanin identification and quantification were achieved using authentic standards. For each sampling stage, extract of peel tissue pooled from four to eight flowers/fruits from three replicate trees of each pomegranate variety was analyzed.

Spectroscopic Measurement of Total Anthocyanins

Total anthocyanin content was determined following Rabino and Mancinelli (1986). Frozen, grounded tissue (0.15 g fresh weight) was extracted in 1.5 mL of methanol containing 1% (v/v) HCl. The extract was centrifuged at 4°C, 14000 rpm for 10 min. Absorption of the extracts at wavelengths of 530 and 657 nm was determined photo metrically (CARY 50 Bio UV-visible Spectrophotometer, Agilent Technologies Inc., Santa Clara, CA, United States). When the absorption value exceeded 1.5, extracts were diluted with acidic methanol for the measurements. The anthocyanin content in the supernatant was calculated using the formula $A530 - 0.25^{*}A657$, allowing for the subtraction of chlorophyll interference. Results are shown as total anthocyanins per gram fresh weight. Each result is an average of three measurements.

Tissue Culture of Fruit Cells

Calli were developed from cultured ovary mesocarp cells of the 'Black' variety and a "red" variety (P.G.372-383). Unfertilized hermaphroditic flowers were collected. The flowers were surface

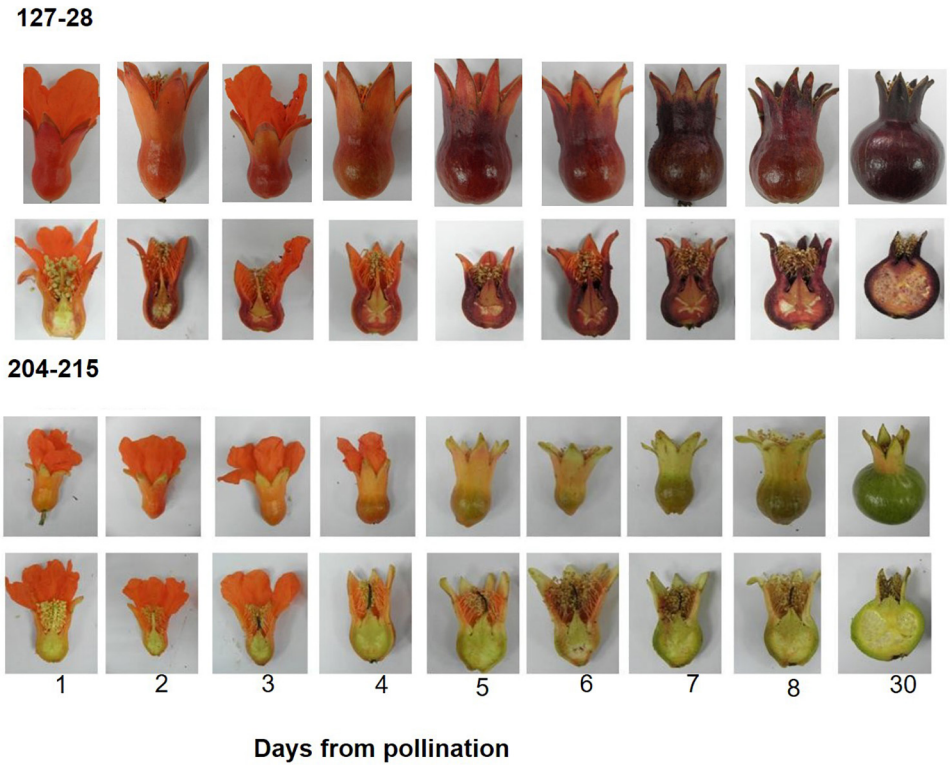

FIGURE 3 | Timing of color accumulation in peel skin following hand pollination in the 'Black' pomegranate (P.G. 127-28) and a "red" variety (P.G.204-215). Young fruitlets were collected every day after pollination for 8 days and then 30 days after pollination (stage 2). Whole and half fruitlets are presented in order to show color development.

sterilized (EtOH 70% for 5 min followed by three washes in sterilized water, sodium hypochlorite 0.25% for 5 min followed by three washes in sterilized water). Then pieces of the ovary mesocarp with ovules attached were excised and placed on a regeneration medium as described in Terakami et al. (2007): MS medium (Murashige and Skoog 1962) supplemented with 0.5 μM a-naphthalene acetic acid [NAA], 5 μM N6-benzyladenine [BA], and 0.3% phytagel (Sigma-Aldrich Corp., Milwaukee, WI, United States) at pH 5.8. The plates were kept in the dark (covered with aluminum foil) for 2 weeks to encourage callus formation, and then were subjected for further experiments. Growth room conditions were 24°C and 16 h light (fluorescent light).

Light Influence on Anthocyanin Accumulation

The effect of light on anthocyanin accumulation in pomegranate peels was studied in tissue culture of fruit cells and on whole fruits grown in the field.

Tissue Culture

Calli formed in the dark for 2 weeks were exposed to light. Half were kept in the dark for another 2 weeks. Then, the calli were photographed.

Whole Fruit

The experiment on intact fruits was conducted in the field. Fruits of three "black" peel accessions and 27 "red" peel accessions of the Newe Ya'ar pomegranate collection were covered with aluminum bags that do not allow light penetration. The fruits were covered in mid-July, 1–3 months before ripening, depending on the ripening date of each accession. The color of mature fruits from covered and uncovered controls was analyzed. Fruit color was evaluated visually and fruits were photographed. Total anthocyanin content was measured spectroscopically in peels of covered vs. uncovered mature fruits. The "black" peel and "red" peel accessions used are listed in **Supplementary Table 1**.

The experiment with the F_2 population included ten "red" peel progenies and ten "black" peel progenies. For each progeny about five young fruitlets were covered with aluminum bags during May through June, and five were marked and left uncovered. Upon ripening fruits were analyzed as described above.

Mapping the "Black" Trait in the F_2 Population

The "black" trait was mapped to the published pomegranate genetic map (Harel Beja et al., 2015). Mapping was performed using JoinMap 3.0 software (Van Ooijen and Voorrips, 2001). The software uses Kosambi mapping functions to translate recombination frequency into map distance. Markers were

grouped at a minimum LOD score of 4.0 and a recombination frequency value of 0.4.

MapQTL 5 software (Van Ooijen, 2004) was used for QTL analysis by interval mapping (IM), MQM, and for permutation analysis (1000 permutation, $p < 0.05$).

Genotyping With Markers Close to the "Black" Trait

Genetic markers that were mapped close to the "black" peel trait were blasted to the WGS of pomegranate cv. 'Dabenzi' (296 Mb) reported by Qin et al. (2017) using BLASTN (Altschul et al., 1990). Six SNP markers and two SSR markers were developed within the genomic region associated with the "black" trait (**Supplementary Table 2**).

The F_2 (n = 83) and F_3 (n = 240) segregating populations were genotyped with SSR markers by fluorescence labeling using 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, United States). The PCR mixture for SSR amplification contained: 30 ng plant genomic DNA, 0.2 pmol primers, 2X Taq red master mix (Apex Bioresearch Products, Genesee Scientific Corporation, El Cajon, CA, United States) in a total volume of 20 μl. PCR conditions: 96°C for 2 min, 30 cycles of 94°C for 15 s, 55°C for 30 s and 72°C for 30 s, followed by 72°C for 45 min.

SNP markers were genotyped by sequencing PCR products with 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, United States).

RNA-Seq of "black" vs. "red" Fruit

To further elucidate the "black" peel trait, RNA-seq analysis of fruit peels of very young fruit, i.e., after pollination and before stage 3, was performed. At this stage "black" fruits had already gained their color while "red" fruits were on the verge of becoming green (losing the red color) (**Figure 2**). Two varieties, 'Black' and 'Nana,' the parents of the F_2 population, as well as twelve representatives of the F_2 population (six with "red" peel and six with "black" peel) were selected for RNA-seq analysis. A sample consisted of the peel skin of ten fruits collected from different branches on the same tree. The tissues were crushed, immediately frozen in liquid nitrogen and kept at −80°C until further analysis. Total RNA was extracted using Norgen Plant/Fungi Total RNA Purification Kit (Norgen Biotek Corporation, Thorold, ON, Canada) including DNase treatment with RNase-free DNase I (Epicenter Biotechnologies, Madison, WI, United States) on the column, as described in the kit protocol, to remove genomic DNA. Total RNA was quantified spectrophotometrically (ND-1000 spectrophotometer, NanoDrop Technologies, Wilmington, DE, United States), and RNA integrity was examined on a 2200 TapeStation instrument using RNA ScreenTape sample buffer and RNA ScreenTapes (Agilent Technologies, Santa Clara, CA, United States). RNA samples for RNA-seq analysis consisted of three biological replicates of 'Black' and 'Nana' (three sets of fruit taken from the same tree). Construction and sequencing of mRNA-seq libraries were performed at the Crown Genomics Institute of the Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science (G-INCPM). Briefly, mRNA-seq libraries were generated from total RNA, using the INCPM mRNA Seq protocol. The indexed libraries were pooled and subjected to sequencing on an Illumina HiSeq4000 instrument (Illumina, San Diego, CA, United States).

Transcriptome Analysis

Raw-reads were subjected to a filtering and cleaning procedure. The SortMeRNA tool was used to filter out rRNA. Next, the FASTX Toolkit[2] (version 0.0.13.2) was used to trim read-end nucleotides with quality scores < 30, using the FASTQ Quality Trimmer, and to remove reads with less than 70% base pairs with a quality score ≤ 30 using the FASTQ Quality Filter. Clean reads were mapped to the reference genome of *Punica granatum*[3] using Tophat2 software v. 2.1 (Kim et al., 2013) with an average mapping rate of 95.8%. Gene abundance estimation was performed using Cufflinks v. 2.2 (Trapnell et al., 2010), combined with gene annotations[4]. Heatmap visualization was performed using R Bioconductor (Gentleman et al., 2004). Gene expression values were computed as FPKM (Fragments per kilo base per million mapped reads). Differential expression analysis was completed using the edgeR R package (Robinson et al., 2010). Genes that were more than twofold differentially expressed with false discovery-corrected statistical significance of at most 0.05 were considered differentially expressed. Venn diagrams were generated using the online tool at bioinformatics.psb.ugent.be/webtools/Venn/.

The gene sequences were used as a query term for a search of the NCBI non-redundant (nr) protein database that was carried out with the DIAMOND program (Buchfink et al., 2014). Homologous sequences were also identified by searching against the *Eucalyptus grandis* genome with the BLAST tool and an *E*-value threshold of 10^{-5}. The search results were imported into Blast2GO version 4.0 for GO assignments. GO enrichment analysis was carried out using Blast2GO program based on Fisher's Exact Test with multiple testing correction of false discovery rate (FDR). KOBAS 3.0 tool[5] was used to detect the statistical enrichment of differential expression genes in KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway and GO.

Sequence Analysis of *ANR*

ANR full length transcript sequences of 'Nana' and 'Black' were cloned and sequenced. Primers used for the full-length transcribed gene were p.g.ANR F1 and p.g.ANR R1 (**Supplementary Table 2**). In addition, a fragment of the gene was amplified with specific primers ANR4-F4 and ANR5-R5 (**Supplementary Table 2**). PCR products were sequenced with 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, United States). *In silico* characterization and phylogenetic analysis of *ANR* were performed using various online applications. Virtual translation of the gene was done

[2]http://hannonlab.cshl.edu/fastx_toolkit/index.html

[3]https://www.ncbi.nlm.nih.gov/assembly/GCA_002201585.1

[4]http://peppergenome.snu.ac.kr/download.php

[5]http://kobas.cbi.pku.edu.cn/kobas3/?t=1

using the Expasy proteomics server[6] (Artimo et al., 2012). NCBI Blastp was used to find homologous proteins and conserved domains[7]. Multiple alignment was done with MultAlin sequence alignment tool[8] (Corpet, 1988).

Statistical Analysis

Means, standard deviations and Wilcoxon/Kruskal–Wallis test (rank sums) analyses were conducted with the JMP program, v. 14.0 (SAS Institute Inc., Cary, NC, United States).

RESULTS

The 'Black' Pomegranate Contains Unusual High Content of Delphinidin and Cyanidin in the Fruit Peel Skin

The 'Black' pomegranate variety found in Israel (P.G.127-28) is characterized by a very distinct deep-purple peel skin ("black," **Figure 2A**). HPLC analysis of the anthocyanins in the peel revealed major differences between the 'Black' variety and three "red" varieties. Delphinidin (both of its forms, mono- and di-glycosidic) was identified at very high levels in the 'Black' peel skin (from here on referred to as peel). Delphinidin derivatives were responsible for the purple color. Delphinidin was not identified at all or in very minute amounts in the peel of the other three "red" varieties that were analyzed (**Figures 2B,C**). The 'Black' peel also accumulated a high level of cyanidin derivatives responsible for the red color (**Figures 2B, C**). The levels of cyanidin in the 'Black' fruit peel can reach up to two orders of magnitude as compared to other pomegranates' peel anthocyanin content (for example, at stage 8, cyanidin level was 1,598,822 peak area/mg tissue in 'Black' and only 2,723 in the "red' variety P.G.100-1). Moreover, the accumulation of anthocyanins in the 'Black' variety started already in the early stage 3 of fruit development (stages as described in Ben-Simhon et al., 2011) contrary to the "red" accessions analyzed in this study that did not accumulate anthocyanins at this stage. As can be seen in **Figure 2A**, 'Black' is "black" from stage 3 on, while the other varieties lose anthocyanins at this stage and regain color again later on in fruit development. The levels of delphinidin and cyanidin were reduced in the 'Black' variety at later stages, but the color of the fruit was not diluted visually. Still, total anthocyanin was 1.2- to 1.5-fold higher in the 'Black' than in the examined "red" varieties at stage 12 (**Figure 2C**). Total anthocyanins in 'Black' ripe fruits consisted of 54% delphinidin, 45% cyanidin and only 2% pelargonidin. In the "red" fruit the major anthocyanin is cyanidin, constituting 80–90% of total anthocyanins. Pelargonidin constitutes 7–18% and delphinidin is either absent or found in very small amounts (0–2%).

It seems that the unique accumulation of delphinidin together with the high levels of cyanidin and delphinidin anthocyanins accumulating from early stages of fruit development is responsible for the special "black" phenotype.

[6]http://www.expasy.org

[7]https://blast.ncbi.nlm.nih.gov/Blast.cgi

[8]http://multalin.toulouse.inra.fr/multalin

The 'Black' Pomegranate Starts Accumulating Color Right After Pollination

A daily follow up of color appearance since flower pollination during the development into young fruit (30 days, stage 2) clearly demonstrated that "black" color accumulation begins very early in fruit development (**Figure 3**). The "black" color already started to accumulate in the sepals of the 'Black' variety 1 day following pollination. From this stage on, the "black" color was maintained throughout fruit development. In the "red" variety, however, a few days after pollination the red color of the sepals faded and was replaced by the green color of the chlorophyll. "red" varieties accumulated anthocyanins at later stages of fruit development toward ripening (**Figure 2**), the rate and precise timing of accumulation varied between the different cultivars. Chlorophyll and carotenoid contents in peels of the 'Black' variety did not vary from the "red" accessions (data not shown). The difference in color can thus be attributed to anthocyanin content. The time of color appearance in 'Black' is another unique trait that differentiates it from other varieties.

Color Accumulation in the 'Black' Pomegranate Does Not Appear to Depend on Light

The effect of light on anthocyanin accumulation in pomegranate peels was studied in tissue culture of fruit cells and on whole fruits grown in the field. In the first experiment, calli developed from cultured ovary mesocarp cells taken from flowers of the 'Black' variety and a "red" accession P.G.372-383 were compared in response to light deprivation. Plates containing 4 weeks-old cultured calli were covered with aluminum foil to avoid light penetration. After 2 weeks the plates were transferred to regular light conditions. As can be seen in **Figure 4A**, the calli of the "red" variety lost their anthocyanins in the dark, while the calli originating from the 'Black' pomegranate remained colored. The "red" pomegranate calli accumulated red color as soon as 1 day after returning to light.

The second experiment was conducted in the field on intact fruits. Fruits of the 'Black' pomegranate, as well as 4 "black" peel varieties and 27 "red" varieties of the Newe Ya'ar pomegranate collection (**Supplementary Table 1**) were covered with aluminum bags that do not allow light penetration. The fruits were covered at stage 6, before red color developed; however, "black" fruits were already colored at that stage. At ripening "black" pomegranate fruits retained their deep purple color regardless of light exposure, while the "red" varieties lost their color in the dark (**Figure 4B**).

The effect of light deprivation was also evaluated quantitatively using spectrophotometric analysis of total anthocyanins in peels of covered vs. uncovered mature fruits. As shown in **Figure 4C**, total anthocyanins in the fruit peel of "red" varieties decreased up to 20-fold in the dark (P.G. 135-36; 5.46 in the light vs. 0.28 in the dark). By contrast, total anthocyanins in the fruit peel of 'Black' pomegranate increased in covered fruit.

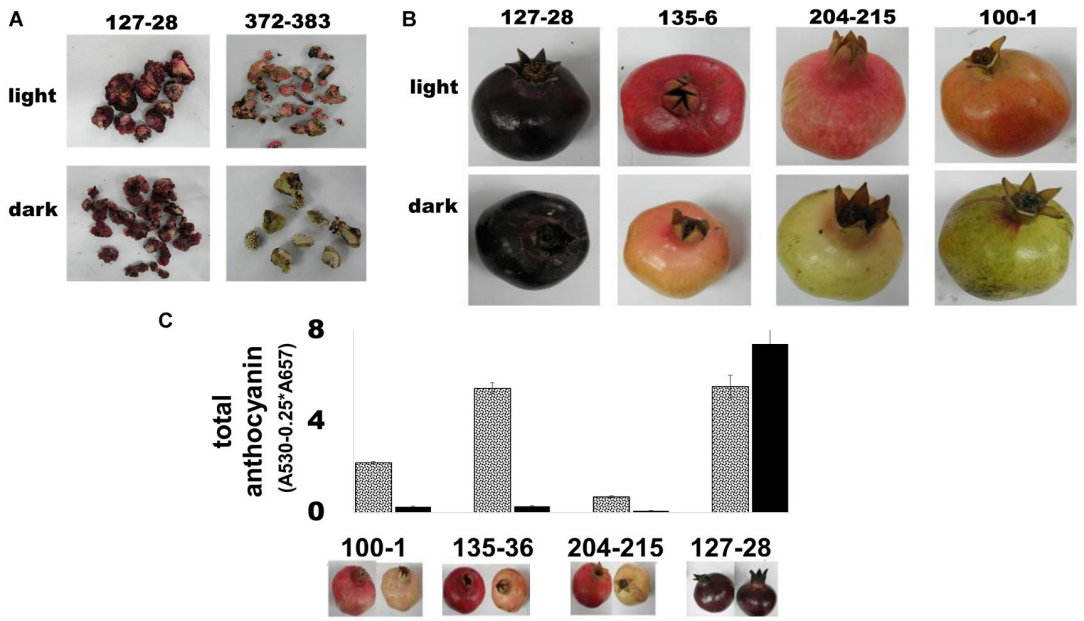

FIGURE 4 | Color accumulation in fruit and tissue culture of the 'Black' pomegranate in the absence of light. **(A)** Light influence on color development in calli derived from the ovary mesocarp of flowers from the 'Black' pomegranate and the "red" variety P.G.372-383. The calli were photographed after 2 weeks in complete darkness **(B)** Light influence on peel color in intact fruit. Fruits at stage 6 were covered with aluminum bags, which prevented any exposure to light. At ripening, covered fruits were compared to uncovered fruits. The figure shows three representative "red" varieties in addition to the 'Black' pomegranate. **(C)** Total amount of anthocyanin in the peel of fully ripe fruit grown in the dark (full black column) was compared with that of control uncovered fruit exposed to light (gray dotted column). Under each column there is an image that represents the uncovered fruit (left) and covered fruit (right).

Interestingly, insensitivity to light appears to be specific to color accumulation in the fruit as the etiolation response of seedlings of the 'Black' pomegranate was similar to that of the "red" pomegranates.

It seems that the 'Black' pomegranate lost its sensitivity to light with respect to color development in the fruit peel.

Inheritance of the "black" Phenotype in Pomegranate

We undertook a mapping approach to identify the mode of inheritance of the "black" phenotype and the gene(s) responsible for it. Initially, mapping studies were based on an F_2 'Nana × Black' population originating from a cross between the 'Black' pomegranate and the "red" variety 'Nana' (P.G.232-243). This population was used earlier to construct a genetic map of pomegranate and for QTL mapping of several important traits (Harel Beja et al., 2015). The F_2 population (204 siblings in total) segregated for the "black" and "red" phenotypes as follows: 23% (47) of the siblings showed a "black" phenotype and 77% (157) showed a "red" phenotype. These results together with the results of a 'Black' self-population in which all the siblings (n = 64) were "black," lead to the conclusion that the "black" phenotype is controlled by a single gene showing a Mendelian inheritance ratio of one gene (1:3 ratio). The peel of the "black" siblings contained a higher level of total anthocyanins compared to the "red" peel siblings, beside one "red" with very high anthocyanin content (**Figure 5A**).

Interestingly, the insensitivity of the "black" fruit color to light deprivation co-segregated with the "black" phenotype in the F_2 population. Statistical analysis (**Figure 5B**) showed no significant difference in total anthocyanins in covered vs. uncovered fruit in the "black" siblings. Among the "red" siblings, the covered fruits had a significantly (p = 0.0246) lower content of total anthocyanins, analyzed by the Wilcoxon/Kruskal–Wallis test (rank sums) and by non-parametric comparisons probability (F = 0.03). The association between peel color and the response to light strongly suggests that the insensitivity to light is controlled by the same gene that controls the "black" phenotype.

The "black" trait was mapped to LG2 on the published 'Nana × Black' F_2 genetic map (Harel Beja et al., 2015) based on SNP genetic markers described in Ophir et al. (2014) (**Figure 6A**). Total anthocyanins content was mapped as a major QTL in the same area on LG2 (on the region of marker c7881) with a LOD score of 19 (**Figure 6B**). Minor QTLs were analyzed by MQM in LG3, LG6, LG9 and LG10 (**Supplementary Table 3**).

F3'5'H, the gene responsible for synthesis of delphinidin, the purple anthocyanin color that characterizes the peel of the "black" phenotype (**Figure 1**), was initially considered as a likely site of the "black" mutation. However, it did not map to any of the above QTLs in the F_2 population. *F3'5'H* (**Supplementary Table 2**) was mapped to LG7 on the 'Nana × Black' F_2 genetic map (Harel Beja et al., 2015). Therefore, it is not the gene that controls the "black" phenotype.

The sequence of genetic markers that were mapped close to the "black" peel trait was blasted using BLASTN on the

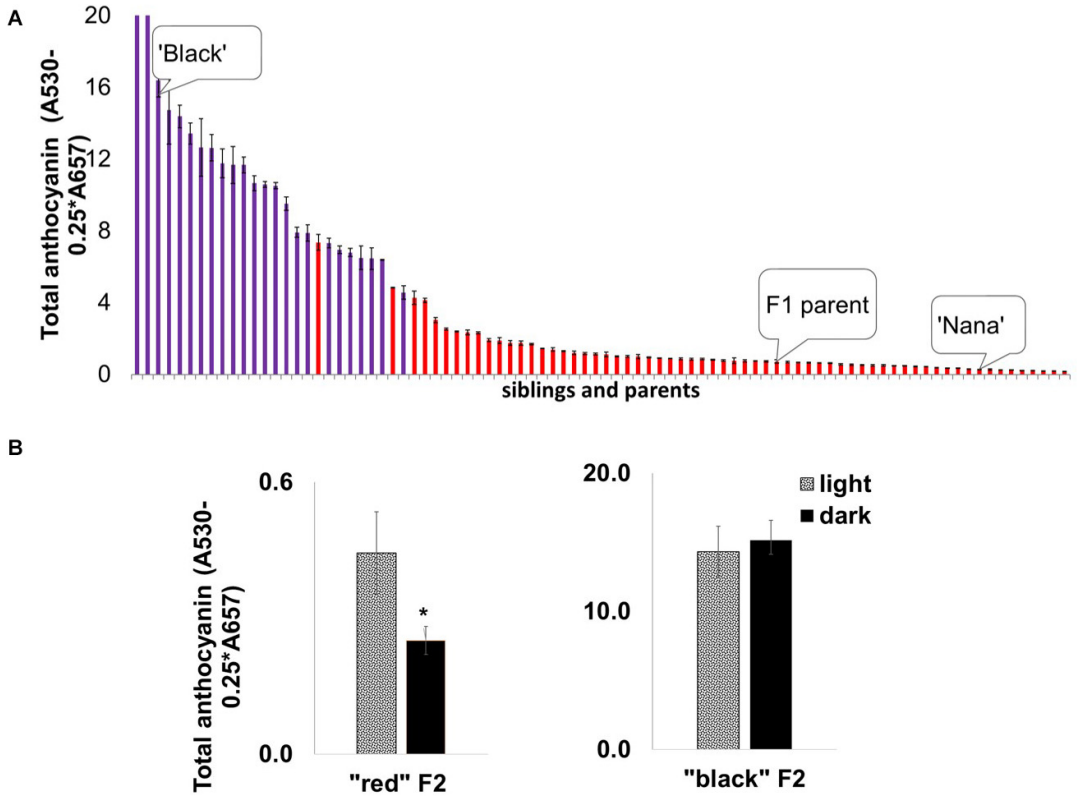

FIGURE 5 | Segregation of anthocyanin in mature fruit peels and the effect of depletion of light on peel color development in the 'Nana × Black' F_2 population. **(A)** Distribution of total anthocyanin amount in the F_2 progeny (n = 82) for "black" peel (purple columns) and "red" peel (red columns). The parents, 'Black' and 'Nana' and the F_1 offspring whose self-pollination created this F_2 population are marked. Each column represented in the graph is an average of three extracts ± standard error. **(B)** Effect of light deprivation on "red" and "black" representatives from the F_2 population: Total amount of anthocyanin in the peel of fruits developed in the dark (covered) compared with the control (light, uncovered fruits). Asterisk indicates a statistically significant difference ($p < 0.05$) analyzed by Wilcoxon/Kruskal–Wallis test (rank sums) and by non-parametric comparisons probability (F = 0.03).

WGS of pomegranate cv. 'Dabenzi' (296 Mb; Qin et al., 2017). This approach revealed a region of 1,263 kb on scaffold36 (MTKT01002534.1) that matched the region between the mapping markers bordering the trait (c8534-c17529, **Figure 6A**). This region contains 172 genes. Among them there are two annotated genes known to be involved in anthocyanin biosynthesis. One of the genes is a transcription factor *bHLH (CDL15_Pgr017019)* and the second gene is *ANR* (*CDL15_Pgr017032*; **Figure 1**). bHLH type proteins are known to be part of the complex that regulates the anthocyanin biosynthetic pathway (Petroni and Tonelli, 2011). ANR acts in the anthocyanin pathway, diverting anthocyanins to epicatechin (proanthocyanidins; **Figure 1**). For each of these genes an SNP and/or an SSR molecular marker was created (**Supplementary Table 2**).

In order to fine map the "black" trait in the c8534-c17529 region, and to study which of the candidate genes is associated to the trait, advanced F_3 populations were constructed (**Table 1**). Within 76 plants of the mapping F_2 population, there were seven recombination events at the associated c8534-c17529 region. However, to distinguish between the candidate genes, two plants that showed recombination between *ANR* and *bHLH* markers were selected. These two plants were selected for self-propagation to create F_3 populations that would show recombination within the "black" genomic region. One of the F_2 plants was heterozygous for the *ANR* marker and homozygous for the "red" allele of the *bHLH* marker (population 1). The other F_2 plant was the other way around, i.e., homozygous for the "red" allele of the *ANR* marker and heterozygous for the *bHLH* marker (population 2). We screened 120 plantlets of F_3 population 1 with the *ANR* marker. Homozygous recombinant plants that were homozygous for the "black" allele of *ANR* and homozygous for the "red" allele of *bHLH* were planted in the orchard. 120 plantlets of F_3 population 2 were screened with

TABLE 1 | The allelic composition of *ANR* and *bHLH* homozygous markers and the matching peel color in two F_3 populations.

F_3 population	*ANR* marker homozygous for "black" allele	*bHLH* marker homozygous for "black" allele	Peel color
F_3 population 1	12	0	"black"
F_3 population 2	0	7	"red"

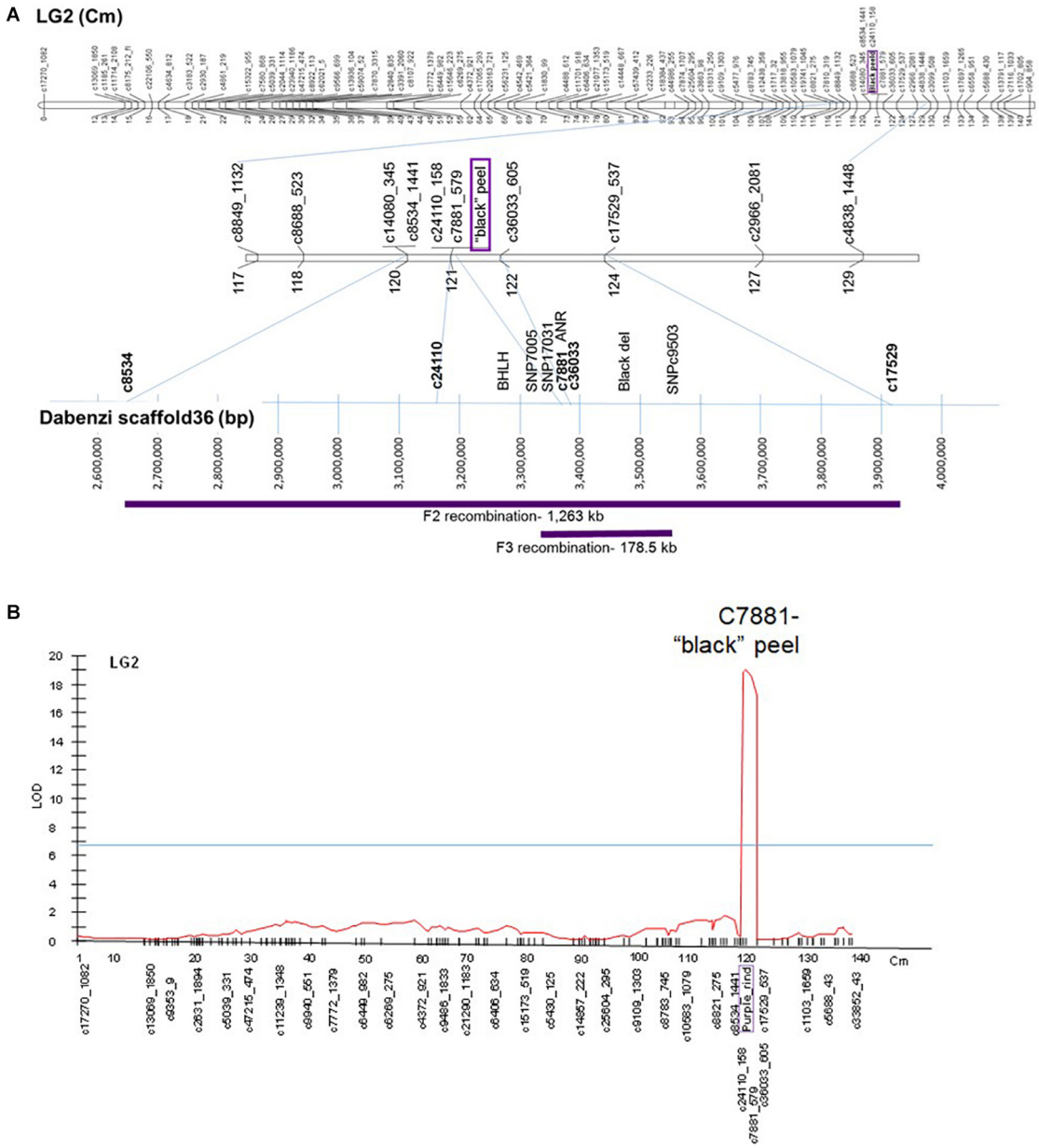

FIGURE 6 | Mapping the "black" trait in pomegranate. **(A)** On top - The "black" trait (in purple) was mapped to LG2 on the published 'Nana × Black' F_2 genetic map (Harel Beja et al., 2015). At the bottom - Positioning of the mapped genetic markers and additional markers on the published sequenced genome ('Dabenzi' scaffold36; Qin et al., 2017). The regions on the genome corresponding to the F_2 and F_3 recombinations are in purple. **(B)** QTL analysis (MQM) of total anthocyanins for the F_2 progeny at LG2.

TABLE 2 | Matching for F_2 and F_3 populations between the "black" trait and neighboring markers positioned on 'Dabenzi' scaffold 36 (Qin et al., 2017) sequence.

Position on Dabenzi scaffold 36	**2655494**	**3162458**	**3274725**	**3327005**	**3348965**	**3360974**	**3389282**	**3392436**	**3407657**	**3505550**	**3918578**
Marker	**c8534**	**c24110**	**bHLH**	**7005**	**17031**	**ANR**	**c7881**	**c36033**	**black del**	**c9503**	**c17529**
F_2 progeny	95	100	100	nd	nd	100	100	100	nd	nd	95
F_3 progeny*	nd	nd	0	0	100	100	100	100	100	91.7	91.7

*The "black" trait area for each population is highlighted in purple. The distances on 'Dabenzi' scaffold 36 between F_2 and F_3 recombinations are 1,263 kb and 178.5 kb, respectively. *F3 population1. nd, not determined. Values are expressed as percentage.*

the *bHLH* marker. Plants that were homozygous for the "red" allele of *ANR* and homozygous for the "black" allele of *bHLH* were planted in the orchard (**Table 1**). Therefore, there are two F_3 populations that differ at the "black" region genotype. Fruit peel color was described as "black" or "red" (**Table 1**). Twelve plants of F_3 population 1, that were homozygous for the "black" allele of *ANR* and homozygous for the "red" allele of *bHLH* had a "black" peel. None of the plants of this genotype had a "red" peel. On the other hand, all the seven plants of F_3 population 2 that were homozygous for the "red" allele of *ANR* and homozygous for the "black" allele of *bHLH* had "red" peel (**Table 1**). The recombination data (**Table 2**) clearly indicates that the "black" phenotype is 100% linked to the *ANR* gene but not to the *bHLH*, suggesting that the *ANR* gene might be the gene that confers the "black" phenotype. Nevertheless, other traits such as plant height and fruit size varied among the F_3 plants. Six additional DNA markers were developed between the *bHLH* and *ANR* markers (**Supplementary Table 2**). Genotype analysis of the F_3 plants from both populations with those markers showed additional 23 recombinations and reduced the site of the gene to a region of 178.5 kb on the draft pomegranate genome of 'Dabenzi' (**Figure 6A** and **Table 2**). This region contains 27 genes based on the sequenced pomegranate genomes (Qin et al., 2017; Yuan et al., 2017; Luo et al., 2020) and annotated genes published by Ophir et al. (2014) (**Supplementary Table 4**). Among those genes, *ANR* is the best candidate since it is known to function at the flavonoid pathway. Furthermore, it was differentially expressed between 'Black' and 'Nana' as is explained below.

Comparative Expression Analysis of Anthocyanin Related Genes in Fruit Peels of 'Black' and 'Nana' Fruits

Expression analysis was performed in order to explore the differences in the expression of the anthocyanin pathway-related genes between 'Black' and 'Nana.' For this purpose, RNA was extracted from fruit peels of fruitlets harvested before stage 3 (**Figure 7C**). At this stage, 'Black' fruits had already gained their color while 'Nana' fruits were on the verge of becoming green (losing the red color). Differences in expression were observed for structural genes of the anthocyanin pathway, including: *chalcone synthase* (*CHS*), *chalcone isomerase* (*CHI*), *flavanone 3-hydroxylase* (*F3H*), *dihydroflavonol reductase* (*DFR*), *flavonoid 3'5'-hydroxylase* (*F3'5'H*), *flavonoid 3'-hydroxylase* (*F3'H*), *LDOX*, *ANR* and *LAR* (**Figure 7A** and **Figure 1**). Surprisingly, the expression of these genes was higher in the peel of 'Nana' as compared to 'Black,' except for the *F3'5'H* gene.

Differential expression levels were observed for several *MYB* and *bHLH* genes (**Figure 7B**). These transcription factors are associated with anthocyanin biosynthesis and metabolism. Nevertheless, none of these genes was mapped to the "black" 178.5 kb region. Moreover, when the expression of the above structural or regulatory genes was analyzed at the same stage of development on fruit bulks from representative "black" or "red" fruits collected from the F_2 population, the linkage between the expression of a specific gene and the "black" trait was lost (**Supplementary Figure 1**).

There were differences in the expression level of 3,799 genes between the 'Black' peels and the 'Nana' peels. However, out of the 27 genes found in the 178.5 region on scaffold36, 23 genes were found to be transcribed in both the 'Black' pomegranate and the 'Nana' peels with no dramatic difference in gene expression except for the *ANR* gene (**Supplementary Table 4**). When the expression of these genes was compared between "red" and "black" representatives of the F_2 population, no correlation was seen except for the *ANR* gene, though the difference was less dramatic.

A Single Mutation in the *ANR* Gene Could Be Responsible for the "black" Phenotype

Comparative genomic sequencing of the putative *ANR* gene of "red" peel and "black" peel varieties (**Figure 8**) revealed a single base deletion unique to "black" peel varieties. This mutation co-segregates (100%) with the "black" phenotype in the F_2 population and among all "black" varieties in the Newe Ya'ar pomegranate collection (**Figure 8A** and **Table 3**). The single base deletion creates a (nonsense) mutation toward the end of the gene, thereby shortening the deduced protein by 25 aa (**Figures 8A,B**) and could potentially inactivate or reduce the activity of the ANR protein.

DISCUSSION

Pomegranate varieties characterized by "black" fruits are known from central and western Asia (Holland et al., 2009; Shams Ardekani et al., 2011; Özgüven et al., 2012; Zhu et al., 2015). Some work has been done concerning anthocyanin or flavonoid levels and composition in the peels of these "black" pomegranates (Shams Ardekani et al., 2011; Zhao et al., 2013; Zhu et al., 2015; Balli et al., 2020), as well as their health benefits (Dana et al., 2015). In agreement with our findings for the Israeli 'Black' pomegranate, the Persian 'Black pomegranate' peel has higher flavonoid content then other pomegranate cultivars (Shams Ardekani et al., 2011). The Chinese "black" peel variety 'Moshilio' was shown to have high levels of delphinidin and cyanidin in the peel compared to other pomegranates (Zhao et al., 2013; Zhu et al., 2015). This study undertook a genetic approach to identify the genetic components that are responsible for the "black" trait. This remarkable feature is characterized by accumulation of delphinidin together with unusual high levels of cyanidin and delphinidin in the fruit peel that accumulate from early stages of fruit development. By establishing and mapping an F_2 population originated by a crossbred with a pomegranate variety that carries the "black" phenotype, it was possible to demonstrate that this phenotype is controlled mainly by a single recessive gene located at LG2 according to the pomegranate genetic map (Harel Beja et al., 2015). The associated region was placed on the pomegranate genomic sequence ('Dabenzi' scaffold36 2,655,494–3,900,000; Qin et al., 2017). Two genes positioned at that region (3,360,974 and 3,274,725 on the 'Dabenzi' scaffold36), *ANR* and *bHLH*, respectively,

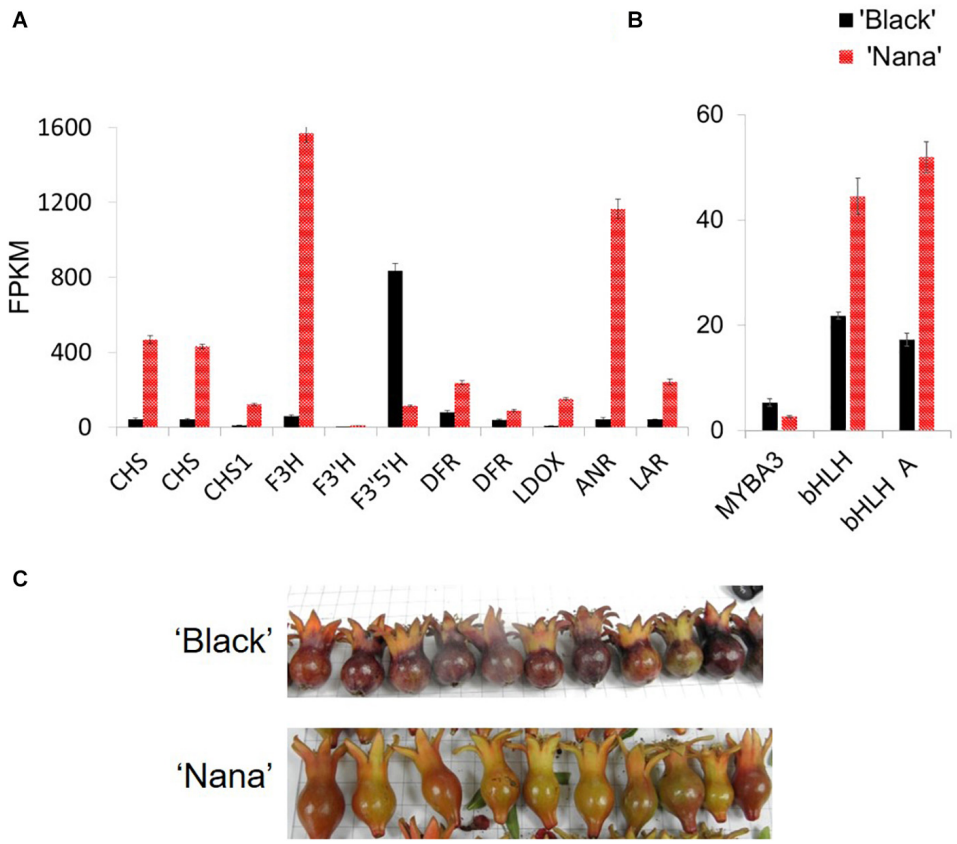

FIGURE 7 | Differential expression of genes from the anthocyanin pathway in 'Black' and 'Nana' young fruitlets. Peels were collected from fruits of 'Black' (black columns) and 'Nana' (red columns). Fruitlets were collected a few days after pollination. Expression analysis is explained in detail in the Materials and Methods section. **(A)** Genes corresponding to anthocyanin pathway. Gene names according to Qin et al. (2017): left to right: CDL15_Pgr005566 (CHS), CDL15_Pgr025723 (CHS), DL15_Pgr026373 (CHS1), CDL15_Pgr020918 (F3H), CDL15_Pgr008828 (F3'H), CDL15_Pgr026644 (F3'5'H), CDL15_Pgr021400 (DFR), CDL15_Pgr021400 (DFR), CDL15_Pgr017842 (LDOX), CDL15_Pgr017032 (ANR), CDL15_Pgr024128 (LAR). **(B)** Regulatory elements. From left to right- MYBA3 (JF747151), bHLH (CDL15_Pgr025), bHLH (JF747152). **(C)** Fruitlets analyzed showing the observed peel color at this stage. FPKM, fragments per kilo base per million mapped reads.

were annotated as genes that encode proteins that might function in the anthocyanin biosynthetic pathway. Fine mapping based on two additional F_3 populations generated from two different heterozygotic progenies of the F_2 population enabled us to reduce the "black" trait to a region of 178.5 kb. In addition, segregation analysis of the two F_3 populations demonstrated that the "black" phenotype segregated with the *ANR* "Black" genotype, hence excluding the *bHLH* as a possible cause for the trait and strongly suggesting *ANR* as the responsible gene. This assumption was further strengthened by surveying 100 genotypes of our pomegranate collection (Holland et al., 2009). The collection includes five different "black" varieties that differ with respect to taste, fruit size, time of ripening and size of the trees. Interestingly, only pomegranate accessions carrying the "black" phenotypes have a deletion of a single T nucleotide which deduces a shorter ANR protein (**Figure 8**). Thus, the single T deletion is in complete linkage with the "black" phenotype in the F_2 population, in two F_3 populations and within the pomegranate collection. Finally, expression analysis performed on fruit peel at an early developmental stage when "black" color accumulation begins, demonstrated that the *ANR* gene expression was significantly reduced in the 'Black' pomegranate as compared to the "red" pomegranate 'Nana.' This significant change in gene expression was only seen for the *ANR*, while the other 27 genes that are present within the 178.5 kb associated region did not show significant change in their expression between 'Black' and 'Nana' or between "red" and "black" phenotypes in the F2 population.

We do not exclude the possibility that other genes in the mapped area could be the genes responsible for the "black" phenotype. There are non-synonymous mutations in other genes in the mapped region (**Supplementary Table 5**). However, we strongly suggest that the mutation in the *ANR* gene is responsible for the "black" phenotype for the following reasons: (1) *ANR* belongs to the anthocyanin pathway and its inactivity can explain anthocyanin accumulation. (2) Expression data, which clearly show bold differences in its expression between 'Black' and 'Nana' (a "red" pomegranate). (3) The genotyping data is consistent with the phenotype in the collection and populations. (4) There is

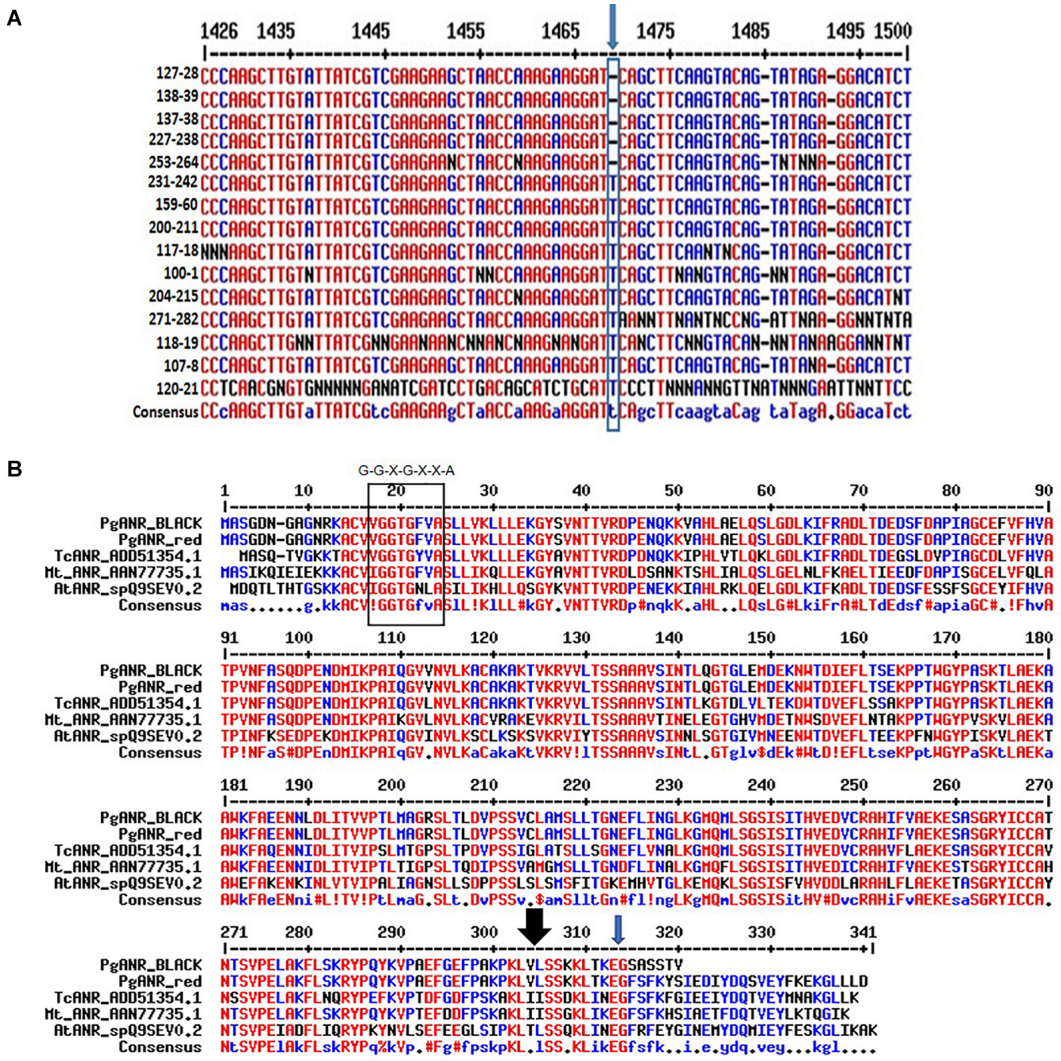

FIGURE 8 | Characterization of the *ANR* mutation. **(A)** The site and nature of the *ANR* mutation in the DNA sequence corresponding to the *ANR* gene. Multiple alignment of partial nucleotide sequence showing the mutation (blue arrow). Details of varieties shown are in **Supplementary Table 1**. **(B)** Multiple alignment of the deduced pomegranate proteins corresponding to ANR. The rectangle at the N terminus of the protein shows the conserved Rossmann dinucleotide-binding domain amino-termini (G-G-X-G-X-X-A). The black arrow at the C terminus of the protein shows the area of a putative substrate binding site. The ANR proteins shown in the alignment are- *Arabidopsis thaliana* AtANR_spQ9SEV, *Medicago truncatula* MtANR_AAN77735.1, *Gossypium hirsutum* GhANR_ABM64802.1 and *Glycine max* G.max_ANR2_NP_001243072.1. http://multalin.toulouse.inra.fr/multalin.

a clear non-synonymous mutation in the *ANR* gene (**Figure 8**) which shows a difference between all "black" and non-black ("red") phenotypes that were analyzed. (5) There are non-synonymous mutations in other genes in the mapped region, but they are either expressed in a very low level or show a similar expression level between "black" and "red' phenotypes.

We hypothesize that the mutation in the *ANR* gene shortened the ANR enzyme (**Figure 8B**), and as a result affected its activity. ANR belongs to a group of flavonoid reductases of the extended SDR-type (NCBI, Blastp, conserved domains: Lu et al., 2020). Extended SDRs have a C-terminal extension of approximately 100 aa that is less conserved (Moummou et al., 2012; Yonekura-Sakakibara et al., 2019). Moreover, a substrate binding domain is located in the C-terminal region, which determines specificity (Petit et al., 2007). The pomegranate ANR protein is 339 aa long, in accordance with other known ANRs (**Figure 8B**). One of the putative substrate binding sites is at the 304 aa position (lysine 304). Interestingly, the "black" ANR is mutated at the C terminal at aa 314. The mutation causes a frameshift that shortens the protein to a predicted open reading frame encoding 319 aa. The predicted changes at the C terminal of the protein may cause different folding due to aa change or shortening of the protein, and as such may affect the binding of the substrate and the activity of the enzyme.

ANR competes for the same substrate with glycosyltransferase enzymes. While glycosyltransferases convert anthocyanidins

into their mono- and di-glycosylated colored forms, ANR diverts the biosynthetic activity toward the production of proanthocyanidins, which are colorless (Xie et al., 2003; Kitamura et al., 2016). Hence, the *ANR* mutation could affect color production by reducing competition over anthocyanidins formation, resulting in overproduction of anthocyanin (**Figure 1**). It was shown that overexpression of *ANR* negatively affected color production in transgenic plants as *Vitis vinifera* (Bogs et al., 2005), *Theobroma cacao* (Liu et al., 2013), *Malus domestica* (Han et al., 2012) and *Fragaria chiloensis* (Salvatierra et al., 2013). The "black" phenotype in pomegranate is the first natural mutation in *ANR* published, suggesting a novel regulatory pathway that controls color production in fruit by partially or completely blocking a competing pathway. Mutations that cause color accumulation by blocking successive biosynthetic pathways were reported in the carotenoid pathway in fruit of the Cucurbitaceae family (Feder et al., 2015; Chayut et al., 2017).

Genes annotated as related to the anthocyanin pathway were expressed at a much lower level in 'Black' pomegranate as compared to the "red" pomegranate 'Nana' (**Figure 7**). These differences in expression level were not found in fruit bulks from representative "black" or "red" fruits of the F_2 population (**Supplementary Figure 1**). This may reflect a feedback mechanism whereby high anthocyanin accumulation in the 'Black' pomegranate generates a negative feedback mechanism which reduces transcription of most of the anthocyanin pathway. We found one exception to this, which is the *F3'5'H* gene (**Figure 7A**.). Intriguingly, this gene is responsible for delphinidin biosynthesis (Tanaka and Brugliera, 2013). Nevertheless, the *F3'5'H* gene was not mapped with the "black" trait, and therefore it is not the gene controlling the "black" phenotype.

TABLE 3 | The ANR mutation is unique to the "black" varieties in the Newe Ya'ar pomegranate collection.

Pomegranate variety	Peel color	*ANR* mutation
100-1	Red	No
118-19	Red	No
120-21	Red	No
271-282	Red	No
107-8	Red	No
117-18	Red	No
200-211	Red	No
204-215	Red	No
231-242	Red	No
159-60	Red	No
127-28 'Black'	Black	Yes
137-38	Black	Yes
138-39	Black	Yes
227-238	Black	Yes
253-264	Black	Yes

Only part of the red varieties analyzed are shown. All "black" varieties are shown.

The reduction of delphinidin and cyanidin levels toward ripening (stage 12; **Figure 2B**) can be explained as a dilution effect caused by low anthocyanin biosynthesis accompanied by a dramatic increase in the size of the fruit (**Figure 2A**). In addition, former experiments indicate that the expression of structural genes involved in the anthocyanin biosynthetic pathway are very low at ripening stage of the 'Black' pomegranate (Ben-Simhon, 2015).

The specific high accumulation of cyanidin and delphinidin in the fruit peel and not in arils of the "black" phenotype is of interest. Since we found only one single gene encoding for ANR, we propose that the anthocyanin pathway in the arils and peels could be controlled differently. This assumption is corroborated by the fact that there is no correlation between the outer skin color of the rind and the color of the arils (Holland et al., 2009). For example, in the arils, the biosynthetic pathway that diverts anthocyanins to proanthocyanidins may not be active and therefore a mutation in *ANR* does not influence color accumulation in the arils.

Another aspect of the "black" phenotype is the insensitivity of color accumulation in response to light exposure. Numerous studies have shown that the biosynthesis of anthocyanins in maturing fruit such as apples or grapes is a light-dependent process (Takos et al., 2006; Azuma et al., 2012). As indicated in **Figure 4**, color accumulation in the calli and peel of 'Black' pomegranate was not affected by light deprivation. Light-independence was specific to color accumulation in the fruit and was not reflected in other light-dependent responses such as in etiolated germinating seedlings. Thus, elongated hypocotyls, shortened roots, and small, closed cotyledons were evident in 'Black' self-seedlings as well as in "red" self-seedlings when exposed to darkness. We have demonstrated by genetics that light-independence is linked to the "black" phenotype (**Figure 5B**). How can a mutation in *ANR* explain such a phenotype? Within the genomic associated region there may be an additional genetic component that controls sensitivity to light and due to its close proximity to the *ANR* gene on the pomegranate genome they co-segregate in the population. Perhaps this gene is the *bHLH* gene which was found in the mapped area (**Figure 6A**). This gene has a mutation that is unique to the "black" phenotype in F_2 siblings as well as in the pomegranate collection. However, as discussed before, we found recombinants in the F_3 populations which show that the *bHLH* gene does not fully segregate with the "black" phenotype. Hence, there is a possibility that the "black" phenotype could be a combination of two or more gene mutations which are closely linked to each other. Light-independence of fruit development, in addition to color, is important for the protection of the fruit and influences its commercial value. Developing fruit color without dependence on light is a most desirable trait that could increase the quality of fruit and enable colored fruit production under light-limiting conditions (e.g., in the shaded part of the tree or under nets). Therefore, deciphering the "black" mutation is of high importance in finding new methodologies for breeding and manipulating color accumulation in fruits.

DATA AVAILABILITY STATEMENT

Data of this project have been deposited with links to BioProject accession number PRJNA694423 in The National Center for Biotechnology Information (NCBI). BioSamples accession numbers: SAMN17522954, SAMN17523170, SAMN17523467, SAMN17523758 for 'Black,' 'Nana,' "red" peel bulks and "black" peel bulks, respectively. The Raw data from the Illumina sequencing have been deposited in NCBI Sequence Read Archive (SRA) under accession numbers: SRR13530002; SRR13530001; SRR13529998; SRR13529997; SRR13529996; SRR13529995; SRR13529994; SRR13529993; SRR13529992; SRR13529991; SRR13530000; SRR13529999.

AUTHOR CONTRIBUTIONS

TT performed the experiments, analyzed the study, and wrote the manuscript. RH-B performed the experiments and analyzed the study. IB-Y, ZB-S, and RY performed the experiments. HB-N performed chemical analyses. RO, AS, and AD-F contributed to bioinformatics. DH designed and conducted the research. All authors contributed to the article and approved the submitted version.

FUNDING

The study was supported by the Chief Scientist of the Israeli Ministry of Agriculture and Rural Development (Grant No. 203-0856-19). Chemical analysis was supported by Keren Kayemeth LeIsrael–Jewish National Fund (KKL-JNF).

ACKNOWLEDGMENTS

The authors thank Mr. Kamel Hatib for maintenance of the pomegranate trees in Newe Ya'ar.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.642019/full#supplementary-material

REFERENCES

Albert, N. W., Davies, K. M., Lewis, D. H., Zhang, H., Montefiori, M., Brendolise, C., et al. (2014). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell* 26, 962–980. doi: 10.1105/tpc.113.122069

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

Arlotta, C., Puglia, G. D., Genovese, C., Toscano, V., Karlova, R., Beekwilder, J., et al. (2020). MYB5-like and bHLH influence flavonoid composition in pomegranate. *Plant Sci.* 298:110563. doi: 10.1016/j.plantsci.2020.110563

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 40, W597–W603. doi: 10.1093/nar/gks400

Azuma, A., Yakushiji, H., Koshita, Y., and Kobayashi, S. (2012). Flavonoid biosynthesis-related genes in grape skin are differentially regulated by temperature and light conditions. *Planta* 236, 1067–1080. doi: 10.1007/s00425-012-1650-x

Balli, D., Cecchi, L., Khatib, M., Bellumori, M., Cairone, F., Carradori, S., et al. (2020). Characterization of arils juice and peel decoction of fifteen varieties of *Punica granatum* L.: a focus on anthocyanins, ellagitannins and polysaccharides. *Antioxidants* 9:238.

Ban, Y., Honda, C., Hatsuyama, Y., Igarashi, M., Bessho, H., and Moriguchi, T. (2007). Isolation and functional analysis of a MYB transcription factor gene that is a key regulator for the development of red coloration in apple skin. *Plant Cell Physiol.* 48, 958–970. doi: 10.1093/pcp/pcm066

Bar-Ya'akov, I., Tian, L., Amir, R., and Holland, D. (2019). Primary metabolites, anthocyanins, and hydrolyzable tannins in the pomegranate fruit. *Front Plant Sci.* 10:620. doi: 10.3389/fpls.2019.00620

Ben-Simhon, Z. (2015). *Studying the Genetic-Molecular Components that are Responsible for Anthocyanin Biosynthesis in Pomegranate (Punica granatum L.)*. Ph.D. dissertation, The Technion-Israel Institute of Technology, Haifa.

Ben-Simhon, Z., Judeinstein, S., Nadler-Hassar, T., Trainin, T., Bar-Ya'akov, I., Borochov-Neori, H., et al. (2011). A pomegranate (*Punica granatum* L.) WD40-repeat gene is a functional homologue of Arabidopsis *TTG1* and is involved in the regulation of anthocyanin biosynthesis during pomegranate fruit development. *Planta* 234, 865–881. doi: 10.1007/s00425-011-1438-4

Ben-Simhon, Z., Judeinstein, S., Trainin, T., Harel-Beja, R., Bar-Ya'akov, I., Borochov-Neori, H., et al. (2015). A "White" anthocyanin-less pomegranate (*Punica granatum* L.) caused by an insertion in the coding region of the Leucoanthocyanidin Dioxygenase (*LDOX*; *ANS*) gene. *PLoS One* 10:e0142777. doi: 10.1371/journal.pone.0142777

Bogs, J., Downey, M., Harvey, J., Ashton, A., Tanner, G., and Robinson, S. (2005). Proanthocyanidin synthesis and expression of genes encoding leucoanthocyanidin reductase and anthocyanidin reductase in developing grape berries and grapevine Leaves. *Plant Physiol.* 139, 652–663. doi: 10.1104/pp.105.064238

Borochov-Neori, H., Judeinstein, S., Harari, M., Bar-Ya'akov, I., Patil, B. S., Lurie, S., et al. (2011). Climate effects on anthocyanin accumulation and composition in the pomegranate (*Punica granatum* L.) fruit arils. *J. Agric. Food. Chem.* 59, 5325–5334. doi: 10.1021/jf2003688

Borochov-Neori, H., Judeinstein, S., Tripler, E., Harari, M., Greenberg, A., Shomer, I., et al. (2009). Seasonal and cultivar variations in antioxidant and sensory quality of pomegranate (*Punica granatum* L.) fruit. *J. Food Compost. Anal.* 22, 189–195. doi: 10.1016/j.jfca.2008.10.011

Borochov-Neori, H., Judeinstein, S., Tripler, E., Holland, D., and Lazarovitch, N. (2014). Salinity effects on colour and health traits in the pomegranate (*Punica granatum* L.) fruit peel. *Int. J. Postharvest. Technol. Innov.* 4:54. doi: 10.1504/IJPTI.2014.064145

Boss, P. K., Davies, C., and Robinson, S. P. (1996). Analysis of the expression of anthocyanin pathway genes in developing *Vitis vinifera* L.cv Shiraz grape berries and the implications for pathway regulation. *Plant Physiol.* 111:1059. doi: 10.1104/pp.111.4.1059

Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12:59. doi: 10.1038/nmeth.3176

Chayut, N., Yuan, H., Ohali, S., Meir, A., Sa'ar, U., Tzuri, G., et al. (2017). Distinct mechanisms of the ORANGE protein in controlling carotenoid flux. *Plant Physiol.* 173:376. doi: 10.1104/pp.16.01256

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 16, 10881–10890. doi: 10.1093/nar/16.22.10881

Cutanda-Perez, M. C., Ageorges, A., Gomez, C., Vialet, S., Terrier, N., Romieu, C., et al. (2009). Ectopic expression of *VlmybA1* in grapevine activates a narrow set of genes involved in anthocyanin synthesis and transport. *Plant Mol. Biol.* 69, 633–648. doi: 10.1007/s11103-008-9446-x

Dana, N., Javanmard, S. H., and Rafiee, L. (2015). Antiangiogenic and antiproliferative effects of black pomegranate peel extract on melanoma cell line. *Respharm. Sci.* 10, 117–124.

Du, C., Wang, P., and Francis, F. (2006). Anthocyanins of pomegranate, *Punica granatum*. *J. Food Sci.* 40, 417–418. doi: 10.1111/j.1365-2621.1975.tb02217.x

Feder, A., Burger, J., Gao, S., Lewinsohn, E., Katzir, N., Schaffer, A. A., et al. (2015). A Kelch domain-containing F-Box coding gene negatively regulates flavonoid accumulation in Muskmelon. *Plant Physiol.* 169:1714.

Gao, J., Shen, L., Yuan, J., Zheng, H., Su, Q., Yang, W., et al. (2020). Correction to: functional analysis of GhCHS, GhANR and GhLAR in colored fiber formation of *Gossypium hirsutum* L. *BMC Plant Biol.* 20:28. doi: 10.1186/s12870-019-2211-2

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5:R80. doi: 10.1186/gb-2004-5-10-r80

Gil, M. I., García-Viguera, C., Artés, F., and Tomás-Barberán, F. A. (1995). Changes in pomegranate juice pigmentation during ripening. *J. Sci. Food Agric.* 68, 77–81. doi: 10.1002/jsfa.2740680113

Han, Y., Vimolmangkang, S., Soria-Guerra, R. E., and Korban, S. S. (2012). Introduction of apple *ANR* genes into tobacco inhibits expression of both *CHI* and *DFR* genes in flowers, leading to loss of anthocyanin. *J. Exp. Bot.* 63, 2437–2447. doi: 10.1093/jxb/err415

Harel Beja, R., Sherman, A., Rubinstein, M., Eshed, R., Bar-Ya'akov, I., Trainin, T., et al. (2015). A novel genetic map of pomegranate based on transcript markers enriched with QTLs for fruit quality traits. *Tree Genet. Genom.* 11:109. doi: 10.1007/s11295-015-0936-0

Harel-Beja, R., Tian, L., Freilich, S., Habashi, R., Borochov-Neori, H., Lahav, T., et al. (2019). Gene expression and metabolite profiling analyses of developing pomegranate fruit peel reveal interactions between anthocyanin and punicalagin production. *Tree Genet. Genom.* 15:22. doi: 10.1007/s11295-019-1329-6

Heber, D., Schulman, R. N., and Seeram, N. P. (2006). *Pomegranates: Ancient Roots to Modern Medicine*. Boca Raton, FL: CRC Press.

Hernández, F., Melgarejo, P., Tomás-Barberán, F. A., and Artés, F. (1999). Evolution of juice anthocyanins during ripening of new selected pomegranate (*Punica granatum*) clones. *Eur. Food Res. Technol.* 210, 39–42. doi: 10.1007/s002170050529

Holland, D., and Bar-Ya'akov, I. (2014). "Pomegranate: aspects concerning dynamics of health beneficial phytochemicals and therapeutic properties with respect to the tree cultivar and the environment," in *Medicinal and Aromatic Plants of the Middle-East*, eds Z. Yaniv and N. Dudai (Dordrecht: Springer Netherlands), 225–239.

Holland, D., Hatib, K., and Bar-Ya'akov, I. (2009). Pomegranate: botany, horticulture, breeding. *Horticult. Rev.* 35, 127–191.

Holton, T., and Cornish, E. (1995). Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* 7, 1071–1083. doi: 10.1105/tpc.7.7.1071

Jaakola, L. (2013). New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends Plant Sci.* 18, 477–483. doi: 10.1016/j.tplants.2013.06.003

Khaksar, G., Sayed-Tabatabaei, B., Arzani, A., Ghobadi, C., and Ebrahimie, E. (2015). Functional analysis of a pomegranate (*Punica granatum* L.) MYB transcription factor involved in the regulation of anthocyanin biosynthesis. *Iran. J. Biotechnol.* 13, 17–25. doi: 10.15171/ijb.1045

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14: R36. doi: 10.1186/gb-2013-14-4-r36

Kitamura, S., Oono, Y., and Narumi, I. (2016). Arabidopsis *pab1*, a mutant with reduced anthocyanins in immature seeds from banyuls, harbors a mutation in the MATE transporter *FFT*. *Plant Mol. Biol.* 90, 7–18. doi: 10.1007/s11103-015-0389-8

Koes, R., Verweij, W., and Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* 10, 236–242. doi: 10.1016/j.tplants.2005.03.002

Liu, Y., Shi, Z., Maximova, S., Payne, M. J., and Guiltinan, M. J. (2013). Proanthocyanidin synthesis in *Theobroma cacao*: genes encoding anthocyanidin synthase, anthocyanidin reductase, and leucoanthocyanidin reductase. *BMC Plant Biol.* 13:202. doi: 10.1186/1471-2229-13-202

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi: 10.1093/nar/gkz991

Luo, X., Li, H., Wu, Z., Yao, W., Zhao, P., Cao, D., et al. (2020). The pomegranate (*Punica granatum* L.) draft genome dissects genetic divergence between soft- and hard-seeded cultivars. *Plant Biotechnol. J.* 18, 955–968. doi: 10.1111/pbi.13260

Mancinelli, A. L. (1985). Light-dependent anthocyanin synthesis: a model system for the study of plant photomorphogenesis. *Bot. Rev.* 51, 107–157. doi: 10.1007/BF02861059

Moummou, H., Kallberg, Y., Tonfack, L., Persson, B., and van der Rest, B. (2012). The Plant short-chain dehydrogenase (SDR) superfamily: genome-wide inventory and diversification patterns. *BMC Plant Biol.* 12:219. doi: 10.1186/1471-2229-12-219

Ophir, R., Sherman, A., Rubinstein, M., Eshed, R., Sharabi Schwager, M., Harel-Beja, R., et al. (2014). Single-nucleotide polymorphism markers from de-novo assembly of the pomegranate transcriptome reveal germplasm genetic diversity. *PLoS One* 9:e88998. doi: 10.1371/journal.pone.0088998

Özgüven, A. I., Yılmaz, C., and Keleş, D. (2012). Pomegranate biodiversity and horticultural management. *Acta Hortic.* 940, 21–28. doi: 10.17660/ActaHortic.2012.940.1

Petit, P., Granier, T., d'Estaintot, B. L., Manigand, C., Bathany, K., Schmitter, J. M., et al. (2007). Crystal structure of grape dihydroflavonol 4-reductase, a key enzyme in flavonoid biosynthesis. *J. Mol. Biol.* 368, 1345–1357. doi: 10.1016/j.jmb.2007.02.088

Petroni, K., and Tonelli, C. (2011). Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci. Int. J. Exp. Plant Biol.* 181, 219–229. doi: 10.1016/j.plantsci.2011.05.009

Porebski, S., Bailey, L. G., and Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* 15, 8–15. doi: 10.1007/Bf02772108

Qin, G., Xu, C., Ming, R., Tang, H., Guyot, R., Kramer, E. M., et al. (2017). The pomegranate (*Punica granatum* L.) genome and the genomics of punicalagin biosynthesis. *Plant J.* 91, 1108–1128. doi: 10.1111/tpj.13625

Rabino, I., and Mancinelli, A. L. (1986). Light, temperature, and anthocyanin production. *Plant Physiol.* 81, 922–924. doi: 10.1104/pp.81.3.922

Robinson, M., McCarthy, D., and Smyth, G. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.

Rouholamin, S., Zahedi, B., Nazarian-Firouzabadi, F., and Saei, A. (2015). Expression analysis of anthocyanin biosynthesis key regulatory genes involved in pomegranate (*Punica granatum* L.). *Sci. Hortic.* 186, 84–88. doi: 10.1016/j.scienta.2015.02.017

Salvatierra, A., Pimentel, P., Moya-León, M. A., and Herrera, R. (2013). Increased accumulation of anthocyanins in *Fragaria chiloensis* fruits by transient suppression of *FcMYB1* gene. *Phytochemistry* 90, 25–36. doi: 10.1016/j.phytochem.2013.02.016

Seeram, N. P., Zhang, Y., Reed, J. D., Krueger, C. G., and Vaya, J. (2006). "Pomegranate phytochemicals," in *Pomegranates: Ancient Roots to Modern Medicine*, eds N. P. Seeram, R. H. Schulman, and D. Heber (New York, NY: Taylor and Francis Group), 3–29.

Shams Ardekani, M. R., Hajimahmoodi, M., Oveisi, M. R., Sadeghi, N., Jannat, B., Ranjbar, A. M., et al. (2011). Comparative antioxidant activity and total flavonoid content of Persian pomegranate (*Punica granatum* L.) Cultivars. *Iran. J. Pharm. Res.* 10, 519–524.

Takos, A. M., Jaffé, F. W., Jacob, S. R., Bogs, J., Robinson, S. P., and Walker, A. R. (2006). Light-induced expression of a *MYB* gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol.* 142, 1216–1232. doi: 10.1104/pp.106.088104

Tanaka, Y., and Brugliera, F. (2013). Flower colour and cytochromes P450. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 368:20120432. doi: 10.1098/rstb.2012.0432

Terakami, S., Matsuta, N., Yamamoto, T., Sugaya, S., Gemma, H., and Soejima, J. (2007). Agrobacterium-mediated transformation of the dwarf pomegranate (*Punica granatum* L. var. nana). *Plant Cell Rep.* 26, 1243–1251. doi: 10.1007/s00299-007-0347-2

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals

unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Tzulker, R., Glazer, I., Bar-Ilan, I., Holland, D., Aviram, M., and Amir, R. (2007). Antioxidant activity, polyphenol content, and related compounds in different fruit juices and homogenates prepared from 29 different pomegranate accessions. *J. Agric. Food Chem.* 55, 9559–9570. doi: 10.1021/jf071413n

Van Ooijen, J. W. (2004). *MapQTL §5, Software for the Mapping of Quantitative Trait Loci in Experimental Populations*. Wageningen: Kyazma B.V.

Van Ooijen, J. W., and Voorrips, R. E. (2001). *JoinMap §3.0, Software for the Calculation of Genetic Linkage Maps*. Wageningen: Plant Research International.

Winkel-Shirley, B. (2001). Flavonoid Biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 126:485.

Xie, D.-Y., Sharma, S., and Dixon, R. (2004). Anthocyanidin reductases from *Medicago truncatula* and *Arabidopsis thaliana*. *Archbiochembiophys* 422, 91–102. doi: 10.1016/j.abb.2003.12.011

Xie, D.-Y., Sharma, S., Paiva, N., Ferreira, D., and Dixon, R. (2003). Role of anthocyanidin reductase, encoded by *BANYULS* in plant flavonoid biosynthesis. *Science* 299, 396–399. doi: 10.1126/science.1078540

Yonekura-Sakakibara, K., Higashi, Y., and Nakabayashi, R. (2019). The origin and evolution of plant flavonoid metabolism. *Front. Plant Sci.* 10:943. doi: 10.3389/fpls.2019.00943

Yuan, Z., Fang, Y., Zhang, T., Fei, Z., Han, F., Liu, C., et al. (2017). The pomegranate *(Punica granatum* L.) genome provides insights into fruit quality and ovule developmental biology. *Plant Biotechnol. J.* 16, 1363–1374. doi: 10.1111/pbi.12875

Zhao, X., Yuan, Z., Fang, Y., Yin, Y., and Feng, L. (2013). Characterization and evaluation of major anthocyanins in pomegranate (*Punica granatum* L.) peel of different cultivars and their development phases. *Eur. Food Res. Technol* 236, 109–117. doi: 10.1007/s00217-012-1869-6

Zhu, F., Yuan, Z., Zhao, X., Yin, Y., and Feng, L. (2015). Composition and contents of anthocyanins in different pomegranate cultivars. *Acta Hortic.* 1089, 35–41. doi: 10.17660/ActaHortic.2015.1089.3

Conflict of Interest: 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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 16 March 2021
doi: 10.3389/fpls.2021.645055

Comprehensive Characterization and Validation of Chromosome-Specific Highly Polymorphic SSR Markers From Pomegranate (*Punica granatum* L.) *cv.* Tunisia Genome

Prakash Goudappa Patil[1], Nripendra Vikram Singh[1], Abhishek Bohra[2], Keelara Puttaswamy Raghavendra[3], Rushikesh Mane[1], Dhananjay M. Mundewadikar[1], Karuppannan Dhinesh Babu[1] and Jyotsana Sharma[1]*

[1] ICAR-National Research Centre on Pomegranate, Solapur, India, [2] ICAR-Indian Institute of Pulses Research, Kanpur, India, [3] ICAR-Central Institute for Cotton Research, Nagpur, India

OPEN ACCESS

Edited by:
Malli Aradhya, Agricultural Research Service, United States Department of Agriculture, United States

Reviewed by:
Ksenija Gasic, Clemson University, United States
Messaoud Mars, Institution de la Recherche et de l'Enseignement Supérieur Agricoles, Tunisia

***Correspondence:**
Prakash Goudappa Patil
patilbt@gmail.com

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

Received: *22 December 2020*
Accepted: *12 February 2021*
Published: *16 March 2021*

Citation:
Patil PG, Singh NV, Bohra A, Raghavendra KP, Mane R, Mundewadikar DM, Babu KD and Sharma J (2021) Comprehensive Characterization and Validation of Chromosome-Specific Highly Polymorphic SSR Markers From Pomegranate (Punica granatum L.) cv. Tunisia Genome. Front. Plant Sci. 12:645055. doi: 10.3389/fpls.2021.645055

The simple sequence repeat (SSR) survey of 'Tunisia' genome (296.85 Mb) identified a total of 365,279 perfect SSRs spanning eight chromosomes, with a mean marker density of 1,230.6 SSRs/Mb. We found a positive trend in chromosome length and the SSR abundance as marker density enhanced with a shorter chromosome length. The highest number of SSRs (60,708) was mined from chromosome 1 (55.56 Mb), whereas the highest marker density (1,294.62 SSRs/Mb) was recorded for the shortest chromosome 8 (27.99 Mb). Furthermore, we categorized all SSR motifs into three major classes based on their tract lengths. Across the eight chromosomes, the class III had maximum number of SSR motifs (301,684, 82.59%), followed by the class II (31,056, 8.50%) and the class I (5,003, 1.37%). Examination of the distribution of SSR motif types within a chromosome suggested the abundance of hexanucleotide repeats in each chromosome followed by dinucleotides, and these results are consistent with 'Tunisia' genome features as a whole. Concerning major repeat types, AT/AG was the most frequent (14.16%), followed by AAAAAT/AAAAAG (7.89%), A/C (7.54%), AAT/AAG (5.23%), AAAT/AAAG (4.37%), and AAAAT/AAAAG (1.2%) types. We designed and validated a total of 3,839 class I SSRs in the 'Tunisia' genome through electronic polymerase chain reaction (ePCR) and found 1,165 (30.34%) SSRs producing a single amplicon. Then, we selected 906 highly variable SSRs (> 40 nt) from the ePCR-verified class I SSRs and *in silico* validated across multiple draft genomes of pomegranate, which provided us a subset of 265 highly polymorphic SSRs. Of these, 235 primers were validated on six pomegranate genotypes through wet-lab experiment. We found 221 (94%) polymorphic SSRs on six genotypes, and 187 of these SSRs had ≥ 0.5 PIC values. The utility of the developed SSRs was demonstrated by analyzing

genetic diversity of 30 pomegranate genotypes using 16 HvSSRs spanning eight pomegranate chromosomes. In summary, we developed a comprehensive set of highly polymorphic genome-wide SSRs. These chromosome-specific SSRs will serve as a powerful genomic tool to leverage future genetic studies, germplasm management, and genomics-assisted breeding in pomegranate.

Keywords: chromosome, genome, hypervariable SSR, pomegranate, polymorphism

INTRODUCTION

Pomegranate (*Punica granatum* L.) is an economically important perennial fruit crop (Patil et al., 2020c). The popularity of pomegranate worldwide stems from its nutritional values, specific organoleptic characteristics, and a variety of health benefits (Filannino et al., 2013; Teixeira da Silva et al., 2013). This crop has originated from Iran and is widely cultivated in drier parts of Southeast Asia, Iran, China, Japan, the United States (California), West Indies, tropical America, and India (Holland and Bar-Ya'akov, 2014). It is considered an excellent fruit crop for arid zones owing to its drought tolerance. Now, it is widely cultivated in Mediterranean, tropical, and subtropical regions (Holland and Bar-Ya'akov, 2008; Chandra et al., 2010). Its adaptation to the Mediterranean climate has led to its wide diffusion and the creation of a multitude of new genetic individuals over time (Tinebra et al., 2021). Therefore, Mediterranean and Middle-East countries are currently the main regions of pomegranate cultivation and production (Jbir et al., 2008; Melgarejo et al., 2009). Globally, India stands first in pomegranate cultivation with an area of 2.46 hundred thousand hectare, and production and productivity of 27.91 hundred thousand metric tonnes and 12 tonnes/ha, respectively (NHB, 2018).

Efforts to improve pomegranate through standard breeding practices have led to the development and release of few improved varieties in India (Jalikop et al., 2005). However, there is an urgent need to improve the efficiency of current breeding programs to deliver higher genetic gains. In this context, appropriate DNA marker technology facilitates the identification of genetic determinants (genes/QTLs) underlying various traits of economic significance (Patil et al., 2020b,c). Among the various marker techniques, microsatellites or simple sequence repeat (SSR) markers represent one of the most informative, abundant, and easy-to-use marker systems for genetic studies and plant breeding programs (Bohra et al., 2011, 2012; Ravishankar et al., 2013; Xu et al., 2013; Abd El-Moghny et al., 2017). Recently, three major classes of SSRs were reported in eggplant genome based on the SSR motifs length and repeats, which included class I (hypervariable: > 30 nt), class II (potentially variable: 20–30 nt), and class III (variable: < 20 nt) types (Portis et al., 2018). Earlier, Temnykh et al. (2001) also highlighted the significance of SSR tract lengths for marker development and breeding in rice. The rationale for such categorization is that the SSRs with greater tract length have shown higher degree of polymorphism in human (Weber, 1990; Xu et al., 2000). Similar research in various crops including rice, pigeonpea have shown significantly higher polymorphism of class I SSRs (> 20 bp) and highly variable SSRs (> 40 bp) (Singh et al., 2012; Dutta et al., 2013; Bohra et al., 2017).

Sequencing of multiple genomes enabled by the diverse array of sequencing technologies has paved the way for large-scale DNA markers. Discovery and characterization of the DNA markers spanning entire genome such as SSRs provides the foundation for trait discovery studies and molecular breeding. Following genome sequencing, genome-wide SSRs were designed for several plant species, including rice (Zhang et al., 2007), soybean (Song et al., 2010), *Brachypodium* (Sonah et al., 2011), maize (Xu et al., 2013), foxtail millet (Pandey et al., 2013), *Brassica* (Shi et al., 2014), cotton (Wang et al., 2015), *Nicotiana* (Wang et al., 2018), peanut (Zhao et al., 2017; Lu et al., 2019), eggplant (Portis et al., 2018), carrot (Uncu and Uncu, 2020), and more recently in pomegranate (Patil et al., 2020c).

In pomegranate, SSRs have been extensively employed to study genetic diversity and understanding population structure and association analysis (Curro et al., 2010; Pirseyedi et al., 2010; Singh et al., 2015). However, majority of these studies have reported low level of SSR polymorphism. Paucity of highly polymorphic and chromosome-specific molecular markers in pomegranate has hampered map-based gene/QTL strategy. In view of this, the present study aimed at developing a comprehensive set of chromosome-specific hypervariable SSR markers in pomegranate, which would be of immense utility for the pomegranate research and breeding. Availability of chromosome-level genome assembly of the pomegranate cultivar such as 'Tunisia' (Luo et al., 2019) provided us with unprecedented opportunity for genome-wide characterization and development of the first set of chromosome-specific highly polymorphic SSR markers.

MATERIALS AND METHODS

Retrieving Genome Sequences

High-quality genome assembly with eight pseudo-chromosome molecules of pomegranate cultivar 'Tunisia' (Luo et al., 2019) was retrieved in FASTA format from the NCBI[1]. Three other draft genomes of pomegranate cultivars Dabenzi, Taishanhong, and AG2017 (Akparov et al., 2017; Qin et al., 2017; Yuan et al., 2018) were also retrieved for validation of identified SSRs using electronic polymerase chain reaction (ePCR).

[1]http://www.ncbi.nlm.nih.gov

Chromosome-Specific Survey for SSR Motifs and Primer Design

Genomic sequences of eight 'Tunisia' pseudo-chromosome molecules, 296.85 Mb (excluding unknown 23.49 Mb sequences), were surveyed for the presence of chromosome-specific perfect, compound, and imperfect SSR repeats using Krait:ultra-fast SSR search module (Du et al., 2018). As criteria, only two to six nucleotide motifs were considered, and the minimum repeat unit was defined as 12 for mononucleotides, six for dinucleotide repeats, four for trinucleotide repeats, three for tetranucleotides and pentanucleotides, and two for hexanucleotides. Compound microsatellites were defined as two microsatellites interrupted by 100 bases.

The chromosome-specific hypervariable SSR motifs (> 30 bp) were first considered from the total list of motifs for each chromosome by using search options provided in Krait software, followed by selecting all the class I SSRs motifs (> 30 bp). Primer designing was performed using Primer module, which is implemented in Krait software. Primers were designed to generate amplicons of 100–400 bp in length with the following parameters: primer length (bp) 18–20, with 19 as the optimum; GC content 40–70%; Tm 52–60°C, with 55 as the optimum. The other parameters used were as that of default program values. All the genome-wide designed class I primers were designated as hypervariable SSR markers 'Tunisia' (HvSSRT).

Creating Chromosome-Specific Marker Distribution Graphs and Physical Maps

Preliminary information was generated for all the SSR loci such as start and end positions on each chromosomes, and their major classes, i.e., classes I, II, and III using Krait software. Circular graph was drawn to depict chromosome-wise distribution of each SSRs by using the software ShinyCircos (Yu et al., 2017). Apart from this, based on physical positions and tract length of class I highly variable SSRs (> 40 nt), chromosome-specific scatter plots were drawn using Microsoft Excel. Finally, based on the information on physical positions of highly variable SSRs, the saturated SSR marker-based physical map of each chromosome was drawn using MapChart v 2.2 software (Voorrips, 2002).

In silico Evaluation of Newly Developed SSRs for Polymorphism Through ePCR

To evaluate amplification efficiency of newly designed class I SSRs (> 30 nt) and to map the designed marker to genomic sequences of eight chromosomes of 'Tunisia,' the Genome-Wide Microsatellite Analyzing Tool Package (GMATA) software (Wang and Wang, 2016) was used to perform an *in silico* amplification by calling the ePCR algorithm (Schuler, 1997). The ePCR result was used to process the marker mapping information. The settings for ePCR were margin 3,000, no gap in primer sequence, no mismatch in primer sequence, allowed size range of 100–1,000, word size (-w) 12, and contiguous word (-f) 1. The output file (.emap) provided the detailed amplification patterns of the markers with calculated amplicon sizes and target positions on chromosomes and identified single-locus and multi-locus markers. Subsequently, highly variable SSRs (> 40 nt) were evaluated on eight chromosomes of 'Tunisia' to identify primers producing single amplification products. Finally, all the identified single-locus SSR primers of 'Tunisia' chromosomes were evaluated across the three draft genome sequences of pomegranate *cv.* Dabenzi, Taishanhong, and AG2017. The approximate amplicon sizes obtained for highly variable SSRs across the four pomegranate genomes using GMATA were used to calculate various marker parameters like number of alleles (Na), number of effective alleles (Ne), major allelic frequency (MAF), Shannon's information index (*I*), observed heterozygosity (*Ho*), expected heterozygosity (*He*), and polymorphism information content (PIC) using GenAlEx v. 6.5 (Peakall and Smouse, 2012) software.

Wet-Lab Validation of Developed SSR Markers

Genomic DNA was extracted from the fresh leaf samples of 30 pomegranate genotypes (**Table 1**) following the modified CTAB method (Ravishankar et al., 2000). For PCR experiment, a total of 235 chromosome-specific highly variable HvSSRT primer pairs were synthesized and initially screened on a subset of six pomegranate genotypes, i.e., 'Bhagawa,' 'Daru 17,' 'Mridula,' P-23, IC 318728, and IC 318790, following touch down and normal PCR conditions for different primers (**Supplementary Table 5**) with Prime-96^{TM} Thermal Cycler (Himedia, India). Amplicons were then scored on fragment analyzer QIAxcel Advanced (Qiagen India Pvt. Ltd.) and analyzed using QIAxcel Screen Gel Software. Subsequently, 16 HvSSRTs showing clear amplifications were selected randomly from eight chromosomes of 'Tunisia' for genetic diversity study in 30 pomegranate genotypes. For PCR experiments, amplification was carried out in 10 μl reaction volume containing 1.0 μl of 10 × PCR buffer, 1 μl (1 mM dNTP mix), 0.5 μl each of forward and reverse primers (10 pmol), 0.2 μl of *Taq* DNA polymerase 5U/ μl (Himedia, India), and 1 μl (10 ng) of template DNA. Touchdown PCR was performed with the following conditions: 94°C for 5 min, followed by 16 cycles of 94°C for 30 s, decrease 0.2°C/cycle from 60°C for 30 s, 72 C for 45 s, followed by 20 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 45 s, and a final extension at 72°C for 5 min. For normal PCR, initial denaturation at 94°C for 5 min, followed by 36 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 2 min, and a final extension at 72°C for 7 min was followed. PCR products were separated on 3% metaphor agarose gels, visualized, and photographed in gel documentation system (Vilbert Dourmet, France).

Genetic Diversity Analysis

The genotypic data of 30 test genotypes were used for estimating the genetic diversity parameters using GenAlEx v. 6.5 (Peakall and Smouse, 2012), the Na, Ne, *I*, *Ho*, *He*, and PIC. Frequency distribution graphs for allele number and PIC values were drawn using Microsoft Excel. The unweighted pair group method with an arithmetic mean (UPGMA)-based neighbor-joining tree (NJ) and principal coordinate analysis (PCoA) was performed using DARwin v. 6.0.13 (Perrier and Jacquemoud-Collet, 2006).

TABLE 1 | Pomegranate genotypes used in the study.

Sl. No.	Genotype name	Type	Origin/source
1	Kalpitiya	Exotic cultivar	Sri Lanka
2	Bassein seedless	Cultivar	India (Karnataka)
3	Bhagawa	Commercial cultivar	India (Maharashtra)
4	Muscat	Exotic cultivar	Oman
5	GR Pink	Exotic cultivar	Russia
6	Ruby	Commercial cultivar	India (Karnataka)
7	P-26	Commercial cultivar	India (Maharashtra)
8	Yercaud HRS	Cultivar	India (Tamil Nadu
9	Ganesh	Commercial cultivar	India (Maharashtra)
10	Arakta	Commercial cultivar	India (Maharashtra)
11	Mridula	Commercial cultivar	India (Maharashtra)
12	P-16	Commercial cultivar	India (Maharashtra)
13	G-137	Commercial cultivar	India (Maharashtra)
14	Gulesha Red	Exotic cultivar	Russia
15	Jodhpur Red	Cultivar	India (Rajasthan)
16	KRS	Local collection	India (Karnataka)
17	P-23	Commercial cultivar	India (Maharashtra)
18	Jyoti	Commercial cultivar	India (Karnataka)
19	Kandhari	Exotic breeding line	Afghanistan
20	P-13	Commercial cultivar	India (Maharashtra)
21	Daru-17	Wild collection	India (Himachal Pradesh)
22	IC 1181	Wild collection	India (Uttaranchal)
23	IC 1194	Wild collection	India (Uttaranchal)
24	IC 318703	Wild collection	India (Himachal Pradesh)
25	IC 318720	Wild collection	India (Himachal Pradesh)
26	IC 318735	Wild collection	India (Himachal Pradesh)
27	IC 318753	Wild collection	India (Himachal Pradesh)
28	IC 318754	Wild collection	India (Himachal Pradesh)
29	IC 318779	Wild collection	India (Himachal Pradesh)
30	IC 318790	Wild collection	India (Himachal Pradesh)

RESULTS

Genome-Wide Discovery of SSRs

The SSRs were surveyed in the available genome assembly of pomegranate cultivar Tunisia for the presence of mono- to hexanucleotides having a tract length of $\geq$ 12 bp. As a result, a total of 365,279 perfect SSRs were identified from the 296.85 Mb of genomic sequences (**Table 2**). A total of 55,836 (15.28%) belonging to compound SSR category were detected. The hexanucleotide repeats were the most abundant (201,501) with a representation of 55.16%, followed by di- (55,437, 15.18%), tetra- (36,455, 9.98%), tri- (29,940, 8.2%), mono- (27,536, 7.54%), and penta (14,410, 3.94%) nucleotide repeats (**Table 2**).

The frequency distribution of different SSR motifs in the 'Tunisia' genome sequence is presented in **Supplementary Figure 1**. Among them, AT/AG had the highest occurrence (14.16%), followed by AAAAAT/AAAAAG (7.89%), A/C (7.54%), AAT/AAG (5.23%), AAAT/AAAG (4.37%), and AAAAT/AAAAG (1.2%) (**Supplementary Figure 1**). Across mono- to hexanucleotide repeats, the major motifs were A, AT, AAT, AAAT, AAAAT, and AAAAAT, of which AT motif had relative abundance of 120.6 loci/Mb in the genome, followed by A (80.3 loci/Mb) and AG (53.7 loci/Mb) motifs. SSRs with CG-rich repeats were rare in the pomegranate genome. We found an inverse relationship between SSR abundance and motif repeat number, and the trend was the most conspicuous for hexa- and tetranucleotide repeats (**Supplementary Figure 4**).

TABLE 2 | Characterization of microsatellites in the pomegranate genome *cv.* Tunisia.

SSR mining			Total	
Examined sequences size (bp)			296,847,911	
Total number of perfect SSRs			365,279	
Total length of perfect SSRs (bp)			5,177,752	
Relative abundance of SSRs (loci/Mb)			1,230.6	
Relative density for SSRs (bp/Mb)			17,443.55	
Total number of compound SSRs			55,836	
Motif	**SSR counts and percentage**	**Length (bp)**	**Relative abundance (loci/Mb)**	**Relative density (bp/Mb)**
Mono	27,536 (7.54%)	419,278	92.77	1,412.52
Di	55,437 (15.18%)	1,067,314	186.76	3,595.72
Tri	29,940 (8.2%)	466,194	100.87	1,570.58
Tetra	36,455 (9.98%)	485,912	122.81	1,637.01
Penta	14,410 (3.94%)	237,540	48.55	800.26
Hexa	201,501 (55.16%)	2,501,514	678.85	8,427.46
Total	365,279 (100%)	5,177,752		

The Intra-Chromosomal Distribution of SSRs

We further analyzed the distribution of SSRs on each chromosome (**Table 3**). We identified a total of 365,279 perfect SSRs in comparison with 415,716 imperfect SSRs with mean marker densities of 1,230.6 and 1,400.5 (SSR/Mb), respectively. The highest number of SSRs (60,708 perfect, 67,141 imperfect) was assigned to the largest chromosome 1 (55.56 Mb) followed by SSRs mapped onto chromosomes 2 (44.57 Mb, 56,038 perfect, 64,041 imperfect) and 4 (40.13 Mb, 51,511 perfect, 58,934 imperfect). Less perfect (35,868, 37,304, and 36,241) as well as imperfect SSRs (41,094, 42,809, and 41,901) were assigned to shorter chromosomes Chm_6 (28.33 Mb), Chm_7 (28.78 Mb), and Chm_8 (27.99 Mb), respectively. Therefore, chromosome length showed relation with the SSR abundance per chromosome in our study. The differences in SSR densities on different chromosomes were significant, ranging from 1,092.77 SSRs/Mb (Chm_1) to 1,294.62 SSRs/Mb (Chm_8), with an average of 1,230.60 SSRs/Mb (**Table 3**). It was also interesting to note that the marker density increased with the reduced chromosomes length (**Supplementary Figure 2A**).

The intra-chromosomal distribution of SSR motif types in the 'Tunisia' genome as a whole (**Supplementary Figure 2B**), reflected the abundance of hexanucleotide repeats and the least occurrence of penta-/mononucleotide repeats. The distribution of motifs within each chromosome followed the pattern observed in the genome assembly as a whole, and hexanucleotides were the most frequent SSR types followed by dinucleotides

TABLE 3 | The chromosome-wise distribution of perfect, compound, and imperfect SSRs.

Chromosome	Total Mb	Perfect								Compound		Imperfect	
		Mono	Di-	Tri-	Tetra-	Penta-	Hexa-	Total	SSRs/Mb	Total	SSRs/Mb	Total	SSR/Mb
*Chm_1	55.56	4,289	8,574	4,842	6,141	2,275	34,587	60,708	1,092.77	9,090	163.62	67,141	1,208.57
Chm_2	44.57	4,278	8,617	4,518	5,459	2,241	30,925	56,038	1,257.43	8,573	192.37	64,041	1,437.01
Chm_3	39.96	3,486	7,027	3,944	4,863	1,933	26,189	47,442	1,187.40	7,138	178.65	53,647	1,342.70
Chm_4	40.13	4,082	8,041	4,187	5,072	2,008	28,121	51,511	1,283.75	8,036	200.27	58,934	1,468.74
Chm_5	31.53	3,053	6,186	3,377	4,006	1,637	21,908	40,167	1,273.87	6,115	193.93	46,149	1,463.58
Chm_6	28.33	2,746	5,419	3,048	3,701	1,427	19,527	35,868	1,266.29	5,506	194.38	41,094	1,450.79
Chm_7	28.78	2,933	5,694	2,951	3,685	1,487	20,554	37,304	1,296.21	5,716	198.62	42,809	1,487.50
Chm_8	27.99	2,669	5,879	3,073	3,528	1,402	19,690	36,241	1,294.62	5,662	202.26	41,901	1,496.81
Total	296.85	27,536	55,437	29,940	36,455	14,410	201,501	365,279	1,230.60	55,836	188.11	415,716	1,400.52

**Chm-, chromosome.*

in every chromosome. The hexa- (55.05%) and dinucleotide (15.24%) exhibited the highest level of variation among chromosomes, within Chm_1 showing the lowest percentage for di- (14.12%) and the highest for hexanucleotides (56.97%), while in Chm_8, we found higher percentages for di- (16.22%) and hexanucleotides (54.33%). Considering frequency distribution on a chromosome-by-chromosome basis, AT was found to be the most abundant dinucleotide motif followed by 'A' mononucleotide repeat. Similarly, AAT was the most prominent among trinucleotide repeat category, while AAAG was the most abundant tetranucleotide repeat (as in the genome as a whole), except for the Chr_7 (AATT). However, penta- and hexanucleotides showed two different motif combinations for each chromosome (**Supplementary Figure 5**).

Chromosome-Specific Distribution of Three Major Classes of SSRs

To study distribution of SSRs on different chromosomes, we characterized all SSR motifs into three major classes, i.e., classes I (> 30 nt), II (20–30 nt), and III (< 20 nt). For each chromosome, the variation in the three classes of perfect SSRs with regard to the number of repeat units is presented in **Table 4**. Total of 337,743 SSRs were considered for classification excluding mononucleotide repeats. The highest number of motifs (301,684; 82.59%) belonged to class III, followed by class II (31,056; 8.50%) and class I (5,003; 1.37%) across all eight chromosomes. The overall distribution graph for three major SSR classes for each chromosome revealed, Chm_1, Chm_2, and Chm_3 with higher number of SSRs for three classes, followed by Chm_4, Chm_5, Chm_7, and Chm_6 (**Supplementary Figure 3A**). Furthermore, the overall distribution of class I SSRs with respect to the number of repeat units for di-, tri-, tetra-, penta-, and hexanucleotides in each chromosomes was examined (**Supplementary Figure 3B**), and visually confirmed (**Figure 1**). As evident from these graphs, the total number of SSRs for the three classes (III, II, and I) and their motif types (di- to hexanucleotides) in each chromosome decreased from inside to outside rings of circos graph. We inferred from the intra-chromosomal distribution of these SSRs that dinucleotides (3,488) were more abundant followed by tri- (1,087), tetra- (214), hexa- (136), and pentanucleotides (78), and the proportion was consistent on all chromosomes (**Table 4**). As observed on all chromosomes, SSR frequency decreased with an increase in the number of repeat units with the exception of hexanucleotides (**Supplementary Figure 3B**).

Designing of Chromosome-Specific Hypervariable SSR Primers

Based on the genomic distribution of class I SSRs across all chromosomes, we successfully designed primers for 3,839 out of the total of 5,003 SSRs (**Supplementary Table 1**). In accordance to class I motif content of various chromosomes, the most primers were designed for Chm_1 (628) followed by Chm_2 (591), Chm_4 (564), and so forth. The majority of these primers were specific to dinucleotide motifs (primers 2,825, 73.59%), followed by trinucleotide repeats (698, 18.18%) (**Table 5**). For experimental validation, we zeroed in on a set of 906 highly variable SSRs targeting ≥ 40 nt tract length from each chromosome through ePCR on four pomegranate genomes (**Supplementary Table 2**). The majority of SSR primers used for validation were dinucleotides (616) or trinucleotides (228).

Distribution of the New SSR Markers on Different Chromosomes

The 906 SSR genomic location was examined (**Supplementary Table 2**) and placed on chromosomes (**Supplementary Figure 6**), of which Chm_1 (159 markers), Chm_2 (142), and Chm_4 (131) had higher number of assigned markers, followed by Chm_3 (122), Chm_5 (107), Chm_7 (85), Chm_8 (83), and Chm_6 (77). Interestingly, scatter plots clearly depicted the physical distance (Mb), intervals between SSRs and their tract lengths on each chromosome. Most of the SSR markers remained in the range of 41–50 nt tract length (612), followed by 51–60 nt (172), 61–70 nt (63), and > 71 nt (59) (**Table 6**). For tract length 41–50 nt, Chm_1 had most markers (111), whereas Chm_8 had the least number of markers (51). Similarly, for tract length 51–60 nt, Chm_1 had most markers (30) and Chm_6 least markers (14). Concerning the SSR tract length 61–70 nt, three chromosomes, Chm_2 (13), Chm_1 (11), and Chm_3 (11), had higher numbers than Chm_8(4). However, with respect to tract length > 71 nt, Chm_5 had most markers (11), followed by Chm_2 (9), Chm_3 (8), or

TABLE 4 | Distribution of three major classes of SSRs in different chromosomes of pomegranate *cv.* Tunisia.

Chromosomes	Class I						Class II	Class III	Total
	Di	Tri	Tetra	Penta	Hexa	Total			
*Chm_1	547	177	34	11	29	798	4,724	50,897	56,419
Chm_2	549	158	31	13	17	768	4,860	46,132	51,760
Chm_3	451	146	21	15	18	651	4,040	39,265	43,956
Chm_4	494	160	26	11	21	712	4,442	42,275	47,429
Chm_5	397	140	31	8	10	586	3,399	33,129	37,114
Chm_6	324	104	26	8	14	476	3,162	29,484	33,122
Chm_7	377	93	21	6	13	510	3,206	30,655	34,371
Chm_8	349	109	24	6	14	502	3,223	29,847	33,572
Total	3,488	1,087	214	78	136	5,003	31,056	301,684	337,743

**Chm-, chromosome.*

Chm_1 and Chm_4 (7), Chm_7 and Chm_8 (6), and Chm_6 (5). It is also interesting to note that Chm_7 had highest track length markers (189 bp), followed by Chm_5 (168 bp), Chm_8 (126 bp), Chm_3 (124 bp), Chm_4 (118 bp), Chm_2 (112 bp), Chm_6 (110 bp), and Chm_1 (88 bp) (**Supplementary Figure 6**).

Construction of a High-Density SSR-Based Physical Map for Pomegranate

The physical position and start positions of 906 HvSSRTs on each chromosome, used to generate a high-density physical map (**Figure 2**) showed that Chm_1 carried the highest number of markers (159), followed by Chm_2 (142), Chm_4 (131), Chm_3 (122), Chm_5 (107), Chm_7 (85), Chm_8 (83), and Chm_6 (77).

ePCR Validation of the Identified SSRs Across Four Genomes

To assess the amplification efficiency and specificity of the SSRs, initially 3,839 class I SSR primer pairs were tested on 'Tunisia' genome through *in silico* analysis. We validated equal proportions of primers across eight chromosomes, and the SSRs produced one to > 3 alleles in 'Tunisia' genome (**Supplementary Table 3a**). A total of 1,165 (30.34%) primers yielded a single amplicon of expected size, whereas 1,263 (32.90%) primers had two alleles, and 1,254 (32.66%) and 157 (4.09%) primers produced three and > 3 alleles, respectively. Furthermore, to validate SSRs with tract lengths of > 40 nt, located on physical map, *in silco* PCR analysis for 906 primers was performed on all the four genome assemblies ('Dabenzi,' 'Taishanhong,' 'AG2017,' and 'Tunisia'). Notably, we validated all 906 (100%) primers in 'Tunisia' genome to that of 853 (94.15%), 832 (91.83%), and 545 (60.15%) in 'Dabenzi,' 'Taishanhong,' and 'AG2017' genomes, respectively (**Supplementary Table 3a**). A total of 289 ('Tunisia'), 277 ('Dabenzi'), 264 ('Taishanhong'), and 183 ('AG2017') SSRs showed single-locus amplification.

To show the informativeness of these chromosome-specific SSRs, we selected a subset of 289 that was validated in all four genomes. The various amplicons detected through ePCR for these 289 primer pairs across the four genomes were recorded to further compute the marker parameters (**Supplementary Table 3b**). Of these, 265 (91.70%) SSRs were polymorphic across the four genomes. A total of 719 alleles were obtained across the all eight chromosomes. The Na per locus ranged from 2 to 4, with an average of 2.49 alleles/loci (**Supplementary Table 4**). The MAF per locus varied between 0.50 and 0.88, with an average of 0.74. With an average of 0.46, the PIC values had a range from 0.25 to 0.80. Out of 289 HvSSRT primer pairs validated, 144 SSRs had PIC values ≥ 0.50. For the four genomes tested, the average Shannon information index was 0.66.

We compared all marker parameters at the chromosome level (**Table 7**). Among eight chromosomes, Chm_8 had lower number of polymorphic markers (25) and higher average value of Ne (1.83), Shannon's information index (I = 0.72), and PIC (0.51). However, Chm_5 also showed lower number of polymorphic markers (26) and lower average value Ne (1.54), I (0.51), and PIC (0.35). With respect to PIC value remaining, all six chromosomes (Chm_1, Chm_2, Chm_3, Chm_4, Chm_6, and Chm_7) had mean PIC values of 0.44–0.48.

PCR Amplification and Polymorphism

For wet-lab validation of the SSRs, we initially screened 235 primer pairs on six pomegranate genotypes. As a result, 225 (94.5%) yielded the amplicons of expected size, whereas 10 remaining primer pairs did not show any amplification. A total of 221 (98.2%) SSRs were polymorphic across six pomegranate genotypes, one marker (HvSSRT_416) remained monomorphic, and three SSRs (HvSSRT_258, HvSSRT_702, and HvSSRT_900) amplified only in one genotype (**Supplementary Table 5**). A representative gel image illustrating the SSR profiles of pomegranate genotypes is presented in **Figures 3A,B**. Using SSR markers, we detected 797 alleles among six genotypes as listed in "MATERIALS AND METHODS" (**Supplementary Table 5**). The Na for loci ranged from 1 to 8, with an average of 3.59. The *Ho* and the *He* for each locus ranged from 0 to 1 (with a mean of 0.58) and 0 to 0.86 (with a mean of 0.61), respectively, and the Shannon's information index ranged from 0.00 to 2.02 with a mean of 1.08. PIC ranged from 0 (for HvSSRT_416) to 0.96 (HvSSRT_891) with a mean value of 0.68. A total of 187 markers had PIC values ≥ 0.5 (**Figure 4B**). It is interesting to

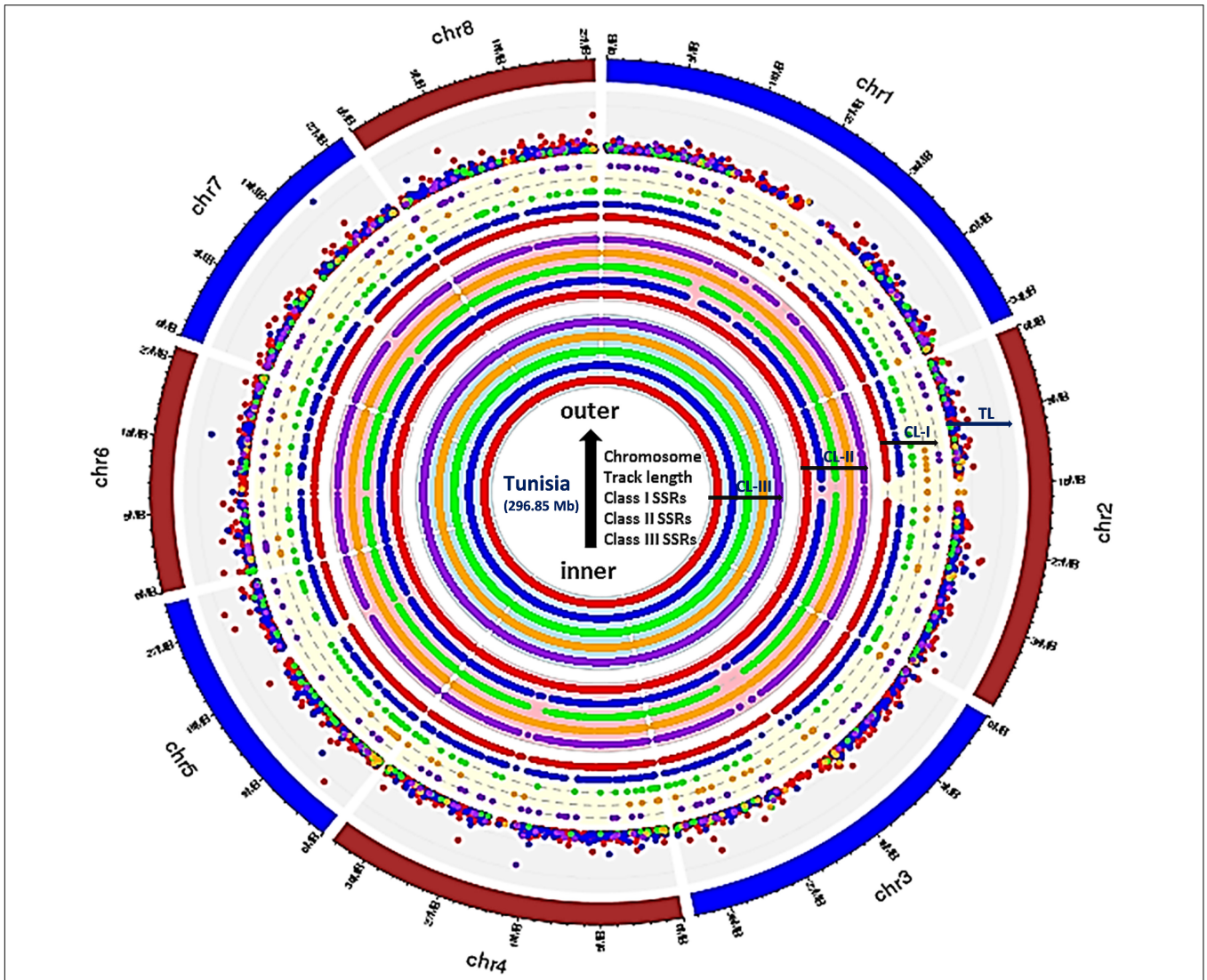

FIGURE 1 | Circos graph depicting the chromosome-wide distribution for three major classes of perfect SSRs (excluding mononucleotides). Inner ring represents class III, whereas middle and outer rings correspond to class II and class I, respectively, and the densely placed colored dots depict the tract length for each motif types. All rings from inside to outside have five subrings representing, di-, tri-, tetra-, penta-, and hexanucleotides with color codes red, blue, green, orange, and purple, respectively. The densely placed spots in the outer ring represented tract length variations for each repeat motif with specific color code. The di (red spots) and tri (blue spots) showed maximum variations which can be exploited for designing chromosome-specific hypervariable SSR markers.

note that 68 SSRs had PIC values $\geq$ 0.80. The frequency analysis of HvSSR markers with respect to number of alleles showed that 78 markers had two alleles while eight alleles were obtained for two markers (**Figure 4A**).

Assessment of the Genetic Relationships

A subset of 16 HvSSRs located on eight chromosomes, selected randomly from each chromosome based on their clear amplification profiles observed on six genotypes, were tested on 30 pomegranate genotypes (**Table 1**), and a total of 34 alleles were detected with an average of 2.13 alleles across the pomegranate genotypes. The *Ho* ranged from 0 to 0.52, with an average number of 0.23. The PIC values ranged from 0.33 to 0.60, with an average of 0.48 (**Supplementary Table 6**). The mean Shannon's information index at 0.69 revealed moderate diversity among the genotypes.

In the NJ tree, all 30 pomegranate genotypes were grouped into two major clusters: cluster 1 comprised 11 while cluster 2, 19 genotypes (**Figure 5A**). Cluster 1 carried 9 wild genotypes with an exceptional placement of one wild genotype (Daru 17) within cluster 2. Cluster 2 had all cultivars (18), with exception of out grouping of two cultivars Muscat and Jodhpur Red in cluster 1. Furthermore, the PCoA assigned 30 genotypes to two major clusters (**Figure 5B**). The principal coordinates (PCos) 1 and 2 explained 21.16 and 12.24%, respectively, of the total variance among the genotypes and accounted for 33.84% of the total variation. Interestingly, PCo 1 clearly separated two clusters into wild and cultivar groups.

TABLE 5 | Description of chromosome-specific class I SSR markers designed for eight chromosomes of Tunisia.

	Number of class I (>30 nt) primers						Highly variable SSRs (≥40 nt) primers					
	Di	Tri	Tetra	Penta	Hexa	Total	Di	Tri	Tetra	Penta	Hexa	Total
*Chm_1	459	112	27	9	21	628	108	35	5	1	10	159
Chm_2	444	106	23	5	13	591	92	43	2	2	3	142
Chm_3	353	95	16	11	9	484	86	25	4	4	3	122
Chm_4	415	106	22	9	12	564	87	39	1	1	3	131
Chm_5	330	88	22	8	9	457	77	26	1	1	2	107
Chm_6	255	63	17	5	11	351	54	17	1	0	5	77
Chm_7	305	62	18	6	7	398	58	21	3	1	2	85
Chm_8	264	66	18	6	12	366	54	22	2	1	4	83
Total	2,825	698	163	59	94	3,839	616	228	19	11	32	906

**Chm-, chromosome.*

TABLE 6 | Classification of 906 highly variable chromosome-specific SSR markers based on their tract lengths.

Tract length (nt)	*Chm_1	Chm_2	Chm_3	Chm_4	Chm_5	Chm_6	Chm_7	Chm_8	Total primers
41–50	111	103	81	92	64	53	57	51	612
51–60	30	17	22	26	25	14	16	22	172
61–70	11	13	11	6	7	5	6	4	63
71–189	7	9	8	7	11	5	6	6	59
Total	159	142	122	131	107	77	85	83	906

**Chm-, chromosome.*

DISCUSSION

Lack of chromosome-wise details on polymorphic molecular markers in pomegranate has greatly hampered the progress of trait discovery and gene cloning. SSR markers have been proven very useful for genetic analysis, trait mapping, and molecular breeding in several tree species. A large number of genome-wide hypervariable SSRs were recently discovered using draft genome sequence of pomegranate *cv.* Dabenzi (Patil et al., 2020c). However, lack of a chromosome-wide profiling of these hypervariable SSR greatly limits their immediate use in genetic studies.

The SSR Content of the Tunisia Genome

The present study aimed to offer highly informative chromosome-wise SSR markers in pomegranate. The genomic sequences containing perfect SSRs accounted for 1.74% (5.18 Mb) of total assembled genome, which is slightly higher than observed in the grapevine (0.67%) genome (Portis et al., 2016). The 'Tunisia' genome revealed 55,836 (15.28%) SSRs of compound type, which is higher than the 'Dabenzi' genome 15,483 (8.92%) (Patil et al., 2020c). Our observation of higher SSR density (1,230.6 SSRs/Mb) in 'Tunisia' genome slightly deviated from a previous report that established a negative relationship between genome size and SSR density (Morgante et al., 2002). Such trends were also witnessed through comparing various plant genomes (Portis et al., 2016). By conducting SSR survey in genome sequences of 16 tree species, Xia et al. (2017) found that SSR densities with the tree species may be greater than > 1,200, as in *Theobroma cacao* (1,446) and others. Furthermore, consistent with the observation of Xia et al. (2017), we found hexanucleotide repeats were the most dominant in 'Tunisia' genome (55.16%), followed by di-, tetra-, tri-, mono-, and pentanucleotides. The frequency distribution for motif category revealed AT/AG had the highest occurrence, followed by AAAAAT/AAAAAG, A/C, AAT/AAG, AAAT/AAAG, and AAAAT/AAAAG in the 'Tunisia' genome. Previously, authors have confirmed the abundance of AT/TA and TTA/TAT/ATT types in the pomegranate genome (Ravishankar et al., 2015). Similarly, Patil et al. (2020c) observed abundance of AT/AT- and AAT/ATT-type motifs in the 'Dabenzi' genome. Besides this, we observed abundance of SSR was inversely proportional to the increase in the motif repeat number, and this trend was pronounced for hexa- and tetranucleotide repeat classes, followed by penta-, tri-, di, and mononucleotide repeats, which is in agreement with other reports (Cavagnaro et al., 2010; Shi et al., 2013; Cheng et al., 2016).

The Intra-Chromosomal Distribution of SSRs

In this study, we found that the number of SSRs and SSR densities vary according to the length of the chromosomes. Consistent with our finding, Portis et al. (2016) also observed greatest number of SSRs (15,212 perfect, 19,282 imperfect) being assigned to the longest LG02 (70.34 Mb) and the lowest (5,181 perfect, 6,518 imperfect) to the shortest LG14 (14.48 Mb) in globe artichoke genome, yielding varying SSR densities of 216.3 to 357.9/Mbp. Similar observations were made in soybean (Ott et al., 2011) and eggplant (Portis et al., 2018). Concerning the intra-chromosomal

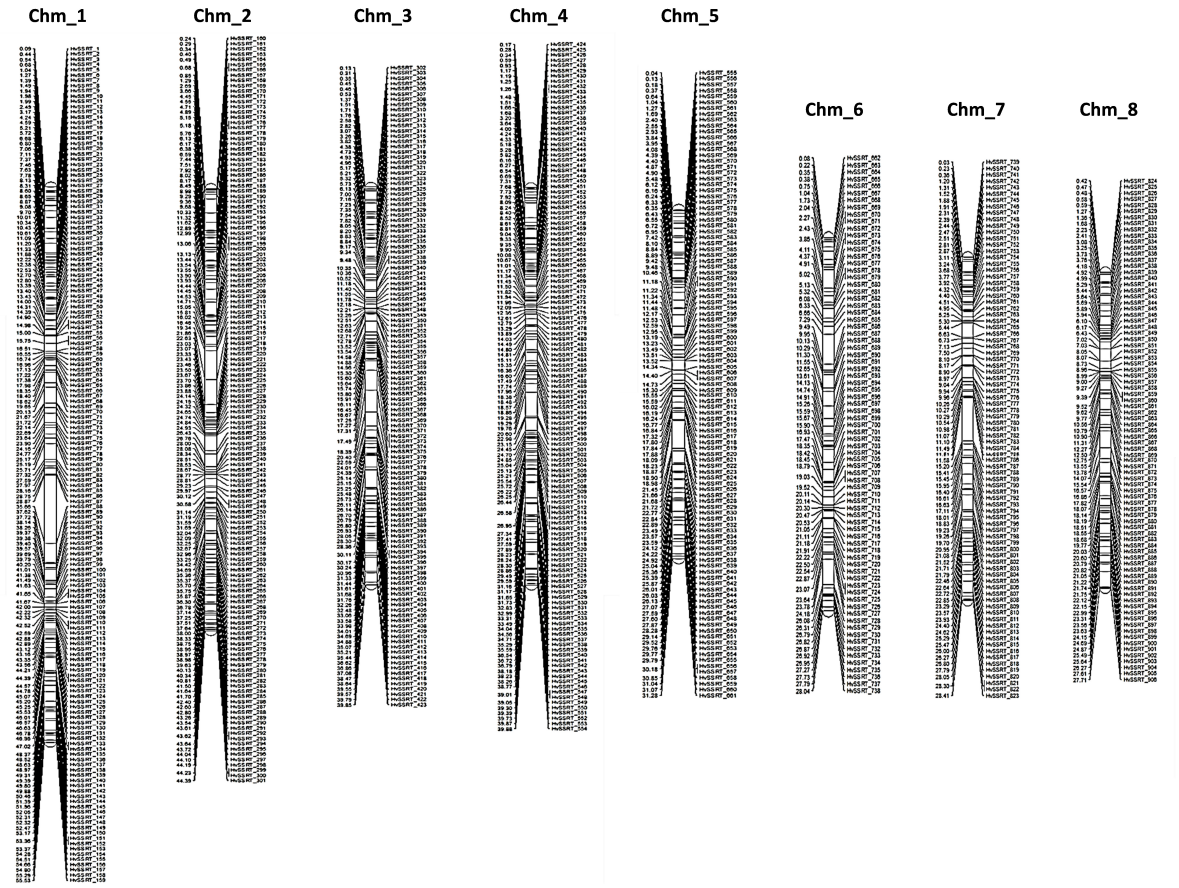

FIGURE 2 | A high-density map of 906 highly variable SSR markers Tunisia (HvSSRT) showing their physical locations on eight chromosomes of pomegranate *cv.* Tunisia.

distribution of motif types in 'Tunisia,' hexanucleotide was the most common SSR type followed by dinucleotides, which is in best agreement with earlier report, where hexamers accounted for 55.16% of all motifs, and pentamers were the least abundant type (2.33%) in tree genomes (Xia et al., 2017).

TABLE 7 | Chromosome-specific marker statistics for 289 highly variable SSR primer pairs assayed through ePCR across the four pomegranate genotypes based on their genome sequences.

Chromosome	TNP	TPP	Na	MAF	Ne	*I*	*Ho*	*He*	PIC
Chm_1	55	48	2.40	0.75	1.71	0.63	0.50	0.38	0.44
Chm_2	42	39	2.52	0.73	1.75	0.68	0.54	0.40	0.47
Chm_3	32	31	2.69	0.72	1.78	0.71	0.56	0.42	0.48
Chm_4	41	40	2.59	0.74	1.75	0.69	0.53	0.41	0.48
Chm_5	35	26	2.17	0.80	1.54	0.51	0.40	0.30	0.35
Chm_6	30	29	2.67	0.73	1.73	0.70	0.53	0.41	0.47
Chm_7	28	27	2.39	0.72	1.73	0.66	0.55	0.41	0.47
Chm_8	26	25	2.54	0.70	1.83	0.72	0.60	0.44	0.51
Total	289	265	2.50	0.74	1.73	0.66	0.53	0.39	0.46

**TNP, total number of primers; TPP, total number of polymorphic primers; Na, number of alleles; MAF, major allele frequency; Ne, number of effective alleles; I, Shannon's Information Index; Ho, observed heterozygosity; He, expected heterozygosity; PIC, polymorphic information content.*

Chromosome-Specific Distribution of Three Major Classes of SSRs

Examination of distribution of SSR types across chromosomes revealed class III SSRs as the most dominant followed by classes II and I. The relationship between SSR frequency and their repeat number in each chromosomes reported here agreed with previous reports in various plant species, i.e., *Brassica*, *Arabidopsis*, and other angiosperm species (Shi et al., 2013), pepper (Cheng et al., 2016) and globe artichoke (Portis et al., 2016). Furthermore, dinucleotides dominated followed by trinucleotide repeats for class I SSRs in each chromosome. These patterns were also reflected from the genome-wide distribution of the three major classes of SSRs.

Designing and Distribution of Chromosome-Specific Hypervariable SSR Markers

We designed primers for 3,839 class I SSR markers of which most markers were specific to Chm_1 (628), while Chm_8

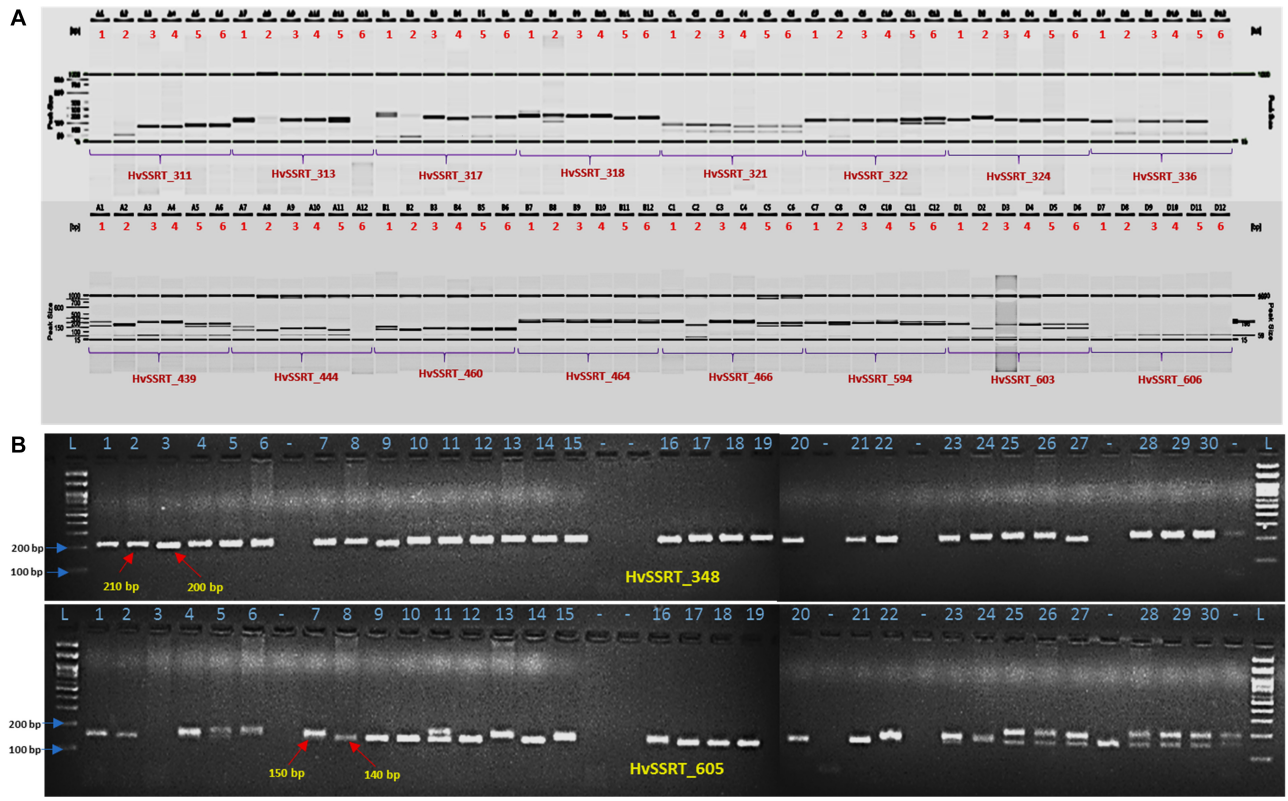

FIGURE 3 | Allelic variations revealed by HvSSRT markers when assayed on six pomegranate genotypes using fragment analyzer **(A)**, HvSSRT markers 348 and 605 on 30 pomegranate genotypes on 3% metaphor gels **(B)** (Lanes 1–6, six sets of pomegranate genotypes, i.e., 'Bhagawa', 'Daru 17,' 'Mridula,' P-23, IC 318728, and IC 318790, L-100 bp DNA ladder for lanes 1–30 pomegranate genotypes as listed in **Table 1**, genotypes in lane –, were excluded from the analysis).

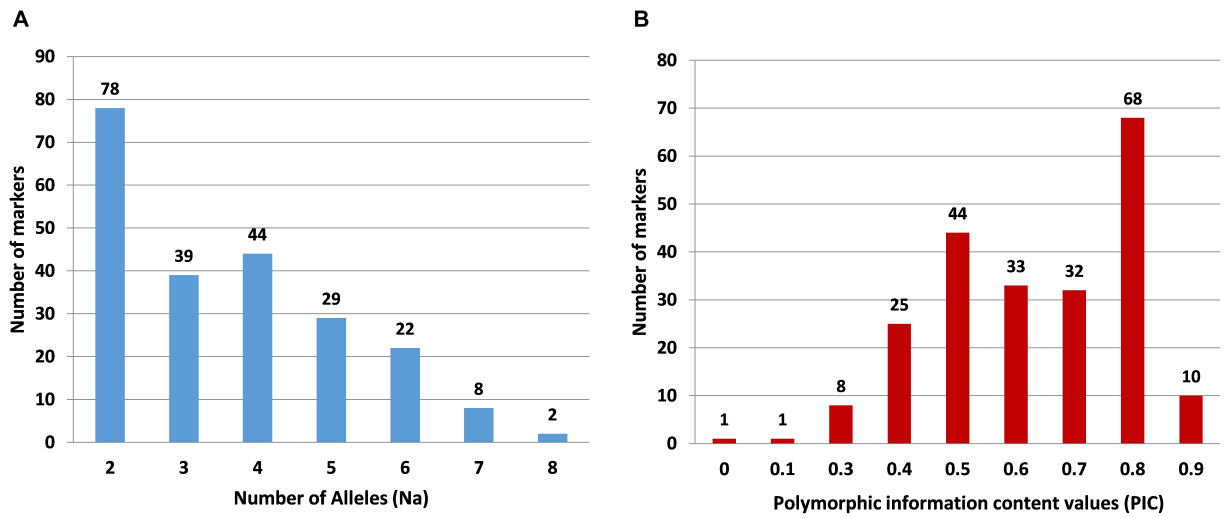

FIGURE 4 | Frequency distribution of SSR markers in six pomegranate genotypes with respect to number of alleles **(A)** and PIC values **(B)**.

(366) and Chm_6 (351) had lower number of markers. Earlier, a high correlation (R^2 = 0.96, P < 0.01) was reported between chromosome length and number of SSRs in eggplant chromosomes, with the longest chromosome (E01) containing the highest number of SSRs and the shortest hosting the least (Portis et al., 2018). The present study suggested that for each chromosome, the distribution of markers decreases with increase in tract length, which was also reflected at whole genome level.

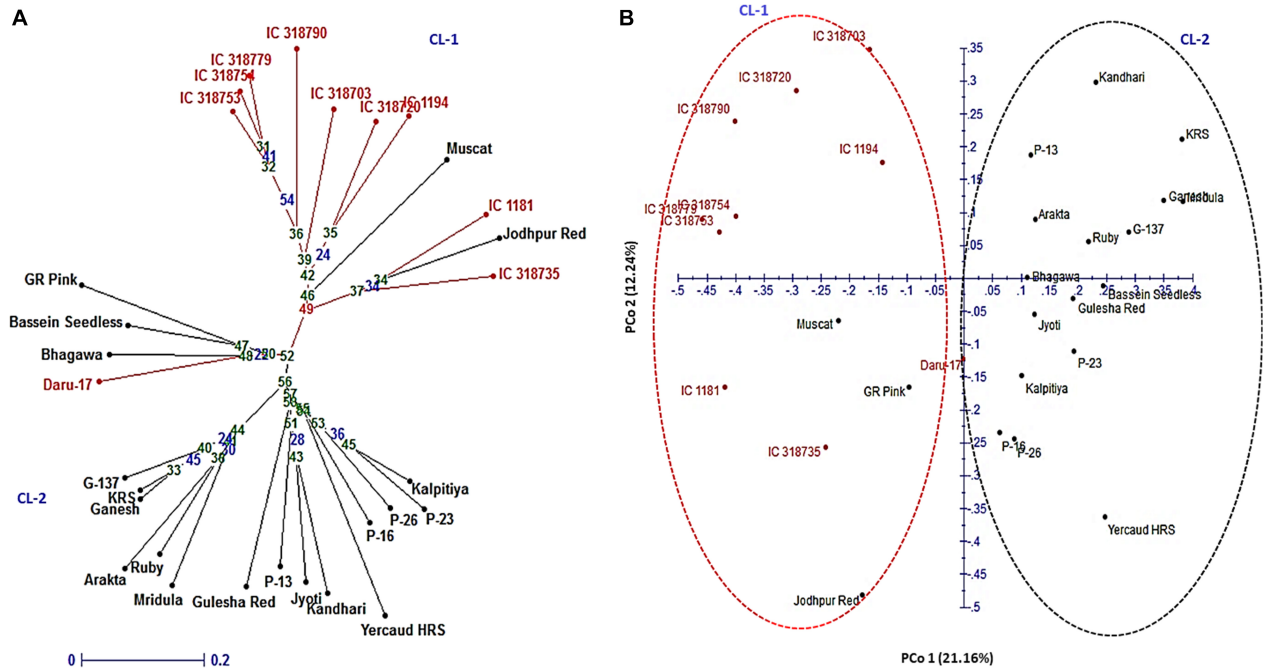

FIGURE 5 | Genetic relationships among 30 pomegranate genotypes based on 16 HvSSRT markers: **(A)** neighbor-joining tree and **(B)** principal coordinate analysis.

Singh et al. (2010) also found a decrease in the number of SSR loci with increase in tract length in the rice genome. In pomegranate, dinucleotides dominated for each chromosome (AT) followed by trinucleotide (AAT) motifs as found over whole 'Tunisia' genome. Portis et al. (2018) also noticed similar distribution of motif types within individual chromosomes of eggplant which was very similar to the pattern found over the whole eggplant genome with di- and trinucleotide repeats being the most frequent.

Construction of a High-Density Physical Map Based on HvSSRTs

A high-density physical map with uniform genomic positions and coverage is necessary for conducting high-resolution gene/QTL mapping (Lu et al., 2019). In pomegranate, however, reports on construction of SSR-based linkage map are currently lacking. Here, we built a saturated physical map of pomegranate using 906 HvSSRT markers. Chm_1 showed maximum number of markers, followed by Chm_2, Chm_4, Chm_3, Chm_5, Chm_7, Chm_8, and Chm_6. This reflected a concordance in numbers of markers with length of chromosomes. The SSR distribution pattern suggested nearly even presence of markers on each chromosome, with slight deviation of spare distribution of markers at the middle of chromosomes as compared with distal ends. This high-density physical map could serve as the reference map for analyzing the high-throughput genotyping data for different types of populations and accelerate mapping and breeding applications of different traits in pomegranate. Several reports have shown the application of SSR-based physical map in fine mapping of reported QTLs (Zhao et al., 2017). Genetic and physical maps based on SSR markers in many crops have served as important genomic resources to study collinearity and synteny (Jamshed et al., 2016; Kirungu et al., 2018). Therefore, we assume that the information generated here could definitely provide benefits to the pomegranate researchers for trait mapping.

ePCR Validation and Identification of Single-Locus SSRs

We performed an *in silico*-simulated PCR to assess SSR polymorphism levels across four pomegranate genome sequences. An initial ePCR with 3,839 class I SSR primer pairs on 'Tunisia' chromosomes identified 1,165 (30.34%) produced single allele of expected size. Subsequently, we selected 906 high variable class (> 40 nt) markers for validation and identified 289 primer pairs producing expected product size with single amplicon. The ePCR technique has been applied to validate *in silico*-discovered DNA markers in different plant species including wheat (Han et al., 2015), sesame (Dossa et al., 2017), carrot (Uncu and Uncu, 2020), cucumber (Liu et al., 2015), bitter melon (Cui et al., 2017), and tobacco (Wang et al., 2018).

Identification of a Set of SSRs With Higher Level of Polymorphism

High level of DNA polymorphism is the most important characteristic of molecular markers. The present study provides a total of 906 (> 40 bp) ePCR-validated SSRs spanning all eight chromosomes of 'Tunisia.' Among these, 289 markers showed single ePCR product in the 'Tunisia' genome. Of these 289 primer pairs, 277 were validated in 'Dabenzi,' 264 in 'Taishanhong,' and 183 in 'AG2017' through e-mapping with 91.70% being polymorphic and having an average PIC value of 0.46. Further

classification of primer pairs based on PIC values led to the identification of 144 SSRs as highly polymorphic (PIC: 0.50 to 0.80), while 121 SSRs (PIC: 0.25–0.46) had moderate level of polymorphism. In our previous study in pomegranate, the 82 HvSSRs identified from draft genome sequence of 'Dabenzi' had average PIC value of 0.28, and importantly, 46 HvSSRs (56.10%) had PIC values $\geq$ 0.4 (Patil et al., 2020c). Taken together, these findings suggest an increase in the SSR polymorphism with an increase in the length of the SSR tracts (Cavagnaro et al., 2010; Bohra et al., 2011).

Wet-Lab Validation of SSRs and Marker Polymorphism Survey

Wet-lab validation of ePCR-derived 235 HvSSRs revealed 98% (221) polymorphism in six pomegranate genotypes. The PIC observed among six genotypes (ranged from 0 to 0.96 with a mean of 0.68) is higher than the range (0–0.91) reported earlier in 12 pomegranate genotypes (Ravishankar et al., 2015). Earlier, we obtained PIC values between 0.12 and 0.63 for 82 polymorphic SSRs from analysis of eight pomegranate genotypes (Patil et al., 2020c). The differences in SSR allele count and PIC values observed in these studies could possibly be due to use of different genotyping platforms, i.e., agarose, metaphor gels, and automated capillary-based systems (Patil et al., 2020c).

SSR-Based Diversity Analysis of a Broader Set of Pomegranate Germplasm

The utility of the new SSRs for pomegranate genetic improvement was evident from the diversity estimation of 30 pomegranate genotypes. Analysis with 16 HvSSRs representing eight chromosomes generated a total of 34 alleles with an average PIC value of 0.48. The current results concur with our earlier findings where we obtained 30 alleles having the PIC values between 0.12 and 0.63, following analysis of 46 pomegranate genotypes using 13 HvSSRs derived from 'Dabenzi' genome with unknown chromosomal regions (Patil et al., 2020c). Our present results suggested presence of moderate level of genetic diversity among 30 pomegranate genotypes. This could be probably due to low resolution of agarose or metaphor gel for SSR separation and scoring, or due to lesser number of polymorphic alleles among the cultivars examined here (Patil et al., 2020a). The neighbor-joining tree based on 16 HvSSRs grouped the 30 pomegranate genotypes into two distinct clusters representing wild and cultivars groups. These patterns are in concordance with previous clustering patterns observed in pomegranate (Patil et al., 2020a,c). Likewise, a previous analysis of 88 pomegranate accessions with 44 SSRs distinctly separated wild accessions from cultivated types including commercial varieties, local types and introduced accessions (Singh et al., 2015). Considering the higher genetic diversity levels, except Muscat (Oman), all the introduced exotic pomegranate accessions like Kalpitiya (Sri Lanka), GR Pink and Gulesha Red (Russia), and Kandhari (Afghanistan) were distributed randomly in the cluster 2 as compared with the cluster 1 that contained only wild accessions. The resulting clustering patterns agreed with geographical distributions and pedigree relationships. For instance, cluster 1 contained Daru type wild accessions belonging to hill regions of North India including Uttarakhand and Himachal Pradesh.

The close proximity of the genotypes G-137, Ganesh, and Arakta was according to their pedigree. Similarly, the genotypes namely P-23, P-26, P-16, and P-13 representing a selection from Muscat grouped together in a single cluster. In PCoA plot, a total of 33.84% of the variation was explained and the PCo 1 explained 21.16% of variability by clearly separating wild and cultivars groups. In a previous PCoA of 42 diverse pomegranates, axis 1 explained higher proportion of the variance (30.8%) to that of axis 2 (17.95%), and axis 1 had clearly separated the two major clusters (Patil et al., 2020a). Overall, a strong agreement was observed between NJ tree and factorial analysis in our study.

CONCLUSION

In the current work, we characterized microsatellites from the survey of entire 'Tunisia' genome and validated SSRs that represent eight different chromosomes. A total of 365,279 perfect SSRs were identified. The study provides a large number of chromosome-specific hypervariable (3,839) and highly variable SSRs (> 40 nt, 906) for genotyping applications in pomegranate. We also provide genomic positions of these new SSRs to enable informed selection of most useful SSRs for future use. We first assessed *in silico* amplification of 906 HvSSRs through ePCR and identified 289 SSRs that amplify a single locus. The ePCR approach applied on four pomegranate genome assemblies elucidated 265 SSRs to be polymorphic. We finally confirmed 235 HvSSRs through wet-lab experiment, and 94% of these SSRs were polymorphic. The immediate application of these markers was shown in a panel of 30 pomegranate genotypes. The new set of HvSSRTs developed here could be useful for future gene discovery, genomics-assisted breeding, and germplasm management in pomegranate.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

PP, NS, and JS designed the research experiments. PP and KR performed the *in silico* analysis to design genome-wide SSR markers. KB and NS contributed in collection of test materials. PP, RM, and DM performed the wet-lab experiments for SSR validation and diversity analysis. PP and AB wrote the original manuscript with assistance of JS. All authors contributed to the article and approved the submitted version.

FUNDING

This work was carried out with the financially assistance of Indian Council of Agricultural Research (ICAR), New Delhi,

India extended through ICAR-National Research Centre on Pomegranate, Solapur (MS) under Institute Project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.645055/full#supplementary-material

Supplementary Figure 1 | Frequency distribution of the different SSR motif types in the pomegranate genome *cv.* Tunisia (**Supplementary Table 7**).

Supplementary Figure 2 | Intra-chromosomal distribution of SSRs. Comparison of perfect SSR number/density **(A)** and frequency of mono- to hexanucleotide motifs **(B)** in the eight chromosomes of pomegranate genome *cv.* Tunisia (**Supplementary Table 7**).

Supplementary Figure 3 | Distribution of SSRs across all chromosomes of pomegranate *cv.* Tunisia, frequency of three major classes of SSRs **(A)** and of di- to hexanucleotide motifs **(B)** (**Supplementary Table 7**).

Supplementary Figure 4 | The distribution of SSR repeat lengths for mono- to hexanucleotide motif types across the genome.

Supplementary Figure 5 | Chromosome-wise relative abundance of most frequent SSR motif types.

Supplementary Figure 6 | Scatter plot depicting chromosome-wise physical distribution of highly variable SSR markers Tunisia (906 HvSSRT) in relation to their track lengths on eight chromosomes of pomegranate *cv.* Tunisia.

Supplementary Table 1 | Description of 3,839 class I SSR primer pairs designed based on 'Tunisia' genome information.

Supplementary Table 2 | Description of 906 highly variable class I SSRs selected for ePCR validation and their physical location and tract lengths across the eight chromosomes of 'Tunisia'.

Supplementary Table 3 | (a) Experimental validation of class I SSRs in 'Tunisia' genome and highly variable SSRs (906) through ePCR or eMapping across the other three draft genomes of pomegranate cultivars Dabenzi, Taishanhong, and AG2017. **(b)** Allele score for 289 highly variable SSRs validated through ePCR on four different pomegranate genomes.

Supplementary Table 4 | Marker statistics computed for 289 HvSSRT markers screened through ePCR on four pomegranate genomes.

Supplementary Table 5 | Marker statistics for 222 HvSSRT primer pairs validated through PCR on six pomegranate genotypes.

Supplementary Table 6 | Genetic diversity statistics of 16 hypervariable SSR markers Tunisia (HvSSRT) obtained from 30 pomegranate genotypes.

Supplementary Table 7 | Input data used for drawing various frequency distribution graphs.

REFERENCES

Abd Ei-Moghny, A. M., Santosh, H. B., Raghavendra, K. P., Sheeba, J. A., Singh, S. B., and Kranthi, K. R. (2017). Microsatellite marker based genetic diversity analysis among cotton (Gossypium hirsutum) accessions differing for their response to drought stress. *J. Plant Biochem. Biotechnol.* 26, 366–370. doi: 10.1007/s13562-016-0395-1

Akparov, Z., Amiraslanov, A., Hajiyeva, S., Abbasov, M., Kaur, K., Hamwieh, A., et al. (2017). *De-novo Sequencing of Pomegranate (Punica granatum L.) Genome*. Heidelberg: EMBL.

Bohra, A., Dubey, A., Saxena, R. K., Penmetsa, R. V., Poornima, K. N., Kumar, N., et al. (2011). Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea. *BMC Plant Biol.* 11:56. doi: 10.1186/1471-2229-11-56

Bohra, A., Jha, R., Pandey, G., Patil, P. G., Saxena, R. K., Singh, I. P., et al. (2017). New hypervariable SSR markers for diversity analysis, hybrid purity testing and trait mapping in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *Front. Plant Sci.* 8:377. doi: 10.3389/fpls.2017.00377

Bohra, A., Saxena, R. K., Gnanesh, B. N., Saxena, K. B., Byregowda, M., Rathore, A., et al. (2012). An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. *Theor. Appl. Genet.* 125, 1325–1338.

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

Chandra, R., Babu, K. D., Jadhav, V. T., and Teixeira da Silva, J. A. (2010). Origin, history and domestication of pomegranate. *Fruit Veg. Cereal Sci. Biotechnol.* 4, 1–6.

Cheng, J., Zhao, Z., Li, B., Qin, C., Wu, Z., Trejo-Saavedra, D. L., et al. (2016). A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in Capsicum. *Sci. Rep.* 6:18919. doi: 10.1038/srep18919

Cui, J., Cheng, J., Nong, D., Peng, J., Hu, Y., He, W., et al. (2017). Genome-wide analysis of simple sequence repeats in bitter gourd (*Momordica charantia*). *Front. Plant Sci.* 8:1103. doi: 10.3389/fpls.2017.01103

Curro, S., Caruso, M., Distefano, G., Gentile, A., and La Malfa, S. (2010). New microsatellite loci for pomegranate, *Punica granatum* (Lythraceae). *Am. J. Bot.* 97, e58–e60.

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

Du, L., Zhang, C., Liu, Q., Zhang, X., and Yue, B. (2018). Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design. *Bioinformatics* 34, 681–683.

Dutta, S., Mahato, A. K., Sharma, P., Ranjeet, S. R., Sharma, T. R., and Singh, N. K. (2013). Highly variable Arhar simple sequence repeat markers for molecular diversity and phylogenetic studies in pigeonpea (*Cajanus cajan* (L.) Millisp.). *Plant Breed.* 132, 1439–1523. doi: 10.1111/pbr.12014

Filannino, P., Azzi, L., Cavoski, I., Vincentini, O., Rizzello, C. G., Gobbetti, M., et al. (2013). Exploitation of the health promoting and sensory properties of organic pomegranate (*Punica granatum* L.) juice through lactic acid fermentation. *Int. Food Microb.* 163, 184–192.

Han, B., Wang, C., Tang, Z., Ren, Y., Li, Y., Zhang, D., et al. (2015). Genome-wide analysis of microsatellite markers based on sequenced database in Chinese spring wheat (*Triticum aestivum* L.). *PLoS One* 10:e0141540. doi: 10.1371/journal.pone.0141540.t006

Holland, D., and Bar-Ya'akov, I. (2008). The pomegranate: new interest in an ancient fruit. *Chronica Hortic.* 48, 12–15.

Holland, D., and Bar-Ya'akov, I. (2014). "Pomegranate: aspects concerning dynamics of health beneficial phytochemicals and therapeutic properties with respect to the tree cultivar and the environment," in *Medicinal and Aromatic Plants of the Middle-East*, eds Z. Yaniv and N. Dudai (Netherlands: Springer), 225–239.

Jalikop, S., Rawal, R., and Kumar, R. (2005). Exploitation of sub-temperate pomegranate Daru in breeding tropical varieties. *Acta Hortic.* 696, 107–112.

Jamshed, M., Jia, F., Gong, J., Palanga, K. K., Shi, Y., Li, J., et al. (2016). Identification of stable quantitative trait loci (QTLs) for fiber quality traits across multiple environments in Gossypium hirsutum recombinant inbred line population. *BMC Genomics* 17:197. doi: 10.1186/s12864-016-2560-2

Jbir, R., Hasnaoui, N., Mars, M., Marrakchi, M., and Trifi, M. (2008). Characterization of Tunisian pomegranate (*Punica granatum* L.) cultivars using amplified fragment length polymorphism analysis. *Sci. Hortic.* 115, 231–237.

Kirungu, J. N., Deng, Y., Cai, X., Magwanga, R. O., Zhou, Z., Wang, X., et al. (2018). Simple sequence repeat (SSR) genetic linkage map of D genome diploid cotton derived from an interspecific cross between *Gossypium davidsonii* and *Gossypium klotzschianum*. *Int. J. Mol. Sci.* 19:204. doi: 10.3390/ijms19010204

Liu, J., Qu, J., Hu, K., Zhang, L., Li, J., Wu, B., et al. (2015). Development of genome wide simple sequence repeat fingerprints and highly polymorphic markers in cucumbers based on next-generation sequence data. *Plant Breed.* 134, 605–611. doi: 10.1111/pbr.12304

Lu, Q., Hong, Y., Li, S., Liu, H., Li, H., Zhang, J., et al. (2019). Genome-wide identification of microsatellite markers from cultivated peanut (*Arachis hypogaea* L.). *BMC Genomics* 20:799. doi: 10.1186/s12864-019-6148-5

Luo, X., Li, H., Wu, Z., Yao, W., Zhao, P., Cao, D., et al. (2019). The pomegranate (*Punica granatum* L.) draft genome dissects genetic divergence between soft- and hard-seeded cultivars. *Plant Biotech. J.* 18, 955–968. doi: 10.1111/pbi.13260

Melgarejo, P., Martínez, J., Hernández, F., Martínez, R., Legua, P., Oncina, R., et al. (2009). Cultivar identification using 18S–28S rDNA intergenic spacer-RFLP in pomegranate (*Punica granatum* L.). *Sci. Hortic.* 120, 500–503.

Morgante, M., Hanafey, M., and Powell, W. (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.* 30, 194–200. doi: 10.1038/ng822

NHB (2018). Available online at: http://nhb.gov.in/PDFViwer.aspx?Enc=3ZOO8K5CzcdC/Yq6HcdIxC0U1kZZenFuNVXacDLxz28, 2nd Advanced estimate (accessed August 20, 2018).

Ott, A., Trautschold, B., and Sandhu, D. (2011). Using microsatellites to understand the physical distribution of recombination on soybean chromosomes. *PLoS One* 6:e22306. doi: 10.1371/journal.pone.0022306

Pandey, G., Misra, G., Kumari, K., Gupta, S., Parida, S. K., Chattopadhyay, D., et al. (2013). Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet *Setaria italica* (L.). *DNA Res.* 20, 197–207.

Patil, P. G., Jamma, S. M., Singh, N. V., Bohra, A., Parashuram, S., Injal, A. S., et al. (2020a). Assessment of genetic diversity and population structure in pomegranate (*Punica granatum* L.) using new hypervariable SSR markers. *Physiol. Mol. Biol. Plants* 26, 1249–1261. doi: 10.1007/s12298-020-00825-y

Patil, P. G., Singh, N. V., Parashuram, S., Bohra, A., Mundewadikar, D. M., Sangnure, V. R., et al. (2020b). Genome wide identification, characterization and validation of novel miRNA-based SSR markers in pomegranate (*Punica granatum* L.). *Physiol. Mol. Biol. Plants* 26, 683–696. doi: 10.1007/s12298-020-00790-6

Patil, P. G., Singh, N. V., Parashuram, S., Bohra, A., Sowjanya, R., Gaikwad, N., et al. (2020c). Genome-wide characterization and development of simple sequence repeat markers for genetic studies in pomegranate (*Punica granatum* L.). *Trees Struct. Func.* 34, 987–998. doi: 10.1007/s00468-020-01975-y

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

Perrier, X., and Jacquemoud-Collet, J. P. (2006). *DARwin Software*. Paris: CIRAD.

Pirseyedi, S. M., Valizadehghan, S., Mardi, M., Ghaffari, M. R., Mahmoodi, P., Zahravi, M., et al. (2010). Isolation and characterization of novel microsatellite markers in pomegranate (*Punica granatum* L.). *Int. J. Mol. Sci.* 11, 2010–2016.

Portis, E., Lanteri, S., Barchi, L., Portis, F., Valente, L., Toppino, L., et al. (2018). Comprehensive characterization of simple sequence repeats in eggplant (*Solanum melongena* L.) genome and construction of a web resource. *Front. Plant Sci.* 9:401. doi: 10.3389/fpls.2018.00401

Portis, E., Portis, F., Valente, L., Moglia, A., Barchi, L., Lanteri, S., et al. (2016). A Genome-wide survey of the microsatellite content of the Globe artichoke genome and the development of a web- based database. *PLoS One* 11:e0162841. doi: 10.1371/journal.pone.0162841

Qin, G., Xu, C., Ming, R., Tang, H., Guyot, R., Kramer, E. M., et al. (2017). The pomegranate (*Punica granatum* L.) genome and the genomics of punicalagin biosynthesis. *Plant J.* 91, 1108–1128. doi: 10.1111/tpj.13625

Ravishankar, K. V., Anand, L., and Dinesh, M. R. (2000). Assessment of genetic relatedness among mango cultivars of India using RAPD markers. *J. Hort. Sci. Biotechnol.* 75, 198–201.

Ravishankar, K. V., Chaturvedi, K., Puttaraju, N., Gupta, S., Pamu, S., and Flachowsky, H. (2015). Mining and characterization of SSRs from pomegranate (L.) by pyrosequencing. *Plant Breed.* 134, 247–254.

Ravishankar, K. V., Raghavendra, K. P., Athani, V., Rekha, A., Srinivas, B. V., and Ananad, L. (2013). Development and characterisation of microsatellite markers for wild banana (*Musa balbisiana*). *J. Hort. Sci. Biotech.* 88, 605–609. doi: 10.1080/14620316.2013.11513013

Schuler, G. D. (1997). Sequence mapping by electronic PCR. *Genome Res.* 7, 541–550.

Shi, J., Huang, S., Fu, D., Yu, J., Wang, X., Hua, W., et al. (2013). Evolutionary dynamics of microsatellite distribution in plants: insight from the comparison of sequenced Brassica, *Arabidopsis* and other Angiosperm species. *PLoS One* 8:e59988. doi: 10.1371/journal.pone.0059988

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.

Singh, H., Deshmukh, R. K., Singh, A., Singh, A. K., Gaikwad, K., Sharma, T. R., et al. (2010). Highly variable SSR markers suitable for rice genotyping using agarose gels. *Mol. Breed.* 25, 359–364. doi: 10.1007/s11032-009-9328-1

Singh, N. K., Gupta, D. K., Jayaswal, P. K., Mahato, A. K., Dutta, S., Singh, S., et al. (2012). The first draft of the pigeonpea genome sequence. *J. Plant Biochem. Biotechnol.* 21, 98–112.

Singh, N. V., Abburi, V., Ramajayam, D., Kumar, R., Chandra, R., Sharma, K., et al. (2015). Genetic diversity and association mapping of bacterial blight and other horticulturally important traits with microsatellite markers in pomegranate from India. *Mol. Genet. Genomics* 290, 1393–1402.

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

Song, Q., Jia, G., Zhu, Y., Grant, D., Nelson, R. T., Hwang, E. Y., et al. (2010). Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. *Crop Sci.* 50, 1950–1960.

Teixeira da Silva, J., Rana, T. S., Narzaryd, D., Vermae, N., Meshramf, D. T., and Ranade, S. A. (2013). Pomegranate biology and biotechnology: a review. *Sci. Hort.* 160, 85–107.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., and McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genet. Res.* 11, 1441–1452.

Tinebra, I., Scuderi, D., Sortino, G., Mazzaglia, A., and Farina, V. (2021). Pomegranate cultivation in mediterranean climate: plant adaptation and fruit quality of 'mollar de elche' and 'wonderful' cultivars. *Agronomy* 11:156.

Uncu, A. O., and Uncu, A. T. (2020). High-throughput simple sequence repeat (SSR) mining saturates the carrot (*Daucus carota* L.) genome with chromosome-anchored markers. *Biotechnol. Biotech. Equipment* 34, 1–9. doi: 10.1080/13102818.2019.1701551

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Heredit.* 93, 77–78. doi: 10.1093/jhered/93.1.77

Wang, Q., Fang, L., Chen, J., Hu, Y., Si, Z., Wang, S., et al. (2015). Genome-wide mining, characterization, and development of microsatellite markers in *Gossypium* species. *Sci. Rep.* 5:10638.

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

Wang, X., Yang, S., Chen, Y., Zhang, S., Zhao, Q., Li, M., et al. (2018). Comparative genome-wide characterization leading to simple sequence repeat marker development for Nicotiana. *BMC Genomics* 19:500. doi: 10.1186/s12864-018-4878-4

Weber, J. L. (1990). Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. *Genomics* 7, 524–530.

Xia, X., Luan, L. L., Qin, G., Yu, L. F., Wang, Z. W., Dong, W. C., et al. (2017). Genome-wide analysis of SSR and ILP markers in trees: diversity profiling, alternate distribution, and applications in duplication. *Sci. Rep.* 7:17902. doi: 10.1038/s41598-017-17203-6

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, X., Peng, M., Fang, Z., and Xu, X. (2000). The direction of microsatellite mutations is dependent upon allele length. *Nat. Genet.* 24, 396–399.

Yu, Y., Ouyang, Y., and Yao, W. (2017). ShinyCircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics* 34, 1229–1231.

Yuan, Z., Fang, Y., Zhang, T., Fei, Z., Han, F., Liu, C., et al. (2018). The pomegranate (*Punica granatum* L.) genome provides insights into fruit quality and ovule developmental biology. *Plant Biotechnol. J.* 16, 1363–1374.

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.

Zhao, C., Qiu, J., Agarwal, G., Wang, J., Ren, X., Xia, H., et al. (2017). Genome-wide discovery of microsatellite markers from diploid progenitor species, *Arachis duranensis* and *A. ipaensis*, and their application in cultivated peanut (*A. hypogaea*). *Front. Plant Sci.* 8:1209. doi: 10.3389/fpls.2017.01209

Conflict of Interest: 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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 August 2021
doi: 10.3389/fpls.2021.647599

Analysis of the Segregation Distortion of *FcRAN1* Genotypes Based on Whole-Genome Resequencing of Fig (*Ficus carica* L.) Breeding Parents

Hidetoshi Ikegami[1], Kenta Shirasawa[2], Hiroshi Yakushiji[3], Shiori Yabe[4], Masaru Sato[2], Takeshi Hayashi[4], Kosuke Tashiro[5] and Hitoshi Nogata[1]*

[1] Fukuoka Agriculture and Forestry Research Center Buzen Branch, Yukuhashi, Japan, [2] Kazusa DNA Research Institute, Kisarazu, Japan, [3] Institute of Fruit Tree and Tea Science, National Agriculture and Food Research Organization, Higashihiroshima, Japan, [4] Institute of Crop Science, National Agriculture and Food Research Organization, Tsukuba, Japan, [5] Faculty of Agriculture, Kyushu University, Fukuoka, Japan

OPEN ACCESS

Edited by:
Giuseppe Ferrara, University of Bari Aldo Moro, Italy

Reviewed by:
Amir Sherman, University of North Carolina at Chapel Hill, United States
Stefano Biricolti, University of Florence, Italy

***Correspondence:**
Hidetoshi Ikegami
ikegami@farc.pref.fukuoka.jp

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

Received: *30 December 2020*
Accepted: *08 June 2021*
Published: *10 August 2021*

Citation:
Ikegami H, Shirasawa K, Yakushiji H, Yabe S, Sato M, Hayashi T, Tashiro K and Nogata H (2021) Analysis of the Segregation Distortion of FcRAN1 Genotypes Based on Whole-Genome Resequencing of Fig (Ficus carica L.) Breeding Parents. Front. Plant Sci. 12:647599. doi: 10.3389/fpls.2021.647599

The common fig (*Ficus carica* L.) has a gynodioecious breeding system, and its sex phenotype is an important trait for breeding because only female plant fruits are edible. During breeding to select for female plants, we analyzed the *FcRAN1* genotype, which is strongly associated with the sex phenotype. In 12 F_1 populations derived from 13 cross combinations, the *FcRAN1* genotype segregation ratio was 1:1, whereas the M119-226 × H238-107 hybridization resulted in an extremely male-biased segregation ratio (178:7 = male:female). This finding suggests that the segregation distortion was caused by some genetic factor(s). A whole-genome resequencing of breeding parents (paternal and maternal lines) identified 9,061 high-impact SNPs in the parents. A genome-wide linkage analysis exploring the gene(s) responsible for the distortion revealed 194 high-impact SNPs specific to Caprifig6085 (i.e., seed parent ancestor) and 215 high-impact SNPs specific to H238-107 (i.e., pollen parent) in 201 annotated genes. A comparison between the annotated genes and the genes required for normal embryo or gametophyte development and function identified several candidate genes possibly responsible for the segregation distortion. This is the first report describing segregation distortion in *F. carica*.

Keywords: segregation distortion, male-biased, genome-wide linkage analysis, embryo, gametophyte

INTRODUCTION

During cross breeding, the import frequency of alleles into gametes often varies in the progeny population or a specific gene combination causes sterility, resulting in a segregation ratio that is inconsistent with Mendelian inheritance (i.e., segregation distortion). Segregation distortions have been detected in a wide range of taxa (Burt and Trivers, 2006) and are increasingly recognized as a potentially powerful evolutionary force (Taylor and Ingvarsson, 2003). Segregation distortion is thought to have diverse causes, some of which involve gametophytic lethality, embryonic lethality, or cross-incompatibility (Lashermes et al., 2001; Candela et al., 2011) or are related to (selfish)

genetic elements that can enhance their own transmission (Dawkins, 1976; Werren et al., 1988). The selfish genetic factors include Cytoplasmic male sterility (Lewis, 1941), B chromosome (Ostergren, 1945), X chromosome drive (Gershenson, 1928; Sturtevant and Dobzhansky, 1936), Y chromosome drive (Hickey and Craig, 1966; Sweeny and Barr, 1978; Wood and Newton, 1991), t haplotype (Dobrovolskaia-Zavadskaia and Kobozieff, 1927), and transposable factors (McClintock, 1950).

Fig (*Ficus carica* L.) is a dioecious or gynodioecious species, and because only female plant fruits are edible, sex is an important target trait during breeding (Stover et al., 2007; Flaishman et al., 2008). The common fig has the following two sexual types: caprifig type (males) with both male and female flowers and fig type (females) with only female flowers. The progeny resulting from a normal fig type (G/G) × caprifig type (G/A) cross should theoretically be 50% fig and 50% caprifig (fig:caprifig = 1:1) (Storey, 1975).

Mori et al. (2017) identified *FcRAN1* and determined that it is highly correlated with the fig sex phenotype on the basis of several analyses, including a genome-wide association study. Additionally, because of a single nucleotide polymorphism (SNP), its alleles are heterozygous (G/A) in males (caprifig type) and homozygous (G/G) in females (fig type) (Storey, 1975). Moreover, *FcRAN1* is the only gene in the fig genome that carries a polymorphism that perfectly corresponds to the sex phenotype (Mori et al., 2017). In our current fig breeding program, we genotyped the *FcRAN1* alleles of some breeding populations to identify sex phenotypes. We observed that individuals with male type alleles are produced at a high rate only in a specific cross combination (M119-226 × H238-107), suggesting that the *FcRAN1* locus is affected by segregation distortion. Clarifying the cause of this distortion may lead to improved sex control and selection during fig breeding, but it may also provide insights into sex ratio transitions and sex chromosome evolution in fig.

In the present study, we examined the segregation distortion in fig breeding populations and performed preliminary analyses to explore the responsible genetic factors. More precisely, we investigated the distribution of SNPs specific to each pollen and seed parent generating segregation distortion in a whole-chromosome pseudomolecule and considered the biological and evolutionary background related to the occurrence of the segregation distortion.

RESULTS

Crossing experiments involving 10 cross combinations of seven seed parents (M119-226, V180-13, Masui Dauphine, H180-2, HIG5, Toyomitsuhime, and Burjassotte Greece) and six pollen parents (H238-107, M238-1, VD238-83, VS238-53, NG60, and CH-13) generated 1,232 seedlings. The progeny population sizes ranged from 12 to 370 (**Table 1**). Three cross combinations involving three seed parents (Violette de Dauphine, Violette de Solies, and Royal Vineyard) and two pollen parents (M106-238 and VC-180) from a previous study (Mori et al., 2017) were also analyzed. In 12 of the 13 cross combinations, the segregation ratio of the male (G/A) and female (G/G) genotypes of *FcRAN1* was close to the theoretical segregation ratio of 1:1, ranging from 3:9 to 192:176 (G/A:G/G; χ^2-test: p = 0.014–1.00). In contrast, for the cross combination (C100) involving the seed parent M119-226 and the pollen parent H238-107, the segregation ratio was 178:7 (G/A:G/G; χ^2-test: p < 2.2e-16) (**Table 1** and **Figure 1**). An examination of pollen conditions using an optical microscope revealed that more than 99% of the pollen grains were clearly stained and had a normal appearance in all parental lines, including H238-107 (**Figure 2**).

To identify the factor(s) responsible for the segregation distortion, a genome-wide linkage analysis (GWLA) was performed using the whole-genome sequence of each seed and pollen parent. However, the seed parent of C100, M119-226, died after seeds were acquired, preventing it from being analyzed. Therefore, the genome sequence of Caprifig6085, which was an ancestral genotype of M119-226 and may have a genomic region associated with segregation distortion, was used for seed parent analysis as an alternative to M119-226 (**Table 1** and **Figure 3**) (Segregation distortion was not observed in the mating with Masui Dauphine and Violette de Solies as parents). The 4,614,159 SNPs generated by Shirasawa et al. (2021, unpublished data) included 9,061 high-impact SNPs in the seed and pollen parents. The GWLA using the SNP panel indicated that 194 high-impact SNPs were specific to Caprifig6085 as the seed parent and 215 high-impact SNPs were specific to H238-107 as the pollen parent. The functional annotation of these SNPs on the basis of a BLAST analysis produced information for 201 SNPs (98 and 103 SNPs for the seed and pollen parents, respectively) (**Supplementary Tables 1**, **2**). The genes with high-impact SNPs included a *LEA* gene (i.e., encoding a late embryogenesis abundant protein). A sequence analysis revealed that the fig *LEA* gene (designated as *FcLEA1*) is a single-copy gene located on chromosome Fc01a, which also includes the *FcRAN1* locus (Fc01a: 181,985–187,609). Moreover, the polymorphic site in *FcLEA1* corresponds to a non-synonymous substitution that changes a leucine to a proline (**Figure 4**).

The mapping of all GWLA SNPs to whole pseudomolecules indicated that high-impact SNPs were widely distributed throughout the genome in the seed and pollen parents, but there were no high-impact SNPs on chromosomes Fc03 and Fc06 of the seed parents (**Figure 5**).

DISCUSSION

It is unclear whether the observed segregation distortion is caused by gametophytic lethality, embryonic lethality, or specific genetic events. Accordingly, in this study, a GWLA was performed assuming that the causative factor is located on the chromosomes of either the seed parent (M119-226) or the pollen parent (H238-107) of the population with distorted segregation. Unfortunately, all M119-226 plants died before their genome sequence data could be obtained, making it impossible to further analyze line M119-226. However, there were four ancestral lines of M119-226, namely Caprifig6085, VC-119, Masui Dauphine, and Violette de Dauphine. Both Masui Dauphine and Violette de Dauphine are not associated with segregation distortion (**Table 1**). Therefore,

TABLE 1 | F-test for the segregation distortion of the *FcRAN1* genotype in the F_1 population.

				***FcRAN1* genotype**			
Progeny ID	**Seed parent**	**Pollen parent**	**Number of plants**	**Male type (A/G)**	**Female type (G/G)**	**n.d.**	***P* value**
C100	MI19-226	H238-107	188	178	7	3	2.2E-16
C101	V180-13	M238-1	370	192	176	2	0.404
C102	V180-13	VD238-83	116	58	55	3	0.778
C95	Masui Dauphine	VS238-53	167	83	83	1	1.000
C94	Masui Dauphine	VD238-83	273	138	135	0	0.856
C98	HI80-2	VD238-83	42	19	22	1	0.639
C78	HIG5	NG60	23	10	11	2	0.827
C79	Toyomitsuhime	NG60	26	6	18	2	0.014
C74	Masui Dauphine	CH-13	15	6	9	0	0.439
C205	Burjassotte Greece	NG60	12	3	9	0	0.083
M2017-1	Violette de Dauphine	MI06-238	71	36	35	0	0.906
M2017-2	Violette de Solies	VC-180	58	20	38	0	0.018
M2017-2	Royal Vineyard	VC-180	62	24	38	0	0.075
Total			1,423	773	636	14	

The results for lines M2017-1, M2017-2, and M2017-3 are from an earlier study by Mori et al. (2017).

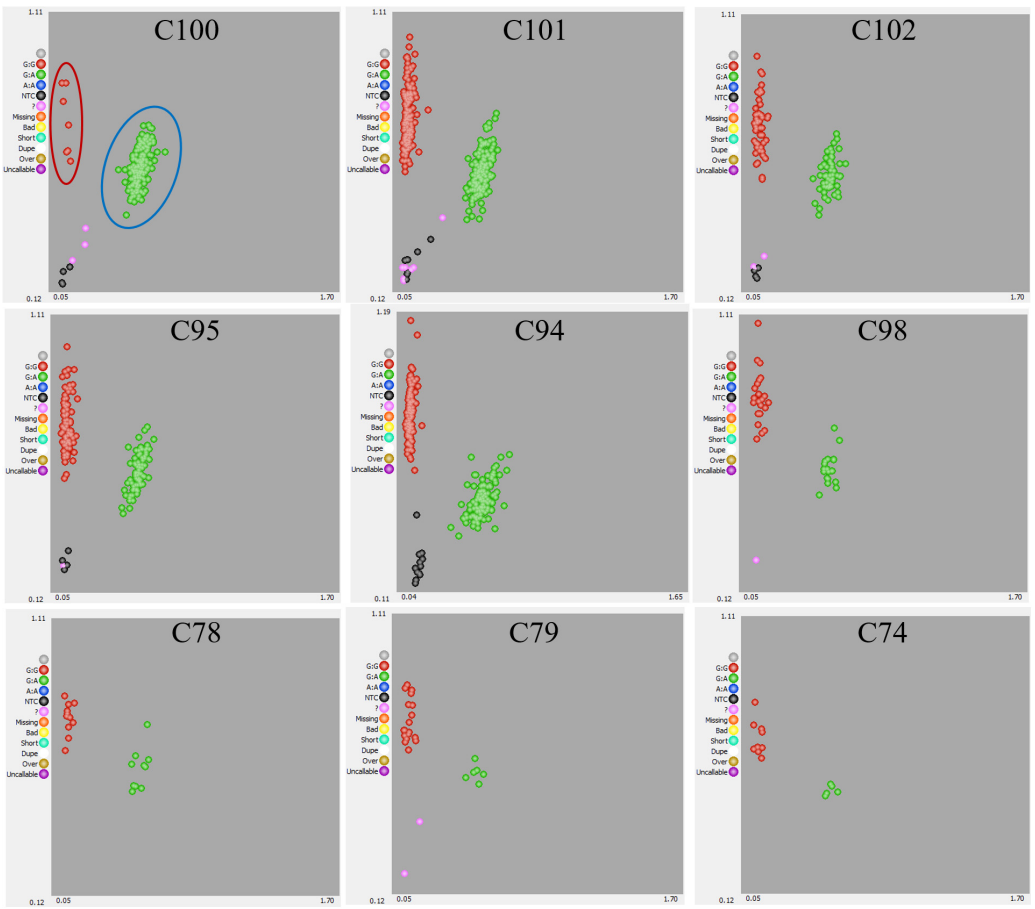

FIGURE 1 | Genotyping results of *FcRAN1* SNP by KASP assay. Genotypic results for 10 progenies derived from 10 cross combination using 7 pollen parent lines. The scatter plot with axes x and y represents allelic discrimination of *FcRAN1* genotypes. The red, green dots represent the A/A homozygous (female fig type), A/G heterozygous (male caprifig type) respectively. NTC, no template controls, or empty.

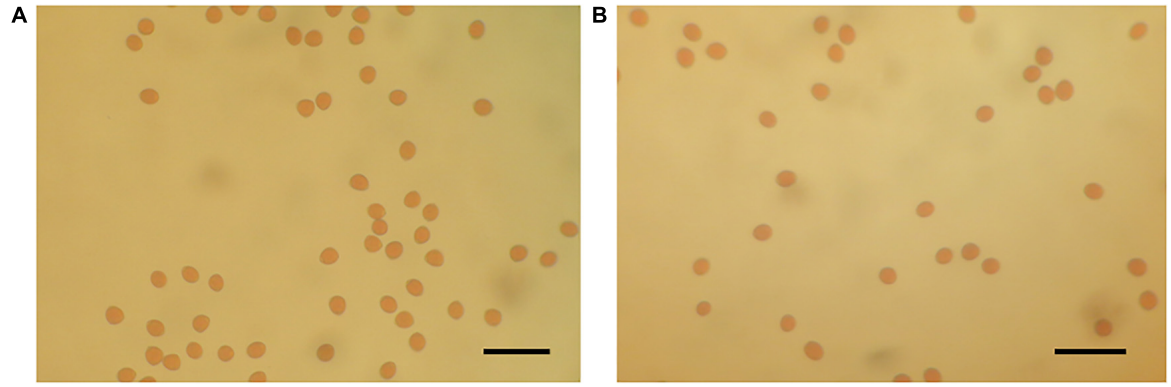

FIGURE 2 | Optical microscope images of stained pollen grains **(A)** H238-107 (distorting line: left) and **(B)** Kibaru (CH13) (non-distorting line: right). Bar = 40 μm.

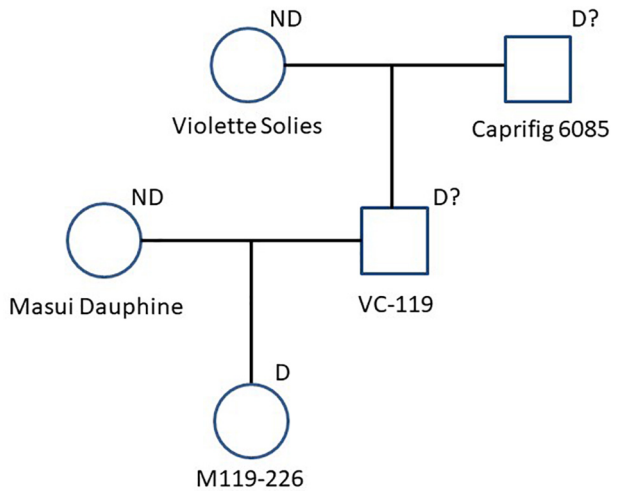

FIGURE 3 | Pedigree chart related to "M119-226." Seed parent (Masui Dauphine) and paternal grand seed parent (Violette Solies) are marked with "ND" as parents that didn't cause segregation distortion. "M119-226" and the other genotypes are marked with "D" or "D?" as parents that caused or could cause segregation distortion.

during the analysis of the seed parents, it was assumed that the Caprifig6085 and VC-119 genomes include a causative factor of segregation distortion. In this study, Caprifig6085 was used as a substitute for M119-226.

Since it is assumed that there are a wide variety of factors that can cause segregation distortion, the number of candidate genes involved is considered to be large corresponding to be the number of factors. For example, in *Arabidopsis thaliana*, 510 *EMBRYO-DEFECTIVE* (*EMB*) genes (Meinke, 2020) and 129 genes associated with female gametophyte deficiency (Pagnussat et al., 2005) have been identified, all of which may contribute to segregation distortion. During the analysis of the seed parents, 98 genes were detected as specific to M119-226. Additionally, we identified several genes similar to the above-mentioned *A. thaliana* genes related to female gametophyte deficiency.

The fertilization of the *A. thaliana* mutant line UNE15, which has a mutation in a gene encoding a LEA protein, is reportedly abnormal because of defective pollen tube guidance (Pagnussat et al., 2005). Accordingly, the female gametophytes of the seed parent M119-226 may be infertile because of a mutation similar to that in UNE15. In addition to the *LEA* gene, other candidate genes may be responsible for segregation distortion, including genes encoding proteins with unknown functions.

If segregation distortion is due to male gametophytic lethality, the causal allele (gene) is likely located on the X chromosomes of the pollen parents because most of the genotyped alleles in the C100 population were the G/A allele derived from the Y chromosome. However, an examination of the pollen used for mating indicated there were no differences in the shapes and appearance of the pollen grains between the pollen parent lines with and without distorted segregation, implying the distortion was at least independent of the pollen physiological state. It remains possible there is a defect in the process after the pollination event (e.g., pollen tube elongation). The TATA element modulatory factor (TMF) (FCA_r2.3chr09: 8,140,218) may affect the segregation distortion. The *A. thaliana* TFIID factor AtTAF6 (i.e., a TATA-binding protein) controls pollen tube growth (Lago et al., 2005). Additionally, TMF binds to the HIV-1 TATA element and inhibits the transcriptional activation by the TATA-binding protein (Garcia et al., 1992).

Although the current study did not determine whether *FcRAN1* is a true sex-determining gene in *F. carica*, a male-biased *FcRAN1* genotype ratio suggests that Y chromosome drive (Voelker, 1972; Sweeny and Barr, 1978; Wood and Newton, 1991; Presgraves et al., 1997) is important (**Figure 6**). For a Y chromosome drive system to be functional, the distorter gene and *FcRAN1* (i.e., potential sex-determining gene) should be at different loci on the same Y chromosome (Fc01a) of the pollen parent. Of the nine genes detected on chromosome Fc01a (**Supplementary Table 2**), the RNase H-encoding gene should be investigated for its potential role in segregation distortion. Simoni et al. (2020) introduced a sex-distorter (I-*Ppo*I) and a CRISPR-based gene drive into mosquito (*Anopheles gambiae*), resulting in stable Y chromosome drive and decreases in the proportion of male individuals in successive generations.

```
FcLEA_Horaishi          1 MATPTSSSKSEPKYAPLPPDENYVVLPLYFPRRRLRHYAISAVCLVLLISAAYAFWPSDP 60
FcLEA_Caprifig6085      1 MATPTSSSKSEPKYAPLPPDENYVVLPLYFPRRRLRHYAISAVCLVLLISAAYAFWPSDP 60
FcLEA_Masui Dauphine    1 MATPTSSSKSEPKYAPLPPDENYVVLPLYFPRRRLRHYAISAVCLVLLISAAYAFWPSDP 60
                                                  *
FcLEA_Horaishi         61 DIKVVRMRLTSVRVHRTPHVAVDVSVSLTVRVRNDDVYSLDYTELDVAVGYRGKWLGHMR 120
FcLEA_Caprifig6085     61 DIKVVRMRLTSVRVHRTPHVAVDVSVSPTVRVRNDDVYSLDYTELDVAVGYRGKWLGHMR 120
FcLEA_Masui Dauphine   61 DIKVVRMRLTSVRVHRTPHVAVDVSVSLTVRVRNDDVYSLDYTELDVAVGYRGKWLGHMR 120

FcLEA_Horaishi        121 SDRGHVRAQGSSYVDAELNLDGVEIFSDVVFLIEDLAKGTVRFDTVTVVKGQLGLLFLHF 180
FcLEA_Caprifig6085    121 SDRGHVRAQGSSYVDAELNLDGVEIFSDVVFLIEDLAKGTVRFDTVTVVKGQLGLLFLHF 180
FcLEA_Masui Dauphine  121 SDRGHVRAQGSSYVDAELNLDGVEIFSDVVFLIEDLAKGTVRFDTVTVVKGQLGLLFLHF 180

FcLEA_Horaishi        181 PLQAKVSCELSINTINQTILRQNCYPVVSSSYIYTCFSS                   219
FcLEA_Caprifig6085    181 PLQAKVSCELSINTINQTILRQNCYPVVSSSYIYTCFSS                   219
FcLEA_Masui Dauphine  181 PLQAKVSCELSINTINQTILRQNCYPVVSSSYIYTCFSS                   219
```

FIGURE 4 | Amino acid sequence alignment of *FcLEA* genes (LC639910). "Horaishi" is the variety of the reference genome. "Caprifig6085" indicates a variety with segregation distortion, and "Masui Dauphine" indicates a representative varietie of the group without segregation distortion. The leucine residue marked with an asterisk (*) in *FcLEA* was mutated to proline to make a non-synonymous substitution. Numbers in left side of sequence indicate the position of the beginning amino acid residue in each protein sequence.

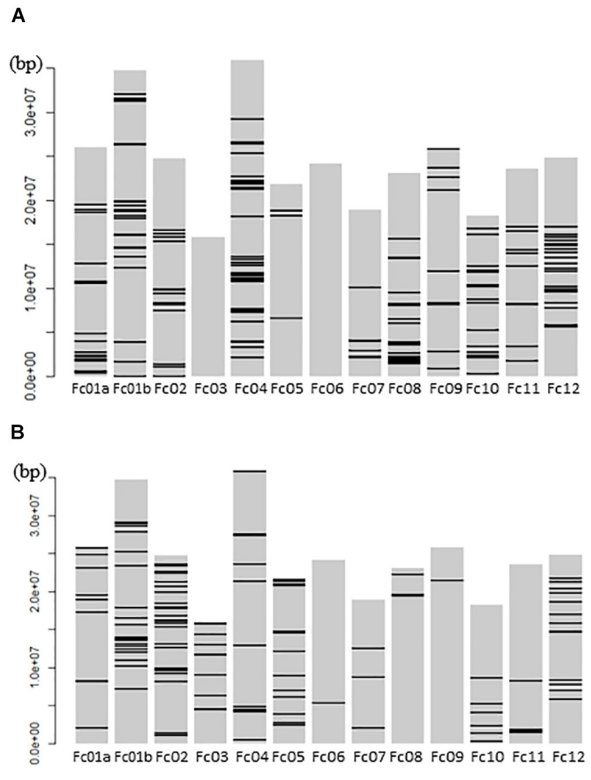

FIGURE 5 | Schematic Diagram showing GWLA SNPs that mapped to the whole chromosome pseudomolecules. **(A)** High impact SNP marker map in pollen parents and **(B)** high impact SNP marker map in seed parents. Horizontal lines on each chromosome indicate the location of SNP markers. The left side number indicate the distance from a telomere point of Fc01a chromosome pseudomolecule.

Thus, a site-specific DNA double-strand break system involving endonucleases, including I-*Ppo*I, is necessary for a functional Y chromosome drive (Ohle et al., 2016).

The *FcRAN1* KASP marker analysis indicated that some plants in the C100 population had the G/A genotype. There are two possible explanations for this observation. First, the effect of the distorter may have been incomplete. Second, specific events (e.g., chromosomal recombination) may have occurred in which the distorter gene was not activated. Future studies should further characterize the genetic background of the distortion via crossing experiments involving multiple seed parents, including H238-107, and the female C100 progeny. Moreover, because the candidate genes discussed herein were selected on the basis of bioinformatics data, their functions will need to be validated through genetic transformation or genome editing.

The high-impact SNPs in the pollen parents were within 5 Mbp from the end of Fc01a (**Figure 5**), suggestive of a relationship between the sex-linked region and the segregation distortion. Because the sex-determining gene (i.e., *FcRAN1* or a nearby gene) basically affects the population sex ratio, the distorter gene and *FcRAN1* are likely located close together and behave similarly.

The biological or evolutionary significance of the segregation distortion observed in figs is unclear. However, the occurrence of segregation distortion can change the genetic composition of the species population and may maintain or increase the genetic diversity of the population. In fact, segregation distortion reportedly leads to stable large-scale genomic differences between males and females in hybrid ants (Kulmuni et al., 2010).

Because the *FcRAN1* genotype is strongly correlated with the sex phenotype, the genotypic distortion could influence the evolution of sexual reproduction and sex chromosomes in various ways (Kozielska et al., 2010). Fig X and Y chromosomes are homomorphic, with no significant differences in their molecular structure (Storey, 1975), indicating that fig sex chromosomes are like autosomal chromosomes and are not well differentiated. Therefore, clarifying the segregation distortion of the *FcRAN1* genotype may help elucidate the differentiation of fig sex and autosomal chromosomes.

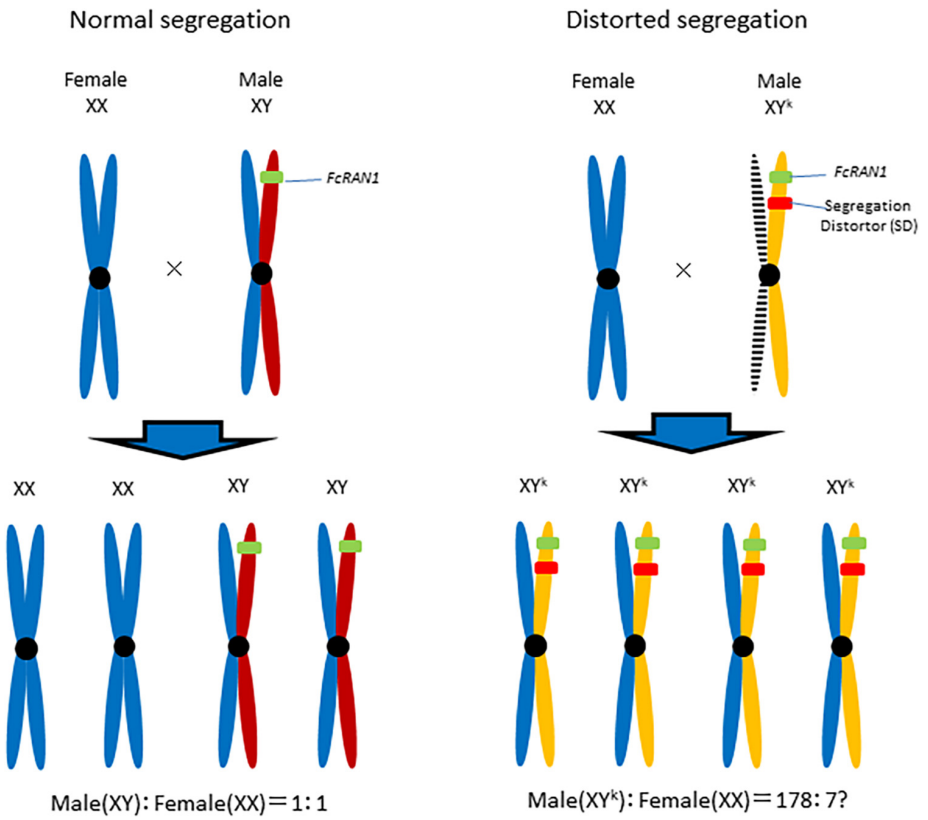

FIGURE 6 | Y chromosome drive model for the role of distorter in Segregation Distortion (SD). Segregation distorter, on the fig Y-chromosome, is expressed during male meiosis to cut a target sequence on the X-chromosome. The shredding of the X-chromosome favors the unaffected Y-carrying gametophyte and results in the production of a male-biased progeny.

MATERIALS AND METHODS

Seven seed parents (M119-226, V180-13, Masui Dauphine, H180-2, HIG5, Toyomitsuhime, and Burjassotte Greece) and six pollen parents (H238-107, M238-1, VD238-83, VS238-53, NG60, and CH13) were used for hybridizations, which were completed between 1989 and 2018 according to Storey (1975). The obtained seeds were sown in a seedbed and the resulting young seedlings were transplanted to 9-cm pots containing dedicated soil. Leaf samples were collected from the seedlings for KASP and CAPS marker analyses as described below.

KASP and CAPS Marker Analyses

Of the 1,232 plants resulting from 10 cross combinations, 1,220 (i.e., from the C100, C101, C102, C95, C94, C98, C78, C79, and C74 populations) were included in the *FcRAN1* KASP marker analysis, which was performed according to the manufacturer-recommended protocol for the LGC Genomics high-throughput genotyping system. The KASP Master mix and consumable materials were supplied by LGC Genomics (Middlesex, United Kingdom). A PCR primer pair, *FcRAN1*_f (5′-AGATCCTTAGTTGATGGGGT-3′) and *FcRAN1*_r (5′-CCTCAAACATGTTTAGACTG-3′), was designed based on the sequences flanking the SNP in the probe sequence: GAAGGTTTAAATTAC[A/G]TGTTGCTAATCCTT. Additionally, the 12 plants in the C205 population were genotyped using the CAPS marker as described by Mori et al. (2017). The SNP genotyping data generated from the KASP marker analysis were visualized using the SNPviewer software (LGC, Biosearch Technologies, Beverly, MA, United States).

Pollen Examination

The pollen of each paternal line was immersed undisturbed in a 600 μl 1% (w/v) iodine solution for 5 min, after which 50 μl of the mixture was added to a microscope slide, covered with a cover glass, and examined using an optical microscope. Pollen grains stained orange or brown were considered to be viable and were counted.

Whole-Genome Resequencing and Data Processing

Total genomic DNA was extracted from the leaves of eight caprifig-type varieties (H238-107, M238-1, VS238-53, M106-238, CH13, VD238-83, NG-60, and VC-180) and eight fig-type varieties (Masui Dauphine, H180-2, HIG5, Toyomitsuhime, Burjassotte Greece, Violette de Dauphine, Violette de Soloes, and Royal Vineyard) using the DNeasy Plant kit (Qiagen, Valencia, CA, United States). Sequencing libraries were constructed using the genomic DNA as previously described

(Shirasawa et al., 2016). Sequences were obtained using the HiSeq 2000 system in the paired-end mode, with a read length of 151 bp. The generated data were processed as described by Shirasawa et al. (2020). High-quality reads were selected by trimming adapters with fastx_clipper (parameter: -a AGATCGGAAGAGC) in the FASTX Toolkit (version 0.0.13)[1] and deleting low-quality bases with PRINSEQ (version 0.20.4) (Schmieder and Edwards, 2011). Reads were aligned to a common fig reference genome sequence (Shirasawa et al., 2021, unpublished data, the data is available upon request) using Bowtie2 (version 2.2.3) (Langmead and Salzberg, 2012). Sequence variants were detected using the mpileup command in SAMtools (version 0.1.19) (Li et al., 2009). High-confidence heterozygous SNPs were selected using VCFtools (version 0.1.12b) (parameters: –minDP5 –minQ 999 –max-missing 0.75) (Danecek et al., 2011). The SNP effects were categorized using SnpEff (version 3.0) The default parameters of SnpEff were used to analyze variant effects (Cingolani et al., 2012). Genes were predicted and annotated according to previousinformation (Shirasawa et al., 2020).

Linkage Analysis

The GWLAs of 10 seed parents (M119-226, V180-13, Masui Dauphine, H180-2, HIG5, Toyomitsuhime, Burjassotte Greece, Violette de Dauphine, Violette de Soloes, and Royal Vineyard) and eight pollen parents (H238-107, M238-1, VS238-53, CH13, VD238-83, NG-60, M106-238, and VC-180) were performed using the corresponding genotype datasets derived from the whole-genome resequencing. In each panel, only SNPs with a minor allele frequency >5% were used for the linkage analysis. Case-control association analyses treating the non-distorting type as the control and the distorting type as the case for high-, moderate-, and low-impact SNP sets were conducted to calculate allele frequencies using PLINK 1.90 beta (Purcell et al., 2007). The sex phenotype dataset was generated in a recent study (Mori et al., 2017). The SNPs with an F_A (frequency of this allele in cases) of 1.0 or 0.5 and an F_U (frequency of this allele in controls) of 0.0 were selected as candidate SNPs responsible for the difference between distorting and non-distorting lines.

Accession Numbers

The *FcLEA* gene sequence has been deposited in DDBJ/EMBL/GenBank under accession number LC639910. Genome sequencing reads of caprifig lines have been deposited in the NCBI Sequence Read Archive under accession numbers DRA011350 and DRA012209.

[1] http://hannonlab.cshl.edu/fastx_toolkit

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

HI and HY designed the study. HI, KS, HY, SY, TH, and HN completed the analyses. HI, KS, HY, SY, MS, TH, KT, and HN interpreted the data. HI, KS, and HY wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by a grant from the Ministry of Education, Culture, Sports, Science and Technology/Japan Society for Promoting Science (KAKENHI grant 19H02952 and 16H04878).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.647599/full#supplementary-material

Supplementary Table 1 | "Type" refers to the genetic segregation type: 1| 0 means that the target sample is homozygous for the reference sequence and the other samples are homozygous for the other sequence, whereas 0.5| 0 means the target sample is heterozygous for the reference sequence and the other samples are homozygous for the other sequence.

Supplementary Table 2 | "Type" refers to the genetic segregation type: 1| 0 means that the target sample is homozygous for the reference sequence and the other samples are homozygous for the other sequence, whereas 0.5| 0 means the target sample is heterozygous for the reference sequence and the other samples are homozygous for the other sequence.

REFERENCES

Burt, A., and Trivers, R. (2006). *Genes in Conflict. The Biology of Selfish Genetic Elements*. Cambridge, MA: Harvard University Press.

Candela, H., Pérez-Pérez, J. M., and Micol, J. L. (2011). Uncovering the post-embryonic functions of gametophytic- and embryonic-lethal genes. *Trends Plant Sci.* 16, 336–345. doi: 10.1016/j.tplants.2011.02.007

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695

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

Dawkins, R. (1976). *The Selfish Gene. Best Books*. Oxford: Oxford University Press.

Dobrovolskaia-Zavadskaia, N., and Kobozieff, N. (1927). Sur la reproduction des souris anoures. *C. R. Soc. Biol.* 97, 116–118.

Flaishman, M., Rodov, V., and Stover, E. (2008). "The fig: botany, horticulture, and breeding," in *Horticultural Reviews*, ed. J. Janick (Hoboken, NJ: John Wiley & Sons), 113–196.

Garcia, J. A., Ignatius Ou, S.-H., Wu, F., Lusis, A. J., Sparkes, R. S., and Gaynor, R. B. (1992). Cloning and chromosomal mapping of a human immunodeficiency virus 1 "TATA" element modulatory factor. *Proc. Natl. Acad. Sci. U.S.A.* 89, 9372–9376. doi: 10.1073/pnas.89.20.9372

Gershenson, S. (1928). A new sex-ratio abnormality in DROSOPHILA OBSCURA. *Genetics* 13, 488–507. doi: 10.1093/genetics/13.6.488

Hickey, W. A., and Craig, G. B. Jr. (1966). Genetic distortion of sex ratio in a mosquito, *Aedes aegypti*. *Genetics* 53, 1177–1196. doi: 10.1093/genetics/53.6.1177

Kozielska, M., Weissing, F. J., Beukeboom, L. W., and Pen, I. (2010). Segregation distortion and the evolution of sex-determining mechanisms. *Heredity* 104, 100–112. doi: 10.1038/hdy.2009.104

Kulmuni, J., Seifert, B., and Pamilo, P. (2010). Segregation distortion causes large-scale differences between male and female genomes in hybrid ants. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7371–7376. doi: 10.1073/pnas.0912409107

Lago, C., Clerici, E., Dreni, L., Horlow, C., Caporali, E., Colombo, L., et al. (2005). The Arabidopsis TFIID factor AtTAF6 controls pollen tube growth. *Dev. Biol.* 285, 91–100. doi: 10.1016/j.ydbio.2005.06.006

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lashermes, P., Combes, M. C., Prakash, N. S., Trouslot, P., Lorieux, M., and Charrier, A. (2001). Genetic linkage map of Coffea canephora: effect of segregation distortion and analysis of recombination rate in male and female meioses. *Genome* 44, 589–596. doi: 10.1139/gen-44-4-589

Lewis, D. (1941). Male sterility in natural populations of hermaphrodite plants the equilibrium between females and hermaphrodites to be expected with different types of inheritance. *New Phytol.* 40, 56–63. doi: 10.1111/j.1469-8137.1941.tb07028.x

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U.S.A.* 36, 344–355. doi: 10.1073/pnas.36.6.344

Meinke, D. W. (2020). Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in Arabidopsis. *New Phytol.* 226, 306–325. doi: 10.1111/nph.16071

Mori, K., Shirasawa, K., Nogata, H., Hirata, C., Tashiro, K., Habu, T., et al. (2017). Identification of RAN1 orthologue associated with sex determination through whole genome sequencing analysis in fig (*Ficus carica* L.). *Sci. Rep.* 7:41124.

Ohle, C., Tesorero, R., Schermann, G., Dobrev, N., Sinning, I., and Fischer, T. (2016). Transient RNA-DNA hybrids are required for efficient double-strand break repair. *Cell* 167, 1001–1013.e7.

Ostergren, G. (1945). Parasitic nature of extra fragment chromosomes. *Bot. Not.* 2, 157–163.

Pagnussat, G. C., Yu, H. J., Ngo, Q. A., Rajani, S., Mayalagu, S., Johnson, C. S., et al. (2005). Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. *Development* 132, 603–614. doi: 10.1242/dev.01595

Presgraves, D. C., Severance, E., and Wilkinson, G. S. (1997). Sex chromosome meiotic drive in stalk-eyed flies. *Genetics* 147, 1169–1180. doi: 10.1093/genetics/147.3.1169

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 6:e17288. doi: 10.1371/journal.pone.0017288

Shirasawa, K., Hirakawa, H., and Isobe, S. (2016). Analytical workflow of double-digest restriction site-associated DNA sequencing based on empirical and in silico optimization in tomato. *DNA Res.* 23, 145–153. doi: 10.1093/dnares/dsw004

Shirasawa, K., Yakushiji, H., Nishimura, R., Morita, T., Jikumaru, S., Ikegami, H., et al. (2020). The Ficus erecta genome aids Ceratocystis canker resistance breeding in common fig (F. carica). *Plant J.* 102, 1313–1322. doi: 10.1111/tpj.14703

Simoni, A., Hammond, A. M., Beaghton, A. K., Galizi, R., Taxiarchi, C., Kyrou, K., et al. (2020). A male-biased sex-distorter gene drive for the human malaria vector *Anopheles gambiae*. *Nat. Biotechnol.* 38, 1054–1060. doi: 10.1038/s41587-020-0508-1

Storey, W. B. (1975). "Figs," in *Advances in Fruit Breeding*, eds J. Janick and J. N. Moore (West Lafayette, IN: Purdue University Press), 568–589.

Stover, E., Aradhya, M., Ferguson, L., and Crisosto, C. H. (2007). The fig: overview of an ancient fruit. *HortScience* 42:1083. doi: 10.21273/hortsci.42.5.1083

Sturtevant, A. H., and Dobzhansky, T. (1936). Geographical distribution and cytology of "Sex Ratio" in Drosophila pseudoobscura and related species. *Genetics* 21, 473–490. doi: 10.1093/genetics/21.4.473

Sweeny, T. L., and Barr, A. R. (1978). Sex ratio distortion caused by meiotic drive in a mosquito, *Culex pipiens* L. *Genetics* 88, 427–446. doi: 10.1093/genetics/88.3.427

Taylor, D. R., and Ingvarsson, P. K. (2003). Common features of segregation distortion in plants and animals. *Genetica* 117, 27–35.

Voelker, R. A. (1972). Preliminary characterization of "sex ratio" and rediscovery and reinterpretation of "male sex ratio" in *Drosophila affinis*. *Genetics* 71, 597–606. doi: 10.1093/genetics/71.4.597

Werren, J. H., Nur, U., and Wu, C.-I. (1988). Selfish genetic elements. *Trends Ecol. Evol.* 3, 297–302.

Wood, R. J., and Newton, M. E. (1991). Sex-ratio distortion caused by meiotic drive in mosquitoes. *Am. Nat.* 137, 379–391.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 20 October 2021
doi: 10.3389/fpls.2021.755358

AGAMOUS Gene as a New Sex-Identification Marker in Fig (*Ficus carica* L.) Is More Efficient Than *RAN1*

*Xu Wang[1], Miaoyu Song[1], Moshe A. Flaishman[2], Shangwu Chen[3] and Huiqin Ma[1]**

[1] Department of Fruit Tree Sciences, College of Horticulture, China Agricultural University, Beijing, China, [2] Department of Fruit Tree Sciences, Agricultural Research Organization, The Volcani Center, Bet Dagan, Israel, [3] College of Food Science and Nutrition Engineering, China Agricultural University, Beijing, China

Fig is an ancient gynodioecious fruit tree with females for commercial fruit production and hermaphrodites (males) sometimes used as pollen providers. An early sex-identification method would improve breeding efficiency. Three *AGAMOUS* (*AG*) genes were recruited from the *Ficus carica* genome using *AG* sequences from *Ficus microcarpa* and *Ficus hispida*. *FcAG* was 5230 bp in length, with 7 exons and 6 introns, and a 744-bp coding sequence. The gene was present in both female and male fig genomes, with a 15-bp deletion in the 7th exon. The other two AG genes (*FcAG2-Gall_Stamen* and *FcAG3-Gall_Stamen*) were male-specific, without the 15-bp deletion (759-bp coding sequence), and were only expressed in the gall and stamen of the male fig fruit. Using the deletion as the forward primer (AG-Marker), male plants were very efficiently identified by the presence of a 146-bp PCR product. The previously reported fig male and female polymorphism gene *RESPONSIVE-TO-ANTAGONIST1* (*RAN1*) was also cloned and compared between male and female plants. Fifteen SNPs were found in the 3015-bp protein-coding sequence. Among them, 12 SNPs were identified as having sex-differentiating capacity by checking the sequences of 27 known male and 24 known female cultivars. A RAN1-Marker of 608 bp, including 6 SNPs, was designed, and a PCR and sequencing-based method was verified with 352 fig seedlings from two hybrid populations. Our results confirmed that the newly established AG-Marker is as accurate as the RAN1-Marker, and provide new clues to understanding *Ficus* sex determination.

Keywords: *FcAG*, *FcRAN1*, molecular marker, sex identification, *Ficus carica* L.

OPEN ACCESS

Edited by:
Inaki Hormaza, Institute of Subtropical and Mediterranean Horticulture La Mayora, Spain

Reviewed by:
Lei Gao, Wuhan Botanical Garden, Chinese Academy of Sciences (CAS), China
Jia-Long Yao, The New Zealand Institute for Plant and Food Research Ltd., New Zealand

****Correspondence:***
Huiqin Ma
hqma@cau.edu.cn

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

Received: *08 August 2021*
Accepted: *29 September 2021*
Published: *20 October 2021*

Citation:
Wang X, Song M, Flaishman MA, Chen S and Ma H (2021) AGAMOUS Gene as a New Sex-Identification Marker in Fig (Ficus carica L.) Is More Efficient Than RAN1. Front. Plant Sci. 12:755358. doi: 10.3389/fpls.2021.755358

INTRODUCTION

Fig (*Ficus carica* L., Moraceae) is regarded as one of the earliest domesticated fruit trees, widely cultivated in the Mediterranean region, and in other parts of the world (Kislev et al., 2006). As with other species in the genus *Ficus*, the fig is gynodioecious. The trees can be divided into female figs and caprifigs (usually called male figs) (Ikegami et al., 2013). The female figs bear edible fruit. Upon ripening, the long-styled female flowers (pistils) develop inside the syconium, making up most of the edible part of the fruit (Stover et al., 2007). The caprifig is not usually edible, but stamens growing inside the syconia produce pollen, and the quality and storability of female fig fruit improve after pollination (Lama et al., 2020; Marcotuli et al., 2020). In addition, there are short-styled female

flowers (galls) in the syconia of caprifigs that can serve as hosts for fig-wasp (*Blastophaga psenes* L.) larvae (Gu and Yang, 2013; Zhang et al., 2020; Wang et al., 2021).

The evolution of fig from monoecious to dioecious was the first step in the double mutation model (Charlesworth and Charlesworth, 1978; Charlesworth, 2016). A recessive homozygous mutation appeared in wild fig trees (caprifigs), which resulted in male flower abortion but retention of female function, resulting in fig trees with all female flowers and significantly improved fruit taste. To date, no dominant mutation has been found that suppresses female function, resulting in all male individuals.

Key genes regulating plant sex differentiation are of high interest in the plant sciences. Examples of these include: the candidate male-determining gene *OGI* and the anther fertility-regulating gene *MeGI* found in persimmon (*Diospyros lotus*) (Akagi et al., 2014); the regulators of male flower development *CYP703* and *GPAT3* identified in *Phoenix dactylifera* (Torres et al., 2018); the feminization-suppressing gene *SyGI* and male-function-maintaining gene *FrBy* reported in kiwifruit (Akagi et al., 2019); a mutation of gene *INP1* in *Vitis vinifera* resulting in male sterility of grape; the transcription factor *YABBY3* in *V. vinifera* found to be associated with female sterility (Massonnet et al., 2020); and the female-suppressor gene *SOFF* and the male-promoting gene *aspTDF1* identified in *Asparagus officinalis* (Harkess et al., 2020).

According to previous reports, the short-styled female flower (gall, G) is dominant over long-styled female flower (pistil, g), and stamen presence (A) is dominant over stamen absence (a) (Storey, 1975; Stover et al., 2007). Crossing between caprifigs can produce fertile offspring—a prerequisite for the shift from monoecy to the development of all female figs. A caprifig with a recessive mutation (GA/ga) will produce 25% recessive homozygous offspring, i.e., female figs (ga/ga) (Mori et al., 2017). The GA and ga alleles always appear in pairs, and there is a linkage in the key genes regulating floral organ differentiation.

The first-draft whole-genome sequence of fig *F. carica* was published in 2017 using female cv. Horaishi (Mori et al., 2017). Restriction-site-associated sequencing technique and genome-wide association analysis were used to identify the candidate sex-determining gene *RESPONSIVE-TO-ANTAGONIST1* (*RAN1*), located on the seq000259 scaffold in the 9.2–12.1 cM region of the Fc01a linkage group. Two missense mutations in *RAN1* were found to be strongly associated with fig sex phenotypes. The *FcRAN1* gene had 9 exons, and a 3015-bp protein-coding sequence. *FcRAN1* was the homolog of a copper-transporting ATPase, and the gene was also named *Heavy Metal ATPase 7* (*HMA7*). In *Arabidopsis thaliana*, *RAN1* was located on the Golgi membrane, and was responsible for transporting copper ions to ethylene receptors and activating them (Hirayama et al., 1999; Woeste and Kieber, 2000; Binder et al., 2010). *FcRAN1* was preliminarily speculated to be the sex-determining *A* gene for figs, and a cleaved amplified polymorphism sequence (CAPS)-based method was established (Mori et al., 2017). Ikten and Yilmaz (2019) verified the CAPS method and found that restriction endonuclease *Pci*I provides 100% sex discrimination for PCR products, while restriction endonuclease *Hpy*CH4IV gave some false-positive results. Nevertheless, restriction endonuclease digestion is time-consuming, the required enzymes are expensive, and well-controlled conditions are needed for successful performance.

The role of *RAN1* in fig sex determination was challenged by Zhang et al. (2020). Whole-genome sequencing and assembly of two *Ficus* species—*F. microcarpa* (monoecious) and *F. hispida* (functional dioecious)—revealed a male-specific sex-determination candidate gene *FhAG2* (floral homeotic protein AGAMOUS) (Zhang et al., 2020). *AtAG* belongs to the C-class of the MADS-box gene family, which has been confirmed to regulate the determination of stamen, carpel and floral meristem (Dreni and Kater, 2014). Transcriptome data showed that *FhAG2* was only highly expressed in the stamens of male *F. hispida* syconia; *FhAG3*, which was not anchored to any chromosome, was also specifically expressed in stamens. The results strongly suggested that *FhAG2* is the critical sex-determination gene (Zhang et al., 2020).

In this study, we tried to establish a convenient, reliable and rapid molecular marker for the identification of male and female figs, without relying on rare and expensive restriction endonucleases. Both *AG* and *RAN1* were cloned and their sequences were compared using known-sex cultivars and hybrid progeny, resulting in the development of two markers. Our results confirmed *FcAG* as a new sex-identification marker, and the possible involvement of *AG* and *RAN1* in fig sex determination is discussed.

MATERIALS AND METHODS

Plant Materials and Preparation of Amplification Templates

In this study, leaf, root, stem, and fruit were collected from 27 male and 24 female fig cultivars. Leaf and fruit of two F1 hybrid populations (group 207, $n = 263$; group 205, $n = 89$) were also used. The detailed list of materials is shown in **Supplementary Table 1**. Pistil, stamen, gall, and peel were separated from female and male fig fruit, and all collected fresh materials were quickly frozen in liquid nitrogen, fully ground, and stored at −80°C. The genomic DNA (gDNA) and total RNA were extracted by the CTAB method as described previously (Cui et al., 2021); concentration and purity were measured in a NanoDrop2000 spectrophotometer (Thermo Scientific, United States). Integrity of gDNA and total RNA was checked by 1% agarose gel (Biowest, Spain) electrophoresis (Reid et al., 2006; Chai et al., 2019), and RNA concentration was normalized (Wang et al., 2019). Then the total RNA was treated with DNase I (D2270A, Takara, Japan) at 37°C for 30 min to remove any contaminating DNA, and mRNA was then isolated from 2 μg total RNA using cellulose oligo (dT) magnetic beads (3806, Takara, Japan); cDNA was synthesized with a cDNA Synthesis Kit (639552, Takara, Japan).

Primer Sequence Design

Specific primers AG-F and AG-R were designed according to the coding sequences (CDSs) of the genus *Ficus* (*F. microcarpa*, *F. hispida,* and *F. carica*) (**Table 1**). AG-Marker was a pair

TABLE 1 | Primers used in this study.

Primer name	Primer sequence (5′→3′)	Length (bp)	Target length (bp)
AG-F	ATGKCGTWCCAAAACAAGGWGAWGA	25	759
AG-R	TTASACTAATTGGAGGGCCATGGRA	25	
AG-Marker-F	CAGGAGGAGGAAGCT	15	146
AG-Marker-R	TTACACTAATTGGAGGGCCA	20	
FcRAN1-F	ATGGCGGCGAGCGTCCGACA	20	3015
FcRAN1-R	CTATTCTACAGTTATTTCTAGTATA	25	
RAN1-Marker-F	ATATCAAGAATGCAATCGAGGA	22	608
RAN1-Marker-R	GCTGAGAAATAGACTAGAGATG	22	
qPCR-F	CCACTGGCAAAGGCAATAGT	20	156
qPCR-R	TTGAACTCCTCTGCCGGGAA	20	
1-SNP-F	TCCTGTCTTTCTCATACGAGTAGTT	25	371
1-SNP-R	TGGCGTCAGATGTTTTCCCC	20	
6SNP-2-F	CTGGAGCTCTTCGTGCTTAC	20	766
6SNP-2-R	AAGCATTGAACTCCTCTGCCG	21	
2-SNP-F	TCACGATGTAAGGGCAGAGG	20	444
2-SNP-R	GTGGTCAGCCTTGGTTTTCTG	21	
Actin-F	GAACCACCAGACAGGACGATG	21	250
Actin-R	CTACCACTGCTGAACGGGAAA	21	

AG primers contain degenerate bases: K(G/T), W(A/T), S(G/C), R(A/G).

of male-specific primers for *FcAG2-Gall_Stamen* and *FcAG3-Gall_Stamen*. FcRAN1 was the full-length gene primer pair based on the *FcRAN1* cDNA sequences of female fig "Horaishi" and "Caprifig6085" reported by Mori et al. (2017), and the quantitative real-time PCR (qRT-PCR) primers of *FcRAN1* were designed from the non-mutated region of the 7th exon. The single nucleotide polymorphism (SNP)-identification primers 6SNP-1 (RAN1-Marker), 1-SNP, 6SNP-2, and 2-SNP were designed with 6, 1, 6, and 2 SNP sites in their amplified fragments, respectively. Actin was used as the housekeeping gene following Freiman et al. (2015).

PCR and Quantitative Real Time-PCR

High-fidelity enzyme (KOD-Plus-Neo, TOYOBO, Japan) was used for all PCRs related to gene cloning, sequencing and SNP-site identification. PCR amplification was carried out in a Model T100 thermal cycler (BioRad, United States). The PCR products were first checked by electrophoresis, then collected with an agarose gel purification and extraction kit (AP-GX-50, Axygen, United States). The fragments were individually ligated into the pClone007 vector (pClone007 Blunt Simple Vector Kit; Tsingke, China), and transformed into *Escherichia coli* DH5α cells; at least 15 single clones were picked and at least 10 clones were plasmid-sequenced, for each replication. Three biological replicates were used if not otherwise indicated. T5 Super PCR Mix (Tsingke) was used for amplification without further sequencing. The PCR products were checked by 1% agarose gel electrophoresis. ChamQ Universal SYBR qPCR Master Mix (Q711, Vazyme, China) was used for the qRT-PCRs carried out with QuantStudio^{TM} 6 (ABI, United States). Relative expression was calculated by the $2^{-\Delta\Delta Ct}$ method (Chai et al., 2018, 2019).

Bioinformatics Tools and Data Analysis

TBtools (Chen et al., 2020) was used for the following bioinformatics analyses: local sequence alignment (BLAST-BLAST GUI Wrapper – Several Sequences to a Big File), visualization of gene structure (Graphics – Bio Sequence Structure Illustrator – Gene Structure View), location of genes on chromosomes (Graphics – Show Genes on Chromosome – Gene Location Visualize), and chromosome collinearity analysis (Graphics – Synteny Visualization – One Step MCScanX and Dual Systeny Plot and Multiple Synteny Plot).

NCBI BLAST[1] was used for DNA sequence alignment, and UniProt BLAST[2] for protein sequence alignment. NCBI ORF Finder[3] was used to find possible protein-coding sequences. Novopro[4] was used for CDS translation. The phylogenetic tree was constructed using Mega software (Mega v. 11[5]) with maximum likelihood (ML) method, the Jones–Taylor–Thornton (JTT) model and 1000 bootstrap replications. DNAMAN software[6] was used for sequence alignment. SnapGene software[7] was used for SNP localization and sequence alignment.

The transcriptome data of fig cv. Purple Peel fruit at different developmental stages, previously produced by our laboratory, were used for data mining (SRA accession: PRJNA723733, NCBI) (Wang et al., 2017; Zhai et al., 2021). Accession numbers for *F. carica* cv. Horaishi and cv. Dottato genomes are GCA_002002945.1 and GCA_009761775.1, respectively, in NCBI GenBank (Mori et al., 2017; Usai et al., 2020); for *F. microcarpa* and *F. hispida* genomes: GWHABKV00000000.1 and GWHALOG00000000, respectively, in CNCB[8] (Zhang et al., 2020); and for the *Morus alba* genome: GCA_012066045.3 in NCBI GenBank (Jiao et al., 2020). The published female fig cv. Dottato genome was used to predict *FcAG, FcAG2-Gall_Stamen* and *FcAG3-Gall_Stamen* gene structure. All nucleotide and protein sequences used in this study are provided in **Supplementary Table 2**.

RESULTS

FcAG Identification and Sequence Alignment

Fifty-three AG-related proteins were predicted using Usai et al.'s (2020) published "Dottato" fig genome sequence, including one AG protein (Gene ID: *FCD_00034093*, **Supplementary Table 3**), one AG isoform X1 and 51 AGAMOUS-LIKE (AGL) proteins. Thirty-one AG-related genes were annotated from fruit transcriptome data of fig cv. Purple Peel, including one AG (Gene ID: *c23673_g1*, 991 bp, 744 bp ORF) and 30 AGLs (**Supplementary Table 3**). These proteins and mRNAs

[1]https://blast.ncbi.nlm.nih.gov/Blast.cgi
[2]https://www.uniprot.org/blast/
[3]https://www.ncbi.nlm.nih.gov/orffinder/
[4]https://www.novopro.cn/tools/translate.html
[5]https://www.megasoftware.net/
[6]https://www.lynnon.com/
[7]https://www.snapgene.com/try-snapgene/
[8]https://ngdc.cncb.ac.cn/

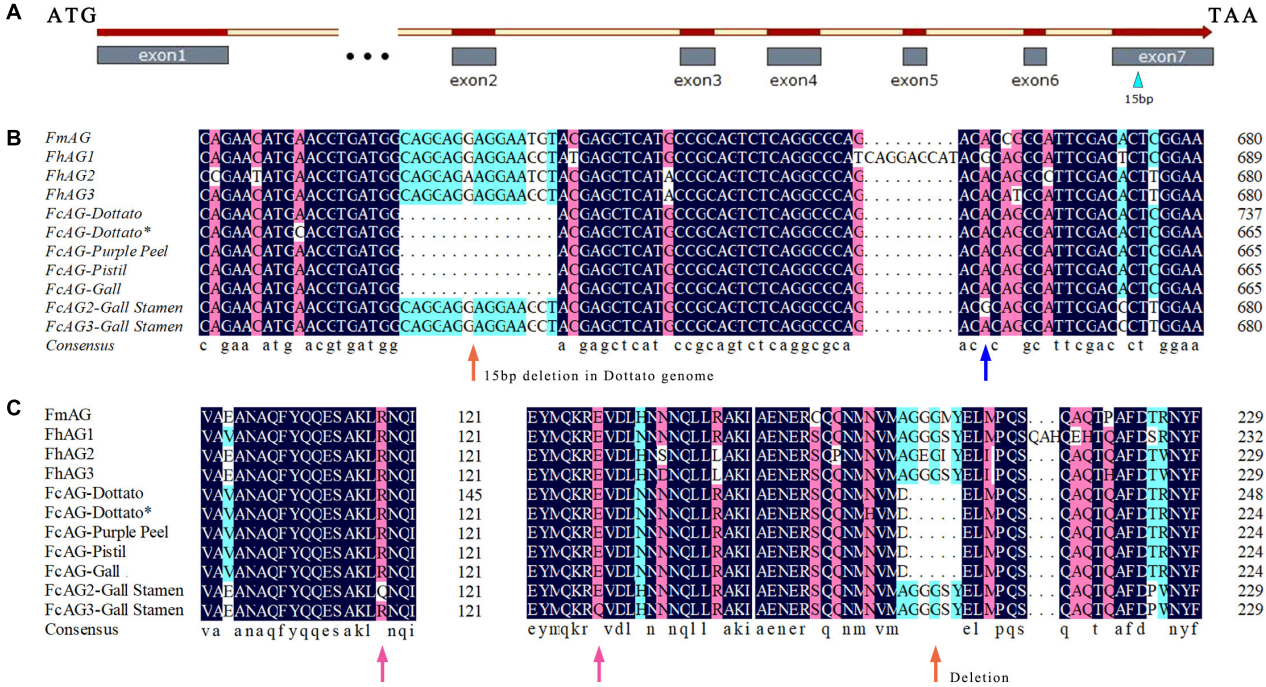

FIGURE 1 | *AG* genes of different *Ficus* species. **(A)** Gene structure of *FcAG*s; the male–female-differentiating 15 bp in the 7th exon is indicated by a blue triangle. **(B)** The 15-bp deletion (orange arrow) is common to the conserved *AG* genes, whereas the 15 bp are present in two other *AG* genes expressed in galls and stamens. Blue arrow indicates base difference. **(C)** Some synonymous and missense mutations (pink arrows) revealed by amino acid sequence prediction of the *AG* genes.

were further screened by aligning with *FmAG*, *FhAG1*, *FhAG2,* and *FhAG3* published by Zhang et al. (2020), and only the single homolog annotated as *AG* gene from the "Dottato" genome and "Purple Peel" transcriptome was identified and named *FcAG*, while the others were clustered as AGLs. Unfortunately, the recruited AG gene was not assembled into the "Dottato" chromosome.

Specific primers AG-F and AG-R were designed between the start and stop coding regions using the *AG* gene of *Ficus* (*F. microcarpa*, *F. hispida,* and *F. carica*) (**Table 1**). cDNA of "Purple Peel" pistil, and the gall and stamen of male fig cv. Syria_Xu were used as the templates for PCR amplification. Only one *AG* gene was found in the pistils and was named *FcAG-Pistil*. Alignment of *FcAG-Pistil* with the full-length DNA of *FcAG-Dottato* revealed 7 exons (**Figure 1A**).

Two other *AG* genes were found to be expressed in both stamens and galls, and were named *FcAG2-Gall_Stamen* and *FcAG3-Gall_Stamen*. Another AG homolog specifically expressed in gall tissues was identified and named *FcAG-Gall*, which had only a few base differences from *FcAG-Pistil*. The main difference between *FcAG2-Gall_Stamen*/*FcAG3-Gall_Stamen* and *FcAG-Pistil* was that the two former genes had the 15 bp in the 7th exon that were absent in *FcAG-Dottato*, *FcAG-Purple Peel*, *FcAG-Pistil,* and *FcAG-Gall* (**Figures 1A,B**). *FcAG2-Gall_Stamen* and *FcAG3-Gall_Stamen* were very similar, with only 9 base differences (**Supplementary Figure 1**), and the different bases led to changes in two amino acids (**Figures 1B,C**).

Clustering Analysis of AGs

The phylogenetic tree of fig AGs was constructed with the *F. microcarpa* and *F. hispida* counterparts and AtAG (*A. thaliana*, accession number: P17839), BnAG1 (*Brassica napus*, Q01540), NtAG1 (*Nicotiana tabacum*, Q43585), SlAG1 (*Solanum lycopersicum*, Q40168), MnAG (*Morus notabilis*, W9SCS9), and VvMADS1 (*V. vinifera*, Q93XH4) collected from UniProt (**Supplementary Table 2**). In addition, we recruited three *AG* genes from the *M. alba* genome and translated them into protein sequences. *MaAG2* and *MaAG3* were different transcript variants of the same gene (**Figure 2A**). The clustering results showed that *Ficus* AG proteins are rather conserved, with the AGs of *F. carica*, *F. microcarpa,* and *F. hispida* being classified into one group. Among this group, FcAG2-Gall_Stamen and FcAG3-Gall_Stamen in male fig clustered closer to the male-specific FhAG2 and FhAG3 of *F. hispida*, and these four AGs composed subgroup I. The remaining AG proteins, except for FmAG and FhAG1, could be classified into subgroup II, sharing the characteristic 15-bp deletion. AGs of the Moraceae family *M. alba* and *M. notabilis* were more clearly distinct from *Ficus* with respect to both protein sequence (**Figure 2A**) and gene structure (**Figure 2B**).

There were 7 exons in the *AG* genes of all tested species except *FcAG-Dottato* (**Figure 2B**). The *FcAG-Dottato* CDS, used for clustering, was extracted from the "Dottato" genome (Usai et al., 2020), where a 71-bp exon insertion was found between the 1st and 2nd exon of other *FcAG*s, and there was a 1-bp insertion in the 3rd exon (**Supplementary Figure 1**). As the two insertions

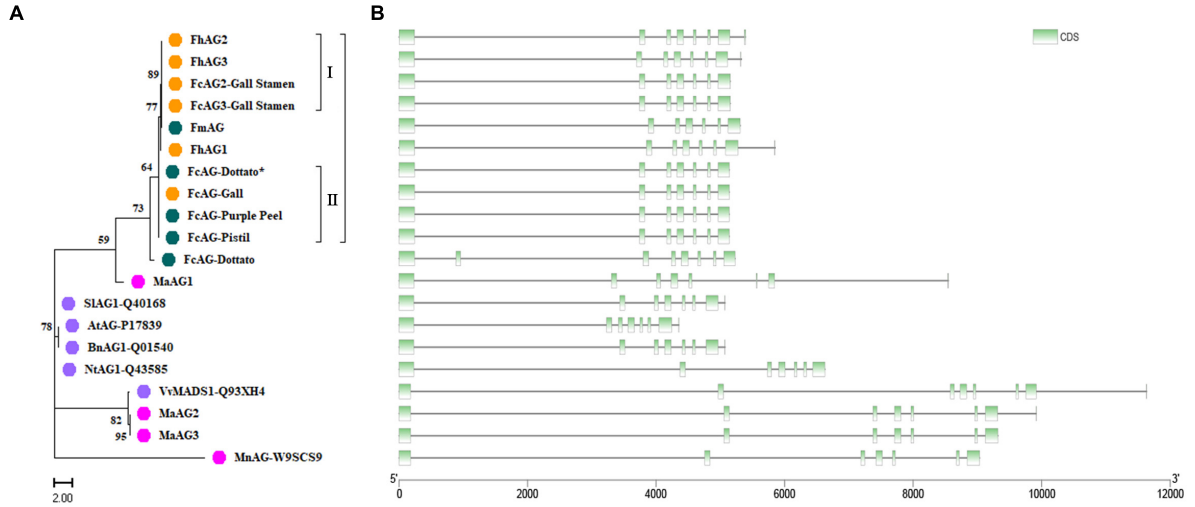

FIGURE 2 | Cluster analysis of *AG* genes and their encoded proteins from various species. **(A)** Phylogenetic tree. Orange – *Ficus* males; green – *Ficus* females; pink – Moraceae plants; purple – other plant species. **(B)** Gene structure. Each gene in panel **B** corresponds to the protein to its left in panel **A**. Green – coding sequence (CDS); black line – intron.

were not present in the other *Ficus AG*s, we speculated that they were errors in the "Dottato" genome assembly. Therefore, we amplified the *AG* gene from "Dottato" fig fruit cDNA and sequenced it. As predicted, only one *AG* gene was amplified in "Dottato" and the sequence was exactly the same as that of *FcAG-Gall*, *FcAG-Pistil,* and *FcAG-Purple Peel*. We named our cloned and sequenced *AG* gene from "Dottato" *FcAG-Dottato**, and used it in the nucleotide and amino acid sequence comparisons (**Figures 1B,C**), the AG phylogenetic tree and gene structure visualization (**Figures 2A,B**).

Expression Analysis of AGs in Fig

FcAG demonstrated spatial and temporal patterns of expression. Highest expression was found in the gall, followed by the pistil and stamen, and low or no expression in the stem, leaf, peel, and roots by semi-quantitative RT-PCR. *FcAG2-Gall_Stamen* and *FcAG3-Gall_Stamen* transcripts were only detected in the male plants. The gall exhibited higher expression than the stamen, and a very light, if any, band was detected in the other tissues of the male fig plant (**Figure 3A**).

The same expression pattern was obtained from the transcriptome data of "Purple Peel," where *FcAG* was only expressed in the pistil, and not in the peel. Along fig fruit development, *FcAG* expression in the pistils decreased when the young fruit formed, with the lowest expression just before the start of ripening, at the S4 stage, then its transcript abundance increased again (**Figure 3B**).

Agamous-Marker for Sex Identification by Amplification

We tested the *F. hispida* sex-specific primers designed by Zhang et al. (2020) for amplification of the DNA and cDNA of fig fruit (three females and three males) from six cultivars (**Figure 3D** and **Supplementary Table 1**). The primers were not found suitable for sex identification of *F. carica*, and no male-specific fragment was obtained (**Supplementary Figure 2**).

As *FcAG2-Gall_Stamen* and *FcAG3-Gall_Stamen* were only found in male fig trees, and they both possessed the 15-bp signature sequence that was absent in *FcAG*, the 15-bp sequence was used as the forward primer (AG-Marker-F, **Table 1**), and 20 bp before the termination site were used as the reverse primer (AG-Marker-R). Using the DNA and cDNA templates of six fig fruit from three male and three female cultivars, amplification resulted in a specific 146-bp fragment in male figs, but not in female figs (**Figure 3C**), thus establishing the AG-Marker for fig sex determination.

Accurate differentiation of male and female fig plants by the AG-Marker was validated using leaf DNA of 42 fig cultivars (25 male and 17 female) (**Supplementary Table 1**). The AG-Marker was also used to screen our two F1 populations; 110 female and 132 male trees (1:1.2) were identified in population 1; 49 female and 36 male trees (1:0.73) were identified in population 2 (**Supplementary Figure 3** and **Supplementary Table 1**). In the 3rd year after sowing, 20.25 and 22.35% of the population 1 and 2, respectively, produced flowers. All these flowering trees were female and actually had been identified as female seedlings using the AG-marker. The result of analyzing F1 populations further verified the effectiveness of the AG-Marker. Moreover, it should be noted that application of the AG-Marker did not require the steps of PCR fragment recovery, enzyme digestion and gel verification of the CAPS method, and avoided the possibility of false-positive results after *Hpy*CH4IV digestion.

FcRAN1 Gene Structure and Expression Analysis

The *FcRAN1* sequence of female cv. Purple Peel (ID: *c40459_g1_i1*) was recruited by using TBtools to blast the

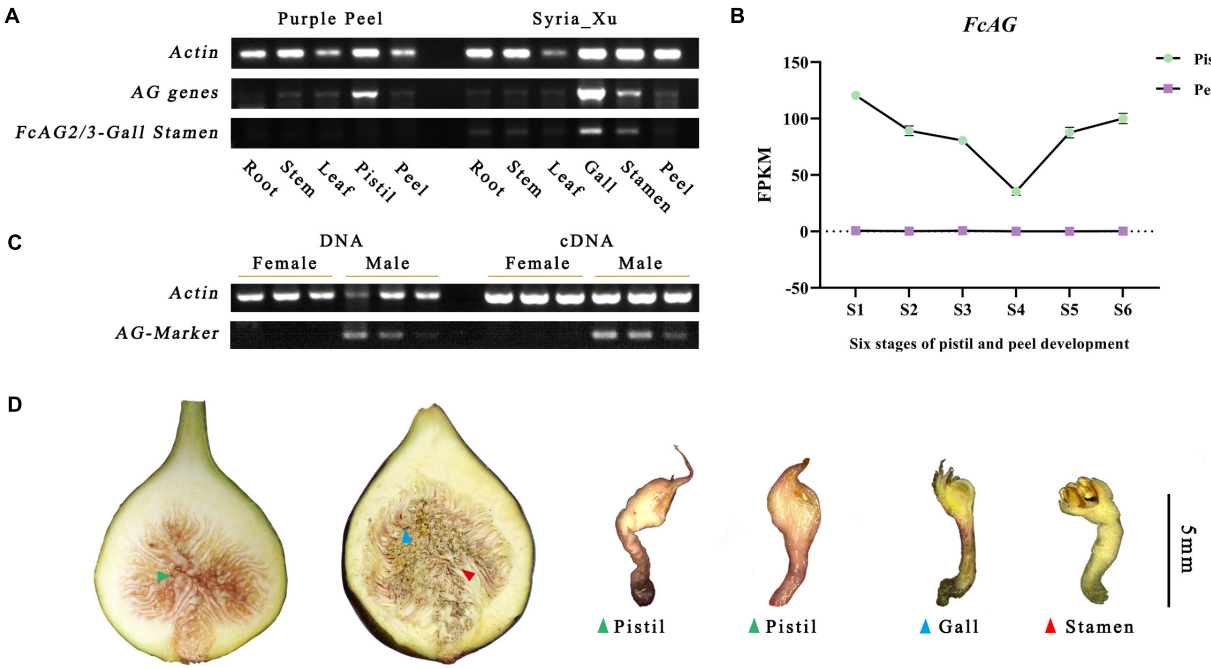

FIGURE 3 | Specific expression of *AG* genes revealed by semi-quantitative RT-PCR and fig fruit transcriptome. **(A)** *AG* genes are widely expressed in fig flowers, but male-specific *AG* genes *FcAG2/3-Gall_Stamen* are only found in fig galls and stamens. **(B)** Transcriptome data of "Purple Peel" fruit reveal that the *FcAG* gene is only expressed in the pistils; no expression was detected in the peel. **(C)** AG-Marker PCR products are only identified in male individuals; both DNA and cDNA can be used as the template. **(D)** Sectional view of a female fig fruit (left) and male fig fruit (right). Both were confirmed by our sex-identification markers.

FcRAN1 protein-coding sequences of female fig "Horaishi" and male fig "Caprifig6085" (Mori et al., 2017) against our previous "Purple Peel" fruit transcriptome data. The protein-coding sequence of *FcRAN1-Purple Peel* was predicted to be 3015 bp, consistent with the PCR amplification and sequencing results obtained from the cDNA template of "Purple Peel." *FcRAN1-Purple Peel* was aligned with *FcRAN1-Horaishi* and had identical sequences, both with 9 exons. However, the protein-coding sequence of *FcRAN1-Dottato* (ID: *FCD_00026676*, **Supplementary Table 2**), extracted from the "Dottato" genome (Usai et al., 2020), was different from that of "Horaishi" and "Purple Peel," with a 75-bp deletion in the 2nd exon for a *FcRAN1-Dottato* protein-coding sequence of 2940 bp, which was located on Chr01 (**Figure 4A**). We speculated that the deletion in *FcRAN1-Dottato* might be due to genome assembly error. Therefore, we amplified and sequenced *FcRAN1-Dottato** from the cDNA of "Dottato" leaves and confirmed that the full-length protein-coding sequence was 3015 bp with 9 exons, like that of "Purple Peel" and "Horaishi."

A phylogenetic tree was constructed with the protein sequences of RAN1 from fig and other plants, i.e., FmRAN1 (*F. microcarpa*, Gene ID: *Fm.01G0001630*), FhRAN1 (*F. hispida*, *Fh.01G0000580*), MaRAN1 (*M. alba*, *XM_010089630.2*), AtRAN1 (*A. thaliana*, Q9S7J8), ZjRAN1 (*Ziziphus jujuba*, A0A6P4BE01), and MnRAN1 (*M. notabilis*, A0A6P4BE01) (**Supplementary Table 2** and **Supplementary Figure 4**). Clustering results demonstrated that the FcRAN1 protein was rather conserved in *F. carica* and could be divided into two groups: FcRAN1 from the female cultivar and that from the male cultivar (green and orange in **Figure 4B**, respectively). We used all FcRAN1 CDSs to match the "Dottato" reference genome and there was no significant difference in the exon structure of the *FcRAN1* genes among different fig cultivars (**Figure 4C**). The gene structure of *FhRAN1* was similar to that of *FcRAN1* except for exon 1, whereas *FmRAN1* was markedly shorter and had 8 exons instead of the 9 exons found in all other *RAN1* genes checked in the present study.

qRT-PCR analysis revealed wide expression of *FcRAN1* in the roots, stems, leaves, flower, and peel of the female fig cv. Purple Peel and male fig cv. Syria_Xu (**Figure 5A**). Highest expression of *FcRAN1* was in the roots, similar to *F. hispida* (Zhang et al., 2020). This result could be related to the roots' prominent role in copper ion absorption and transport. Data mining of the "Purple Peel" fruit transcriptome showed that *FcRAN1* is continuously expressed in the developing pistils and peel, and its expression in peels at the S4 stage was markedly higher than at the other stages (**Figure 5B**).

Male fig *FcRAN1* was cloned from the leaf cDNA of cv. Syria_Xu. The sequence was aligned with *FcRAN1-Horaishi*, *FcRAN1-Caprifig6085*, *FcRAN1-Purple Peel*, and *FcRAN1-Dottato**. Multiple SNP sites were identified in the male fig *FcRAN1* (**Figure 5C**). Among them, there were missense mutations at seq000259_9855, seq000259_9876, seq000259_9900, seq000259_12314, seq000259_12722, and seq000259_12743 sites. The mutations included a transition between positive and negative charges, such as 9855 and

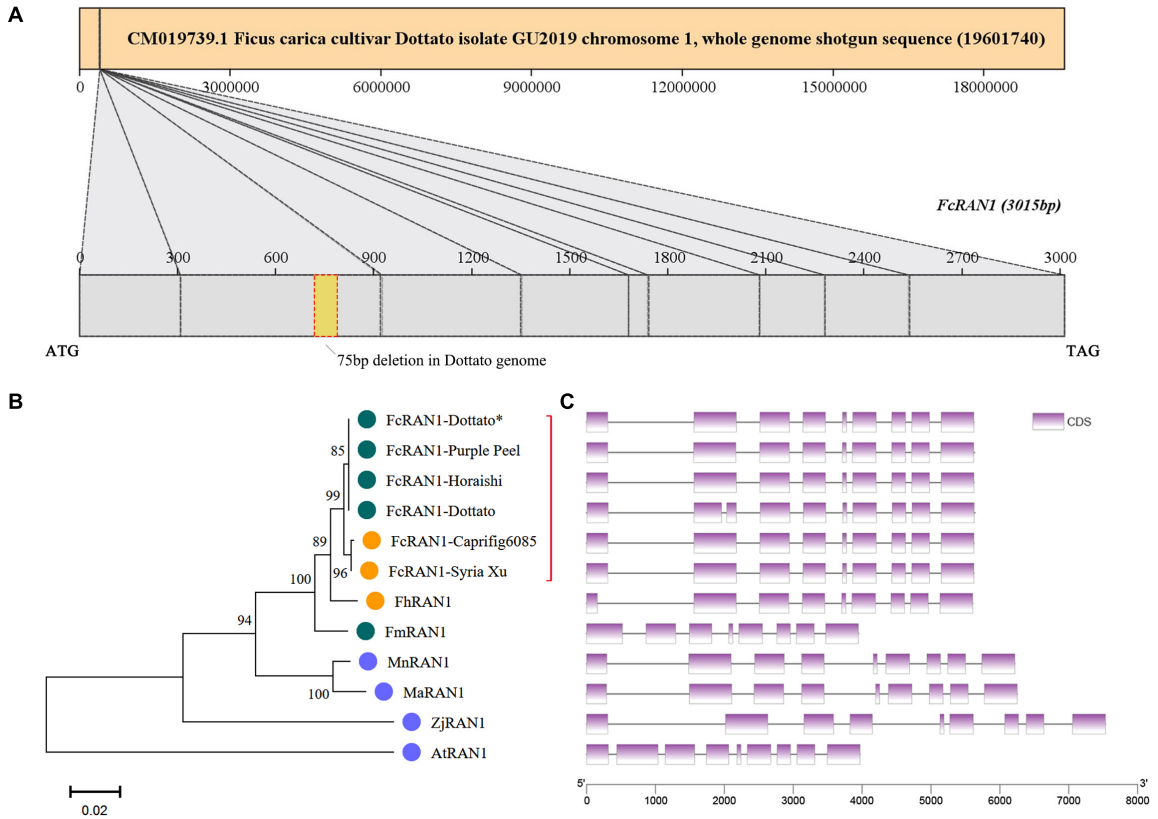

FIGURE 4 | Chromosome localization of *FcRAN1* and clustering of *RAN1* from several species. **(A)** Position of *FcRAN1* on the chromosome and the fragment-deletion annotation error in the published "Dottato" genome. **(B)** Clustering of *RAN1*-encoded proteins of various species. Red bracket highlights close clustering of RAN1s from different fig cultivars. **(C)** Gene structure of *RAN1* in several species. Each gene in panel **C** corresponds to the protein to its left in panel **B**.

9876; substitution of amino acids with the same chemical properties, such as 12314; change in polarity and non-polarity of amino acids resulting in changes in side-chain hydrophobicity, such as 12722 and 12743. Synonymous mutations were also identified at seq000259_8998, seq000259_9166, seq000259_9823, seq000259_10377, seq000259_10431, seq000259_11760, seq000259_12388, seq000259_12646, and seq000259_12722 sites, which did not change the properties of the amino acids (**Supplementary Figure 4**).

Improved Sex-Identification Method Using *RAN1*

Four pairs of primers (6-SNP-1, 1-SNP, 6-SNP-2, and 2-SNP) were designed to detect the 15 exon-located SNPs in *FcRAN1* (**Table 1** and **Figure 6A**). Specific fragments were amplified and sequenced using 51 known-sex fig cultivars (24 female and 27 male). SNP13-9823, SNP12-9855, and SNP9-10377 were found to not be sex-specific (**Figure 6B**). In addition to SNP15-8998, SNP11-9876, and SNP6-12314, which were reported by Mori et al. (2017), 9 new SNPs were found with stable sex polymorphism (**Figure 6B**). In other words, heterozygous double-base peaks characterized the males, while a homozygous single-base peak tagged the females.

Because 6 out of 12 sex-discriminating SNPs were located in the 2nd exon, the 6-SNP-1 primers whose 608-bp amplicon covered these 6 SNPs were renamed RAN1-Marker (**Figure 6A**). The advantage of this primer pair was that the amplified fragment was located completely within the exon, and therefore the PCR template could be either cDNA or DNA, making it an ideal and stable molecular marker.

The effectiveness of RAN1-Marker was further verified using seedlings from the two F1 populations. In population 1, 110 and 132 seedlings were identified as female and male, respectively, and in population 2, 49 and 36 seedlings were identified as female and male, respectively (**Supplementary Table 1**). The results were identical to those using the AG-Marker, proving that both methods are equally accurate for fig plant sex identification. However, the AG-Marker method was more convenient due to its one-step PCR compared to our RAN1-Marker-based method, which was nevertheless improved.

Chromosomal Localization of AGs and *RAN1*

FmAG and *FhAG1* were located on their chromosome (Chr) 01 by Zhang et al. (2020); all *F. microcarpa* and *F. hispida* plants

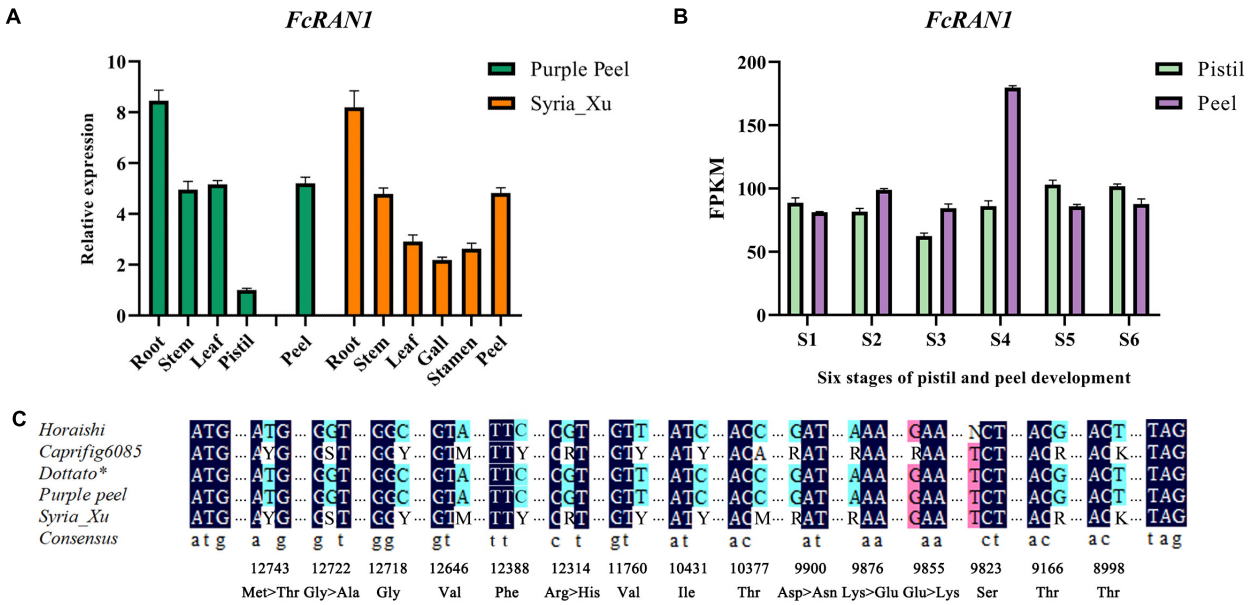

FIGURE 5 | Expression analysis of *FcRAN1* and its SNPs. **(A)** Results of qRT-PCR for *FcRAN1* expression in different tissues of "Purple Peel" (female) and "Syria_Xu" (male) fig trees. **(B)** *FcRAN1* in the transcriptome data of "Purple Peel" fruit. **(C)** Preliminary verification of 15 SNP sites, of which sites seq000259_10377, seq000259_9855, and seq000259_9823 were considered invalid. Degenerate bases: Y, C/T; S, G/C; M, A/C; R, A/G; K, G/T; N, base absence. Dark blue indicates identical bases in both males and females, light blue marks true SNP and pink indicates false SNP.

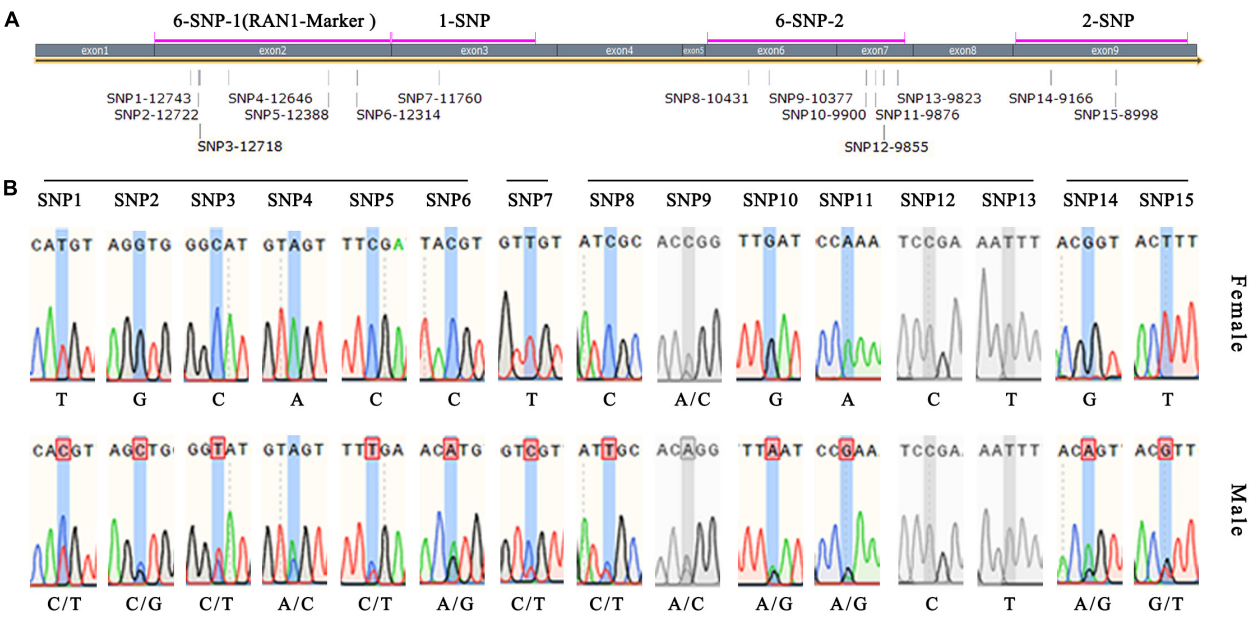

FIGURE 6 | Verification of *FcRAN1* SNP-specific primers. **(A)** Amplification interval of the SNP identification primers and number of covered SNP sites. **(B)** Sex-identification capacity of 15 SNP sites validated by 51 fig cultivars of known sex. Gray color indicates invalid SNPs.

have this Chr01 *AG*, which we refer to as the conserved *AG* gene. Our study revealed that the conserved *AG* is also present in female and male fig trees (*FcAG-Pistil* and *FcAG-Gall*). In the "Dottato" genome, *FcAG-Dottato* was annotated but not localized to any chromosome.

In addition to the conserved *AG* gene, there were two other *AG* genes (*FhAG2* and *FhAG3*) in the *F. hispida* genome which were highly homologous with *FcAG2-Gall_Stamen* and *FcAG3-Gall_Stamen* of male fig trees (**Figure 2A**). *FhAG2* was located on Chr12 and *FhAG3* was not localized to any chromosome.

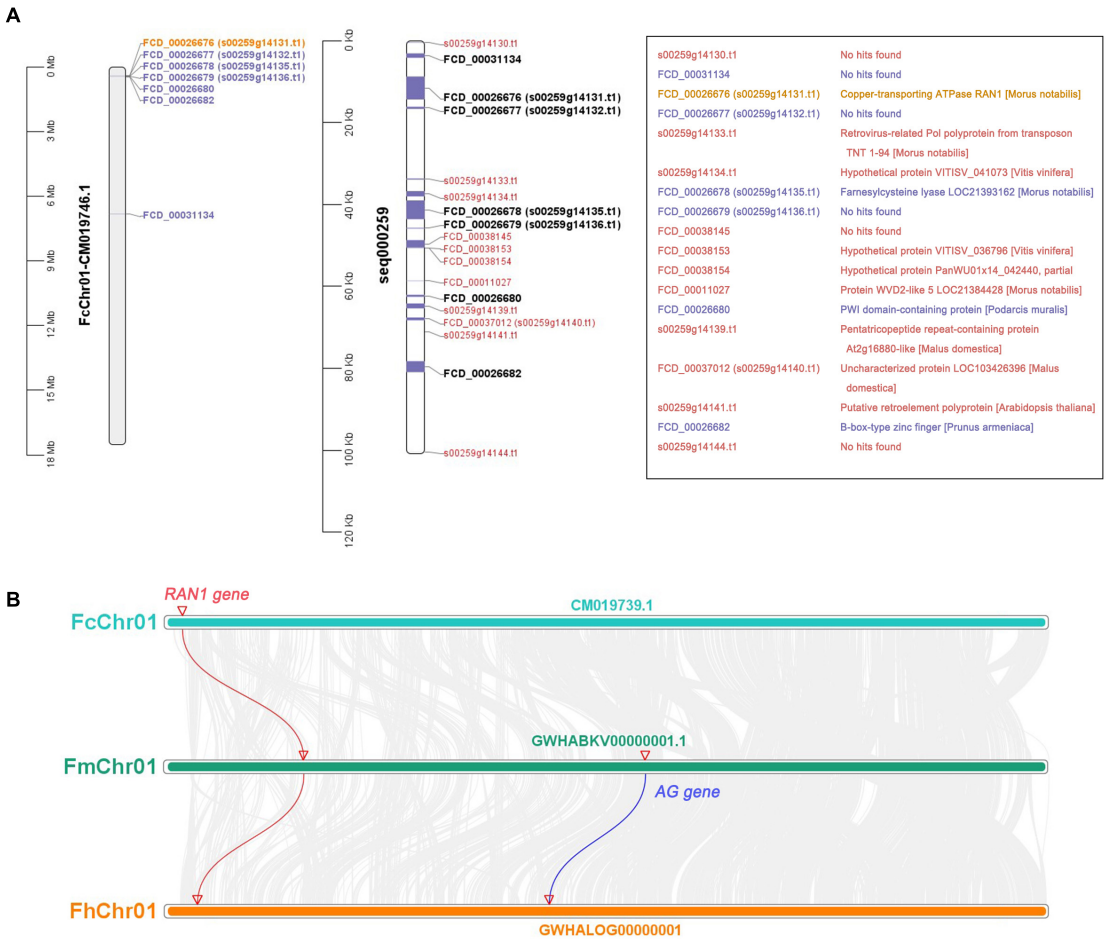

FIGURE 7 | Gene annotation of 0–100 kb in seq000259 scaffold and chromosome collinearity analysis. **(A)** Position of genes on FcChr01 of cv. Dottato (left) and seq000259 scaffold of cv. Horaishi (middle). Orange, purple, and black genes can be matched to Chr01; red genes lack chromosome location information. Gene annotation information is shown in the box (right). **(B)** Chromosome collinearity of *RAN1* and *AG* from three *Ficus* species (*F. carica*, *F. microcarpa*, and *F. hispida*).

FhAG2 and *FhAG3* are regarded as two different *AG* genes (Zhang et al., 2020). The *FcAG2-Gall_Stamen* and *FcAG3-Gall_Stamen* genes, which we cloned from the male fig trees, differed by only a few bases, and were likely to be a pair of alleles (**Supplementary Figure 1**).

The CDSs of protein-coding genes were extracted from the "Dottato" reference genome using TBtools and the gene-annotation file (**Supplementary Table 4**). They were then aligned with the 100-kb sex-linkage region of the seq000259 scaffold reported by Mori et al. (2017) (**Figure 7A** middle and right). Twelve protein-coding genes were screened out—seven mapped to FcChr01 (**Figure 7A** left and middle), and five were not assembled due to lack of chromosome annotation. Six genes were found in the 401567- to 473625-bp region of FcChr01, distanced from *FCD_ 00031134*. Among the six genes, we confirmed that *FcRAN1* (*FCD_00026676* or *s00259g14131.t1*) had accurate sex-discriminating capability. Two genes were not previously reported: *FCD_ 00026682* was annotated as a B-box-type zinc finger protein in the "Dottato" genome, and *FCD_00026680* was similar to the PWI domain-containing protein and *At2g29210* of *Arabidopsis* in UniProt. Results suggested that *FcRAN1* is very likely involved in fig sex determination in this region.

We also checked the genomes of *F. microcarpa* and *F. hispida* and found both *FmRAN1* and *FhRAN1* to be located on Chr01 (**Figure 7B**). The colocalization results for the conserved *AG* and *RAN1* genes in both *F. microcarpa* and *F. hispida* indicated that the *FcAG-Dottato* gene is probably located on FcChr01 as well. In a recent report, *MaRAN1* and *MaAG1* were also localized to the same chromosome (MaChr04) in *M. alba* genome-annotation files (Jiao et al., 2020).

DISCUSSION

Sex identification of economically valuable dioecious plants is as important as in animal husbandry. In practice, only female fig trees are vegetatively propagated and grown for fig production. As pollination is required by some types of female fig cultivars, and has a positive effect on fruit quality and storability, selected male fig cultivars are used to provide pollen (Lama et al., 2020;

Marcotuli et al., 2020). Sex identification and understanding the mechanism of sex evolution of fig trees are of obvious value to breeding and crop improvement. Therefore, the establishment of a simple and effective method for sex identification of fig seedlings will significantly benefit fig breeding.

Agamous Genes in Fig Sex Determination

Floral organs directly reflect the sex of plants. In the ABCDE model of plant floral organ formation, the C-type functional AG of the MADS-box family is involved in the formation and differentiation of carpels, stamens, and floral meristems (Dreni and Kater, 2014). In addition, AG subfamily members constitute the D-type functional genes of the MADS-box transcriptional factors, which regulate ovule development. In grape, a missense mutation of one amino acid in VviAGL11 leads to a seedless phenotype (Royo et al., 2018).

In our study, a conserved *AG* gene was found in both male and female plants, i.e., *FcAG-Gall* expressed in the gall of male tree syconia and *FcAG-Pistil* in the pistil of female tree syconia. Previous reports have shown that *AG* is the critical gene regulating both carpel and stamen development (Wellmer et al., 2004; Serwatowska et al., 2014). We found that the conserved *FcAG* has higher expression in the early and late stages of fig syconium development, possibly related to carpel differentiation and maturation (**Figure 3B**). In *Arabidopsis*, *AG* regulates the expression of *DAD1*, and subsequently promotes the massive synthesis of jasmonic acid that results in anther development (Ito et al., 2007; Jibran et al., 2017). Coexpression of the conserved *AG* and the male-specific *AG* may result in the presence of both male and female flowers in the syconia growing on male trees, namely stamens and galls.

A recent publication recruited two fig *AG* genes (seq00824 and seq00026) by blasting *MnAG* against the "Horaishi" draft genome. Expression of both genes was reported to be much higher in breba (spring fruit) than in the main crop (autumn fruit) (Marcotuli et al., 2020). However, only one conserved *AG* gene was found in our study. Due to the genetic distance between *M. notabilis* and *F. carica* and the revealed phylogenetic difference between *FcAG*s and the mulberry *AG*s (**Figure 2A**), there could be errors when using *MnAG* to search for *FcAG*. We therefore blasted the "Horaishi" genome using the CDS of *FcAG-Dottato**, and recruited two genes, seq000824 and seq004379. When they were compared with the gDNA sequence of *FcAG-Dottato*, they were found to cover exon 2 to exon 7, and exon 1 of the conserved *AG*, respectively.

Using the published mulberry (*M. alba*) genome (Jiao et al., 2020), we found three *AG* genes, *MaAG1* (ID: XM_024164781.1-0), *MaAG2* (ID: XM_024174096.1-0), and *MaAG3* (ID: XM 010111053.2-0). Gene structure comparison revealed bigger differences between *MaAG1* and *MaAG2/MaAG3* than between *MaAG2* and *MaAG3*, the latter two found to be two different transcripts of the same gene (*M. alba*_G0002364). Both *MaAG1*and *MaRAN1* genes were located on MaChr04. We speculate that *MaAG1* is the conserved *AG* gene, like *FmAG* and *FhAG1*.

RAN1 Could Be Involved in Fig Sex Determination

RAN1 was cloned and found to encode a copper-transporting P-type ATPase, which belongs to the P1B-Type Heavy Metal ATPase family (Colangelo and Guerinot, 2006). RAN1 is involved in transporting copper ions from the cytoplasm to the Golgi apparatus and plays an essential role in the biogenesis and activation of the ethylene receptors, ETR1 (ETHYLENE RESISTANCE 1), ERS1 (ETHYLENE RESPONSE SENSOR 1), ETR2, EIN4 (ETHYLENE INSENSITIVE 4), and ERS2, and in maintaining copper homeostasis in *Arabidopsis* seedlings (Binder et al., 2010). With an early requirement for *RAN1* in the ethylene pathway, mutation of this gene, which hindered ethylene-binding activity in ethylene signal transduction, gave a phenotype with disrupted development and ripening (Hirayama et al., 1999).

The hormone ethylene regulates floral organ development. When the ethylene signal-transduction process is activated, expression of *ACS11* relieves the inhibitory effect of *WIP1* on *ACS2* (Martin et al., 2009; Hu et al., 2017). *ACS2* promotes ethylene synthesis through positive feedback, increases cellular ethylene levels, promotes pistil formation and inhibits stamen development (Boualem et al., 2008). This regulation mechanism suggests that fine regulation of ethylene level could lead to the formation of all female flowers. Nevertheless, the mechanism underlying the discriminatory capacity of *RAN1* SNPs has yet to be elucidated, and further gene-function studies are required to provide more clues.

Functional Genes in Plant Sex Determination

About 6% of flowering plant species have individuals of separate sexes (Ming et al., 2011), including a number of horticulturally important crops, such as persimmon (*D. lotus*), kiwifruit (*Actinidia chinensis*), wild grape (*V. vinifera* subsp. *sylvestris*), asparagus (*A. officinalis*), date palm (*P. dactylifera*), red bayberry (*Morella rubra*), willow (*Salix viminalis*), and poplar (*Populus trichocarpa*), together with wild strawberry (*Fragaria virginiana*) and papaya (*Carica papaya*), from which hermaphrodite types have been selected for commercial growing. Sex-linked regions of a few dozen to a few hundred kilobase pairs have been positioned on specific chromosomes of dioecious plants, yet little is known about the specific genes determining sex and their evolutionary history (Henry et al., 2018). Findings in recent years have suggested that more than one gene/transcription factor is involved in plant sex determination. Consequently, it is logical to propose the existence of multiple sex-identification markers with genes involved in plant sex determination (Leite Montalvão et al., 2021).

Two sex-determining genes—*Shy Girl* (*SyGI*) and *Friendly Boy* (*FrBy*)—have been identified from kiwifruit. *SyGI* negatively regulates cytokinin signaling. *FrBy* belongs to the MTR1 family, which contributes to tapetum degradation *via* programmed cell death (Akagi et al., 2014). In asparagus, two independent sex-determining genes, *SOFF* and *aspTDF*, were both localized to Chr5 (Moreno et al., 2018). *TDF1* was found to be a R2R3 MYB transcription factor, in which a single nucleotide mutation

leads to male sterility (Harkess et al., 2020). In *V. vinifera* subsp. *sylvestris*, the sex-determination region was located on Chr2, and traces of purifying selection were found with a trehalose phosphatase, an exostosin and a WRKY transcription factor in male alleles (Picq et al., 2014).

Eleven sex-associated markers have been identified in red bayberry, six of them located in a sex-determining region. Using the female phenotype locus W, a homozygous "super female" was generated that produces all female red bayberry offspring in the F2 generation (Wang et al., 2020). Edible fig (*F. carica*) and some other species of *Ficus* are dioecious; the female plants only produce female flowers in the syconia, suggesting repression of male-determining genes. *AG* and *RAN1* both had the capacity to differentiate between males and females, but their biological functions in sex determination have not been validated, and the existence of other markers cannot be excluded.

CONCLUSION

The identification of male and female individuals of dioecious plants is of great value to agriculture, the environment and crop breeding. In the present study, one *AG* gene was identified in the female fig tree which was specifically expressed in the pistil. Three *AG* genes were found in the male fig: *FcAG-Gall* was specifically expressed in the gall, whereas *FcAG2-Gall_Stamen* and *FcAG3-Gall_Stamen* were expressed in the gall and stamen. A highly efficient AG-Marker-based method was successfully established for fig sex identification, and the previously reported RAN1-Marker method was improved. Our results provide new markers and new clues to the sex-determination system of *F. carica*.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

XW and HM designed the experiments. XW and MS conducted the experiments and analyzed the results. HM, MF, and SC revised the manuscript. All authors prepared, read, and approved the manuscript for publication.

FUNDING

This work was supported by National Natural Science Foundation of China project NSFC [31372007].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.755358/full#supplementary-material

Supplementary Figure 1 | Full-length alignment of *Ficus AG* genes.

Supplementary Figure 2 | PCR amplification results of *Ficus hispida* sex-specific primers using *Ficus carica* DNA as the template.

Supplementary Figure 3 | AG-Marker validation results for two F1 populations. "M" marks male individuals and "F" marks female individuals.

Supplementary Figure 4 | Alignment of *RAN1* amino acid sequences from different species. Blue box indicates SNP site.

Supplementary Table 1 | List of fig cultivars and sex-identification results. Sheet 1 includes all fig cultivars and populations tested in this study. Light green marks female and light orange marks male. Sheet 2 presents the statistical data pertaining to the results.

Supplementary Table 2 | Coding sequences (CDS) and protein sequences used in this study.

Supplementary Table 3 | AG-related sequences screened from "Purple Peel" fig transcriptome and "Dottato" fig protein files. Red bold, the only *AG* gene screened.

Supplementary Table 4 | Protein-coding gene sequences extracted from the "Dottato" genome.

REFERENCES

Akagi, T., Henry, I. M., Tao, R., and Comai, L. (2014). Plant genetics. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* 346, 646–650. doi: 10.1126/science.1257225

Akagi, T., Pilkington, S. M., Varkonyi-Gasic, E., Henry, I. M., Sugano, S. S., Sonoda, M., et al. (2019). Two Y-chromosome-encoded genes determine sex in kiwifruit. *Nat. Plants* 5, 801–809. doi: 10.1038/s41477-019-0489-6

Binder, B. M., Rodríguez, F. I., and Bleecker, A. B. (2010). The copper transporter RAN1 is essential for biogenesis of ethylene receptors in Arabidopsis. *J. Biol. Chem.* 285, 37263–37270. doi: 10.1074/jbc.M110.170027

Boualem, A., Fergany, M., Fernandez, R., Troadec, C., Martin, A., Morin, H., et al. (2008). A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science* 321, 836–838. doi: 10.1126/science.1159023

Chai, L., Chai, P., Chen, S., Flaishman, M. A., and Ma, H. (2018). Transcriptome analysis unravels spatiotemporal modulation of phytohormone-pathway expression underlying gibberellin-induced parthenocarpic fruit set in San Pedro-type fig (*Ficus carica* L.). *BMC Plant Biol.* 18:100. doi: 10.1186/s12870-018-1318-1

Chai, P., Dong, S., Chai, L., Chen, S., Flaishman, M., and Ma, H. (2019). Cytokinin-induced parthenocarpy of San Pedro type fig (*Ficus carica* L.) main crop: explained by phytohormone assay and transcriptomic network comparison. *Plant Mol. Biol.* 99, 329–346. doi: 10.1007/s11103-019-00820-2

Charlesworth, B., and Charlesworth, D. (1978). A model for the evolution of dioecy and gynodioecy. *Am. Nat.* 112, 975–997. doi: 10.1086/283342

Charlesworth, D. (2016). Plant sex chromosomes. *Annu. Rev. Plant Biol.* 67, 397–420. doi: 10.1146/annurev-arplant-043015-111911

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). Tbtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant.* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Colangelo, E. P., and Guerinot, M. L. (2006). Put the metal to the petal: metal uptake and transport throughout plants. *Curr. Opin. Plant Biol.* 9, 322–330. doi: 10.1016/j.pbi.2006.03.015

Cui, Y., Zhai, Y., Flaishman, M., Li, J., Chen, S., Zheng, C., et al. (2021). Ethephon induces coordinated ripening acceleration and divergent coloration responses in fig (*Ficus carica* L.) flowers and receptacles. *Plant Mol. Biol.* 105, 347–364. doi: 10.1007/s11103-020-01092-x

Dreni, L., and Kater, M. M. (2014). MADS reloaded: evolution of the AGAMOUS subfamily genes. *New Phytol.* 201, 717–732. doi: 10.1111/nph.12555

Freiman, Z. E., Rosianskey, Y., Dasmohapatra, R., Kamara, I., and Flaishman, M. A. (2015). The ambiguous ripening nature of the fig (*Ficus carica* L.) fruit:

a gene-expression study of potential ripening regulators and ethylene-related genes. *J. Exp. Bot.* 66, 3309–3324. doi: 10.1093/jxb/erv140

Gu, D., and Yang, D. R. (2013). Utilisation of chemical signals by inquiline wasps in entering their host figs. *J. Insect Physiol.* 59, 1065–1068. doi: 10.1016/j.jinsphys.2013.08.005

Harkess, A., Huang, K., van der Hulst, R., Tissen, B., Caplan, J. L., Koppula, A., et al. (2020). Sex determination by two Y-linked genes in garden asparagus. *Plant Cell* 32, 1790–1796. doi: 10.1105/tpc.19.00859

Henry, I. M., Akagi, T., Tao, R., and Comai, L. (2018). One hundred ways to invent the sexes: theoretical and observed paths to dioecy in plants. *Annu. Rev. Plant Biol.* 69, 553–575. doi: 10.1146/annurev-arplant-042817-040615

Hirayama, T., Kieber, J. J., Hirayama, N., Kogan, M., Guzman, P., Nourizadeh, S., et al. (1999). RESPONSIVE-TO-ANTAGONIST1, a Menkes/Wilson disease-related copper transporter, is required for ethylene signaling in *Arabidopsis*. *Cell* 97, 383–393. doi: 10.1016/s0092-8674(00)80747-3

Hu, B., Li, D., Liu, X., Qi, J., Gao, D., Zhao, S., et al. (2017). Engineering non-transgenic gynoecious cucumber using an improved transformation protocol and optimized CRISPR/Cas9 system. *Mol. Plant.* 10, 1575–1578. doi: 10.1016/j.molp.2017.09.005

Ikegami, H., Habu, T., Mori, K., Nogata, H., Hirata, C., Hirashima, K., et al. (2013). De novo sequencing and comparative analysis of expressed sequence tags from gynodioecious fig (*Ficus carica* L.) fruits: caprifig and common fig. *Tree Genet. Genomes* 9, 1075–1088. doi: 10.1007/s11295-013-0622-z

Ikten, H., and Yilmaz, Y. (2019). Validation of gender specific CAPS marker in Turkish fig (*Ficus carica* L.) collection and F1 progenies. *Notulae Botanicae Horti Agrobotanici Cluj Napoca* 47, 867–871. doi: 10.15835/nbha47311499

Ito, T., Ng, K. H., Lim, T. S., Yu, H., and Meyerowitz, E. M. (2007). The homeotic protein AGAMOUS controls late stamen development by regulating a jasmonate biosynthetic gene in *Arabidopsis*. *Plant Cell* 19, 3516–3529. doi: 10.1105/tpc.107.055467

Jiao, F., Luo, R., Dai, X., Liu, H., Yu, G., Han, S., et al. (2020). Chromosome-level reference genome and population genomic analysis provide insights into the evolution and improvement of domesticated mulberry (*Morus alba*). *Mol. Plant.* 13, 1001–1012. doi: 10.1016/j.molp.2020.05.005

Jibran, R., Tahir, J., Cooney, J., Hunter, D. A., and Dijkwel, P. P. (2017). *Arabidopsis* AGAMOUS regulates sepal senescence by driving jasmonate production. *Front. Plant Sci.* 8:2101. doi: 10.3389/fpls.2017.02101

Kislev, M. E., Hartmann, A., and Bar-Yosef, O. (2006). Early domesticated fig in the Jordan Valley. *Science* 312, 1372–1374. doi: 10.1126/science.1125910

Lama, K., Peer, R., Shlizerman, L., Meir, S., Doron-Faigenboim, A., Sadka, A., et al. (2020). Tissue-specific organic acid metabolism in reproductive and non-reproductive parts of the fig fruit is partially induced by pollination. *Physiol. Planta* 168, 133–147. doi: 10.1111/ppl.12941

Leite Montalvão, A. P., Kersten, B., Fladung, M., and Müller, N. A. (2021). The diversity and dynamics of sex determination in dioecious plants. *Front. Plant Sci.* 11:580488. doi: 10.3389/fpls.2020.580488

Marcotuli, I., Mazzeo, A., Colasuonno, P., Terzano, R., Nigro, D., Porfido, C., et al. (2020). Fruit development in *Ficus carica* L.: morphological and genetic approaches to fig buds for an evolution from monoecy toward dioecy. *Front. Plant Sci.* 11:1208. doi: 10.3389/fpls.2020.01208

Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., et al. (2009). A transposon-induced epigenetic change leads to sex determination in melon. *Nature* 461, 1135–1138. doi: 10.1038/nature08498

Massonnet, M., Cochetel, N., Minio, A., Vondras, A. M., Lin, J., Muyle, A., et al. (2020). The genetic basis of sex determination in grapes. *Nat. Commun.* 11:2902. doi: 10.1038/s41467-020-16700-z

Ming, R., Bendahmane, A., and Renner, S. S. (2011). Sex chromosomes in land plants. *Annu. Rev. Plant Biol.* 62, 485–514. doi: 10.1146/annurev-arplant-042110-103914

Moreno, R., Castro, P., Vrána, J., Kubaláková, M., Cápal, P., García, V., et al. (2018). Integration of Genetic and Cytogenetic Maps and Identification of Sex Chromosome in Garden Asparagus (Asparagus officinalis L.). *Front. Plant Sci.* 9:1068. doi: 10.3389/fpls.2018.01068

Mori, K., Shirasawa, K., Nogata, H., Hirata, C., Tashiro, K., Habu, T., et al. (2017). Identification of RAN1 orthologue associated with sex determination through whole genome sequencing analysis in fig (*Ficus carica* L.). *Sci. Rep.* 7:41124. doi: 10.1038/srep41124

Picq, S., Santoni, S., Lacombe, T., Latreille, M., Weber, A., Ardisson, et al. (2014). A small XY chromosomal region explains sex determination in wild dioecious V. vinifera and the reversal to hermaphroditism in domesticated grapevines. *BMC Plant Biol.* 14:229. doi: 10.1186/s12870-014-0229-z

Reid, K. E., Olsson, N., Schlosser, J., Peng, F., and Lund, S. T. (2006). An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biol.* 6:27. doi: 10.1186/1471-2229-6-27

Royo, C., Torres-Pérez, R., Mauri, N., Diestro, N., Cabezas, J. A., Marchal, C., et al. (2018). The major origin of seedless grapes is associated with a missense mutation in the MADS-Box gene *VviAGL11*. *Plant Physiol.* 177, 1234–1253. doi: 10.1104/pp.18.00259

Serwatowska, J., Roque, E., Gómez-Mena, C., Constantin, G. D., Wen, J., Mysore, K. S., et al. (2014). Two eu AGAMOUS genes control C-function in Medicago truncatula. *PLoS One* 9:e103770. doi: 10.1371/journal.pone.0103770

Storey, W. B. (1975). "Figs," in *Advances in Fruit Breeding*, eds J. Janick and J. N. Moore (West Lafayette, IN: Purdue University Press), 568–588.

Stover, E., Aradhya, M., Ferguson, L., and Crisosto, C. H. (2007). The fig: overview of an ancient fruit. *HortScience* 42, 1083–1087. doi: 10.21273/HORTSCI.42.5.1083

Torres, M. F., Mathew, L. S., Ahmed, I., Al-Azwani, I. K., Krueger, R., Rivera-Nuñez, D., et al. (2018). Genus-wide sequencing supports a two-locus model for sex-determination in *Phoenix*. *Nat. Commun.* 9:3969. doi: 10.1038/s41467-018-06375-y

Usai, G., Mascagni, F., Giordani, T., Vangelisti, A., Bosi, E., Zuccolo, A., et al. (2020). Epigenetic patterns within the haplotype phased fig (*Ficus carica* L.) genome. *Plant J.* 102, 600–614. doi: 10.1111/tpj.14635

Wang, R., Yang, Y., Jing, Y., Segar, S. T., Zhang, Y., Wang, G., et al. (2021). Molecular mechanisms of mutualistic and antagonistic interactions in a plant-pollinator association. *Nat. Ecol. Evol.* 5, 974–986. doi: 10.1038/s41559-021-01469-1

Wang, Y., Jia, H. M., Shen, Y. T., Zhao, H. B., Yang, Q. S., and Zhu, C. Q. (2020). Construction of an anchoring SSR marker genetic linkage map and detection of a sex-linked region in two dioecious populations of red bayberry. *Hortic. Res.* 7:53. doi: 10.1038/s41438-020-0276-6

Wang, Z., Cui, Y., Vainstein, A., Chen, S., and Ma, H. (2017). Regulation of Fig (Ficus carica L.) Fruit Color: Metabolomic and Transcriptomic Analyses of the Flavonoid Biosynthetic Pathway. *Front. Plant Sci.* 8:1990. doi: 10.3389/fpls.2017.01990

Wang, Z., Song, M., Li, Y., Chen, S., and Ma, H. (2019). Differential color development and response to light deprivation of fig (*Ficus carica* L.) syconia peel and female flower tissues: transcriptome elucidation. *BMC Plant Biol.* 19:217. doi: 10.1186/s12870-019-1816-9

Wellmer, F., Riechmann, J. L., Alves-Ferreira, M., and Meyerowitz, E. M. (2004). Genome-wide analysis of spatial gene expression in Arabidopsis flowers. *Plant Cell* 16, 1314–1326. doi: 10.1105/tpc.021741

Woeste, K. E., and Kieber, J. J. (2000). A strong loss-of-function mutation in RAN1 results in constitutive activation of the ethylene response pathway as well as a rosette-lethal phenotype. *Plant Cell* 12, 443–455. doi: 10.1105/tpc.12.3.443

Zhai, Y., Cui, Y., Song, M., Vainstein, A., Chen, S., and Ma, H. (2021). Papain-like cysteine protease gene family in fig (*Ficus carica* L.): genome-wide analysis and expression patterns. *Front. Plant Sci.* 12:681801. doi: 10.3389/fpls.2021.681801

Zhang, X., Wang, G., Zhang, S., Chen, S., Wang, Y., Wen, P., et al. (2020). Genomes of the banyan tree and pollinator wasp provide insights into fig-wasp coevolution. *Cell* 183, 875.E–889.E. doi: 10.1016/j.cell.2020.09.043

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

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
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@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