NEXT GENERATION AGRICULTURE: UNDERSTANDING PLANT LIFE FOR FOOD, HEALTH AND ENERGY

EDITED BY: Domenico De Martinis, Eugenio Benvenuto, Nicola Colonna, Briardo Llorente and Edward Rybicki
PUBLISHED IN: Frontiers in Plant Science

frontiers Research Topics

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

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

NEXT GENERATION AGRICULTURE: UNDERSTANDING PLANT LIFE FOR FOOD, HEALTH AND ENERGY

Topic Editors:
Domenico De Martinis, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Italy
Eugenio Benvenuto, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Italy
Nicola Colonna, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Italy
Briardo Llorente, Macquarie University, Australia
Edward Rybicki, University of Cape Town, South Africa

Table of Contents

frontiers
in Plant Science

Check for
updates

Editorial: Next Generation Agriculture: Understanding Plant Life for Food, Health and Energy

Domenico De Martinis[1]*, Edward P. Rybicki[2], Nicola Colonna[1], Eugenio Benvenuto[1] and Briardo Llorente[3,4]

[1] ENEA Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Rome, Italy,
[2] Biopharming Research Unit, Department of Molecular and Cell Biology, University of Cape Town, Cape Town, South Africa,
[3] ARC Center of Excellence in Synthetic Biology, Macquarie University, Sydney, NSW, Australia, [4] CSIRO Synthetic Biology Future Science Platform, Sydney, NSW, Australia

Editorial on the Research Topic

Next Generation Agriculture: Understanding Plant Life for Food, Health and Energy

Current global population growth and the associated increasing demands on farming, together with the threat of climate change and the need for better environmental protection, pose formidable challenges for the agriculture of the future. Crop productivity is already reaching high capacity but will have to increase perhaps by as much as 100% to sustain a world population of nearly 10 billion people by 2050. These global challenges are also not evenly distributed. While developed countries face the needs derived from an aging population, highly urbanized areas, and stagnation of cultivated land, developing nations are blooming in terms of population growth, the building of infrastructure, and the use of land for agriculture. Less developed countries, on the other hand, also face demographic growth but generally lack modern infrastructure and efficient farming practices required for the expansion of agricultural production. Climate change also afflicts different regions of the world with varying intensity and in very diverse ways, generating complex effects on agricultural systems, which remain hitherto unpredictable. This Research Topic provides a perspective of how and where agriculture will be conducted in the future, what will be cultivated, and for what purpose.

Studies published in this topic provide an overview of how different technologies and research areas may converge in agricultural innovation. Biotechnology could improve "intrinsically" crop performance (Fulgosi and Vojta), make plant crops more climate-resilient (Soto et al.; Gonzalo et al.), productive (Mayta et al.; Grossi et al.), and nutritious (Pannico et al.; Yuan and Li; Ferrero et al.), as well as less dependent on the use of agrochemicals (Fabian et al.), and tailored for biofuel production (Gangwar and Shankar).

Approaches dedicated to highly technologically intensive disciplines could be of great support to agriculture. Artificial Intelligence could play a game-changing role in the automation of agricultural practices (Jin et al.; Fabris et al.), freeing farmers from the workload of traditional agriculture and from the variability that still shapes the output of agricultural production.

Future agriculture is expected to produce food, energy, pharmaceuticals, and other high-value commodities, and may take place beyond traditional cultivated lands. In the past, humanity learned to claim agricultural land by draining and terracing. Now, climate change and scarcity of arable land

might lead future agriculture to respond to production needs (Pourkheirandish et al.) by moving to more extreme environments (Pappalardo et al.), taking advantage of algae cultivation (Molino et al.), and merging technologies to create greenhouses to enable sustainable agriculture in desert areas (Lefers et al.). Biocontainment approaches to enable molecular farming in large-scale field conditions are discussed in this issue (Clark and Maselko). Concurrently, we have to consider that, in the near future, more crops would be grown also in urban environments, with little or no ground available. Urban farming in cities might help reduce the environmental impact of agriculture while contributing to sustainably achieving food security.

Agriculture with no-land is already happening and will eventually evolve into agricultural systems to support human on earth and beyond, on the International Space Station[1], and for space exploration[2].

This editorial work has been implemented in the timeframe Fall 2019–Spring 2020 on the rampage of the COVID-19 pandemic. Counting the accepted papers only, this Research Topic involved 47 laboratories, 116 authors, and 39 referees from 24 countries worldwide (Argentina, Brazil, Mexico, Cuba, USA, Australia, Japan, South Korea, Taiwan, China, India, Israel, Saudi Arabia, South Africa, Turkey, Cyprus, Croatia, Italy, Spain, Portugal, Czech Republic, Germany, UK, Poland). Working on future knowledge during a global emergency situation with so many people in lockdown has not been easy, both mentally and practically. Now that the emergency sanitary phase seems to be controlled, a lesson about our ability to continuously provide society with basic needs, such as energy, water, and food, has been clearly highlighted. The issue of food security that has been largely neglected in recent years in developed countries has returned to the fore and governments' agenda. The primary sector must continue to ensure healthy and sufficient food for all, even during crises like the one we have experienced. When the movement of people and goods is stopped, it is of paramount importance to have local food production capacity to cope with the unavailability of external sources. The agriculture of the future will have to sustain a world population with different needs and opportunities and provide resources also in case of a global crisis while at the same time reducing environmental impact in the general upheaval of climatic conditions.

To achieve this, advances at the frontiers of plant science will become essential. Knowledge efforts in the field of physical–chemical life sciences applied to the agricultural production sector must be supported to increase the resilience of the global agri-food system. These include expanding the use of plants for the production of novel materials, complex chemicals, pharmaceuticals and biologics, and bioenergy, as well as understanding how to improve our farming practices toward a circular economy. Next-generation agriculture will certainly shape our future. It will take advantage of Smart and Molecular Farming, Data, and Artificial Intelligence; it will move into the cities and go vertical, occur in extreme environments, and support the human conquest of extraterrestrial space.

AUTHOR CONTRIBUTIONS

The authors contributed equally to the work.

[1] Meals Ready to Eat: Expedition 44 Crew Members Sample Leafy Greens Grown on Space Station NASA Newsletters https://www.nasa.gov/mission_pages/station/research/news/meals_ready_to_eat
[2] Ground Demonstration of Plant Cultivation Technologies for Safe Food Production in Space https://eden-iss.net/

Expression of a Chloroplast-Targeted Cyanobacterial Flavodoxin in Tomato Plants Increases Harvest Index by Altering Plant Size and Productivity

Martín L. Mayta[1‡], Rocío C. Arce[1‡], Matias D. Zurbriggen[1†], Estela M. Valle[1], Mohammad-Reza Hajirezaei[2], María I. Zanor[1]* and Néstor Carrillo[1]*

[1] Instituto de Biología Molecular y Celular de Rosario (IBR-UNR/CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario (UNR), Rosario, Argentina, [2] Leibniz Institute of Plant Genetics and Crop Plant Research, Stadt Seeland, Germany

Tomato is the most important horticultural crop worldwide. Domestication has led to the selection of highly fruited genotypes, and the harvest index (HI), defined as the ratio of fruit yield over total plant biomass, is usually employed as a biomarker of agronomic value. Improvement of HI might then result from increased fruit production and/or lower vegetative growth. Reduction in vegetative biomass has been accomplished in various plant species by expression of flavodoxin, an electron shuttle flavoprotein that interacts with redox-based pathways of chloroplasts including photosynthesis. However, the effect of this genetic intervention on the development of reproductive organs has not been investigated. We show herein that expression of a plastid-targeted cyanobacterial flavodoxin in tomato resulted in significant reduction of plant size affecting stems, leaves, and fruit. Decreased size correlated with smaller cells and was accompanied by higher pigment contents and photosynthetic activities per leaf cross-section. Flavodoxin accumulated in green fruit but declined with ripening. Significant increases in HI were observed in flavodoxin-expressing lines due to the production of higher fruit number per plant in smaller plants. Therefore, overall yields can be enhanced by increasing plant density in the field. Metabolic profiling of ripe red fruit showed that levels of sugars, organic acids, and amino acids were similar or higher in transgenic plants, indicating that there was no trade-off between increased HI and fruit metabolite contents in flavodoxin-expressing plants. Taken together, our results show that flavodoxin has the potential to improve major agronomic traits when introduced in tomato.

Keywords: tomato, flavodoxin, chloroplasts, transgenic plants, harvest index

INTRODUCTION

Harvest index (HI) is defined as the ratio of grain, fruit, or tuber yield to total plant biomass (Gur et al., 2010) and reflects the ability of a sink tissue to capitalize on the availability of photosynthates to increase the yield of harvestable product. As such, HI has been regarded as a reference parameter to evaluate the progress of breeding programs aimed at improving yield potential. Indeed, the so-called "green revolution" that took place in the middle of the twentieth century largely stemmed

from major increases in HI resulting from the development of dwarf varieties of rice and wheat with diminished leaf biomass coupled to similar or higher grain yields (Khush, 2001). These dwarfing traits were found to result from mutations of genes involved in gibberellin synthesis and signaling (Hedden, 2003). Crossing of the mutant lines with high-yielding varieties led to new cultivars displaying a greater proportion of photoassimilates partitioned into the grain (Langridge, 2014).

Intensive research has been carried out to identify genes affecting vegetative and reproductive growth in a way that favors high HI. Association mapping in rice (Li et al., 2012) and rapeseed (Luo et al., 2015) indicated that HI is a complex multigenic trait affected by both environmental and genetic determinants. HI can be improved by favoring nutrient transport from leaves to harvestable organs and/or by decreasing vegetative growth (Luo et al., 2015). An early example of the latter approach was provided by tobacco plants over-expressing a phytochrome gene that exhibited impaired shade avoidance causing proximity-dependent dwarfing and higher HI (Robson et al., 1996). Although many quantitative trait loci associated with HI have been reported in various plant species, the specific genes and molecular mechanisms determining the magnitude of this index are largely unknown.

While a higher HI has a general relevance for crop yield enhancement, its agronomic value is particularly important in the case of species that are grown for industrial purposes (e.g. processing tomatoes) as a strategy to obtain smaller and more compact plants with equivalent or higher fruit production that are expected to render increased yields per planted surface (Gur et al., 2010).

Decreases in plant and leaf size have been observed in creeping bentgrass and *Arabidopsis* lines expressing a chloroplast-located flavodoxin (Fld; Li et al., 2017; Su et al., 2018). Fld is an electron shuttle flavoprotein found in cyanobacteria and some marine algae, which mediates essentially the same electron transfer reactions as the iron-sulfur protein ferredoxin (Fd; Pierella Karlusich et al., 2014). Fd transcript and protein levels are down-regulated by most environmental stresses (Pierella Karlusich et al., 2014, and references therein), and under such conditions Fld expression is induced to take over the activities of its functional counterpart and allow growth and reproduction of the microorganism in the adverse situation (Zurbriggen et al., 2008; Pierella Karlusich et al., 2014). Fld-encoding genes are absent from plant genomes (Pierella Karlusich et al., 2015), but introduction of a plastid-targeted Fld in transgenic plants resulted in increased tolerance to multiple sources of biotic and abiotic stress (Tognetti et al., 2006; Tognetti et al., 2007; Zurbriggen et al., 2008; Zurbriggen et al., 2009; Coba de la Peña et al., 2010; Li et al., 2017; Rossi et al., 2017).

Abbreviations: Chl, chlorophyll; PS, photosystem; dpg, days post-germination; Fd, ferredoxin; Fld, flavodoxin; HI, harvest index; LEF, linear electron flow; MDHAR, monodehydroascorbate reductase; ^{1}H-NMR, proton nuclear magnetic resonance; NPQt, non-photochemical quenching; PETC, photosynthetic electron transport chain; ROS, reactive oxygen species; SDS-PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; TCA, tricarboxylic acid, TSP, 3-(trimethylsilyl) propionic-2,2,3,3-d4 acid; WT, wild-type.

In this study we transformed tomato plants with DNA sequences encoding a cyanobacterial Fld directed to chloroplasts (*Slpfld* lines, for *Solanum lycopersicum* **p**lastidic **Fld**) or the **c**ytosol (*Slcfld* lines), and evaluated vegetative and reproductive growth to determine if tomato HI could be increased by this genetic intervention. Mature-sized Fld was detected in leaves and fruit, but its levels declined with fruit ripening, in parallel with the general decline of total soluble protein. Lines expressing plastid-targeted Fld displayed a number of distinct phenotypic features compared to wild-type (WT) and *Slcfld* siblings, including smaller plants, leaves, and fruits; more flowers per inflorescence; increased fruit number; and higher HI. Biochemical analysis and metabolic profiling revealed that *Slpfld* fruit contained higher levels of soluble solids and similar or increased contents of sugars, amino acids, and organic acids relative to their WT counterparts. The results indicate that the chloroplast Fld approach constitutes a promising strategy to generate novel tomato lines displaying increased HI without affecting fruit metabolite contents.

MATERIALS AND METHODS

Generation of Transgenic Tomato Lines

The *pfld*- and *cfld*-harboring pCAMBIA2200 plasmids (Tognetti et al., 2006; see **Supplementary Figure S1A**) were used to direct expression of Fld from *Anabaena* PCC7119 in the chloroplasts or the cytosol, respectively, of tomato plants (*S. lycopersicum* cv Moneymaker) by standard *Agrobacterium*-mediated procedures (Tauberger et al., 2000). A total of 22 *Slpfld* and 10 *Slcfld* transformants were obtained exhibiting detectable levels of Fld in leaf extracts. Typical examples are shown in **Supplementary Figure S2**. Homozygous lines were selected by evaluating resistance to 100 μg ml^{-1} kanamycin and by measuring Fld levels in the progeny of self-pollinated T2 transformants, using known amounts of purified recombinant Fld as reference (**Supplementary Figure S1B**). Leaf contents of the flavoprotein were analyzed by immunoblotting up to the T5 generation to ensure that the transgene was neither lost nor silenced during seed propagation.

Plants were germinated in soil and grown at 200 μmol photons m^{-2} s^{-1}, 25°C, 40%/90% humidity with a 16/8-h light/dark photoperiod (growth chamber conditions) on randomly distributed 3-L pots. Watering was carried out daily to field capacity until harvest at 120 days post-germination (dpg).

Determination of Cell Size and Number

Discs (0.5 cm in diameter) were punched from the interveinal region of the third leaflet from the fourth fully expanded leaf of several independent plants at 30 dpg (**Figure 1A**) and fixed in 96% (v/v) ethanol, followed by incubation in 85% (w/v) lactic acid for clearing. Four pictures of different regions in each disc were used to calculate cell area and at least 100 cells were counted. Cell number was estimated using leaf and cell areas. Image analysis was performed with ImageJ (Rasband, 1997–2008, http://rsb. info.nih.gov/ij/).

Fruit samples were analyzed at the breaker stage. Four thin (0.5–1 mm) transverse sections of one fruit from the first truss in
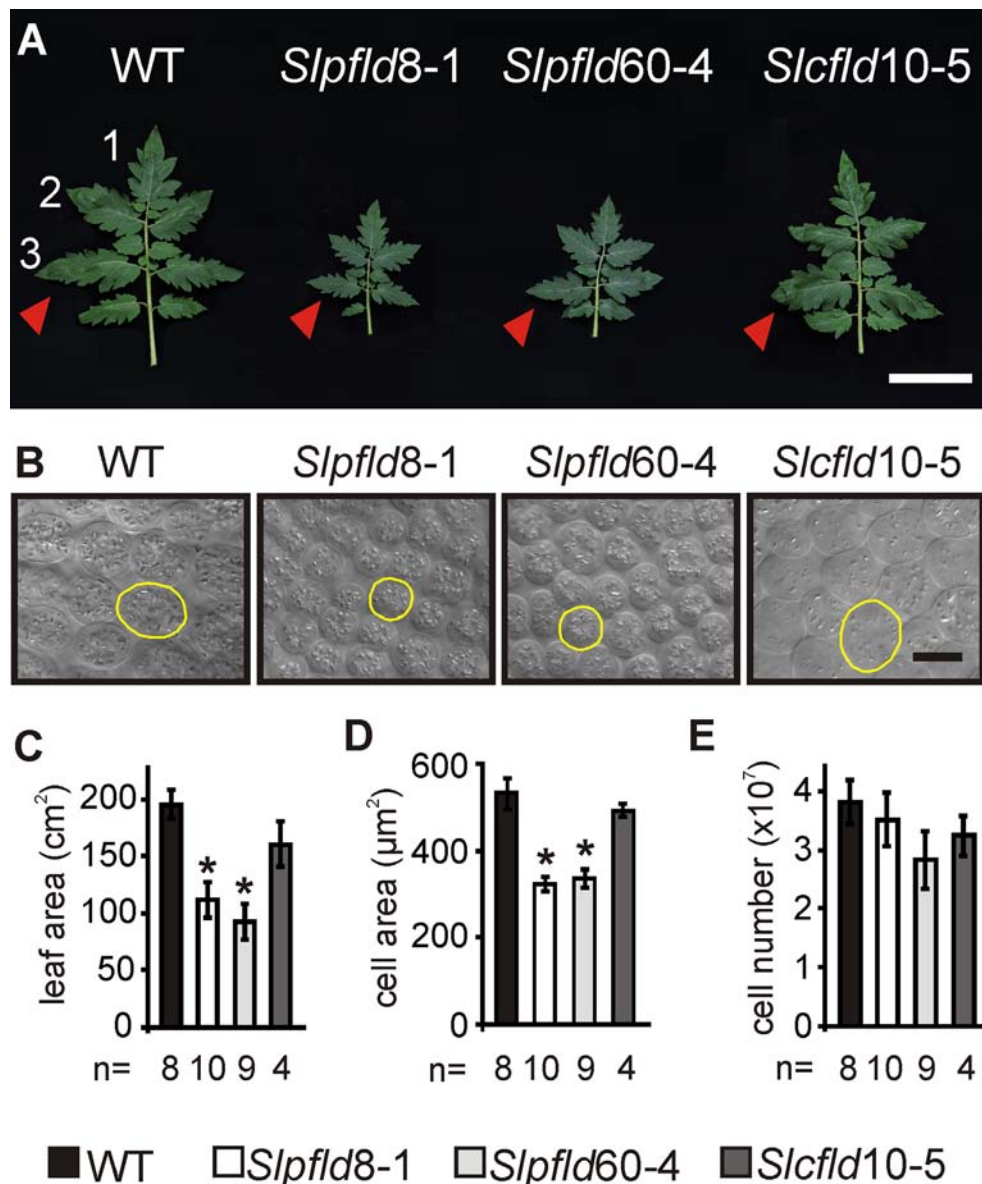
FIGURE 1 | Flavodoxin expression in chloroplasts altered tomato leaf development. **(A)** Phenotypes of the fourth fully expanded leaves (counting from the bottom) of plants at 30 days post-germination. Bar = 10 cm. Arrowheads indicate the third leaflets used to extract tissue samples. **(B)** Representative micrographs from palisade parenchyma cells. Contour of typical cells are shown in yellow. Bar = 20 μm. **(C)** Leaf area was determined by image analysis. **(D)** Cell area and **(E)** cell number were calculated from clarified leaf tissue (see *Materials and Methods*). Data reported are means ± SEM of *n* biological replicates, as shown below each line. Asterisks indicate statistically significant differences ($P < 0.05$) determined using one-way ANOVA and Tukey's multiple comparison test.

six different plants per line (24 sections per line) were hand-cut from the equatorial section of the pericarp with a razor blade. The tissue was fixed in 10% (v/v) formaldehyde, 5% (v/v) acetic acid, and 52% (v/v) ethanol, vacuum-infiltrated twice for 15 min, and incubated overnight. The fixation solution was subsequently replaced by ethanol for 2 h, and pericarp tissues stored in 70% (v/v) ethanol until processing. For cell size studies, samples were stained by incubation with 0.5% (w/v) toluidine blue in 0.1% (w/v) Na_2CO_3 for 30 s. Samples were rinsed with water to prevent staining of the internal cell layers and mounted in 30% (v/v) glycerol. Stained sections were photographed using a

camera attached to a dissecting microscope (Leica Microsystems, Switzerland). Cell layers were counted four times (technical replicates) in each fruit section from the exocarp to endocarp avoiding the vascular bundles, as described by Mu et al. (2017). Cell size was calculated using ImageJ (Rasband, 1997–2008, http://rsb.info.nih.gov/ij/).

Phenotyping

For seed viability and germination analysis, seeds were cultured under growth chamber conditions on 0.8% (w/v) agar plates

containing half-strength Murashige-Skoog basal salts (Sigma). Germination was recorded from the day the radicle broke through the seed coat. Groups of 30 seeds were used for each germination test, and the assay was repeated four times (independent experiments). Time to leaf emergence was determined for the emergence of the first and second node in 12 to 13 plants of each line germinated at the same time and grown in soil. Inflorescence and flower counting and tagging were performed every 2 days. Fruit ripening stages were determined by epicarp color change and by pressing it gently. They were classified as immature green (~10 days post-anthesis), mature green (~50 days post-anthesis), breaker, and ripe red, when they changed to a dark red color and soft texture. For biochemical and physiological measurements, fruit of equivalent developmental stages were employed, irrespective of their days from anthesis. To determine yield and HI, ripe red fruit were collected daily up to 120 dpg. At this stage, all remaining fruit were harvested irrespective of their ripening stage, and used for calculations. Dry weight was recorded after 2 weeks of incubation at 65°C. Vegetative weight (leaves and stems) was determined after fruit harvest, and HI was calculated as the ratio between total fruit yield and total above-ground biomass (fruit plus vegetative). Data shown were averaged from three independent experiments carried out during a 2-year period.

Leaf Pigment Contents and Photosynthetic Measurements

Chlorophyll (*Chl*) and carotenoid levels were determined spectrophotometrically after extraction with 96% (v/v) ethanol (Lichtenthaler, 1987). Chlorophyll fluorescence measurements were performed using a MultispeQ-Beta device controlled by the PhotosynQ platform software (Kuhlgert et al., 2016). Measurements were performed on leaves from the fifth node in two fully expanded leaflets using six independent plants per line at 45 dpg.

Metabolite Quantifications

For determination of fruit metabolites, two ripe red fruit were sampled from the first truss of at least three individual plants (biological replicates) from each genotype. Metabolite profiling was performed by proton nuclear magnetic resonance (¹H-NMR) spectroscopy according to published procedures (Sorrequieta et al., 2013; López et al., 2015). Briefly, pericarp tissue of ripe red fruit was obtained by removing the epicarp, locule tissues, and seeds, immediately frozen in liquid nitrogen and stored at −80°C until analysis of the primary metabolite composition by ¹H-NMR. Frozen samples were ground in liquid nitrogen using a Retsch MM400 mixer mill until obtaining a homogeneous and fine powder, which was rapidly dissolved in 0.3 ml of 1 M cold sodium phosphate buffer (pH 7.4) prepared in D_2O to obtain a mixture containing about 30% by weight of D_2O. The mixture was centrifuged at 20,000g for 15 min at 4°C and the supernatant filtered to remove any insoluble material. Internal standard [1 mM TSP: 3-(trimethylsilyl) propionic-2,2,3,3-d4 acid] was added to the resulting transparent soluble fraction, and the solution was subjected to spectral analysis at 600.13 MHz on a Bruker Avance II spectrometer. Proton spectra were acquired at 298 K by

adding 512 transients of 32 K data points with a relaxation delay of 5 s. A 1D-NOESY pulse sequence was utilized to remove the water signal. The 90° flip angle pulse was always ~10 μs. Proton spectra were referenced to the TSP signal (δ = 0 ppm), and their intensities were scaled to that of TSP. Spectral assignment and identification of specific metabolites was established by fitting the reference proton nuclear magnetic spectroscopy spectra of several compounds using the software Mixtures, developed *ad hoc* as an alternative to commercial programs (Abriata, 2012). Further confirmation of the assignments for some metabolites was obtained by acquisition of new spectra after addition of authentic standards.

Analytical Procedures

Fld levels in the various tissues were estimated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting (Tognetti et al., 2006). Total protein extracts were prepared by grinding 100 mg tissue powder in 200 μL protein extraction buffer [0.2 M Tris-HCl, pH 6.8, 3 M urea, 1% (v/v) glycerol, 8% (w/v) SDS, 0.5 mM dithiothreitol, 5% (v/v) β-mercaptoethanol]. The composition of this buffer allows a more efficient protein extraction (Steiner et al., 2016). Samples were vortexed, incubated for 20 min at 80°C and centrifuged at 13,000g for 15 min. Supernatants were subjected to SDS-PAGE on 15% polyacrylamide gels and transferred to nitrocellulose membranes. Gel loading was carried out on the basis of fresh weight (FW) to avoid major changes in protein patterns and levels among the various tissues (see below). Membranes were washed three times for 15 min each with 5% (w/v) skim milk in 0.01% (v/v) Tween phosphate-buffered saline (TPBS; 8 mM Na_2HPO_4, 2 mM KH_2PO_4 pH 7.4, 137 mM NaCl, 2.7 mM KCl) and incubated for 1 h with polyclonal antibodies raised in rabbits against *Anabaena* Fld (diluted 1:300 in TPBS). Following washing with TPBS (three times × 15 min), membranes were incubated with rabbit anti-IgG immunoglobulins conjugated to alkaline phosphatase (Bio-Rad), in a 1:3,000 ratio in TPBS. After washing with TPBS (three times × 15 min), membranes were finally incubated in phosphatase solution (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$) supplemented with 0.01% (w/v) 5-bromo-4-chloro-3-indolyl phosphate and 0.01% (w/v) nitroblue tetrazolium until color development.

Total protein concentrations were measured in cleared leaf and fruit extracts as described by Simonian (2002), using bovine serum albumin as standard. The concentration of purified recombinant fld was determined by the absorption of bound flavin mononucleotide ($ε_{454}$ = 8.8 $mM^{−1}$ $cm^{−1}$).

Concentrations of total soluble solids were measured in duplicate as described by Zanor et al. (2009) using a portable MA871 Digital Brix refractometer in a random sample of six fruit per line. Results were expressed in Brix degrees.

Statistical Analyses

Data were analyzed using one-way ANOVA and multiple range tests as specified in each experiment. Significant differences refer to statistical significance at $p < 0.05$.

RESULTS

Expression of a Plastid-Targeted Fld Decreases Tomato Plant Size

To generate tomato plants expressing a plastid-targeted Fld (*Slpfld* lines), the coding region of the *Anabaena* PCC7119 *fld* gene was fused in-frame to the 3' end of a DNA sequence encoding the chloroplast transit peptide of pea Fd-NADP$^+$ reductase (Tognetti et al., 2006; see *Materials and Methods*). The fused gene was cloned under the control of the constitutive cauliflower mosaic virus (CaMV) 35S promoter (**Supplementary Figure S1A**). A construct lacking the transit peptide sequence was also prepared to generate plants in which the expressed flavoprotein accumulated in the cytosol (*Slcfld* lines; **Supplementary Figure S1A**). The presence of Fld in foliar tissue was evaluated by SDS-PAGE and immunoblot analysis. Most of the flavoprotein was recovered as mature-sized peptides in *Slpfld* lines (**Supplementary Figure S1B**), indicating that it was imported by chloroplasts as already shown for Fld-expressing tobacco plants (Tognetti et al., 2006; Ceccoli et al., 2012). Homozygous lines were selected by segregation analysis and confirmed by proportional increases in leaf Fld contents. Lines *Slpfld*8-1, *Slpfld*60-4, and *Slcfld*10-5, belonging to the T5 generation and displaying high levels of Fld in chloroplasts or the cytosol, were used for phenotypic characterization.

Fld expression in *Slpfld* and *Slcfld* lines did not affect seed viability, germination rates, and time to leaf setting (**Supplementary Figure S3**), indicating that there was no retardation of vegetative development in the transformants. However, *Slpfld* plants exhibited smaller leaves and leaflets and shorter rachis compared to WT and *Slcfld* siblings (**Figures 1A, C**; **Supplementary Figure S4**; **Supplementary Table S1**), in agreement with the leaf phenotypes displayed by chloroplast Fld-expressing creeping bentgrass (Li et al., 2017) and *Arabidopsis* (Su et al., 2018). The number of leaflets per compound leaf and their overall architecture were instead unchanged (**Figure 1A**). Decreases in internodal distances of ~30% accounted for stem shortening in *Slpfld* plants relative to WT counterparts (**Supplementary Table S1**). Overall size reduction was accompanied by significant decreases in FW and dry weight of the aerial parts (**Supplementary Table S1**). Stem diameter and relative water contents were not affected by the presence of the flavoprotein (**Supplementary Table S1**).

Leaf area reduction in *Slpfld* plants resulted from a decrease in mesophyll cell size, without significant changes in cell number (**Figures 1B, D, E**). Epidermal cells were also significantly smaller (**Supplementary Figure S5**). As in tobacco (Tognetti et al., 2006; Ceccoli et al., 2012; Mayta et al., 2018), *Slpfld* plants contained higher pigment contents per leaf area (**Figure 2**). *Chl a* and *Chl b* levels were 17% and 39% higher than those determined in WT and *Slcfld* leaves (**Figures 2A–C**). Marginal increases in carotenoids were also observed in a number of experiments, albeit without statistical significance (**Figure 2D**).

Higher *Chl* contents in *Slpfld* plants were reflected at the level of photosynthetic activities. The quantum yield of photosystem (PS) II (Φ_{PSII}) per leaf cross-section, which provides an estimation
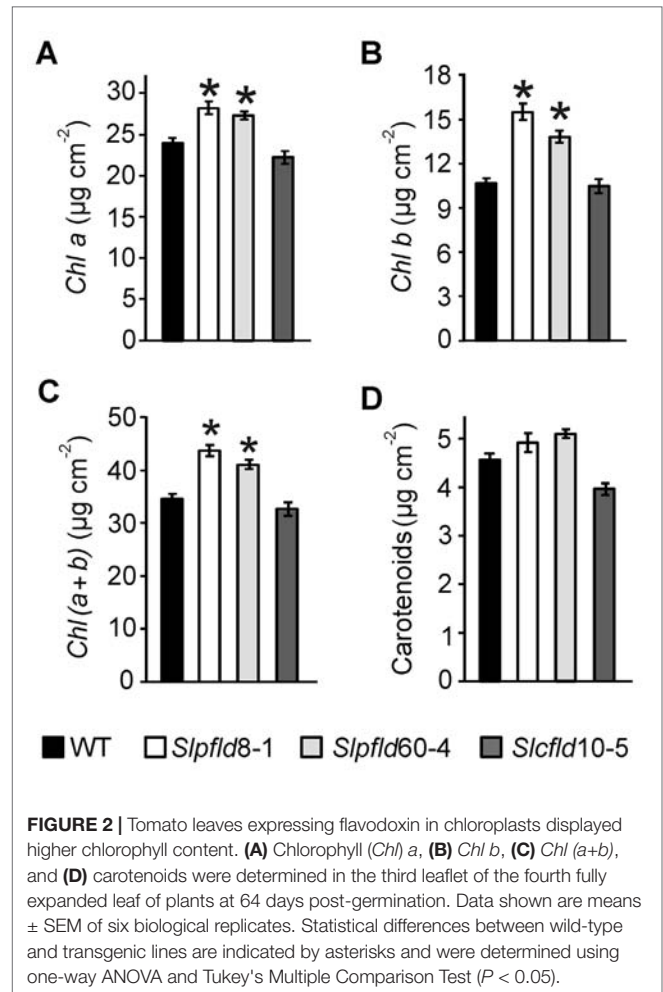
FIGURE 2 | Tomato leaves expressing flavodoxin in chloroplasts displayed higher chlorophyll content. **(A)** Chlorophyll (*Chl*) *a*, **(B)** *Chl b*, **(C)** *Chl (a+b)*, and **(D)** carotenoids were determined in the third leaflet of the fourth fully expanded leaf of plants at 64 days post-germination. Data shown are means ± SEM of six biological replicates. Statistical differences between wild-type and transgenic lines are indicated by asterisks and were determined using one-way ANOVA and Tukey's Multiple Comparison Test ($P < 0.05$).

of the electron flow through PSII (Baker, 2008), and the rate of linear electron flow (LEF), were significantly enhanced in *Slpfld* plants at 45 dpg compared to the WT and *Slcfld* genotypes (**Figures 3A, B**). Other relevant photosynthetic parameters such as the coefficient of photochemical quenching *qP* and the fraction of open PSII reaction centers *qL* were also higher in *Slpfld* transformants (**Figures 3C, D**), whereas the magnitude of non-photochemical quenching (NPQt), which reflects the ability of the photosynthetic electron transport chain (PETC) to dissipate light energy into various processes (Baker, 2008), did not vary significantly among genotypes (**Figure 3E**). Similarly, the F_v'/F_m' parameter, which is regarded as a measure of PSII integrity, was not affected by Fld expression in chloroplasts or cytosol (**Figure 3F**).

Plants Expressing Chloroplast-Located Fld Produce a Higher Number of Smaller Fruit

Flowering time was retarded in plants expressing a chloroplast Fld relative to WT and *Slcfld* siblings, as indicated by a delay of 10 to 12 days to the blossoming of the first flower (**Supplementary**
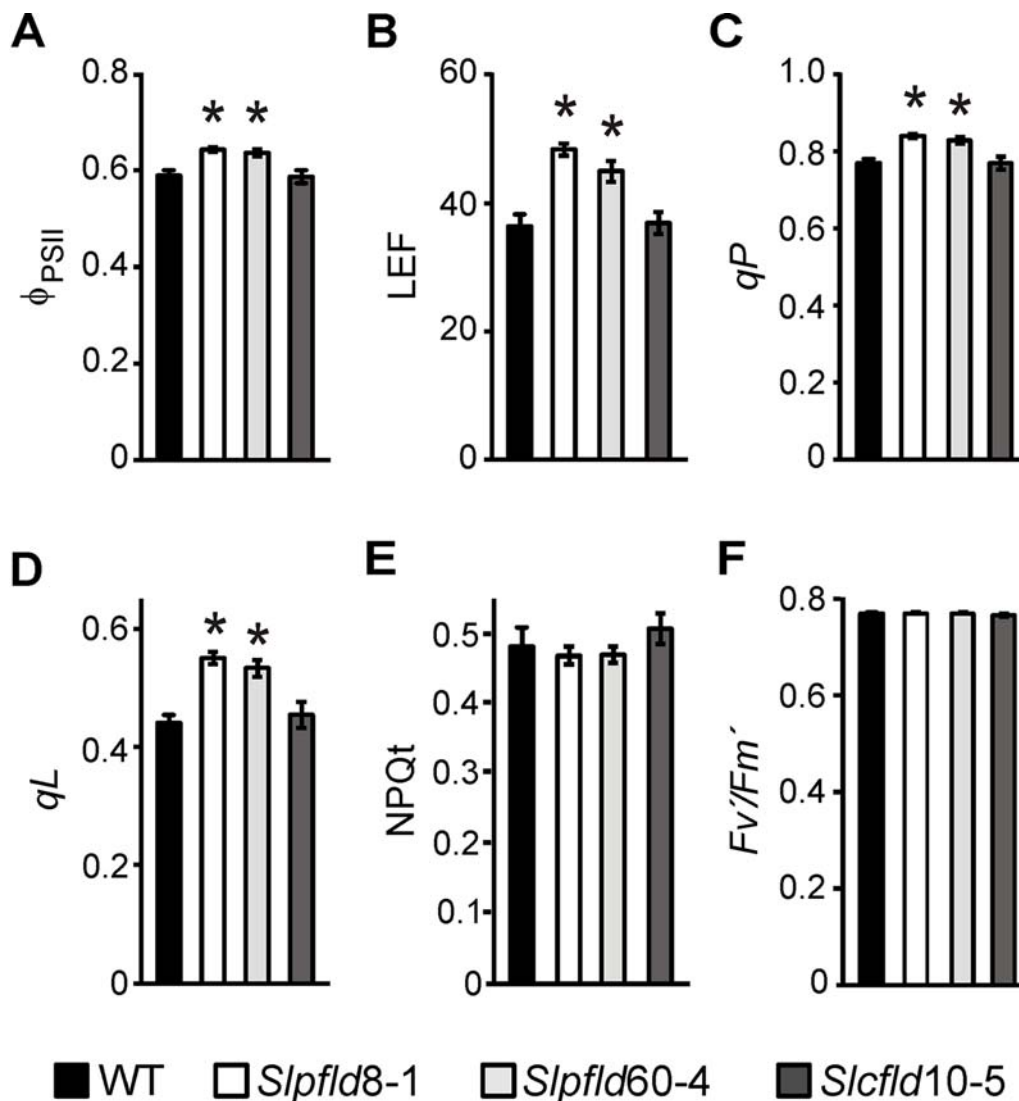
FIGURE 3 | Tomato leaves expressing chloroplast flavodoxin exhibited higher photosynthetic activities. Measurements of quantum yield of photosystem (PS) II (ϕ_{PSII}, **A**), rate of linear electron transport (LEF, **B**), photochemical quenching (qP, **C**), fraction of open PSII reaction centers (qL, **D**), non-photochemical quenching (NPQt, **E**), and PSII integrity (F_v'/F_m', **F**) were carried out in the third leaflet of the fourth fully expanded leaf of plants at 45 days post-germination. They correspond to the means ± SEM of six biological samples with six technical replicates each. Statistical differences between wild-type and transgenic lines are indicated by asterisks and were determined using one-way ANOVA and Tukey's multiple comparison test ($P < 0.05$).

Figures S6A, B). As a result, *Slpfld* plants had normally developed two more leaves on average at the time of flower setting (**Supplementary Figure S6C**). The number of tomato inflorescences and their architecture depend on the cultivar and are dramatically affected by environmental factors (Gur et al., 2010). Tomato trusses typically bear five to six flowers organized in a zigzag branch (Lippman et al., 2008). While the total inflorescences produced by *Slpfld* plants did not differ from the other two lines, they developed ~50% more flowers per truss (**Figure 4A**; **Supplementary Table S2**). Delay in flowering might cause additional branching of the inflorescence, as previously reported (Lippman et al., 2008; Park et al., 2012), but the mechanism by which chloroplast-targeted Fld exerts this effect is at present unknown.

The presence of extra flowers resulted in more prolific fruit production per plant (**Figure 4B**; **Supplementary Table S2**). Interestingly, the delay in flowering was partially compensated by accelerated ripening in *Slpfld* plants, with both color break and fruit ripening occurring 6 to 7 days (counting from anthesis) earlier than WT and *Slcfld* siblings (**Supplementary Table S2**). To estimate yield and HI, fruits were collected as they ripened up to 120 dpg. At this time, all tomatoes remaining in the plant were harvested irrespective of their maturation stage.

Fld activity in chloroplasts depends on interaction with endogenous redox partners, most conspicuously the PETC (Pierella Karlusich et al., 2014). The redox chemistry of chloroplasts closely resembles that of the phototrophic microorganisms in which the flavoprotein is normally found, but Fld expression
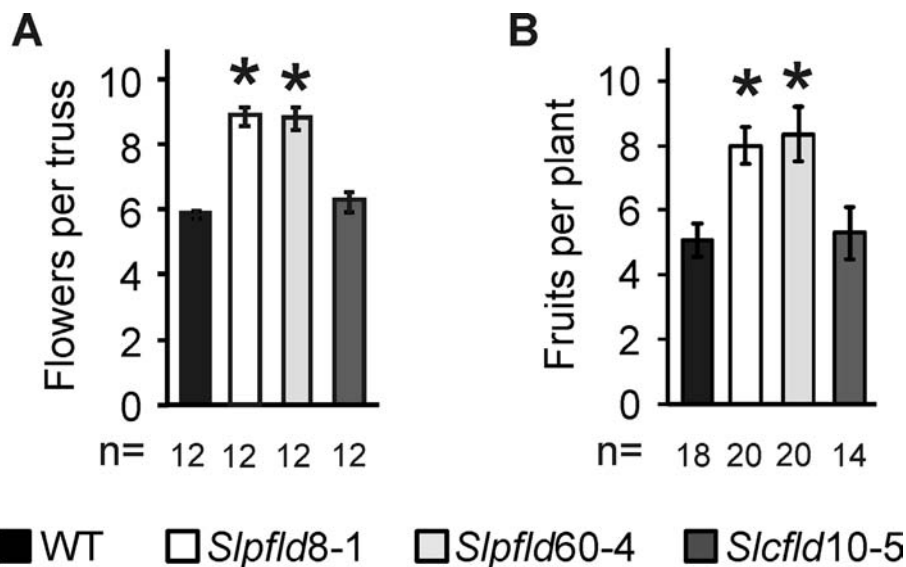
FIGURE 4 | Flavodoxin expression in chloroplasts enhanced flower number and fruit production. **(A)** Flowers were counted in the three to five initial trusses from each plant. **(B)** Total fruit numbers per plant included fully ripe, breaker, and green fruit collected up to 120 days post-germination. Data are means ± SEM of *n* biological replicates, with *n* indicated below each line. Statistically significant differences are shown by asterisks and were determined using one-way ANOVA and Tukey's multiple comparison test ($P < 0.05$).

levels and possible interaction(s) in non-photosynthetic plastids such as those present in red fruit remain unknown. It should be borne in mind, however, that fruits stay green during a large part of their development, before the conversion of chloroplasts into chromoplasts (Marano and Carrillo, 1991; Marano et al., 1993; Muñoz and Munné-Bosch, 2018).

Immunoblot analyses were used to estimate Fld accumulation in fruit tissues. Since ripening proceeded at a different pace in the various genotypes (**Supplementary Table S2**), fruit at equivalent ripening stages (different days from anthesis) were used by employing the classification described in *Materials and Methods*. Levels of total soluble protein were lower in immature green fruits compared to leaves, and declined further in all lines as ripening progressed (**Supplementary Figures S7 and S8**). Gels for Fld detection were therefore loaded on the basis of FW, considering that unlike protein levels, the fraction of dry matter did not change significantly between tomato ripening stages (Matsuda and Kubota, 2010; Radzevičius et al., 2016). **Figure 5A** shows that mature-sized Fld was expressed in the green pericarp of immature *Slpfld* and *Slcfld* fruit. At more advanced ripening stages, chloroplast Fld levels declined together with total protein, to become barely detectable in ripe red fruit (**Figure 5A**; **Supplementary Figure S8B**). Contents of the cytosol-targeted flavoprotein were instead maintained up to the mature green stage (**Supplementary Figure S8B**), resulting in a relative Fld enrichment within total soluble proteins (**Supplementary Figure S8C**). Full ripening led to down-regulation of cytosolic Fld levels to those of *Slpfld* plants (**Supplementary Figure S8**).

Fld is therefore most likely functional in the green developmental stages, and fruit of *Slpfld* plants were smaller on average than those of the WT (**Figures 5B, D**), resembling the

leaf phenotype. As in leaves, this effect was caused by a reduction in cell size (**Figures 5C, E**), while the number of cell layers in a transverse section of the pericarp was similar between WT and *Slpfld* fruit (**Figure 5F**), resulting in transformants with a thinner pericarp (**Figure 5B**).

The compromise between higher fruit number and smaller fruit size and weight resolved in moderately increased fruit yields in *Slpfld* plants relative to those of WT siblings, but the small differences failed to show statistical significance (**Figure 6A**). In turn, the combination of similar total fruit weight per plant with lower vegetative biomass (**Supplementary Table S1**) led to a ~30% increase in the HI of *Slpfld* lines compared to WT and *Slcfld* genotypes (**Figure 6B**).

Metabolite Profiling of Ripe Red Tomato Fruit

While chloroplast Fld increased HI and fruit production, this improvement could be detrimental to fruit quality and nutrient contents. Determination of soluble solids measured in Brix degrees provides a fast and reliable indicator of fruit quality. As shown in **Figure 7A**, *Slpfld* plants displayed a moderate but statistically significant increase in pericarp soluble solids, indicating that reduction of photosynthetically active tissue in these lines was not translated into lower sugar accumulation in the sink organ. Indeed, fruit dry weight and relative water content were similar in all lines (**Supplementary Table S2**).

Quantitative metabolite profiling was performed in ripe red fruit of all lines to further assess the effects of Fld expression on nutrient contents of the marketable product (**Figure 7B**). With the conspicuous exception of sucrose, which declined in *Slpfld* fruit relative to WT siblings, soluble sugars displayed similar
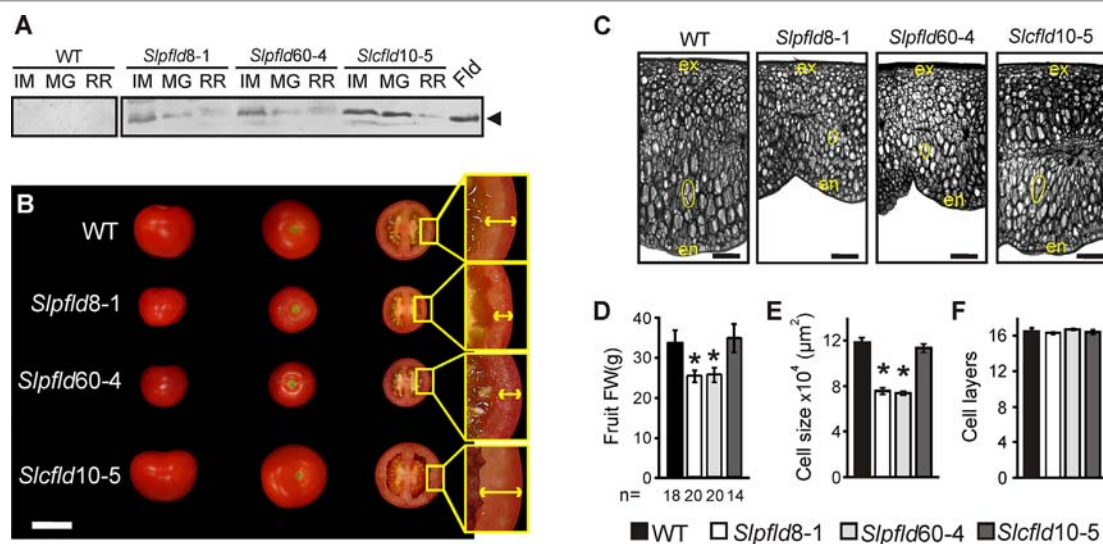
FIGURE 5 | Plastid-located flavodoxin (Fld) affected tomato fruit development. **(A)** Fld expression at different fruit stages: immature green (IM), mature green (MG), and ripe red (RR). Cleared extracts corresponding to 5 mg FW were loaded in each lane, resolved by 15% sodium dodecyl sulfate–polyacrylamide gel electrophoresis and analyzed by immunoblot using Fld antisera, as described in *Materials and Methods*. The two membranes were assayed together and over-reacted to reveal unspecific staining. Purified Fld (0.8 pmol) is shown in the extreme right. MW: molecular weight marker. **(B)** Phenotypes of representative fruits from each line. Bar = 5 cm. Insets show pericarps delineated by arrows. **(C)** Pericarp sections from representative breaker fruits stained with toluidine blue. The inner epidermis is denoted as "en" and the outer epidermis as "ex." Typical cells are contoured in yellow to illustrate size differences. Bar = 1 mm. **(D)** Average fruit weight, *n* indicates the number of fruit assayed. Cell size **(E)** and the number of cell layers of the pericarp **(F)** were calculated from six biological replicates using sections as those depicted in panel (C). Data presented correspond to means ± SEM. Statistically significant differences are indicated by asterisks and were determined using one-way ANOVA and Tukey's multiple comparison test ($P < 0.05$).
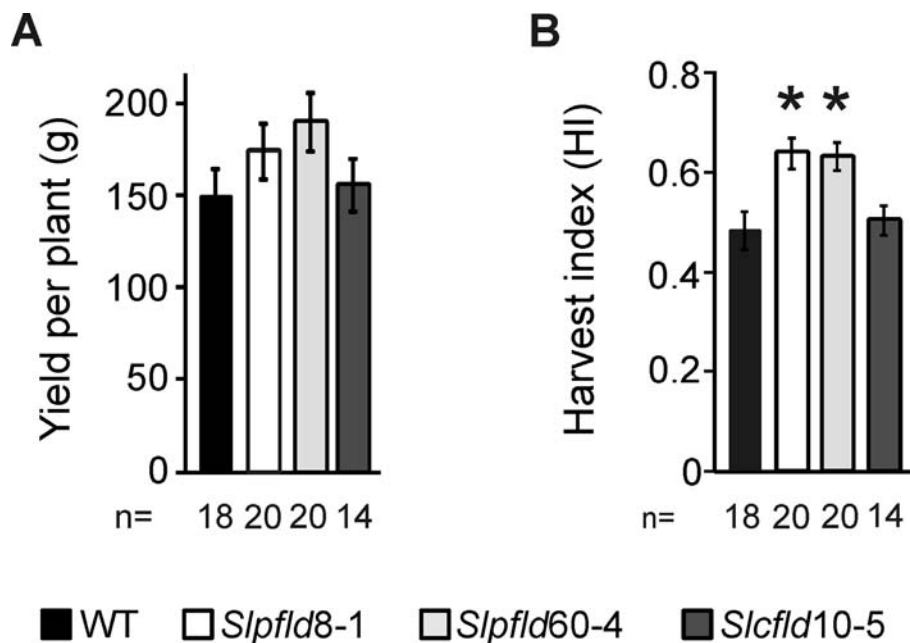
FIGURE 6 | Chloroplast flavodoxin expression increased fruit harvest index. **(A)** Yield was estimated as the sum of all fruit harvested up to 120 days post-germination regardless of the ripening stage. **(B)** Harvest index was calculated as described in *Materials and Methods*. Data reported are means ± SEM of *n* biological replicates, as shown below each line. Statistically significant differences are indicated by asterisks and were determined using one-way ANOVA and Tukey's multiple comparison test ($P < 0.05$).
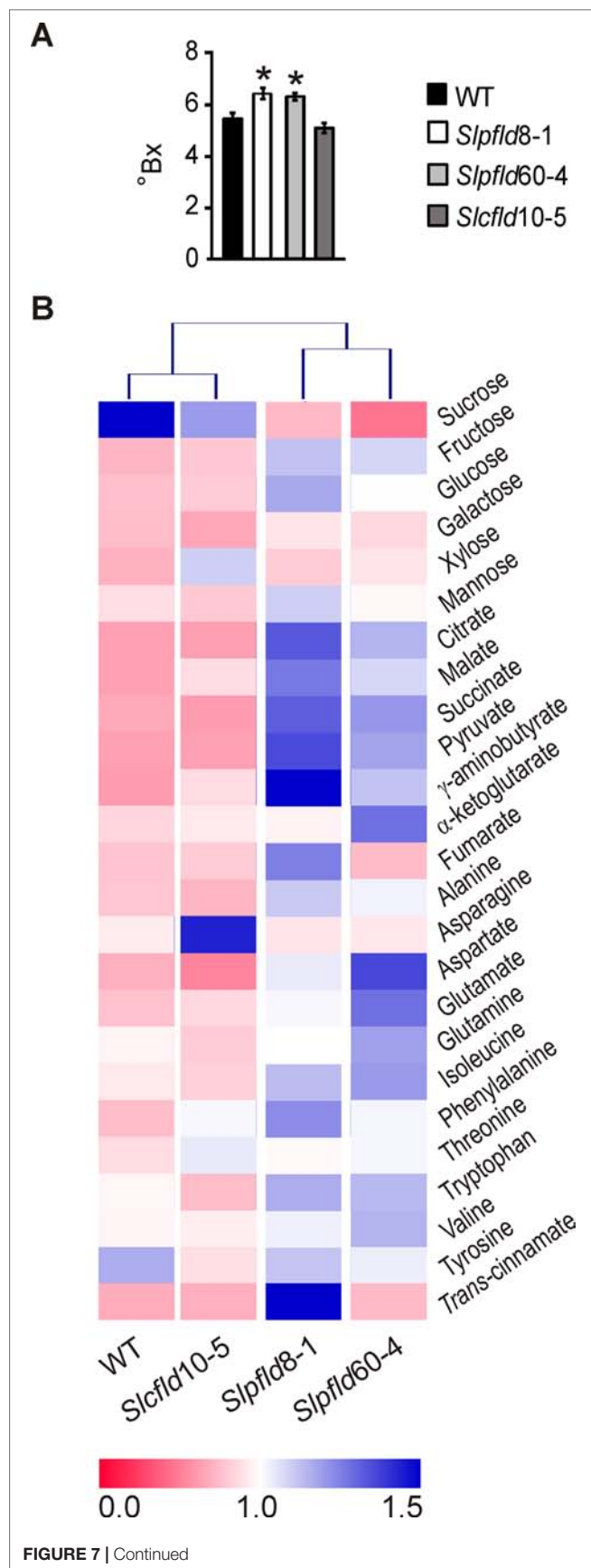
FIGURE 7 | Flavodoxin expression in plastids led to higher soluble solids and metabolite contents in ripe red tomato fruits. **(A)** Brix Index (°Bx) was determined in at least two fruit from six different plants per genotype (12 fruit per line) using a portable refractometer as described in *Materials and Methods*. Data reported are means ± SEM. Statistically significant differences are indicated by asterisks and were determined using one-way ANOVA and Tukey's multiple comparison test ($P < 0.05$). **(B)** Heat map of metabolites assayed in the different lines. Values for each compound were normalized with respect to the mean from all lines, and rectangles represent the normalized amount of each metabolite using a false-color scale. Quantitative metabolite profiling was performed in two ripe red fruit of at least three individual plants by ^{1}H-nuclear magnetic resonance spectroscopy as described in *Materials and Methods*. Heat map was designed with Multiple experiment Viewer (MeV).

(galactose, xylose, mannose) or increased (glucose, fructose) levels in both lines expressing chloroplast Fld (**Supplementary Table S3**). Intermediates of central metabolism such as pyruvate, citrate, succinate, malate, and γ-aminobutyrate also exhibited higher contents in *Slpfld* fruit, whereas fumarate and α-ketoglutarate levels did not show significant differences among genotypes (**Supplementary Table S3**). Similarly, accumulation of the 11 proteinogenic amino acids measured was not affected by Fld presence (**Figure 7B**; **Supplementary Table S3**). In line with the moderate differences in Fld accumulation (**Figure 5A**; **Supplementary Figure S1B**), the two *Slpfld* lines showed a similar trend, without statistically significant differences, for most metabolites, whereas fruit from *Slcfld* plants in which Fld accumulated in the cytosol displayed metabolic profiles that more closely resemble those of their WT counterparts (**Figure 7B**; **Supplementary Table S3**). The only conspicuous exception was *trans*-cinnamic acid, which contents showed a major increase only in *Slpfld*8-1 plants, but remained at WT levels in *Slpfld*60-4 siblings (**Supplementary Table S3**).

The collected results indicate that the increased HI of plants expressing a chloroplast-located Fld was accompanied by similar or higher contents of sugars and other metabolites in ripe fruit, underscoring the potential value of the introduced trait.

DISCUSSION

Tomato domestication has been conducted over the centuries to select varieties with increased fruit weight and number. These traits are genetically controlled through a few *loci* associated with carpel anatomy and cell proliferation (Chakrabarti et al., 2013), but phytohormones and environmental and/or metabolic conditions, most conspicuously photosynthetic activity in source tissues, also affect fruit development (Ariizumi et al., 2013). Searching for traits that can reduce plant size while increasing HI is a most relevant objective and accordingly, limited vegetative growth is generally a desirable trait in crops (Gur et al., 2010).

Expression of a cyanobacterial Fld targeted to chloroplasts of various species led to significant decreases in vegetative growth, but the effects of this intervention on the development of reproductive organs were not reported (Li et al., 2017; Su et al., 2018). We addressed herein this question by expressing a plastid-targeted Fld in a commercial variety of tomato, and

found a similar reduction in plant size (**Figure 1**; **Supplementary Figures S3** and **S4**; **Supplementary Table S1**). Decreased leaf area correlated with lower cell size (**Figure 1**), and was accompanied by higher *Chl* contents and photosynthetic activities per leaf cross-section (**Figures 2** and **3**). Fld was expressed in green fruit but its levels decreased with the progress of fruit ripening (**Figure 5A**; **Supplementary Figure S8**). Since expression of the flavoprotein was driven by a constitutive promoter (**Supplementary Figure S1A**), down-regulation of Fld contents with fruit ripening presumably involved post-transcriptional and/or post-translational mechanisms, including changes in the rates of protein synthesis and/or degradation (responsible for total protein decrease) and in the case of plastid-targeted Fld, alterations of import capacity during the transition leading to chromoplast formation (Sadali et al., 2019).

Chloroplast Fld exerted opposite effects on flowering time and fruit ripening (**Supplementary Figure S6**; **Supplementary Table S2**). Flowering time and the subsequent processes of flower patterning and fruit development are regulated by a different suite of genes, although common players do exist (Zhu and Helliwell, 2011). Association of flowering time to redox poise has been reported (Shim and Imazumi, 2015), and it is tempting to speculate that chloroplast Fld could affect this balance through its electron shuttling activity, but further research will be required to address this issue.

Plants accumulating Fld in plastids produced a higher number of smaller fruit, leading to increased HI (**Figures 5** and **6**; **Supplementary Table S2**) without detrimental effects on metabolite contents (**Figure 7**; **Supplementary Table S3**). Actually, ripe red fruit of *Slpfld* lines contained increased levels of soluble sugars (glucose and fructose) and organic acids (mainly citrate and malate). The higher hexose contents could indicate an increase of invertase activity in ripe red fruit, together with an incomplete or non-cyclic operation of the tricarboxylic acid (TCA) cycle, as reflected by the higher citrate and malate contents. In the partial TCA cycle, one branch produces citrate while the other synthesizes malate (Igamberdiev and Eprintsev, 2016). The metabolite composition of fruit influences its flavor, which is determined by several factors including the sugar/acid ratio as an important determinant of taste (Zanor et al., 2009). Then, the metabolomics approach suggests a tastier tomato.

It is remarkable that fruit yield was maintained and sugar contents increased in plants expressing chloroplast Fld despite significant mass decreases in source tissue. Higher photosynthetic activity per leaf area (**Figure 3**) might contribute to this phenotype by providing higher source strength. While the possibility that the plastid-targeted flavoprotein affected source-sink assimilate partitioning cannot be ruled out, genes differentially expressed by chloroplast Fld in tobacco failed to reveal any obvious trait associated to this system (Bermúdez et al., 2008; Bermúdez et al., 2014; Pierella Karlusich et al., 2017).

Modification of HI as a key agronomical goal has been accomplished through various strategies. Most attempts have relied on crosses with wild relatives to select favorable alleles and on engineering hormone accumulation and signaling. Many introgression lines of *Solanum pennellii* did show high HI (Schauer et al., 2006), usually accompanied by decreased overall yields (Gur

et al., 2010). On the other hand, brassinosteroid metabolism has been targeted to modify several traits in tomato including HI. Overexpression of the brassinosteroid receptor *SlBRI1* resulted in increased plant size and leaf area, with little or no effect on fruit yield (Nie et al., 2017). As a consequence, the overall HI was decreased in those plants, partly compensated by accelerated fruit ripening and improved quality. *DWARF*, the key brassinosteroid biosynthetic gene in tomato, has also been overexpressed, leading to significant increases in plant height and biomass, and lower expansion diameter (Li et al., 2016). The major decline in HI was partially compensated by early flowering and accelerated fruit ripening, and the authors predicted higher yield per planted surface due to the compact architecture of *DWARF*-expressing plants (Li et al., 2016).

In a different approach, improvement of HI was obtained by silencing the expression of a chloroplast DnaJ chaperone involved in assimilate partitioning into fruit (Bermúdez et al., 2014). The increase in HI was gained through higher ripe fruit weight per plant without modification of the aerial biomass (Bermúdez et al., 2014). Under the conditions employed in that trial, the HI of WT control plants was very low (~0.1), compared to 0.48 in our assay (**Figure 4E**). Despite the more stringent background, *Slpfld* plants HI did increase to ~0.63, even higher than those reported for DnaJ-silenced plants that were close to 0.5 (Bermúdez et al., 2014). Moreover, calculations based on the horizontal expansion diameters of the *Slpfld* plants (**Supplementary Figure S3B**), as done by Li et al. (2016), suggest that major improvements in absolute fruit yield per planted surface could be gained by increasing plant density per square meter in the field.

Tomato plants did not represent the only reported case of leaf size decrease upon expression of a plastid-located Fld. As indicated, creeping bentgrass (Li et al., 2017) and *Arabidopsis* (Su et al., 2018) displayed a similar phenotype. Interestingly, development of vegetative tissues in Fld-expressing tobacco (Ceccoli et al., 2012) and *Medicago truncaluta* (Coba de la Peña et al., 2010) lines did not differ significantly from those of their WT siblings, suggesting that the effect of the flavoprotein on growth might have some degree of species specificity. Cell size, however, was actually reduced in leaves of *pfld* tobacco plants (Mayta et al., 2018), prompting for a more detailed study of the developmental features displayed by these plants in the absence of stress.

The mechanisms by which chloroplast-targeted Fld can modulate organ development are presently unknown. Su et al. (2018) proposed that size reduction might reflect the lower efficiency of Fld, compared to Fd, as electron carrier during photosynthesis (Nogués et al., 2004). Alternatively, Fld could modulate redox processes and down-regulate accumulation of reactive oxygen species (ROS), as observed in Fld-expressing plants exposed to adverse environmental situations (Tognetti et al., 2006, Tognetti et al., 2007; Zurbriggen et al., 2008; Zurbriggen et al., 2009; Li et al., 2017; Rossi et al., 2017). Oxidative bursts have been detected during both leaf and fruit transitions (Muñoz and Munné-Bosch, 2018), and proposed to provide signaling cues required for these developmental programs. The cellular origin(s) of the observed ROS build-up are still unclear but they most likely involve chloroplasts and mitochondria (Muñoz and Munné-Bosch, 2018).

Ascorbate is a canonical plant antioxidant, and manipulation of its metabolism has been used to modify yield and HI in cherry

tomatoes. The levels of ascorbate oxidase, which oxidizes ascorbate to monodehydroascorbate, and of monodehydroascorbate reductase (MDHAR), which participates in ascorbate regeneration, were reduced using RNAi techniques (Garchery et al., 2013; Truffault et al., 2016). Knocked-down ascorbate oxidase plants showed improved yield (Garchery et al., 2013), while the opposite effect was observed in siblings with impaired MDHAR activity (Truffault et al., 2016). The results revealed a strong correlation between antioxidant levels (in this case, ascorbate) and yield. Fruit-specific decreases of ascorbate oxidase and MDHAR activities obtained by expressing the RNAi sequences under control of a fruit promoter had no consequences in yield, indicating that the antioxidant capacity of leaves was the key factor determining the yield phenotype (Truffault et al., 2018). Moreover, application of this same strategy to Moneymaker tomatoes failed to show any fruit change (Truffault et al., 2018), underscoring the importance of the genotype in the determination of yield and HI. Cherry tomatoes produce small fruit and exhibit low HI values compared to Moneymaker, indicating that source-sink relationships and metabolite allocation must be necessarily different in the two cultivars.

MDHAR isoforms are distributed in various cellular compartments, whereas ascorbate oxidase is an apoplastic enzyme. Then, to the best of our knowledge this is the first report in which genetic manipulations of a chloroplast redox shuttle modify growth of a sink tissue and HI. The connection between plastid function and organ development has been recognized only recently (Andriankaja et al., 2012; Van Dingenen et al., 2016). Chloroplasts communicate information to nuclei as a response to environmental and developmental stimuli, a process known as retrograde signaling (Hernández-Verdeja and Strand, 2018). Several potential operating signals originating from these organelles have been proposed, including ROS and the redox status of the PETC and the chloroplast stroma (Exposito-Rodriguez et al., 2017), all of which can be affected by Fld presence (Pierella Karlusich et al., 2014; Rossi et al., 2017; Mayta et al., 2018). Organ size, on the other hand, has been related to various cellular processes including endoreduplication (Kawade and Sukaya, 2017), phytochromes (Husaineid et al., 2007), and modulation *via* proteasomal activity (Sonoda et al., 2009; Nguyen et al., 2013). Noteworthy, the ubiquitin-proteasome system has been reported to modulate plastid-nuclear bidirectional communication (Hirosawa et al., 2017), and most proteasomal components were up-regulated by chloroplast Fld presence in tobacco plants grown under normal conditions (Pierella Karlusich et al, 2017), suggesting that the effects of the flavoprotein could be mediated by selective protein degradation. Then, our working hypothesis is that, by productively interacting with the PETC and other oxido-reductive pathways of the chloroplast, Fld affects retrograde signaling involved in organ development, presumably mediated by proteasomal function, ploidy, receptors, etc. Research is currently underway to address these possibilities.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

MM, MZ, EV, M-RH, MIZ, and NC conceived the original research plans. MM, RA, MZ, EV, M-RH, MIZ, and NC designed the experiments. MM, RA, MZ, EV, M-RH, and MIZ performed the experiments. MM, RA, MZ, EV, M-RH, MIZ, and NC analyzed the data. MM, RA, MZ, EV, M-RH, MIZ, and NC wrote the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01432/full#supplementary-material

REFERENCES

Abriata, L. A. (2012). Utilization of NMR spectroscopy to study biological fluids and metabolic processes: two introductory activities. *Concepts Magn. Reson.* 40A, 171–178. doi: 10.1002/cmr.a.21235

Andriankaja, M., Dhondt, S., De Bodt, S., Vanhaeren, H., Coppens, F., De Milde, L., et al. (2012). Exit from proliferation during leaf development in *Arabidopsis thaliana*: a not-so-gradual process. *Dev. Cell* 22, 64–78. doi: 10.1016/j.devcel.2011.11.011

Ariizumi, T., Shinozaki, Y., and Ezura, H. (2013). Genes that influence yield in tomato. *Breed. Sci.* 63, 3–13. doi: 10.1270/jsbbs.63.3

Baker, N. R. (2008). Chlorophyll fluorescence: a probe of photosynthesis *in vivo*. *Annu. Rev. Plant Biol.* 59, 89–113. doi: 10.1146/annurev.arplant.59.032607.092759

Bermúdez, L., Urias, U., Milstein, D., Kamenetzky, L., Asis, R., Fernie, A. R., et al. (2008). A candidate gene survey of quantitative trait loci affecting chemical composition in tomato fruit. *J. Exp. Bot.* 59, 2875–2890. doi: 10.1093/jxb/ern146

Bermúdez, L., de Godoy, F., Baldet, P., Demarco, D., Osorio, S., Quadrana, L., et al. (2014). Silencing of the tomato Sugar Partitioning Affecting protein (SPA) modifies sink strength through a shift in leaf sugar metabolism. *Plant J.* 77, 676–687. doi: 10.1111/tpj.12418

Ceccoli, R. D., Blanco, N. E., Segretin, M. E., Melzer, M., Hanke, G. T., Scheibe, R., et al. (2012). Flavodoxin displays dose-dependent effects on photosynthesis and stress tolerance when expressed in transgenic tobacco plants. *Planta* 236, 1447–1458. doi: 10.1007/s00425-012-1695-x

Chakrabarti, M., Zhang, N., Sauvage, C., Munos, S., Blanca, J., Canizares, J., et al. (2013). A cytochrome P450 regulates a domestication trait in cultivated tomato. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17125–17130. doi: 10.1073/pnas.1307313110

Coba de la Peña, T., Redondo, F. J., Manrique, E., Lucas, M. M., and Pueyo, J. J. (2010). Nitrogen fixation persists under conditions of salt stress in transgenic *Medicago truncatula* plants expressing a cyanobacterial flavodoxin. *Plant Biotechnol. J.* 8, 954–965. doi: 10.1111/j.1467-7652.2010.00519.x

Exposito-Rodriguez, M., Laissue, P. P., Yvon-Durocher, G., Smirnoff, N., and Mullineaux, P. M. (2017). Photosynthesis-dependent H_2O_2 transfer from chloroplasts to nuclei provides a high-light signalling mechanism. *Nat. Commun.* 8, 49. doi: 10.1038/s41467-017-00074-w

Garchery, C., Gest, N., Do, P. T., Alhagdow, M., Baldet, P., Menard, G., et al. (2013). A diminution in ascorbate oxidase activity affects carbon allocation and improves yield in tomato under water deficit. *Plant Cell Environ.* 36, 159–175. doi: 10.1111/j.1365-3040.2012.02564.x

Gur, A., Osorio, S., Fridman, E., Fernie, A. R., and Zamir, D. (2010). *hi2-1*, A QTL which improves harvest index, earliness and alters metabolite accumulation of processing tomatoes. *Theor. Appl. Genet.* 121, 1587–1599. doi: 10.1007/s00122-010-1412-8

Hedden, P. (2003). The genes of the Green Revolution. *Trends Genet.* 19, 5–9. doi: 10.1016/s0168-9525(02)00009-4

Hernández-Verdeja, T., and Strand, Å. (2018). Retrograde signals navigate the path to chloroplast development. *Plant Physiol.* 176, 967–976. doi: doi.org/10.1104/pp.17.01299

Hirosawa, Y., Ito-Inaba, Y., and Inaba, T. (2017). Ubiquitin–proteasome-dependent regulation of bidirectional communication between plastids and the nucleus. *Front. Plant Sci.* 8, 310. doi: 10.3389/fpls.2017.00310

Husaineid, S. S. H., Kok, R. A., Schreuder, M. E. L., Hanumappa, M., Cordonnier-Pratt, M.-M., Pratt, L. H., et al. (2007). Overexpression of homologous phytochrome genes in tomato: exploring the limits in photoperception. *J. Exp. Bot.* 58, 615–626. doi: 10.1093/jxb/erl253

Igamberdiev, A. U., and Eprintsev, A. T. (2016). Organic acids: the pools of fixed carbon involved in redox regulation and energy balance in higher plants. *Front. Plant Sci.* 7, 1042. doi: 10.3389/fpls.2016.01042

Kawade, K, and Tsukaya, Hs (2007). Probing the stochastic property of endoreduplication in cell size determination of *Arabidopsis thaliana* leaf epidermal tissue. *PloS One* 12, e0185050. doi: doi.org/10.1371/journal.pone.0185050

Khush, G. S. (2001). Green revolution: the way forward. *Nat. Rev. Genet.* 2, 815–822. doi: 10.1038/35093585

Kuhlgert, S., Austic, G., Zegarac, R., Osei-bonsu, I., Hoh, D., Chilvers, M. I., et al. (2016). MultispeQ Beta: a tool for phenotyping connected to the open PhotosynQ network Subject Areas. *R. Soc. Open Sci.* 3, 160592. doi: 10.1098/rsos.160592

Langridge, P. (2014). Reinventing the green revolution by harnessing crop mutant resources. *Plant Physiol.* 166, 1682–1683. doi: 10.1104/pp.114.252601

Li, X.-J., Chen, X.-J., Guo, X., Yin, L.-L., Ahammed, G. J., Xu, C.-J., et al. (2016). *DWARF* overexpression induces alteration in phytohormone homeostasis, development, architecture and carotenoid accumulation in tomato. *Plant Biotechnol. J.* 14, 1021–1033. doi: 10.1111/pbi.12474

Li, X., Yan, W., Agrama, H., Jia, L., Jackson, A., Moldenhauer, K., et al. (2012). Unraveling the complex trait of harvest index with association mapping in rice (*Oryza sativa* L.). *PloS One* 7, e29350. doi: 10.1371/journal.pone.0029350

Li, Z., Yuan, S., Jia, H., Gao, F., Zhou, M., Yuan, N., et al. (2017). Ectopic expression of a cyanobacterial flavodoxin in creeping bentgrass impacts plant development and confers broad abiotic stress tolerance. *Plant Biotechnol. J.* 15, 433–446. doi: 10.1111/pbi.12638

Lichtenthaler, H. K. (1987). Chlorophylls and carotenoids: pigments of photosynthetic biomembranes. *Methods Enzymol.* 148, 350–382. doi: 10.1016/0076-6879(87)48036-1

Lippman, Z. B., Cohen, O., Alvarez, J. P., Abu-Abied, M., Pekker, I., Paran, I., et al. (2008). The making of a compound inflorescence in tomato and related Nightshades. *PloS Biol.* 6, e288. doi: 10.1371/journal.pbio.0060288

López, M. G., Zanor, M. I., Pratta, G. R., Stegmayer, G., Boggio, S. B., Conte, M., et al. (2015). Metabolic analyses of interspecific tomato recombinant inbred lines for fruit quality improvement. *Metabolomics* 11, 1416–1431. doi: 10.1007/s11306-015-0798-3

Luo, X., Ma, C., Yue, Y., Hu, K., Li, Y., Duan, Z., et al. (2015). Unravelling the complex trait of harvest index in rapeseed (*Brassica napus* L.) with association mapping. *BMC Genomics* 16, 379. doi: 10.1186/s12864-015-1607-0

Marano, M. R., and Carrillo, N. (1991). Chromoplast formation during tomato fruit ripening. No evidence for plastid DNA methylation. *Plant Mol. Biol.* 16, 11–19. doi: 10.1007/bf00017913

Marano, M. R., Serra, E. C., Orellano, E. G., and Carrillo, N. (1993). The path of chromoplast development in fruits and flowers. *Plant Sci.* 94, 1–17. doi: 10.1016/0168-9452(93)90002-H

Matsuda, R., and Kubota, C. (2010). Variation of total soluble protein content in fruit among six greenhouse tomato cultivars. *HortScience* 45, 1645–1648. doi: 10.21273/HORTSCI.45.11.1645

Mayta, M. L., Lodeyro, A. F., Guiamet, J. J., Tognetti, V. B., Melzer, M., Hajirezaei, M. R., et al. (2018). Expression of a plastid-targeted flavodoxin decreases chloroplast reactive oxygen species accumulation and delays senescence in aging tobacco leaves. *Front. Plant Sci.* 9, 1039. doi: 10.3389/fpls.2018.01039

Mu, Q., Huang, Z., Chakrabarti, M., Illa-Berenguer, E., Liu, X., Wang, Y., et al. (2017). Fruit weight is controlled by *Cell Size Regulator* encoding a novel protein that is expressed in maturing tomato fruits. *PloS Genet.* 13, e1006930. doi: 10.1371/journal.pgen.1006930

Muñoz, P., and Munné-Bosch, S. (2018). Photo-oxidative stress during leaf, flower and fruit development. *Plant Physiol.* 176, 1004–1014. doi: 10.1104/pp.17.01127

Nguyen, H. M., Schippers, J. H., Gõni-Ramos, O., Christoph, M. P., Dortay, H., van der Hoorn, R. A., et al. (2013). An upstream regulator of the 26S proteasome modulates organ size in *Arabidopsis thaliana*. *Plant J.* 74, 25–36. doi: 10.1111/tpj.12097

Nie, S., Huang, S., Wang, S., Cheng, D., Liu, J., Lv, S., et al. (2017). Enhancing brassinosteroid signaling *via* overexpression of tomato (*Solanum lycopersicum*) SlBRI1 improves major agronomic traits. *Front. Plant Sci.* 8, 1386. doi: 10.3389/fpls.2017.01386

Nogués, I., Tejero, J., Hurley, J. K., Paladini, D., Frago, S., Tollin, G., et al. (2004). Role of the C-terminal tyrosine of ferredoxin-nicotinamide adenine dinucleotide phosphate reductase in the electron transfer processes with its protein partners ferredoxin and flavodoxin. *Biochemistry* 43, 6127–6137. doi: 10.1021/bi049858h

Park, S. J., Jiang, K., Schatz, M. C., and Lippman, Z. B. (2012). Rate of meristem maturation determines inflorescence architecture in tomato. *Proc. Natl. Acad. Sci. U.S.A.* 109, 639–644. doi: 10.1073/pnas.1114963109

Pierella Karlusich, J. J., Ceccoli, R. D., Graña, M., Romero, H., and Carrillo, N. (2015). Environmental selection pressures related to iron utilization are involved in the loss of the flavodoxin gene from the plant genome. *Genome Biol. Evol.* 7, 750–767. doi: 10.1093/gbe/evv031

Pierella Karlusich, J. J., Lodeyro, A. F., and Carrillo, N. (2014). The long goodbye: the rise and fall of flavodoxin during plant evolution. *J. Exp. Bot.* 65, 5161–5178. doi: 10.1093/jxb/eru273

Pierella Karlusich, J. J., Zurbriggen, M. D., Shahinnia, F., Sonnewald, S., Sonnewald, U., Hosseini, S. A., et al. (2017). Chloroplast redox status modulates genome-wide plant responses during the non-host interaction of tobacco with the hemibiotrophic Bacterium xanthomonas campestris pv. vesicatoria. *Front. Plant Sci.* 8, 1158. doi: 10.3389/fpls.2017.01158

Radzevičius, A., Viškelis, J., Karkleienė, R., and Viskelis, P. (2016). Determination of tomato quality attributes using near infrared spectroscopy and reference analysis. *Zemdirbyste-Agriculture* 103, 443–448. doi: 10.13080/z-a.2016.103.012

Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, https://imagej.nih.gov/ij/, 1997-2018.

Robson, P. R. H., McCormac, A. C., Irvine, A. S., and Smith, H. (1996). Genetic engineering of harvest index in tobacco through overexpression of a phytochrome gene. *Nat. Biotechnol.* 14, 995–998. doi: 10.1038/nbt0896-995

Rossi, F. R., Krapp, A. R., Bisaro, F., Maiale, S. J., Pieckenstain, F. L., and Carrillo, N. (2017). Reactive oxygen species generated in chloroplasts contribute to tobacco leaf infection by the necrotrophic fungus *Botrytis cinerea*. *Plant J.* 92, 761–773. doi: 10.1111/tpj.13718

Sadali, N., Sowden, R., Ling, Q., and Jarvis, R. (2019). Differentiation of chromoplasts and other plastids in plants. *Plant Cell Rep.* 38, 803–818. doi: 10.1007/s00299-019-02420-2

Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., et al. (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* 24, 447–454. doi: 10.1038/nbt1192

Shim, J. S., and Imaizumi, T. (2015). Circadian clock and photoperiodic response in Arabidopsis: from seasonal flowering to redox homeostasis. *Biochemistry* 54, 157–70. doi: 10.1021/bi500922q.

Simonian, M. H. (2002). Spectrophotometric determination of protein concentration. *Curr. Protoc. Cell Biol.* 15, A.3B.1–A.3B.7. doi: 10.1002/0471143030.cba03bs15

Sonoda, Y., Sako, K., Maki, Y., Yamazaki, N., Yamamoto, H., Ikeda, A., et al. (2009). Regulation of leaf organ size by the Arabidopsis RPT2a 19S proteasome subunit. *Plant J.* 60, 68–78. doi: 10.1111/j.1365-313X.2009.03932.x

Sorrequieta, A., Abriata, L., Boggio, S., and Valle, E. (2013). Off-the-vine ripening of tomato fruit causes alteration in the primary metabolite composition. *Metabolites* 3, 967–978. doi: 10.3390/metabo3040967

Steiner, E., Livne, S., Kobinson-Katz, T., Tal, L., Pri-Tal, O., Mosquna, A., et al. (2016). The putative O-linked N-acetylglucosamine transferase SPINDLY inhibits class I TCP proteolysis to promote sensitivity to cytokinin. *Plant Physiol.* 171, 1485–1494. doi: 10.1104/pp.16.00343

Su, J., Yang, L., Zhu, Q., Wu, H., He, Y., Liu, Y., et al. (2018). Active photosynthetic inhibition mediated by MPK3/MPK6 is critical to effector-triggered immunity. *PloS Biol.* 16, e2004122. doi: 10.1371/journal.pbio.2004122

Tauberger, E., Fernie, A. R., Emmermann, M., Renz, A., Kossmann, J., Willmitzer, L., et al. (2000). Antisense inhibition of plastidial phosphoglucomutase provides compelling evidence that potato tuber amyloplasts import carbon from the cytosol in the form of glucose-6-phosphate. *Plant J.* 23, 43–53. doi: 10.1046/j.1365-313x.2000.00783.x

Tognetti, V. B., Palatnik, J. F., Fillat, M. F., Melzer, M., Hajirezaei, M. R., Valle, E. M., et al. (2006). Functional replacement of ferredoxin by a cyanobacterial flavodoxin in tobacco confers broad-range stress tolerance. *Plant Cell* 18, 2035–2050. doi: 10.1105/tpc.106.042424

Tognetti, V. B., Zurbriggen, M. D., Morandi, E. N., Fillat, M. F., Valle, E. M., Hajirezaei, M.-R., et al. (2007). Enhanced plant tolerance to iron starvation by functional substitution of chloroplast ferredoxin with a bacterial flavodoxin. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11495–11500. doi: 10.1073/pnas.0704553104

Truffault, V., Gest, N., Garchery, C., Florian, A., Fernie, A. R., Gautier, H., et al. (2016). Reduction of MDHAR activity in cherry tomato suppresses growth and yield and MDHAR activity is correlated with sugar levels under high light. *Plant Cell Environ.* 39, 1279–1292. doi: 10.1111/pce.12663

Truffault, V., Riqueau, G., Garchery, C., Gautier, H., and Stevens, R. G. (2018). Is monodehydroascorbate reductase activity in leaf tissue critical for the maintenance of yield in tomato? *J. Plant Physiol.* 222, 1–8. doi: 10.1016/j.jplph.2017.12.012

Van Dingenen, J., Blomme, J., Gonzalez, N., and Inzé, D. (2016). Plants grow with a little help from their organelle friends. *J. Exp. Bot.* 67, 6267–6281. doi: 10.1093/jxb/erw399

Zanor, M. I., Rambla, J.-L., Chaïb, J., Steppa, A., Medina, A., Granell, A., et al. (2009). Metabolic characterization of *loci* affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents. *J. Exp. Bot.* 60, 2139–2154. doi: 10.1093/jxb/erp086

Zhu, Q.-H., and Helliwell, C. A. (2011). Regulation of flowering time and floral patterning by miR172. *J. Exp. Bot.* 62, 487–495. doi: 10.1093/jxb/erq295

Zurbriggen, M. D., Carrillo, N., Tognetti, V. B., Melzer, M., Peisker, M., Hause, B., et al. (2009). Chloroplast-generated reactive oxygen species play a major role in localized cell death during the non-host interaction between tobacco and *Xanthomonas campestris* pv. *vesicatoria*. *Plant J.* 60, 962–973. doi: 10.1111/j.1365-313X.2009.04010.x

Zurbriggen, M. D., Tognetti, V. B., Fillat, M. F., Hajirezaei, M.-R., Valle, E. M., and Carrillo, N. (2008). Combating stress with flavodoxin: a promising route for crop improvement. *Trends Biotechnol.* 26, 531–537. doi: 10.1016/j.tibtech.2008.07.001

Combating Micronutrient Deficiency and Enhancing Food Functional Quality Through Selenium Fortification of Select Lettuce Genotypes Grown in a Closed Soilless System

Antonio Pannico[1†], Christophe El-Nakhel[1†], Marios C. Kyriacou[2], Maria Giordano[1], Silvia Rita Stazi[3], Stefania De Pascale[1] and Youssef Rouphael[1]*

[1] Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy, [2] Department of Vegetable Crops, Agricultural Research Institute, Nicosia, Cyprus, [3] Department of Chemical and Pharmaceutical Sciences (DSCF), University of Ferrara, Ferrara, Italy

Selenium (Se) is an essential trace element for human nutrition and a key component of selenoproteins having fundamental biological and nutraceutical functions. We currently examined lettuce biofortification with Se in an open-gas-exchange growth chamber using closed soilless cultivation for delivering Se-rich food. Morphometric traits, minerals, phenolic acids, and carotenoids of two differently pigmented Salanova cultivars were evaluated in response to six Se concentrations (0–40 µM) delivered as sodium selenate in the nutrient solution. All treatments reduced green lettuce fresh yield slightly (9%), while a decrease in red lettuce was observed only at 32 and 40 µM Se (11 and 21% respectively). Leaf Se content increased in both cultivars, with the red accumulating 57% more Se than the green. At 16 µM Se all detected phenolic acids increased, moreover a substantial increase in anthocyanins (184%) was recorded in red Salanova. Selenium applications slightly reduced the carotenoids content of green Salanova, whereas in red Salanova treated with 32 µM Se violaxanthin + neoxanthin, lutein and β-cryptoxanthin spiked by 38.6, 27.4, and 23.1%, respectively. Lettuce constitutes an ideal target crop for selenium biofortification and closed soilless cultivation comprises an effective tool for producing Se-enriched foods of high nutraceutical value.

Keywords: anthocyanins, carotenoids profile, hydroponics, *Lactuca sativa* L., mineral composition, nutrient solution management, phenolic acids, sodium selenate

INTRODUCTION

Selenium (Se) is considered a non-essential mineral nutrient for higher plants (Sors et al., 2005; Pilon-Smits and Quinn, 2010; Malagoli et al., 2015), nevertheless several studies demonstrate the effectiveness of Se at low concentrations in improving photo-oxidative stress tolerance, delaying senescence and stimulating plant yield (Hartikainen, 2005; Lyons et al., 2009). The anti-oxidative function of Se is related to the increased activity of antioxidant enzymes including lipoxygenase,

superoxide dismutase, catalase, ascorbate peroxidase, and glutathione peroxidase with the consequent decrease of lipid peroxidation, as well as to the enhanced synthesis of antioxidant molecules such as phenols, carotenoids, flavonoids, and anthocyanins in Se treated-plants (Djanaguiraman et al., 2005; Hawrylak-Nowak, 2008; Ramos et al., 2010; Ardebili et al., 2015).

While Se is considered merely beneficial to plants (Pilon-Smits et al., 2009; Vatansever et al., 2017; Chauhan et al., 2019), it is deemed essential for animal and human nutrition as it constitutes the key component of selenoenzymes and selenoproteins with fundamental biological functions (Rayman, 2002). Low dietary intake of Se has been associated with serious human illnesses, such as cardiovascular diseases, viral infections and certain types of cancer (Rayman, 2000; Combs, 2001; Finley, 2005). Selenium deficiency has been estimated to affect up to one billion people worldwide (Jones et al., 2017). Most serious consequences have been reported in China, the UK, Eastern Europe, Africa, and Australia (Chen et al., 2002; Lyons et al., 2004), in areas with arable soils of low Se bioavailability that inevitably limits Se entry into the food supply chain.

The Recommended Dietary Allowance (RDA) of Se for adult men and women is 55 µg day^{-1} (Johnson et al., 2003), however, Burk et al. (2006) have found that Se supplementation of 200 µg day^{-1}, reduces the risk of prostate, lung and colon cancer. Plants constitute a potentially significant source of this element for human diet through biofortification. Biofortification is the process that increases the bioavailable content of targeted elements in edible plant parts through agricultural intervention or genetic selection (White and Broadley, 2005). In this perspective, recent works have demonstrated that Se fertilization increases the content of this element in a wide range of crops including rice (Chen et al., 2002), wheat (Lyons et al., 2004), radish (Pedrero et al., 2006; Schiavon et al., 2016), spinach (Ferrarese et al., 2012), potato (Turakainen et al., 2004), bean (Hermosillo-Cereceres et al., 2011), soybean (Yang et al., 2003), pea (Jerše et al., 2018), tomato (Schiavon et al., 2013), rocket (Dall'Acqua et al., 2019), lamb's lettuce (Hawrylak-Nowak et al., 2018), and lettuce (Businelli et al., 2015; Esringu et al., 2015; Smolen et al., 2016a; Silva et al., 2017; Silva et al., 2018a). Se fertilization is a relatively low-cost approach to the prophylaxis of consumers against nutrient deficiency. Several countries, such as Finland, Malawi, Australia, and New Zealand, have supported this strategy through biofortification programs, demonstrated to boost Se content in human tissue and body fluids of the population (Arthur, 2003; Eurola et al., 2004; Chilimba et al., 2012), as well as Brazil, where studies were performed on upland rice (Reis et al., 2018), rice (Andrade et al., 2018) and cowpea (Silva et al., 2018b; Silva et al., 2019).

Higher plant roots uptake Se mainly as selenate and selenite. Selenate is transported across the plasma membrane of root cells, using the assimilation pathways of sulfate *via* the enzyme sulfate permease (Terry et al., 2000; Hawkesford and Zhao, 2007), while selenite is transported *via* phosphate transporters (Li et al., 2008). The selectivity of these transporters is species-dependent and affected by soil sulfate concentration, salinity, pH and redox potential (Combs, 2001; White et al., 2004); moreover, the

different types of sulphate transporters (SULTR1;1, SULTR1;2, SULTR2;1) may have different selectivity for selenium and sulfur (Dall'Acqua et al., 2019). Nevertheless, selenate is more soluble, less phytotoxic and easily transported and accumulated in crops compared to selenite (Lyons et al., 2005; Smrkolj et al., 2005; Hawrylak-Nowak, 2013).

Regarding the bioactive value of Se, several studies have demonstrated its role in plant secondary metabolism by increasing tocopherol, flavonoids, phenolic compounds, ascorbic acid and vitamin A (Hartikainen et al., 2000; Xu et al., 2003; Ríos et al., 2008; Businelli et al., 2015), noting that plant secondary metabolites are health promoting phytochemicals that prevent a range of human diseases and are used as well as medicinal active ingredients (El-Nakhel et al., 2019). However, at high concentrations Se is phytotoxic, inhibiting growth and modifying the nutritional characteristics of plants (Hartikainen et al., 2000). Selenium phytotoxicity is attributable to non-specific incorporation of selenocysteine (SeCys) and selenomethionine (SeMet) which replace their sulphur analogues compounds in plant proteins (Ellis and Salt, 2003).

Vegetables are widely used in biofortification studies, including lettuce (*Lactuca sativa* L.), which is the most produced and consumed leafy vegetable in the world (Baslam et al., 2013; Hawrylak-Nowak, 2013). It has attained a central role in human nutrition as it combines palatable organoleptic properties with a rich content of nutraceutical compounds (phenolic acids, carotenoids, flavonoids, and vitamins B9, C, and E) and a low content of dietary fats, which makes lettuce an attractive low-calorie food (Kim et al., 2016). Moreover, since lettuce is generally eaten raw, more nutrients are retained compared to cooked foods, including Se that has been has been shown to diminish in concentration after food processing, such as boiling, baking or grilling (Dumont et al., 2006; Sager, 2006). Being also one of the most easily cultivated vegetables both in soil and in hydroponic systems, lettuce can be considered therefore a promising candidate for Se biofortification.

Several biofortification techniques have been proposed, such as soil/substrate dosing with Se, foliar spray with Se solution and hydroponic cultivation with Se enriched nutrient solution (Smrkolj et al., 2007; Puccinelli et al., 2017; Wiesner-Reinhold et al., 2017). The technique choice should consider, among other aspects, the possible run-off of Se fertilizers resulting in Se accumulation in groundwater. In this respect, hydroponic cultivation, especially in closed-loop systems, has several advantages: (i) environmental spread of Se is minimized, (ii) Se uptake is higher than other methods, as the constant exposure of the roots with the fortified nutrient solution and the absence of micronutrient-soil interactions maximize uptake efficiency and accumulation in edible plant parts, (iii) product quality is standardized through precise management of the concentration and composition of nutrient solution, (iv) very small amounts of selenium are needed, and no modification of conventional closed soilless cultivation technique is required thus ensuring no additional cost (Puccinelli et al., 2017; Wiesner-Reinhold et al., 2017; Rouphael and Kyriacou, 2018).

Taking into account these considerations, the effects of sodium selenate application were evaluated in this present work at six

different doses on two lettuce cultivars of different pigmentation (green and red) cultivated in a closed soilless system. The aim of this study was to identify the appropriate Se concentration in the nutrient solution in order to maximize the accumulation of selenium and enhance the nutraceutical characteristics (lipophilic and hydrophilic antioxidant molecules), by creating a dual enrichment of lettuce, without causing important loss of yield in lettuce.

MATERIALS AND METHODS

Growth Chamber Conditions, Plant Material and Experimental Design

Two butterhead lettuce (*L. sativa* L. var. capitata) cultivars with different leaf pigmentation, green Salanova® "Descartes" and red Salanova® "Klee" (Rijk Zwaan, Der Lier, The Netherlands), were cultivated in a 28 m² open-gas-exchange growth chamber (7.0 m × 2.1 m × 4.0 m, width × height × depth) situated at the experimental station of the University of Naples Federico II, Italy.

The lighting of the growth chamber was provided by High Pressure Sodium lamps (Master SON-T PIA Plus 400W, Philips, Eindhoven, The Netherlands) with a photosynthetic photon flux density (PPFD) of 420 ± 10 μmol m^{-2} s^{-1}, measured at leaf height using a spectral radiometer (MSC15, Gigahertz-Optik, Turkenfeld, Germany). Day/night temperatures of 24/18°C were established with a 12 h photoperiod and a relative air humidity of 60–80% respectively. The experiment was carried out at ambient CO_2 concentration (390 ± 20 ppm), while air exchange and dehumidification were guaranteed by two HVAC systems. Plants were grown in nutrient film technique (NFT) established on rigid polyvinyl chloride (PVC) gullies (14.5 cm wide, 8 cm deep and 200 cm long), with a 1% slope. The gullies were 60 cm above ground level and each of them was fed by a separate 25 L plastic reservoir tank containing the nutrient solution (NS). Continuous recirculation (1.5 L min^{-1}) of the NS was provided by a submerged pump (NJ3000, Newa, Loreggia, PD, Italy) in each reservoir tank. Twenty-day-old lettuce seedlings were transplanted in rockwool cubes (7 × 7 × 7cm, Delta, Grodan, Roermond, The Netherlands) and transferred into the gullies with an intra-row and inter-row spacing of 15 and 43 cm respectively, corresponding to a density of 15.5 plants m^{-2}. Each gully was covered with PVC lid in order to avoid NS evaporation. The NS was a modified Hoagland formulation prepared with osmotic water containing: 8.0 mM N–NO$_3^-$, 1.5 mM S, 1.0 mM P, 3.0 mM K, 3.0 mM Ca, 1.0 mM Mg, 1.0 mM NH$_4^+$, 15 μM Fe, 9 μM Mn, 0.3 μM Cu, 1.6 μM Zn, 20 μM B, and 0.3 μM Mo, with electrical conductivity (EC) 1.4 dS m^{-1} and pH 6.0 ± 0.1.

The experimental design was a randomized complete-block factorial design (6 × 2) with six selenium concentrations in the nutrient solution (0, 8, 16, 24, 32, or 40 μM as sodium selenate, from Sigma-Aldrich, St. Louis, MO, USA) and two lettuce cultivars (green or red butterhead Salanova), with three replicates. Each experimental plot consisted of six plants.

Growth Analysis and Biomass Determination

Plants were harvested at nineteen days after transplant (DAT). Number of leaves and fresh weight of the aerial plant parts were determined, then leaf area was measured by an area meter (LI-COR 3100C, Biosciences, Lincoln, Nebraska, USA).

Leaf dry weight was determined on an analytical balance (Denver Instruments, Denver, Colorado, USA) after sample desiccation in a forced-air oven at 70°C to constant weight (around 72 h). Leaf dry matter was determined according to the official method 934.01 of the Association of Official Analytical Chemists.

Collection of Samples for Mineral and Nutritional Quality Analyses

Part of the dried leaf tissue of green and red Salanova plants was used for macro-mineral and selenium analyses. For the identification and quantification of phenolic acids and carotenoid compounds by HPLC-DAD, fresh samples of three plants per experimental unit were instantly frozen in liquid nitrogen and stored at −80°C before lyophilizing them in a Christ, Alpha 1–4 (Osterode, Germany) freeze drier.

Mineral Analysis by Ion Chromatography and ICP-OES and Consumer Safety of Se-Enriched Butterhead Lettuce

Leaf soluble cations and anions were determined by liquid ion exchange chromatography (ICS 3000 Dionex Sunnyvale, CA, USA) with conductimetric detection, as described previously by Rouphael et al. (2017b). Briefly, 250 mg of dried sample ground at 0.5 mm in a Wiley Mill (IKA, MF 10.1, Staufen, Germany) were suspended in 50 ml of ultrapure water (Milli-Q, Merck Millipore, Darmstadt, Germany) and stirred in shaking water bath (ShakeTemp SW22, Julabo, Seelbach, Germany) at 80° C for 10 min. The mixture was centrifuged at 6,000 rpm for 10 min (R-10M, Remi Elektrotechnik Limited, India), then filtered through a 0.45 μm syringe filter (Phenomenex, Torrance, CA, USA). Chromatographic separation of Na, K, Mg, and Ca was achieved in isocratic mode (20 mM methanesulphonic acid) on an IonPac CS12A analytical column (4 × 250 mm, Dionex Sunnyvale, CA, USA) equipped with an IonPac CG12A precolumn (4 × 250 mm, Dionex Sunnyvale, CA, USA) and a self-regenerating suppressor CERS500 (4 mm, Dionex Sunnyvale, CA, USA). Nitrate, phosphate, and sulphate were detected in gradient mode (1mM-50mM KOH) on an IonPac ATC-HC anion trap (9×75 mm, Dionex Sunnyvale, CA, USA), and an AS11-HC analytical column (4 × 250 mm, Dionex Sunnyvale, CA, USA) equipped with an AG11-HC precolumn (4 × 50 mm, Dionex Sunnyvale, CA, USA) and a self-regenerating suppressor AERS500 (4 mm, Dionex Sunnyvale, CA, USA). Ions were expressed as g kg^{-1} dry weight (dw) and nitrate was expressed as mg kg^{-1} fresh weight (fw) on the basis of each sample's original dw.

In addition to macro-minerals analysis, Se content was also measured in green and red Salanova leaf tissue. Each sample was subjected to a first phase of acid digestion performed using a commercial high-pressure laboratory microwave oven (Mars plus CEM, Italy) operating at an energy output of 1,800 W. Approximately 300 mg of each dry sample was inserted directly into a microwave-closed vessel. Two milliliters of 30% (m/m) H_2O_2, 0.5 ml of 37% HCl and 7.5 ml of HNO_3 69% solution were added to each vessel. The heating program was performed in one step: temperature was ramped linearly from 25 to 180°C in 37 min, then held at 180°C for 15 min. After the digestion procedure and subsequent cooling, samples were transferred into a Teflon beaker and total volume was made up to 25 ml with Milli-Q water. The digest solution was then filtered through DISMIC 25HP PTFE syringe filter of pore size 0.45 mm (Toyo Roshi Kaisha, Ltd., Japan) and stored in a screw cap plastic tube (Nalgene, New York). Blanks were prepared in each lot of samples. All experiments were performed in triplicate. The reagents of super pure grade, used for the microwave-assisted digestions, were: hydrochloric acid (36% HCl), nitric acid (69% HNO_3), and hydrogen peroxide (30% H_2O_2) (Merck, Darmstadt, Germany). High-purity water (18 MΩ cm^{-1}) from a Milli-Q water purification system (Millipore, Bedford, USA) was used for the dilution of the standards, for preparing samples throughout the chemical process, and for final rinsing of the acid-cleaned vessels, glasses, and plastic utensils. For this work, tomato leaves (SRM 1573a) were used as external certified reference material. Selenium quantification was performed using an Inductively Coupled Plasma Optical Emission Spectrometer (ICP-OES) with an axially viewed configuration (8,000 DV, PerkinElmer, Shelton, CT, USA) equipped with an Hydride Generation system for Se quantification at 196.06 nm. Twenty-five ml of digested material was pre-reduced by concentrated HCl (5 ml, superpure grade) followed by heating at 90°C for 20 min. After pre-reduction, the solution was diluted to 50 ml in polypropylene vial with deionized water (18 MΩ cm^{-1}). In order to determine the Se concentration calibration standards were prepared, treated in same way before dilution. Selenium content in lettuce leaves was expressed as mg kg-1 dw.

The green vegetables hazard quotient (HQ$_{gv}$) was calculated according to the United States Environmental Protection Agency (USEPA) Protocol using the following formula:

$$HQ_{gv} = (ADD / RfD)$$

where ADD is the average daily dose of selenium (μg Se day^{-1}) and RfD represents the recommended dietary tolerable upper intake level of selenium (μg Se day^{-1}) assessed equal to 400 μg day^{-1} (Johnson et al., 2003), referring to the risk to human health of a 70-kg adult resulting from Se intake through the consumption of a 50-g portion of fresh lettuce.

Phenolic Acids and Anthocyanins Identification and Quantification

Four hundred mg of lyophilized samples were solubilized in a solution of methanol/water/formic acid (50/45/5, v/v/v, 12 ml) as described by Llorach et al. (2008) to determine phenolic acids

as hydroxycinnamic derivatives. The suspensions were sonicated for 30 min and then subjected to centrifugation (2,500 g for 30 min at 4°C). After a second centrifugation of supernatants at 21,100 g for 15 min at 4°C, samples were filtered through 0.22 μm cellulose filters (Phenomenex). A reversed phase C18 column (Prodigy, 250 × 4.6 mm, 5 μm, Phenomenex, Torrance, CA) equipped with a C18 security guard (4.0 × 3.0 mm, Phenomenex) was utilized for the separation of hydroxycinnamic derivatives and anthocyanins. Twenty μL of each extract were injected and the following elution gradient was built based on solvent (A) water formic acid (95:5, v/v) and (B) methanol: (0/5), (25/40), (32/40) in min/%B. The flow rate was 1 ml min^{-1}. The LC column was installed onto a binary system (LC-10AD, Shimadzu, Kyoto, Japan), equipped with a DAD (SPD-M10A, Shimadzu, Kyoto, Japan) and a Series 200 autosampler (Perkin Elmer, Waltham, MA). Chlorogenic and chicoric acids at 330 nm were used for the calibration curves of hydroxycinnamic derivatives. Identification of caffeoyl-meso-tartaric acid and caffeoyl-tartaric acid was performed by LC-MS/MS experiments.

The chromatographic profiles of reference curves and samples were recorded in multiple reaction monitoring mode (MRM) by using an API 3000 triple quadrupole (ABSciex, Carlsbad, CA). Negative electrospray ionization was used for detection and source parameters were selected as follows: spray voltage −4.2 kV; capillary temperature: 400°C, dwell time 100 ms, nebulizer gas and cad gas were set to 10 and 12 respectively (arbitrary units). Target compounds [M–H]− were analyzed using mass transitions given in parentheses: chicoric acid (m/z 473 311, 293), chlorogenic acid (m/z 353 191), caffeoyl tartaric acid (m/z 311 179, 149, retention time 15.8 min), caffeoyl-meso-tartaric acid (m/z 311 179, 149, retention time 17.8 min). The concentration of phenolic acids was reported as mg 100 g^{-1} of dw.

Anthocyanins were also measured within the same LC-DAD chromatographic runs, at 520 nm and the concentration calculated by using cyanidin as reference standard to calculate the concentration. The results were reported as μg of cyanidin equivalent per g of dw.

Carotenoids Identification and Quantification

One gram of lyophilized samples was used to determine carotenoids content following the method of Vallverdú-Queralt et al. (2013) with slight modifications. Samples were solubilized in ethanol/hexane (4:3, v/v, 2.5 ml) with 1% BHT, vortexed at 22°C for 30 s and sonicated for 5 min in the dark. Then, the solution was centrifuged (2500 g, 4°C, 10 min) and filtered through 0.45 μm nylon syringe filters (Phenomenex, Torrance, CA, USA). The extracts were dried in N and the dried extracts were dissolved in 1% BHT in chloroform. Twenty μl of each sample was injected onto a C18 column (Prodigy, 250 × 4.6 mm, 5 μm, Phenomenex, Torrance, C A, USA) with a C18 security guard (4.0 × 3.0 mm, Phenomenex). Two mobile phases were used: (A) acetonitrile, hexane, methanol, and dichloromethane (4:2:2:2, v/v/v/v) and (B) acetonitrile. Carotenoids were eluted at 0.8 ml min^{-1} through the following gradient of solvent B (t in [min]/[%B]): (0/70), (20/60), (30/30), (40/2). Carotenoids were quantified by a binary LC-10AD

system connected to a DAD (SPD-M10A, Shimadzu, Kyoto, Japan) equipped with a Series 200 auto-sampler (Perkin Elmer, Waltham, MA, USA). Violaxanthin, neoxanthin, β-cryptoxanthin, lutein and β-carotene were used as reference standards. Identification of the peaks was achieved by comparison of UV-vis spectra and retention times of eluted compounds with pure standards at 450 nm. Three separate sets of calibration curves were built; each set was injected three times in the same day (intraday assay) and three times in three different days (interday assay). The accuracy was reported as the discrepancies between the calibration curves performed intraday and interday and the results were expressed as relative standard deviation RSD (%). A recovery test was performed spiking two samples with two known amounts of carotenoids (50 and 100 μg ml^{-1} final concentration) and taking into account the overestimation due to the target analytes already present in the samples. Except for violaxanthin + neoxanthin which was expressed as μg violaxanthin equivalent per g dw, the concentration of the target carotenoids was expressed as μg g^{-1} of dw.

Statistics

All morphometric, nutritional and functional quality data were subjected to analysis of variance (two-way ANOVA) using IBM SPSS 20 software package (www.ibm.com/software/analytics/spss). Cultivar means were compared by t-Test. Duncan's multiple range test was performed for comparisons of the selenium treatment means. In order to determine the interrelationship among the morphometric, nutritional and functional quality traits in respect to the experimental treatments, a Principal Component Analysis (PCA) was performed using the appropriate function PCA from the SPSS 20 software package.

RESULTS AND DISCUSSION

Advanced Integrative Simultaneous Analysis of Morpho-Physiological Traits

Genetic material is the main pre-harvest factor that strongly affects the biometric characteristics as well as the biosynthesis, the composition and accumulation of bioactive compounds (Kim et al., 2016). For most of the measured agronomic parameters no significant interaction between the two tested factors, lettuce cultivar (C) and Se concentration in the nutrient solution (Se), was recorded, except for leaf area and fresh yield (**Table 1**). In particular, green Salanova had higher leaf number, shoot dry biomass and leaf dry matter content (%). Regarding the effect of Se concentration in the nutrient solution, increasing Se concentration to 24 μM resulted in non-significant differences in shoot dry biomass with the control (0 μM) and 16 μM treatments; whereas increasing Se concentration from 0 to 40 μM yielded a significant increase in leaf dry matter content, with the highest values observed at 40 μM (5.7%) (**Table 1**). Leaf number was not affected by the addition of Se to the nutrient solution.

TABLE 1 | Growth parameters, fresh biomass, dry biomass and leaf dry matter content of green and red Salanova lettuce grown hydroponically in a Fitotron open-gas-exchange growth chamber under six Se concentrations applied in the nutrient solution.

Source of variance	Leaf area (cm² plant⁻¹)	Leaf number (no. plant⁻¹)	Fresh biomass (g plant⁻¹)	Dry biomass (g plant⁻¹)	Dry matter (%)
Cultivar (C)					
Green Salanova	1,193 ± 16.5	59 ± 0.79 a	78.55 ± 1.13	4.32 ± 0.05 a	5.48 ± 0.06 a
Red Salanova	1,147 ± 21.8	55 ± 0.69 b	76.95 ± 1.65	3.96 ± 0.06 b	5.19 ± 0.06 b
t-test	ns	***	ns	***	***
Selenium (μM Se) (S)					
0	1,253 ± 27.8	57 ± 1.26	84.33 ± 1.71	4.26 ± 0.15 ab	5.06 ± 0.07 d
8	1,141 ± 18.0	56 ± 1.37	76.69 ± 1.47	4.04 ± 0.06 b	5.28 ± 0.06 bc
16	1,192 ± 25.6	57 ± 1.46	80.04 ± 0.95	4.15 ± 0.08 ab	5.18 ± 0.10 cd
24	1,186 ± 8.3	57 ± 1.02	80.46 ± 1.84	4.37 ± 0.06 a	5.33 ± 0.08 bc
32	1,121 ± 37.7	56 ± 2.15	74.87 ± 1.46	4.03 ± 0.13 b	5.44 ± 0.06 b
40	1,127 ± 49.8	60 ± 2.23	70.09 ± 2.35	4.01 ± 0.19 b	5.71 ± 0.09 a
	**	ns	***	*	***
C x S					
Green Salanova × 0 μM Se	1,207 ± 29.6 ab	59 ± 1.02	86.29 ± 1.47 a	4.48 ± 0.12	5.19 ± 0.07
Green Salanova × 8 μM Se	1,126 ± 21.2 bcd	58 ± 0.85	75.72 ± 2.88 cd	4.07 ± 0.10	5.38 ± 0.07
Green Salanova × 16 μM Se	1,236 ± 22.6 ab	59 ± 1.76	79.30 ± 1.85 bcd	4.26 ± 0.10	5.38 ± 0.09
Green Salanova × 24 μM Se	1,201 ± 6.2 ab	57 ± 1.02	78.08 ± 1.71 bcd	4.48 ± 0.02	5.50 ± 0.02
Green Salanova × 32 μM Se	1,169 ± 66.9 bc	58 ± 3.00	76.90 ± 2.42 bcd	4.23 ± 0.20	5.53 ± 0.10
Green Salanova × 40 μM Se	1,219 ± 59.8 ab	64 ± 0.93	74.99 ± 0.97 cd	4.41 ± 0.09	5.88 ± 0.12
Red Salanova × 0 μM Se	1,299 ± 29.5 a	55 ± 1.19	82.37 ± 2.93 ab	4.05 ± 0.23	4.94 ± 0.08
Red Salanova × 8 μM Se	1,157 ± 30.4 bc	53 ± 1.47	77.67 ± 1.26 bcd	4.01 ± 0.08	5.17 ± 0.06
Red Salanova × 16 μM Se	1,147 ± 27.5 bc	55 ± 1.35	80.78 ± 0.76 abc	4.03 ± 0.11	4.99 ± 0.09
Red Salanova × 24 μM Se	1,172 ± 9.3 bc	57 ± 2.04	82.84 ± 2.87 ab	4.27 ± 0.08	5.17 ± 0.09
Red Salanova × 32 μM Se	1,074 ± 18.1 cd	53 ± 2.92	72.84 ± 0.86 d	3.82 ± 0.05	5.35 ± 0.05
Red Salanova × 40 μM Se	1,036 ± 21.8 d	55 ± 1.11	65.20 ± 1.67 e	3.61 ± 0.07	5.54 ± 0.03
	**	ns	*	ns	ns

ns,*,**, *** Non-significant or significant at $P \leq 0.05$, 0.01, and 0.001, respectively. In the absence of interaction, cultivar means were compared by t-Test and Se application means by Duncan's multiple-range test (P = 0.05). Different letters within each column indicate significantly different means. All data are expressed as mean ± SE, n = 3.

Leaf area and fresh biomass incurred significant interaction of the tested factors (**Table 1**), as the dose effect of Se on these two morphometric traits was cultivar-dependent. In the red cultivar, a reduction of the leaf area was observed with increasing Se dose, amounting to about 11% reduction in the range of 8–32 μM Se and up to 19% at the higher Se dose (40 μM) compared to the control treatment; whereas no significant differences were recorded in the green cultivar. Cultivars/genotypes may develop different Se-tolerance and response mechanisms depending on the concentration and time of exposure. This was the case in the current experiment, since fresh yield decreased in both cultivars with increasing Se concentration in the nutrient solution although the red-pigmented butterhead lettuce was less affected than the green-pigmented cultivar especially at mild and moderate Se concentrations (i.e. 8 to 24 μM) (**Table 1**). In red Salanova, fresh yield was not affected by the addition of Se up to a concentration of 24 μM, whereas the addition of 32 μM and especially 40 μM induced a reduction in the fresh biomass of 11 and 21%, respectively, compared to the 0, 8, 16, and 24 μM treatments. Finally, a significant decrease in green Salanova fresh biomass (about 10%) was observed in response to Se application without significant differences between the five Se treatments (**Table 1**).

Several studies demonstrate the beneficial or toxic effects on morphometric traits of lettuce depending on the interaction of cultivar and application level (Ríos et al., 2008; Rios et al., 2010a; Ramos et al., 2011; Hawrylak-Nowak, 2013). Ramos et al. (2011) studied the influence of 15 μM of selenate and 15 μM of selenite concentrations in the nutrient solution on the yield of 30 lettuce accessions grown hydroponically. The authors reported that just 5 of 30 accessions treated with 15 μM of selenate showed an increase in fresh biomass compared to the control. Contrarily, Hawrylak-Nowak (2013) confirmed a decrease in both leaf area and fresh biomass of green lettuce cv. Justyna grown hydroponically and supplied with 10 μM of selenate, while in another similar work on green lettuce cv. Vera, a reduction of dry biomass was observed only at 8 μM selenate dose (Ramos et al., 2010), both of which findings are in line with our current ones on green Salanova. Additional studies conducted by Ríos et al. (2008; 2010a) also reported a decrease of dry biomass in hydroponically grown green lettuce (cv. Philipus) treated continuously with nutrient solution containing 80 μM Se compared to the control treatment.

The cultivar-dependent response to supplemental Se observed in our experiment, where the red-pigmented Salanova showed better tolerance to selenate compared to the green one, was in agreement with the study on red lettuce cv. Veneza Roxa by Silva et al. (2018a), where no significant reduction in shoot fresh weight was observed with selenate concentrations ranging from 10 to 40 μM. Considering the above, it appears that the beneficial or toxic effect of Se on plant growth and crop productivity may vary in relation to different interacting variables, including the Se concentration, time of exposure and cultivation system (Pedrero and Madrid, 2009). In the light of this finding, additional studies should focus on elucidating the cultivar × application dose × cultivation system (soilless versus soil) interaction in order to select optimal combinations to ensure balance between yield and biofortification.

Nitrate Content, Mineral Composition, Selenium Biofortification, and Consumer Safety

Nitrate content in plants grown for human consumption is extremely important, since a high intake of this nutrient may harm human health due to its potential transformation to nitrite and nitrogenous compounds that can cause serious pathological disorders, such as methaemoglobinaemia and blue baby syndrome (Colla et al., 2018). In addition, it should be taken into account that lettuce is considered a nitrate hyper-accumulator; hence the European Commission (Commission Regulation no. 1258/2011) has set as maximum limit for nitrate concentration in lettuce at 4,000 and 5,000 mg kg^{-1} fw for harvest occurring from April 1 to September 30 and from October 1 to March 31, respectively. In respect to the effect of Se concentration in the nutrient solution, the green cultivar had a higher nitrate content (1,810 mg kg^{-1} fw) than the red one (1,272 mg kg^{-1} fw), however both values were by far below EU regulation limits (**Table 2**). In fact, it is well established that nitrate accumulation in lettuce, aside from the cultivation management, depends mainly on genotypic factors (Burns et al., 2010; Burns et al., 2011; López et al., 2014). In the current study, nitrate content was influenced by both tested factors and the cultivar × Se interaction (**Table 2**). In green Salanova a significant reduction of nitrate content was observed at 8 μM (15%), 32 μM (16%), and 40 μM Se (32%) compared to the control, while no significant Se effect was found regarding this parameter in red Salanova (**Table 2**). The reduction of nitrate content prompted by selenate could be associated to the antagonistic relation of these two anions (Rios et al., 2010a). Moreover, Nowak et al. (2004) have demonstrated that Se affects the nitrate reductase enzyme, increasing its activity in plants. In addition, the reduction in foliar nitrate could be related to a greater assimilation rate of this anion due to a higher amino acid synthesis driven by enhanced nitrate reductase activity. In fact, Se toxicity in plants may be due to the formation of non-specific selenoproteins; in particular, the replacement of cysteine (Cys) with SeCys in non-specific selenoproteins would invoke a higher demand of amino acids for the synthesis of functional proteins, which would elicit the removal of these malformed selenoproteins (Van Hoewyk, 2013). Our data reflect a nitrate reduction observed in previous works, where selenate has been applied on green-pigmented lettuce at different concentrations (Lee et al., 2008; Rios et al., 2010a; Ríos et al., 2010b).

The growth and development of plants depends on the equilibrium of the mineral elements, as stress occurs in the presence of nutritional imbalances (Salt et al., 2008). Minerals are also essential for human health and lettuce is considered a good source of them (Baslam et al., 2013; Kim et al., 2016). Irrespective of Se concentration in the nutrient solution, green Salanova recorded the higher potassium and calcium content, while red Salanova showed the higher quantity of magnesium and sulphate (**Table 2**). As previously reported in literature,

TABLE 2 | Nitrate, phosphate, sulphate, potassium (K), calcium (Ca), magnesium (Mg) and sodium (Na) concentrations of green and red Salanova lettuce grown hydroponically in a Fitotron open-gas-exchange growth chamber under six Se concentrations applied in the nutrient solution.

Source of variance	Nitrate (mg kg⁻¹ fw)	Phosphate (g kg⁻¹ dw)	Sulphate (g kg⁻¹ dw)	K (g kg⁻¹ dw)	Ca (g kg⁻¹ dw)	Mg (g kg⁻¹ dw)	Na (g kg⁻¹ dw)
Cultivar (C)							
Green Salanova	1,810 ± 69	14.9 ± 0.37	5.7 ± 0.93	59.50 ± 1.19	6.13 ± 0.09 a	2.25 ± 0.03	0.36 ± 0.012
Red Salanova	1,272 ± 25	14.3 ± 0.37	14.8 ± 2.31	54.81 ± 0.67	5.21 ± 0.11 b	2.62 ± 0.04	0.39 ± 0.029
t-test	***	ns	***	**	***	***	ns
Selenium (µM Se) (S)							
0	1,660 ± 175	16.3 ± 0.55 a	2.9 ± 0.36	58.57 ± 3.00	5.73 ± 0.35 a	2.41 ± 0.06	0.37 ± 0.039
8	1,480 ± 112	15.5 ± 0.21 ab	3.9 ± 0.63	54.75 ± 1.12	5.62 ± 0.14 ab	2.31 ± 0.06	0.32 ± 0.010
16	1,680 ± 149	15.5 ± 0.06 ab	6.5 ± 1.38	58.71 ± 1.72	6.00 ± 0.29 a	2.52 ± 0.10	0.36 ± 0.013
24	1,704 ± 168	14.7 ± 0.15 b	10.5 ± 2.67	60.04 ± 1.56	5.68 ± 0.18 a	2.47 ± 0.11	0.35 ± 0.011
32	1,487 ± 111	13.1 ± 0.43 c	17.7 ± 4.14	58.18 ± 1.19	5.80 ± 0.23 a	2.51 ± 0.13	0.44 ± 0.076
40	1,234 ± 64	12.5 ± 0.28 c	20.0 ± 3.12	52.69 ± 0.84	5.21 ± 0.28 b	2.39 ± 0.13	0.42 ± 0.033
	***	***	***	*	*	*	ns
C x S							
Green Salanova × 0 µM Se	2,011 ± 168 a	16.9 ± 1.01	2.1 ± 0.24 f	63.49 ± 4.54 a	6.34 ± 0.37	2.40 ± 0.11 bc	0.44 ± 0.043 bc
Green Salanova × 8 µM Se	1,718 ± 68 b	15.3 ± 0.37	2.5 ± 0.06 f	56.89 ± 0.69 cd	5.86 ± 0.14	2.17 ± 0.03 d	0.34 ± 0.005 cd
Green Salanova × 16 µM Se	2,011 ± 30 a	15.5 ± 0.06	3.4 ± 0.18 ef	62.52 ± 0.36 ab	6.49 ± 0.23	2.31 ± 0.06 bcd	0.37 ± 0.023 bcd
Green Salanova × 24 µM Se	2,074 ± 46 a	14.9 ± 0.31	4.5 ± 0.09 e	63.38 ± 0.94 a	6.04 ± 0.10	2.22 ± 0.04 cd	0.35 ± 0.017 cd
Green Salanova × 32 µM Se	1,681 ± 148 b	13.7 ± 0.74	9.4 ± 0.45 d	57.83 ± 1.56 bcd	6.29 ± 0.08	2.22 ± 0.02 cd	0.31 ± 0.011 d
Green Salanova × 40 µM Se	1,366 ± 36 c	12.9 ± 0.21	12.3 ± 1.05 c	52.91 ± 1.15 d	5.74 ± 0.07	2.15 ± 0.03 d	0.36 ± 0.016 bcd
Red Salanova × 0 µM Se	1,309 ± 36 cd	15.7 ± 0.34	3.6 ± 0.18 ef	53.66 ± 0.39 cd	5.11 ± 0.30	2.42 ± 0.07 bc	0.29 ± 0.010 d
Red Salanova × 8 µM Se	1,242 ± 41 cd	15.6 ± 0.34	5.3 ± 0.15 e	52.62 ± 1.10 d	5.37 ± 0.12	2.44 ± 0.04 b	0.30 ± 0.012 d
Red Salanova × 16 µM Se	1,349 ± 10 cd	15.5 ± 0.09	9.6 ± 0.15 d	54.89 ± 0.29 cd	5.51 ± 0.37	2.73 ± 0.05 a	0.35 ± 0.015 bcd
Red Salanova × 24 µM Se	1,334 ± 54 cd	14.6 ± 0.06	16.4 ± 0.51 b	56.70 ± 0.31 cd	5.32 ± 0.17	2.72 ± 0.01 a	0.36 ± 0.016 bcd
Red Salanova × 32 µM Se	1,293 ± 47 cd	12.6 ± 0.28	26.1 ± 1.56 a	58.53 ± 2.14 abc	5.30 ± 0.11	2.80 ± 0.03 a	0.57 ± 0.112 a
Red Salanova × 40 µM Se	1,103 ± 45 d	12.0 ± 0.40	27.6 ± 0.39 a	52.47 ± 1.48 d	4.68 ± 0.33	2.64 ± 0.15 a	0.48 ± 0.038 ab
	*	ns	***	*	ns	**	ns

ns,*,**, *** Non-significant or significant at P ≤ 0.05, 0.01, and 0.001, respectively. In the absence of interaction, cultivar means were compared by t-Test and Se application means by Duncan's multiple-range test (P = 0.05). Different letters within each column indicate significantly different means. All data are expressed as mean ± SE, n = 3.

lettuce mineral content is quite variable depending on head type, leaf color and cultivar (Kim et al., 2016). However, regardless Se concentration in the nutrient solution and lettuce cultivar, our results particularly, potassium, calcium and magnesium were proximate to those reported by Blasco et al. (2012) on lettuce grown in controlled environment conditions.

Neither cultivar nor Se treatment had significant effect on Na accumulation in leaf tissue (avg. 0.37 g kg⁻¹ dw), whereas phosphate and calcium were highly influenced by cultivar and Se concentration with no significant interaction between the two tested factors (**Table 2**). Averaged over cultivar, phosphate content decreased significantly (about 15%) in response to Se treatments from 24 to 40 µM compared to the 0 to 16 µM treatments. In addition, the calcium content at 40 µM Se was significantly lower than the control (9%) (**Table 2**). Our findings, are in line with those of Rios et al. (2013) who reported a 9% decrease in calcium concentration at a Se dose of 40 µM compared to the control and a similar reduction in phosphate content was also observed by the same authors in response to Se concentration ranging from 20 to 120 µM.

Leaf contents in potassium, magnesium and sulphate were influenced by cultivar and Se treatments with significant C × Se interaction (**Table 2**). In green Salanova, a significant reduction of K was observed at Se 8 µM (10%) and 40 µM (17%) compared to the control (**Table 2**). Likewise, a 10% decrease in Mg content was noted with respect to the control, both at 8 and 40 µM Se. On

the contrary, in the red cultivar potassium content spiked by 9% at Se 32 µM and magnesium content by about 12% increase when Se treatment ranged between 16-40 µM, compared to the control treatment (**Table 2**). The lowest K and Mg contents observed in green Salanova at 40 µM Se application coincide with the results obtained by Rios et al. (2013) at the same dose of selenate on Philipus green lettuce cultivar. Similarly, Smoleń et al. (2016b) found a decrease in potassium content by about 9% in green butterhead lettuce leaves treated with selenium combined with iodine. On the other hand, the increase of K and Mg recorded in red Salanova treated with Se was in disagreement with other scientific literature where the authors found no variation in these two macroelements content after selenate applications (Wu and Huang, 1992; Silva et al., 2018a).

Furthermore, sulphate content increased significantly and linearly in both cultivars with selenate concentration ranging from 2.10 to 12.30 mg kg⁻¹ dw in green Salanova and from 3.63 to 27.60 mg kg⁻¹ dw in red Salanova (**Table 2**). These data imply a synergic relationship between selenate and sulphate. Selenium is chemically similar to sulfur, therefore plants absorb and metabolize Se *via* S uptake and assimilation pathway (Sors et al., 2005; Pilon-Smits and Quinn, 2010). Selenate is assimilated by plants through a process of active transport, which is driven by sulphate transporters (SULTR) (Dall'Acqua et al., 2019). SULTR mediate the movement of the sulfate in the vascular bundles, thus both selenate and sulphate are actively accumulated in

the plant cells against their electrochemical gradient (Terry et al., 2000; Dall'Acqua et al., 2019). Our results are confirmed by White et al. (2004) who found that selenate applications promoted the accumulation of sulphate in the shoots of the model plant *Arabidopsis thaliana*. Similar findings were found in lettuce by several authors (Ramos et al., 2011; Hawrylak-Nowak, 2013; Rios et al., 2013; Silva et al., 2018a), and in particular Rios and co-workers (Ríos et al., 2008) reported an increase in S content in lettuce shoots with Se concentrations up to 40 μM. The first stage in the S-assimilation process consists of the activation of the enzyme ATP-sulfurylase, which produces adenosine phosphosulfate from sulfate and ATP (Pilon-Smits et al., 1999). Then, activated selenate is reduced *via* selenite to selenide and assimilated into SeCys and SeMet. These Se-amino acids can replace their S-analogues, amino acids Cys and Met in proteins (Sors et al., 2005; Van Hoewyk, 2013). In this sense, selenate applications can increase the ATP-sulfurylase activity and consequently a greater presence of selenate could imply increased production of Se and S end products (Ríos et al., 2008). Furthermore, despite the highest SULTR expression and sulphate translocation from roots to the shoots, certain S amino acids tend to decrease as the Se dosage increases. In *Eruca sativa* a lower leaf content of Cys and glutathione was found when plants were treated with Se concentrations equal to or higher than 10 μM (Dall'Acqua et al., 2019). It is conceivable that the lower accumulation of S-compounds may be due to the interference of Se with the S flow through the assimilation pathway, consequently reducing sulphate demand and eliciting a higher accumulation of this anion in the leaves.

The effectiveness of a selenium biofortification program is strongly related with the capacity of the candidate crop to assimilate and accumulate this element in the edible parts of the plant. In the current study Se leaf content increased with selenate application rate (**Figure 1**). Comparing cultivars, red leaf lettuce accumulated on average 57% more Se than green one. Selenium leaf content was influenced by cultivar and Se treatments with highly significant interaction between the two studied factors. In particular, Se concentration peaked in green Salanova at 40 μM dose (128.43 mg kg^{-1} dw), while in red Salanova it peaked at 32 and 40 μM (116.67 and 128.20 mg kg^{-1} dw of Se, respectively). Anyhow, Se leaf content was significantly higher than the control treatment in treatments ≥ 16 μM dose for both cultivars. Our results are in agreement with previous studies on red and green-pigmented lettuce (Ramos et al., 2010; Hawrylak-Nowak, 2013; Silva et al., 2018a) demonstrating the actual feasibility of using lettuce crop in Se biofortification programs.

In the Mediterranean basin, dietary habits vary according to geographical area, but overall the well-known Mediterranean diet is mainly based on cereals, fruit, vegetables, dairy products and meat. The daily intakes of food groups considered part of the Mediterranean diet are: 219 g of cereals, 247 g of fresh and dried fruit, 226 g of vegetables and legumes, 327 g of dairy products and 136 g of meat and fish (Couto et al., 2011). These food intakes, multiplied by the average Se concentration of the individual groups, correspond to a total Se intake of around 80 μg day^{-1} per capita. Considering that the RDA of this trace element stipulated for adults is 55 μg day^{-1} (Johnson et al., 2003),

FIGURE 1 | Effects of genotype and selenium concentration in the nutrient solution on selenium biofortification of green and red Salanova lettuce grown hydroponically in a Fitotron open-gas-exchange growth chamber under six Se concentrations applied in the nutrient solution. Different letters indicate significant differences according to Duncan's test (P < 0.05). The values are means of three replicates. Vertical bars indicate ± SE of means.

it can be deduced that Se deficiency has a very low incidence in the Mediterranean area. In other countries, such as Brazil, it was found that the Se intake is only 25 μg day^{-1}, so about 30 μg Se day^{-1} must be integrated to reach the minimum recommended dose (Silva et al., 2019). The average serving of leafy vegetables, including lettuce, is about 50 g fw (Voogt et al., 2010). In our experiment, Se daily intake and percentage of RDA-Se for Se intake through consumption of 50 g portions of fresh green and red Salanova lettuce were influenced by cultivar and Se treatments with significant C × Se interaction (**Table 3**). Se daily intake increased significantly and linearly in both cultivars with selenate concentration ranging from 2 to 377 μg day^{-1} in green Salanova and from 4 to 355 μg day^{-1} in red Salanova (**Table 3**). Consequently, the RDA-Se varies with the same trend reaching a peak at 40 μM dose in both cultivars (685 and 646%, respectively for the green and red Salanova, respectively). Our RDA-Se values observed at the lowest Se dose (8 μM), were comparable with those found by Smoleń et al. (2019) on six varieties of lettuce biofortified with selenium combined with iodine at the 6.3 μM Se dose. Particularly, the iceberg varieties Krolowa and Maugli showed the lowest values (23.8 and 27.1%, respectively), while the green butterhead Cud Voorburgu and the red lettuce Lollo rossa reached the highest percentage (44.7 and 44.8%, respectively) which were comparable with the values found in green and red Salanova at the 8 μM Se dose (57 and 45%, respectively). Taking into account the Se biofortification target, 50 g fw day^{-1} of green and red Salanova at 16 μM Se dose provide 50 and 106 μg Se day^{-1} respectively (91 and 193% of the RDA), then in countries like Brazil, the RDA can be satisfied by consuming only 15 g fw day^{-1} of red Salanova or 30 g fw day^{-1} of green Salanova. On the other hand, in order to assess the risks

TABLE 3 | Selenium daily intake, percentage of recommended daily allowance for Selenium (RDA-Se) and hazard quotient (HQ_{gv}) for Se intake through consumption of 50 g portions of fresh green and red Salanova lettuce by adult humans (70 kg body weight) grown hydroponically in a Fitotron open-gas-exchange growth chamber under six Se concentrations applied in the nutrient solution.

Source of variance	Se intake with 50 g fw of lettuce (µg day⁻¹)	RDA-Se with 50 g fw of lettuce (%)	HQ_{gv} with 50 g fw of lettuce
Cultivar (C)			
Green Salanova	113 ± 31	205 ± 56	0.28 ± 0.1
Red Salanova	166 ± 33	302 ± 60	0.42 ± 0.1
t-test	ns	ns	**
Selenium (µM Se) (S)			
0	3 ± 0.5	5 ± 0.8	0.01 ± 0.0
8	28 ± 4.1	51 ± 7.4	0.07 ± 0.0
16	78 ± 14	142 ± 26	0.20 ± 0.0
24	136 ± 27	247 ± 49	0.34 ± 0.1
32	226 ± 41	410 ± 74	0.56 ± 0.1
40	366 ± 12	665 ± 21	0.91 ± 0.0
	***	***	***
C × S			
Green Salanova × 0 µM Se	2 ± 0.5 h	4 ± 0.8 h	0.00 ± 0.0 h
Green Salanova × 8 µM Se	31 ± 4.3 gh	57 ± 7.8 gh	0.08 ± 0.0 gh
Green Salanova × 16 µM Se	50 ± 1.0 fg	91 ± 1.8 fg	0.12 ± 0.0 fg
Green Salanova × 24 µM Se	77 ± 5.4 ef	139 ± 10 ef	0.19 ± 0.0 ef
Green Salanova × 32 µM Se	139 ± 22 d	253 ± 40 d	0.35 ± 0.1 d
Green Salanova × 40 µM Se	377 ± 24 a	685 ± 44 a	0.94 ± 0.1 a
Red Salanova × 0 µM Se	4 ± 0.3 h	7 ± 0.6 h	0.01 ± 0.0 h
Red Salanova × 8 µM Se	25 ± 7.3 gh	45 ± 13 gh	0.06 ± 0.0 gh
Red Salanova × 16 µM Se	106 ± 14 de	193 ± 25 de	0.27 ± 0.0 de
Red Salanova × 24 µM Se	195 ± 12 c	354 ± 22 c	0.49 ± 0.0 c
Red Salanova × 32 µM Se	312 ± 19 b	567 ± 34 b	0.78 ± 0.0 b
Red Salanova × 40 µM Se	355 ± 0.8 a	646 ± 1.4 a	0.89 ± 0.0 a
	***	***	***

*ns,**, *** Nonsignificant or significant at P ≤ 0.01, and 0.001, respectively. In the absence of interaction, cultivar means were compared by t-Test and Se application means by Duncan's multiple-range test (P = 0.05). Different letters within each column indicate significantly different means. All data are expressed as mean ± SE, n = 3.*

to human health, the green vegetables hazard quotient (HQ_{gv}) was calculated according to the United States Environmental Protection Agency (USEPA) Protocol, where HQ_{gv} values below 1.00 indicate that the vegetable is safe for consumption by human beings. In the current study HQ_{gv} increased with selenate application rate ranging from 0.00 to 0.94 in green Salanova and from 0.01 to 0.89 in red Salanova, therefore the 50 g daily portion of biofortified lettuce can be considered safe since the values of HQ_{gv} are less than 1 in all treatments (**Table 3**). In particular, in lettuce at 16 µM Se dose, the HQ_{gv} values are very low (0.12 and 0.27, respectively for green and red Salanova), indicating that even if the standard 50 g portion was tripled, these vegetables would not be in any case detrimental to human health.

Target Phenolic Compounds and Carotenoids Profiles

HPLC analysis revealed in both cultivars the presence of four main caffeic acid derivatives (**Table 4**). Chicoric acid was the most abundant phenolic acid detected in both cultivars (101.44

and 105.99 mg 100 g⁻¹ dw, respectively for the green and the red cultivar), chlorogenic acid (88.02 mg 100 g⁻¹ dw) and caffeoyl-meso-tartaric acid (41.08 mg 100 g⁻¹ dw) were higher in red Salanova, while caffeoyl-tartaric acid (17.77 mg 100 g⁻¹ dw) was higher in green Salanova compared to the red cultivar (**Table 4**). The sum of detected phenolic acids was higher in the red-pigmented cultivar with respect to the green one (239.52 and 139.10 mg 100 g⁻¹ dw, respectively). The content of phenolic acids varies according to the type of lettuce (Kim et al., 2016). Our results are consistent with the literature in which red cultivars have more phenolic acids than green ones (Llorach et al., 2008; Kim et al., 2016). The presence of chlorogenic acid, chicoric acid, and caffeoyl tartaric acid was also detected in seven different lettuce cultivars previously studied by Rouphael et al. (2017a). All phenolic acids were affected by cultivar and Se treatments with significant cultivar × Se interaction (**Table 4**). In green Salanova, caffeoyl-tartaric acid increased by 69% and 46% respectively at Se doses of 16 and 24 µM, but decreased by 75% at 32 µM, while in red Salanova the highest content was obtained at 16 µM (105%) compared to the control. Chorogenic acid in the green cultivar decreased by 57% at Se 32 µM but increased by 143% at the most concentrated Se dose, while in the red cultivar the content increased at 8, 16, 24 and 40 µM with the highest value recorded at 16 µM (191.64 mg 100 g⁻¹ dw). Similarly, chicoric acid in the green cultivar increased at Se doses of 8, 16, 24 and 40 µM with the highest value recorded at 16 µM (148.53 mg 100 g⁻¹ dw), but decreased by 67% at 32 µM; conversely, in the red cultivar chicoric acid content increased by 32% at 16 µM but decreased at Se doses 8, 24, 32 and 40 µM (**Table 4**). In red Salanova, caffeoyl-meso-tartaric acid increased by 270%, 84% and 89%, respectively, by adding in the nutrient solution 16, 24, and 40 µM of Se compared to the control treatment, while no significant differences were found for this phenolic acid in green Salanova. In the green cultivar, the sum of detected phenolic acids was significantly higher at 8, 16, 24, and 40 µM with the highest value observed at 24 µM (194.55 mg 100 g⁻¹ dw), but decreased by 67% at 32 µM, while in red cultivar the sum of phenolic acids increased by 112% at 16 µM and decreased at Se doses of 8, 32, and 40 µM compared to the control (**Table 4**).

Our results showed irregular variation of phenolic acids content in both cultivars, as the concentrations of these hydrophilic antioxidant molecules varied with Se concentration without a clear trend. Furthermore, this pattern is consistent with what was found by Schiavon et al. (2016) in radish and by D'Amato et al. (2018) in rice sprouts, but is in disagreement with Ríos et al. (2008) who reported a rise in the total phenol content of lettuce as the Se dose applied increased. On the other hand, the presence of Se constitutes an abiotic stress similar to that caused by other heavy metals. Plants react to their presence by activating the phenylpropanoid pathway (Wang et al., 2016) to produce phenolic compounds that can chelate metals and inhibit enzymes such as xanthine oxidase in an effort to prevent the production of Reactive Oxygen Species (ROS) (Ríos et al., 2008).

Anthocyanins are one of the phenolic phytochemical subclasses (Harborne and Williams, 2001) encompassing water-soluble pigments responsible for the red pigmentation

TABLE 4 | Phenolic acids composition, total phenolic acids and anthocyanins of green and red Salanova lettuce grown hydroponically in a Fitotron open-gas-exchange growth chamber under six Se concentrations applied in the nutrient solution.

Source of variance	Caffeoyl tartaric acid (mg 100 g⁻¹ dw)	Chlorogenic acid (mg 100 g⁻¹ dw)	Chicoric acid (mg 100 g⁻¹ dw)	Caffeoyl meso tartaric acid (mg 100 g⁻¹ dw)	Σ phenolic acids (mg 100g⁻¹ dw)	Anthocyanins (μg cyanidin eq. g⁻¹ dw)
Cultivar (C)						
Green Salanova	17.77 ± 1.86	13.94 ± 1.51	101.44 ± 9.27	5.96 ± 0.49	139.10 ± 12.42	n.d.
Red Salanova	4.43 ± 0.42	88.02 ± 11.71	105.99 ± 12.20	41.08 ± 5.11	239.52 ± 26.73	13.28 ± 1.45
t-test	***	***	ns	***	**	–
Selenium (μM Se) (S)						
0	9.99 ± 2.76	30.76 ± 9.46	116.65 ± 16.75	14.59 ± 3.50	171.99 ± 25.77	8.76 ± 0.23 d
8	11.79 ± 3.48	45.34 ± 14.22	92.41 ± 8.03	16.16 ± 4.47	165.70 ± 8.33	8.73 ± 0.37 d
16	17.56 ± 4.43	103.47 ± 39.47	160.34 ± 15.88	44.45 ± 17.67	325.81 ± 68.1	24.85 ± 2.58 a
24	13.97 ± 4.48	51.68 ± 15.67	114.71 ± 15.29	23.31 ± 8.09	203.68 ± 6.37	16.10 ± 0.96 b
32	3.42 ± 0.34	28.67 ± 11.01	45.83 ± 7.90	17.21 ± 6.82	95.13 ± 25.33	11.48 ± 0.56 c
40	9.88 ± 2.89	45.96 ± 10.37	92.33 ± 9.29	25.39 ± 7.73	173.56 ± 8.48	9.78 ± 0.39 cd
	***	***	***	***	***	***
C x S						
Green Salanova × 0 μM Se	16.15 ± 0.27 c	9.76 ± 0.97 g	85.40 ± 3.40 d	6.85 ± 0.23 d	118.17 ± 3.82 g	n.d.
Green Salanova × 8 μM Se	19.30 ± 1.98 c	13.71 ± 1.46 g	109.83 ± 4.00 c	6.36 ± 0.19 d	149.19 ± 6.72 f	n.d.
Green Salanova × 16 μM Se	27.23 ± 2.09 a	15.30 ± 1.18 fg	124.90 ± 1.53 c	6.33 ± 0.70 d	173.75 ± 2.52 def	n.d.
Green Salanova × 24 μM Se	23.60 ± 2.67 b	16.99 ± 0.64 fg	148.53 ± 4.47 b	5.43 ± 0.70 d	194.55 ± 7.59 cd	n.d.
Green Salanova × 32 μM Se	4.00 ± 0.37 e	4.18 ± 0.66 h	28.35 ± 1.47 f	2.21 ± 0.41 d	38.74 ± 2.31 h	n.d.
Green Salanova × 40 μM Se	16.32 ± 0.45 c	23.73 ± 0.62 f	111.63 ± 7.62 c	8.55 ± 0.39 d	160.23 ± 8.46 ef	n.d.
Red Salanova × 0 μM Se	3.84 ± 0.06 e	51.76 ± 2.26 e	147.89 ± 20.38 b	22.32 ± 1.10 c	225.82 ± 20.25 b	8.76 ± 0.23
Red Salanova × 8 μM Se	4.27 ± 0.12 de	76.98 ± 2.90 c	75.00 ± 1.79 de	25.96 ± 1.93 c	182.21 ± 5.41 de	8.73 ± 0.37
Red Salanova × 16 μM Se	7.89 ± 0.63 d	191.64 ± 3.96 a	195.78 ± 1.65 a	82.57 ± 10.34 a	477.87 ± 7.83 a	24.85 ± 2.58
Red Salanova × 24 μM Se	4.34 ± 0.72 de	86.38 ± 4.79 b	80.90 ± 2.22 de	v ± 2.69 b	212.80 ± 7.87 bc	16.10 ± 0.96
Red Salanova × 32 μM Se	2.84 ± 0.32 e	53.16 ± 2.48 e	63.31 ± 2.10 e	32.21 ± 2.69 bc	151.52 ± 4.75 f	11.48 ± 0.56
Red Salanova × 40 μM Se	3.43 ± 0.19 e	68.18 ± 6.56 d	73.04 ± 1.02 de	73.04 ± 3.80 b	186.89 ± 10.49	9.78 ± 0.39
	***	***	***	***	***	–

*ns, **, *** Nonsignificant or significant at P ≤ 0.01, and 0.001, respectively. In the absence of interaction, cultivar means were compared by t-Test and Se application means by Duncan's multiple-range test (P = 0.05). Different letters within each column indicate significantly different means. All data are expressed as mean ± SE, n = 3. n.d. not detectable.*

in lettuce (Kim et al., 2016). Consequently, these pigments were not detected in green Salanova but exclusively in the red cultivar with an average concentration of 13.28 μg cyanidin eq. g⁻¹ dw (**Table 4**). Anthocyanins have many physiological effects on plants and humans, such as antioxidation, protection against ultraviolet damage and the prevention and treatment of various diseases (Hamilton, 2004). Anthocyanins in red Salanova, were found to be significantly affected by selenate applications; in particular they increased by 184%, 84%, and 31% respectively at Se doses of 16, 24, and 32 μM compared to the control (**Table 4**). Our results are in accordance with Liu et al. (2017), where anthocyanins in red lettuce cv. Purple Rome increased significantly at moderate doses of Se, while they were lower and comparable to the control at higher Se doses. In their study, the authors showed that the Se influence on accumulation and molecular regulation of anthocyanins synthesis was mainly due to the expression levels of the flavanone 3-hydroxylase (F3H) and UDP-glycose flavonoid glycosyl transferase (UFGT) genes that played a key role in anthocyanins biosynthesis. The F3H and UFGT genes were significantly up-regulated by moderate Se treatments compared to the control (Liu et al., 2017).

Carotenoids are essential lipid-soluble pigments that have antioxidant properties and are found in all photosynthetic organisms (Gross, 1991). These compounds play significant roles in the prevention of chronic ailments, such as cancer, cardiovascular disease, diabetes and osteoporosis, owing to their potent antioxidant, immunomodulatory, gap-junction communication, photoprotective, neuroprotective, and vitamin A activity (Saini et al., 2015). Carotenoids are classified into two groups, xanthophylls which include neoxanthin, violaxanthin, lutein, zeaxanthin, and β-cryptoxanthin, and carotenes which include β-carotene, α-carotene, and lycopene. In human diet, neoxanthin, violaxanthin, lutein, and β-carotene are primarily obtained from dark green or red vegetables. Specifically in lettuce, higher carotenoids content has been found in red leaf cultivars compared to green ones (Nicolle et al., 2004). This finding is in agreement with our results where red Salanova had a significantly higher content of all the target carotenoids detected compared to green Salanova. The sum of all detected carotenoids was 133% higher in the red cultivar compared to the green one (**Table 5**). As in the case of phenolic compounds, the content in target carotenoids was affected by both cultivar and Se treatments with significant however cultivar × Se interaction (**Table 5**). In green Salanova, all detected carotenoids decreased in response to selenate applications compared to the control (**Table 5**), whereas in red Salanova this trend was differentiated. violaxanthin + neoxanthin, lutein,

TABLE 5 | Composition of carotenoids profile of green and red Salanova lettuce grown hydroponically in a Fitotron open-gas-exchange growth chamber under six Se concentrations applied in the nutrient solution.

Source of variance	Violaxanthin + neoxanthin (µg violaxanthin eq. g⁻¹ dw)	Lutein (µg g⁻¹ dw)	β-Cryptoxanthin (µg g⁻¹ dw)	β-carotene (µg g⁻¹ dw)
Cultivar (C)				
Green Salanova	507.39 ± 14.1	207.62 ± 8.55	370.60 ± 13.8	165.62 ± 6.53
Red Salanova	993.13 ± 28.8	600.36 ± 15.3	989.43 ± 26.4	337.14 ± 11.8
t-test	***	***	***	***
Selenium (µM Se) (S)				
0	733.14 ± 53.0	421.04 ± 62.6	717.66 ± 107	296.43 ± 37.0
8	633.57 ± 95.3	357.59 ± 81.6	587.32 ± 127	252.25 ± 51.7
16	774.82 ± 117	421.51 ± 101	699.87 ± 165	272.02 ± 57.1
24	762.72 ± 123	385.30 ± 88.9	645.43 ± 138	215.09 ± 29.6
32	850.46 ± 148	461.27 ± 113	784.17 ± 176	239.98 ± 33.2
40	746.85 ± 118	377.20 ± 81.1	645.67 ± 119	232.51 ± 23.2
	***	***	***	***
C × S				
Green Salanova × 0 µM Se	614.93 ± 5.54 d	282.15 ± 3.01 e	478.51 ± 3.85 e	214.60 ± 5.39 e
Green Salanova × 8 µM Se	421.46 ± 7.09 f	175.52 ± 3.87 g	305.07 ± 5.49 h	136.91 ± 2.42 h
Green Salanova × 16 µM Se	513.05 ± 3.29 e	195.75 ± 4.01 fg	331.35 ± 6.79 gh	145.04 ± 3.10 gh
Green Salanova × 24 µM Se	489.24 ± 7.10 e	186.75 ± 2.57 fg	337.96 ± 8.31 gh	149.09 ± 2.93 gh
Green Salanova × 32 µM Se	520.97 ± 4.26 e	209.40 ± 5.19 f	390.71 ± 2.76 f	166.05 ± 4.61 fg
Green Salanova × 40 µM Se	484.69 ± 2.68 e	196.11 ± 3.01 fg	379.99 ± 6.92 fg	182.06 ± 2.73 f
Red Salanova × 0 µM Se	851.34 ± 6.70 c	559.94 ± 17.4 cd	956.81 ± 21.7 c	378.27 ± 10.1 ab
Red Salanova × 8 µM Se	845.68 ± 19.1 c	539.67 ± 10.4 d	869.57 ± 32.3 d	367.60 ± 8.28 b
Red Salanova × 16 µM Se	1,036.59 ± 11.4 b	647.27 ± 15.1 b	1,068.38 ± 25.7 b	399.01 ± 13.9 a
Red Salanova × 24 µM Se	1,036.19 ± 17.1 b	583.85 ± 7.42 c	952.89 ± 8.83 c	281.09 ± 3.13 d
Red Salanova × 32 µM Se	1,179.95 ± 20.8 a	713.14 ± 0.18 a	1,177.62 ± 26.2 a	313.91 ± 4.53 c
Red Salanova × 40 µM Se	1,009.02 ± 26.4 b	558.28 ± 10.9 cd	911.34 ± 16.9 cd	282.95 ± 11.8 d
	***	***	***	***

ns, *** Non-significant or significant at P ≤ 0.001, respectively. In the absence of interaction, cultivar means were compared by t-Test and Se application means by Duncan's multiple-range test (P = 0.05). Different letters within each column indicate significantly different means. All data are expressed as mean ± SE, n = 3.

and β-cryptoxanthin increased in red Salanova with increasing selenate application levels, reaching their highest levels at the 32 µM Se dose, whereas β-carotene in the 24–40 µM Se dose range was on average 23% lower than the control. Regarding the green cultivar, our results are in agreement with what has been found in the literature on lettuce (Hawrylak-Nowak, 2013), rice (D'Amato et al., 2018), and *Arabidopsis* (Sams et al., 2011), where a reduction of the total carotenoids content was observed following the application of sodium selenate. Pertinent to these results is previous work on *Arabidopsis* that has demonstrated that the presence of selenate may down-regulate phytoene synthase, a major enzyme involved in the biosynthesis of carotenoids (Sams et al., 2011). On the other hand, the increase in xanthophylls (violaxanthin + neoxanthin, lutein and β-cryptoxanthin) found in red Salanova in response to Se doses up to the 32 µM could be associated to a dissimilar activation of molecular and physiological mechanisms in this cultivar, which differently influence the biosynthesis and accumulation of secondary metabolites, such as xanthophylls. Moreover, in our experiment, it was noted that the presence of selenate had contrasting effects on various classes of secondary metabolites.

Principal Component Analysis

A comprehensive overview of the nutritional and functional quality profiles determined by ion chromatography and HPLC-DAD on red and green butterhead Salanova lettuce in response to Se concentration in the nutrient solution was obtained through Principal Component Analysis (PCA; **Figure 2**). The principle component (PC1) accounted for 51.1% of the cumulative variance, while PC2, and PC3 explained 23.4 and 8.2%, respectively of the total variance (**Table 6**). PC1 correlated positively to the four target carotenoids, caffeoyl-meso-tartaric and chlorogenic acid, magnesium, and sulphate content. PC1 correlated negatively to agronomical traits (shoot biomass and leaf number), as well as to nitrate, calcium, and potassium content. PC2 positively correlated to fresh yield, chicoric acid, total phenolic acids, and phosphate content; and negatively to leaf dry matter and Se content (**Table 6**). Furthermore, the loading matrix indicated the correlations among the examined quanti-qualitative traits, wherein two variables at an angle < 90° were positively correlated, whereas an angle > 90° designated negatively correlated variables. In our experiment, variation in chlorogenic and anthocyanin contents were most closely aligned with β-carotene content, whereas variation in total phenolics did not correlate to nitrate content (**Figure 2**).

The effectiveness of PCA in interpreting cultivar differences across multiple nutritional and functional quality characters in response to several pre-harvest factors (e.g., nutrient solution management, biofortification, plant biostimulants) has been previously demonstrated (Colonna et al., 2016; Cardarelli et al., 2017; El-Nakhel et al., 2019). This was also the case in our study,
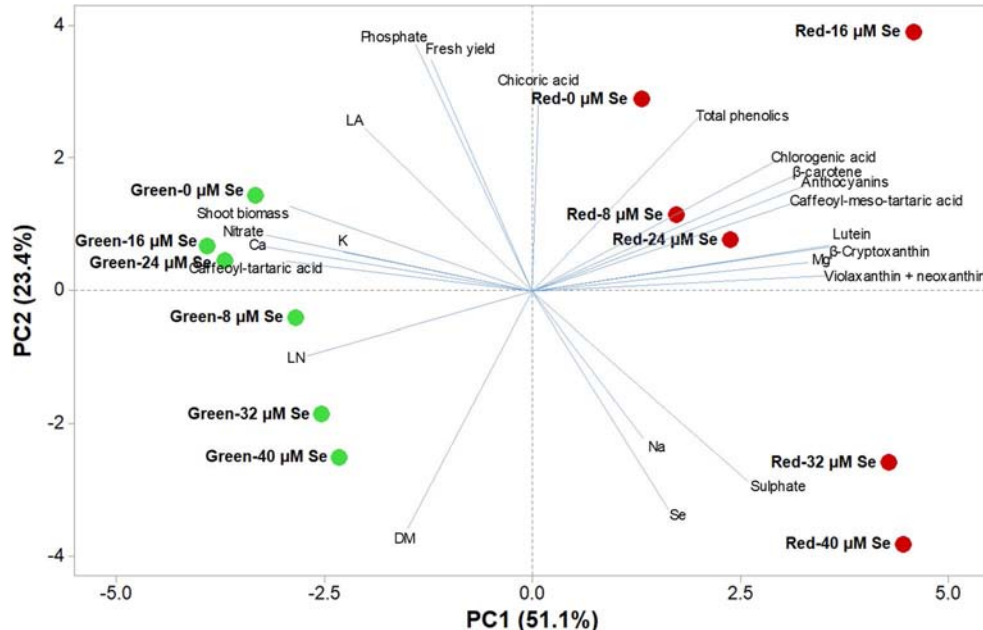
FIGURE 2 | Principal component loading plot and scores of principal component analysis (PCA) of growth parameters (leaf area: LA and leaf number: LN), fresh yield, shoot dry biomass mineral concentrations (Nitrate, phosphate, sulphate, K, Ca, Mg, and Na), lipophilic and hydrophilic antioxidant molecules (target phenolic acids and total phenolics, anthocyanins, ascorbic acid and target carotenoids) in green and red butterhead lettuce Salanova grown under six different concentrations of selenium (Se) added as sodium selenate (0, 8, 16, 24, 32, and 40 µM).

TABLE 6 | Eigen values, relative and cumulative proportion of total variance, and correlation coefficients for growth parameters, mineral profile, nutritional and functional traits of Salanova butterhead lettuce with respect to the three principal components.

Principal components	PC1	PC2	PC3
Eigen value	11.7	5.3	1.8
Percentage of variance	51.1	23.4	8.2
Cumulative variance	51.1	74.5	82.7
Eigen vectors[a]			
Lutein	**0.957**	0.160	0.168
β-Cryptoxanthin	**0.956**	0.156	0.148
Violaxanthin + neoxanthin	**0.954**	0.057	0.240
Mg	**0.889**	0.101	0.363
Anthocyanins	**0.882**	0.370	−0.044
Ca	**−0.858**	0.154	0.236
Caffeoyl-meso-tartaric acid	**0.858**	0.315	−0.113
Nitrate	**−0.855**	0.198	0.362
β-carotene	**0.850**	0.410	−0.049
Caffeoyl-tartaric acid	**−0.790**	0.109	−0.024
Shoot biomass	**-0.781**	0.300	−0.007
Chlorogenic acid	**0.781**	0.452	−0.206
LN	**−0.724**	−0.219	−0.347
Sulphate	**0.697**	**−0.657**	0.108
Phosphate	−0.374	**0.860**	0.187
DM	−0.399	**−0.820**	−0.293
Fresh yield	−0.323	**0.808**	0.216
Se	0.440	**−0.755**	−0.187
Chicoric acid	0.019	**0.672**	−0.374
Total phenolics	0.535	**0.609**	−0.298
LA	−0.540	0.571	−0.218
K	**−0.608**	0.139	**0.676**
Na	0.359	−0.507	0.586

[a]Boldface factor loadings are considered highly weighed.
[b]LN, leaf number; DM, dry matter; LA, leaf area.

since the score plot of the PCA highlighted crucial information on the nutritional and functional quality of the tested butterhead cultivars exposed to different Se concentrations in the nutrient solution. The PCA clearly divided the two tested cultivars along PC1 with red-pigmented lettuce on the positive side and the green one on the negative side. Accordingly, green-pigmented lettuce distinguished for fresh and dry biomass, nitrate and mineral profile (Ca, phosphate and K contents); whereas the red-pigmented cultivar was superior in target lipophilic and hydrophilic antioxidant molecules as well as in total phenolic acids (**Figure 2**). Particularly, the red-pigmented lettuce treated with 8, 16, and 24 µM Se, positioned in the upper right quadrant of the PCA score plot, delivered premium quality and high concentration of hydrophilic and lipophilic antioxidants (**Figure 2**). Red Salanova at the highest two doses of Se was characterized by high content of Se and sulphate. Green butterhead lettuce grown under 0, 16, and 24 µM Se was positioned in the upper left quadrant, characterized overall by higher plant growth parameters (leaf area, fresh yield and shoot dry biomass) and mineral composition (phosphate, K, and Ca). Finally, the lower left quadrant depicted high Se concentration treatments of green lettuce, which yielded the lowest nutritional and functional quality traits of all 12 treatments except from a high percentage of leaf dry matter content (**Figure 2**). The PCA performed in the present study configured an integrated view of yield and quality traits quantitated by ion chromatography and HPLC. It thus enabled the interpretation of variation patterns in these traits with respect to the genetic material and Se biofortification applications studied.

CONCLUSIONS

As demand for functional foods with beneficial effects on human health is rising, selenium biofortification of lettuce facilitated in closed soilless cultivation is presently demonstrated as an effective, low-cost method to produce Se-enriched food of high nutritional value. Our findings indicate that shoot dry biomass, mineral composition, as well as phenolic acids and carotenoids were strongly affected by genotype, with the red cultivar proved to have higher nutritional and functional quality than the green one. Our results demonstrated that the application of 16 μM Se in the nutrient solution improved the phenolic acids content in both cultivars, especially in red Salanova, which was also distinguished by a substantial increase in anthocyanins content (184%). In green Salanova, Se applications slightly reduced the overall carotenoids content, while in the red cultivar 16 and 32 μM Se doses triggered an increase in violaxanthin + neoxanthin, lutein and β-cryptoxanthin. Therefore, we can deduce that the optimal Se dose is 16 μM, as it improves the nutraceutical characteristics in both cultivars with a slight and acceptable reduction in fresh marketable yield (8%) recorded only in green Salanova. Selenium leaf content increased significantly with the sodium selenate application rate in both cultivars. Moreover, the 16 μM treatment yielded sufficient Se leaf content to satisfy 91% and 193% of RDA of this trace element by consuming respectively 50 g fw of green and red Salanova, without any toxic effect to humans, since the amount does not exceed the maximum allowable intake.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

AP wrote the first draft of the manuscript, followed the statistical analysis, and contributed to results data interpretation. CE-N carried out the Fitotron experiment and wrote the first draft of the manuscript. MG and SS performed the mineral analysis and data interpretation. MK and SP were involved in data analysis, data interpretation, and editing the manuscript. YR coordinated the whole project, provided the intellectual input, set up the experiment, and corrected the manuscript.

REFERENCES

Andrade, F. R., da Silva, G. N., Guimarães, K. C., Barreto, H. B. F., de Souza, K. R. D., and Guilherme, L. R. G. (2018). Selenium protects rice plants from water deficit stress. *Ecotoxicol. Environ. Saf.* 164, 562–570. doi: 10.1016/j.ecoenv.2018.08.022

Ardebili, Z. O., Ardebili, N. O., Jalili, S., and Safiallah, S. (2015). The modified qualities of basil plants by selenium and/or ascorbic acid. *Turk. J. Bot.* 39, 401–407. doi: 10.3906/bot-1404-20

Arthur, J. R. (2003). Selenium supplementation: does soil supplementation help and why? *Proc. Nutr. Soc* 62, 393–397. doi: 10.1079/pns2003254

Baslam, M., Morales, F., Garmendia, I., and Goicoechea, N. (2013). Nutritional quality of outer and inner leaves of green and red pigmented lettuces (*Lactuca sativa* L.) consumed as salads. *Sci. Hortic.* 151, 103–111. doi: 10.1016/j.scienta.2012.12.023

Blasco, B., Ríos, J. J., Sánchez-Rodríguez, E., Rubio-Wilhelmi, M. M., Leyva, R., Romero, L., et al. (2012). Study of the interactions between iodine and mineral nutrients in lettuce plants. *J. Plant Nutr.* 35 (13), 1958–1969. doi: 10.1080/01904167.2012.716889

Burk, R. F., Norsworthy, B. K., Hill, K. E., Motley, A. K., and Byrne, D. W. (2006). Effects of chemical form of selenium on plasma biomarkers in a high-dose human supplementation trial. *Cancer Epidemiol. Biomarkers* 15, 804–810. doi: 10.1158/1055-9965.epi-05-0950

Burns, I. G., Zhang, K., Turner, M. K., and Edmondson, R. (2010). Iso-osmotic regulation of nitrate accumulation in lettuce. *J. Plant Nutr.* 34, 283–313. doi: 10.1080/01904167.2011.533328

Burns, I. G., Zhang, K., Turner, M. K., Meacham, M., Al-Redhiman, K., and Lynn, J. (2011). Screening for genotype and environment effects on nitrate accumulation in 24 species of young lettuce. *J. Sci. Food Agr.* 91, 553–562. doi: 10.1002/jsfa.4220

Businelli, D., D'Amato, R., Onofri, A., Tedeschini, E., and Tei, F. (2015). Se-enrichment of cucumber (*Cucumis sativus* L.), lettuce (*Lactuca sativa* L.) and tomato (*Solanum lycopersicum* L. Karst) through fortification in pre-transplanting. *Sci. Hortic.* 197, 697–704. doi: 10.1016/j.scienta.2015.10.039

Cardarelli, M., Rouphael, Y., Pellizzoni, M., Colla, G., and Lucini, L. (2017). Profile of bioactive secondary metabolites and antioxidant capacity of leaf exudates from eighteen Aloe species. *Ind. Crops Prod.* 108, 44–51. doi: 10.1016/j.indcrop.2017.06.017

Chauhan, R., Awasthi, S., Srivastava, S., Dwivedi, S., Pilon-Smits, E. A., and Dhankher, O. P. (2019). Understanding selenium metabolism in plants and its role as a beneficial element. *Crit. Rev. Environ. Sci. Technol.* 49, 1937–1958. doi: 10.1080/10643389.2019.1598240

Chen, L., Yang, F., Xu, J., Hu, Y., Hu, Q., and Zhang, Y. (2002). Determination of selenium concentration of rice in China and effect of fertilization of selenite and selenate on selenium content of rice. *J. Agr. Food Chem.* 50, 5128–5130. doi: 10.1021/jf0201374

Chilimba, A. D., Young, S. D., Black, C. R., Meacham, M. C., Lammel, J., and Broadley, M. R. (2012). Agronomic biofortification of maize with selenium (Se) in Malawi. *Field Crop Res.* 125, 118–128. doi: 10.1016/j.fcr.2011.08.014

Colla, G., Kim, H. J., Kyriacou, M. C., and Rouphael, Y. (2018). Nitrate in fruits and vegetables. *Sci. Hortic.* 237, 221–238. doi: 10.1016/j.scienta.2018.04.01

Colonna, E., Rouphael, Y., Barbieri, G., and De Pascale, S. (2016). Nutritional quality of ten leafy vegetables harvested at two light intensities. *Food Chem.* 199, 702–710. doi: 10.1016/j.foodchem.2015.12.068

Combs, G. F. (2001). Selenium in global food systems. *Br. J. Nutr.* 85, 517–547. doi: 10.1079/bjn2000280

Couto, E., Boffetta, P., Lagiou, P., Ferrari, P., Buckland, G., and Overvad, K. (2011). Mediterranean dietary pattern and cancer risk in the EPIC cohort. *Br. J. Cancer* 104, 1493. doi: 10.1038/bjc.2011.106

Dall'Acqua, S., Ertani, A., Pilon-Smits, E. A., Fabrega-Prats, M., and Schiavon, M. (2019). Selenium biofortification differentially affects sulfur metabolism and accumulation of phytochemicals in two rocket species (*Eruca Sativa* Mill. and *Diplotaxis Tenuifolia*) grown in hydroponics. *Plants* 8, 68. doi: 10.3390/plants8030068

D'Amato, R., Fontanella, M. C., Falcinelli, B., Beone, G. M., Bravi, E., and Marconi, O. (2018). Selenium biofortification in rice (*Oryza sativa* L.) sprouting: Effects on Se yield and nutritional traits with focus on phenolic acid profile. *J. Agric. Food Chem.* 66, 4082–4090. doi: 10.1021/acs.jafc.8b00127

Djanaguiraman, M., Devi, D. D., Shanker, A. K., Sheeba, J. A., and Bangarusamy, U. (2005). Selenium–an antioxidative protectant in soybean during senescence. *Plant Soil* 272, 77–86. doi: 10.1007/s11104-004-4039-1

Dumont, E., Vanhaecke, F., and Cornelis, R. (2006). Selenium speciation from food source to metabolites: a critical review. *Anal. Bioanal. Chem.* 385, 1304–1323. doi: 10.1007/s00216-006-0529-8

Ellis, D. R., and Salt, D. E. (2003). Plants, selenium and human health. *Curr. Opin. Plant Biol.* 6 (3), 273–279. doi: 10.1016/s1369-5266(03)00030-x

El-Nakhel, C., Pannico, A., Kyriacou, M. C., Giordano, M., De Pascale, S., and Rouphael, Y. (2019). Macronutrient deprivation eustress elicits differential secondary metabolites in red and green-pigmented butterhead lettuce grown in closed soilless system. *J. Sci. Food Agric.* doi: 10.1002/jsfa.9985

Esringu, A., Ekinci, M., Turan, M., Usta, S., Dursun, A., and Ercisli, S. (2015). Selenium supplementation affects the growth, yield and selenium accumulation in lettuce (*Lactuca sativa* L.). *Comptes rendus l'Académie bulgare Des. Sci.* 68, 801–810.

Eurola, M., Hietaniemi, V., Kontturi, M., Tuuri, H., Kangas, A., Niskanen, M. et al. (2004). Selenium content of Finnish oats in 1997-1999: effect of cultivars and cultivation techniques. *Agric. Food Sci.* 13, 46-53. doi: 10.2137/1239099041837941

Ferrarese, M., Sourestani, M., Quattrini, E., Schiavi, M., and Ferrante, A. (2012). Biofortification of spinach plants applying selenium in the nutrient solution of floating system. *Vegetable Crops Res. Bull.* 76, 127–136. doi: 10.2478/v10032-012-0009-y

Finley, J. W. (2005). Proposed criteria for assessing the efficacy of cancer reduction by plant foods enriched in carotenoids, glucosinolates, polyphenols and selenocompounds. *Ann. Bot.* 95, 1075–1096. doi: 10.1093/aob/mci123

Gross, J. (1991). *Pigments in vegetables: chlorophylls and carotenoids.* Springer Science Business Media, New York. doi: 10.1007/978-1-4615-2033-7_2

Hamilton, S. J. (2004). Review of selenium toxicity in the aquatic food chain. *Sci. Total Environ.* 326, 1–31. doi: 10.1016/j.scitotenv.2004.01.019

Harborne, J. B., and Williams, C. A. (2001). Anthocyanins and other flavonoids. *Nat. Prod. Rep.* 18, 310–333. doi: 10.1002/chin.199922282

Hartikainen, H. (2005). Biogeochemistry of selenium and its impact on food chain quality and human health. *J. Trace Elem. Med. Bio.* 18, 309–318. doi: 10.1016/j.jtemb.2005.02.009

Hartikainen, H., Xue, T., and Piironen, V. (2000). Selenium as an anti-oxidant and pro-oxidant in ryegrass. *Plant Soil* 225, 193–200. doi: 10.1023/A:102651292

Hawrylak-Nowak, B. (2008). Enhanced selenium content in sweet basil (*Ocimum basilicum* L.) by foliar fertilization. *Vegetable Crops Res. Bull.* 69, 63–72. doi: 10.2478/v10032-008-0021-4

Hawrylak-Nowak, B. (2013). Comparative effects of selenite and selenate on growth and selenium accumulation in lettuce plants under hydroponic conditions. *Plant Growth Regul.* 70, 149–157. doi: 10.1007/s10725-013-9788-5

Hawrylak-Nowak, B., Dresler, S., Rubinowska, K., Matraszek-Gawron, R., Woch, W., and Hasanuzzaman, M. (2018). Selenium biofortification enhances the growth and alters the physiological response of lamb's lettuce grown under high temperature stress. *Plant Physiol. Bioch.* 127, 446–456. doi: 10.1016/j.plaphy.2018.04.018

Hawkesford, M. J., and Zhao, F. J. (2007). Strategies for increasing the selenium content of wheat. *J. Cereal Sci.* 46, 282–292. doi: 10.1016/j.jcs.2007.02.006

Hermosillo-Cereceres, M. A., Sanchez-Chavez, E., Balandran-Quintana, R. R., Mendoza-Wilson, A. M., Guevara-Aguilar, A., and Munoz-Marquez, E. (2011). Ionome variation in bean plant growth under different Se forms and application rates. *J. Food Agric. Environ.* 9, 374–378. doi: 10.1234/4.2011.2287

Jeřše, A., Maršić, N. K., Kroflič, A., Germ, M., Šircelj, H., and Stibilj, V. (2018). Is foliar enrichment of pea plants with iodine and selenium appropriate for production of functional food? *Food Chem.* 267, 368–375. doi: 10.1016/j.foodchem.2018.02.112

Johnson, L. J., Meacham, S. L., and Kruskall, L. J. (2003). The antioxidants-vitamin C, vitamin E, selenium, and carotenoids. *J. Agromedicine* 9, 65–82. doi: 10.1300/j096v09n01_07

Jones, G. D., Droz, B., Greve, P., Gottschalk, P., Poffet, D., and McGrath, S. P. (2017). Selenium deficiency risk predicted to increase under future climate change. *Proc. Natl. Acad. Sci.* 114, 2848–2853. doi: 10.1073/pnas.1611576114

Kim, M. J., Moon, Y., Tou, J. C., Mou, B., and Waterland, N. L. (2016). Nutritional value, bioactive compounds and health benefits of lettuce (*Lactuca sativa* L.). *J. Food Compos. Anal.* 49, 19–34. doi: 10.1016/j.jfca.2016.03.004

Lee, S. J., Kang, H. M., and Kim, I. S. (2008). Effect of sodium selenate supplied condition by fertigation on the growth and content of minerals, ascorbic acid, nitrate, and selenium of some western vegetables. *J. Bio-Environ. Control* 17, 43–50.

Li, H. F., McGrath, S. P., and Zhao, F. J. (2008). Selenium uptake, translocation and speciation in wheat supplied with selenate or selenite. *New Phytol.* 178, 92–102. doi: 10.1111/j.1469-8137.2007.02343.x

Liu, D., Li, H., Wang, Y., Ying, Z., Bian, Z., and Zhu, W. (2017). How exogenous selenium affects anthocyanin accumulation and biosynthesis-related gene expression in purple lettuce. *Pol. J. Environ. Stud.* 26, 2. doi: 10.15244/pjoes/66707

Llorach, R., Martínez-Sánchez, A., Tomás-Barberm, F. A., Gil, M. I., and Ferreres, F. (2008). Characterisation of polyphenols and antioxidant properties of five lettuce varieties and escarole. *Food Chem.* 108, 1028–1038. doi: 10.1016/j.foodchem.2007.11.032

López, A., Javier, G. A., Fenoll, J., Hellín, P., and Flores, P. (2014). Chemical composition and antioxidant capacity of lettuce: comparative study of regular-sized (Romaine) and baby-sized (Little Gem and Mini Romaine) types. *J. Food Compos. Anal.* 33, 39–48. doi: 10.1016/j.jfca.2013.10.001

Lyons, G. H., Genc, Y., Soole, K. M., Stangoulis, J. C. R., Liu, F., and Graham, R. D. (2009). Selenium increases seed production in Brassica. *Plant Soil* 318, 73–80. doi: 10.1007/s11104-008-9818-7

Lyons, G. H., Ortiz-Monasterio, I., Stangoulis, J., and Graham, R. (2005). Selenium concentration in wheat grain: is there sufficient genotypic variation to use in breeding? *Plant Soil* 269, 369–380. doi: 10.1007/s11104-004-0909-9

Lyons, G. H., Stangoulis, J. C., and Graham, R. D. (2004). Exploiting micronutrient interaction to optimize biofortification programs: the case for inclusion of selenium and iodine in the HarvestPlus program. *Nutr. Rev.* 62, 247–252. doi: 10.1111/j.1753-4887.2004.tb00047.x

Malagoli, M., Schiavon, M., Dall'Acqua, S., and Pilon-Smits, E. A. (2015). Effects of selenium biofortification on crop nutritional quality. *Front. Plant Sci.* 6, 280. doi: 10.3389/fpls.2015.00280

Nicolle, C., Carnat, A., Fraisse, D., Lamaison, J. L., Rock, E., and Michel, H. (2004). Characterisation and variation of antioxidant micronutrients in lettuce (*Lactuca sativa folium*). *J. Sci. Food Agr.* 84, 2061–2069. doi: 10.1002/jsfa.1916

Nowak, J., Kaklewski, K., and Ligocki, M. (2004). Influence of selenium on oxidoreductive enzymes activity in soil and in plants. *Soil Biol. Biochem.* 36, 1553–1558. doi: 10.1016/j.soilbio.2004.07.002

Pedrero, Z., and Madrid, Y. (2009). Novel approaches for selenium speciation in foodstuffs and biological specimens: a review. *Anal. Chim. Acta* 634, 135–152. doi: 10.1016/j.aca.2008.12.026

Pedrero, Z., Madrid, Y., and Cámara, C. (2006). Selenium species bioaccessibility in enriched radish (*Raphanus sativus*): a potential dietary source of selenium. *J. Agr. Food Chem.* 54, 2412–2417. doi: 10.1021/jf052500n

Pilon-Smits, E. A., Hwang, S., Lytle, C. M., Zhu, Y., Tai, J. C., and Bravo, R. C. (1999). Overexpression of ATP sulfurylase in Indian mustard leads to increased selenate uptake, reduction, and tolerance. *Plant Physiol.* 119, 123–132. doi: 10.1104/pp.119.1.123

Pilon-Smits, E. A., and Quinn, C. F. (2010). Selenium metabolism in plants, in *Cell biology of metals and nutrients* (Berlin, Heidelberg: Springer), 225–241. doi: 10.1007/978-3-642-10613-2_10

Pilon-Smits, E. A., Quinn, C. F., Tapken, W., Malagoli, M., and Schiavon, M. (2009). Physiological functions of beneficial elements. *Curr. Opin. Plant Biol.* 12, 267–274. doi: 10.1016/j.pbi.2009.04.009

Puccinelli, M., Malorgio, F., Rosellini, I., and Pezzarossa, B. (2017). Uptake and partitioning of selenium in basil (*Ocimum basilicum* L.) plants grown in hydroponics. *Sci. Hortic.* 225, 271–276. doi: 10.1016/j.scienta.2017.07.014

Ramos, S. J., Faquin, V., Guilherme, L. R. G., Castro, E. M., Ávila, F. W., and Carvalho, G. S. (2010). Selenium biofortification and antioxidant activity in lettuce plants fed with selenate and selenite. *Plant Soil Environ.* 56, 584–588. doi: 10.17221/113/2010-pse

Ramos, S. J., Rutzke, M. A., Hayes, R. J., Faquin, V., Guilherme, L. R. G., and Li, L. (2011). Selenium accumulation in lettuce germplasm. *Planta* 233, 649–660. doi: 10.1007/s00425-010-1323-6

Rayman, M. P. (2000). The importance of selenium to human health. *Lancet* 356, 233–241. doi: 10.1016/s0140-6736(00)02490-9

Rayman, M. P. (2002). The argument for increasing selenium intake. *Proc. Nutr. Soc* 61, 203–215. doi: 10.1079/pns2002153

Reis, H. P. G., de Queiroz Barcelos, J. P., Junior, E. F., Santos, E. F., Silva, V. M., and Moraes, M. F. (2018). Agronomic biofortification of upland rice with selenium

and nitrogen and its relation to grain quality. *J. Cereal Sci.* 79, 508–515. doi: 10.1016/j.jcs.2018.01.004

Ríos, J. J., Blasco, B., Cervilla, L. M., Rubio-Wilhelmi, M. M., Rosales, M. A., and Sanchez-Rodriguez, E. (2010b). Nitrogen-use efficiency in relation to different forms and application rates of Se in lettuce plants. *J. Plant Growth Regul.* 29, 164–170. doi: 10.1007/s00344-009-9130-7

Ríos, J. J., Blasco, B., Cervilla, L. M., Rubio-Wilhelmi, M. M., Ruiz, J. M., and Romero, L. (2008). Regulation of sulphur assimilation in lettuce plants in the presence of selenium. *Plant Growth Regul.* 56, 43. doi: 10.1007/s10725-008-9282-7

Rios, J. J., Blasco, B., Leyva, R., Sanchez-Rodriguez, E., Rubio-Wilhelmi, M. M., and Romero, L. (2013). Nutritional balance changes in lettuce plant grown under different doses and forms of selenium. *J. Plant Nutr.* 36, 1344–1354. doi: 10.1080/01904167.2013.790427

Rios, J. J., Blasco, B., Rosales, M. A., Sanchez-Rodriguez, E., Leyva, R., and Cervilla, L. M. (2010a). Response of nitrogen metabolism in lettuce plants subjected to different doses and forms of selenium. *J. Sci. Food Agr.* 90, 1914–1919. doi: 10.1002/jsfa.4032

Rouphael, Y., Colla, G., Giordano, M., El-Nakhel, C., Kyriacou, M. C., and De Pascale, S. (2017b). 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

Rouphael, Y., and Kyriacou, M. C. (2018). Enhancing quality of fresh vegetables through salinity eustress and biofortification applications facilitated by soilless cultivation. *Front. Plant Sci.* 9, 1254. doi: 10.3389/fpls.2018.01254

Rouphael, Y., Kyriacou, M. C., Vitaglione, P., Giordano, M., Pannico, A., and Colantuono, A. (2017a). Genotypic variation in nutritional and antioxidant profile among iceberg lettuce cultivars. *Acta Sci. Pol. Hortoru. Cultus* 16, 37–45. doi: 10.24326/asphc.2017.3.4

Sager, M. (2006). Selenium in agriculture, food, and nutrition. *Pure Appl. Chem.* 78, 111–133. doi: 10.1351/pac200678010111

Saini, R. K., Nile, S. H., and Park, S. W. (2015). Carotenoids from fruits and vegetables: Chemistry, analysis, occurrence, bioavailability and biological activities. *Food Res. Int.* 76, 735–750. doi: 10.1016/j.foodres.2015.07.047

Salt, D. E., Baxter, I., and Lahner, B. (2008). Ionomics and the study of the plant ionome. *Annu. Rev. Plant Biol.* 59, 709–733. doi: 10.1146/annurev.arplant.59.032607.092942

Sams, C. E., Panthee, D. R., Charron, C. S., Kopsell, D. A., and Yuan, J. S. (2011). Selenium regulates gene expression for glucosinolate and carotenoid biosynthesis in *Arabidopsis. J. Am. Soc Hortic. Sci.* 136, 23–34. doi: 10.21273/jashs.136.1.23

Schiavon, M., Berto, C., Malagoli, M., Trentin, A., Sambo, P., and Dall'Acqua, S. (2016). Selenium biofortification in radish enhances nutritional quality *via* accumulation of methyl-selenocysteine and promotion of transcripts and metabolites related to glucosinolates, phenolics, and amino acids. *Front. Plant Sci.* 7, 1371. doi: 10.3389/fpls.2016.01371

Schiavon, M., Dall'Acqua, S., Mietto, A., Pilon-Smits, E. A. H., Sambo, P., and Masi, A. (2013). Selenium fertilization alters the chemical composition and antioxidant constituents of tomato (*Solanum lycopersicon* L.). *J. Agric. Food Chem.* 61, 10542–10554. doi: 10.1021/jf4031822

Silva, E. D. N. D., Aureli, F., D'Amato, M., Raggi, A., Cadore, S., and Cubadda, F. (2017). Selenium bioaccessibility and speciation in selenium-enriched lettuce: Investigation of the selenocompounds liberated after *in vitro* simulated human digestion using two-dimensional HPLC-ICP-MS. *J. Agr. Food Chem.* 65, 3031–3038. doi: 10.1021/acs.jafc.7b01188

Silva, E. D. N. D., Cidade, M., Heerdt, G., Ribessi, R. L., Morgon, N. H., and Cadore, S. (2018a). Effect of selenite and selenate application on mineral composition of lettuce plants cultivated under hydroponic conditions: Nutritional balance overview using a multifaceted study. *J. Brazil. Chem. Soc* 29, 371–379. doi: 10.21577/0103-5053.20170150

Silva, V. M., Boleta, E. H. M., Lanza, M. G. D. B., Lavres, J., Martins, J. T., and Santos, E. F. (2018b). Physiological, biochemical, and ultrastructural characterization of selenium toxicity in cowpea plants. *Envir. Exp. Bot.* 150, 172–182. doi: 10.1016/j.envexpbot.2018.03.020

Silva, V. M., Boleta, E. H. M., Martins, J. T., Mendes dos Santos, F. L., da Rocha Silva, A. C., and Alcock, T. D. (2019). Agronomic biofortification of cowpea with selenium: effects of selenate and selenite applications on selenium and

phytate concentrations in seeds. *J. Sci. Food Agric.* 29, 371-379. doi: 10.1002/jsfa.9872

Smolen, S., Skoczylas, Ł., Rakoczy, R., Ledwozyw-Smolen, I., and Kopec, A. (2016a). Iodine and selenium biofortification of lettuce (*Lactuca sativa* L.) by soil fertilization with various compounds of these elements. *Acta Sci. Pol. Hortic. Cult.* 15, 69-91.

Smoleń, S., Kowalska, I., Czernicka, M., Halka, M., Kęska, K., and Sady, W. (2016b). Iodine and selenium biofortification with additional application of salicylic acid affects yield, selected molecular parameters and chemical composition of lettuce plants (*Lactuca sativa* L. var. capitata). *Front. Plant Sci.* 7, 1553. doi: 10.3389/fpls.2016.01553

Smoleń, S., Kowalska, I., Kováčik, P., Halka, M., and Sady, W. (2019). Biofortification of six varieties of lettuce (*Lactuca sativa* L.) with iodine and selenium in combination with the application of salicylic acid. *Front. Plant Sci.* 10:143. doi: 10.3389/fpls.2019.00143

Smrkolj, P., Osvald, M., Osvald, J., and Stibilj, V. (2007). Selenium uptake and species distribution in selenium-enriched bean (*Phaseolus vulgaris* L.) seeds obtained by two different cultivations. *Eur. Food Res. Technol.* 225, 233–237. doi: 10.1007/s00217-006-0409-7

Smrkolj, P., Pograjc, L., Hlastan-Ribič, C., and Stibilj, V. (2005). Selenium content in selected Slovenian foodstuffs and estimated daily intakes of selenium. *Food Chem.* 90, 691–697. doi: 10.1016/j.foodchem.2004.04.028

Sors, T. G., Ellis, D. R., and Salt, D. E. (2005). Selenium uptake, translocation, assimilation and metabolic fate in plants. *Photosynth. Res.* 86, 373–389. doi: 10.1007/s11120-005-5222-9

Terry, N., Zayed, A. M., De Souza, M. P., and Tarun, A. S. (2000). Selenium in higher plants. *Annu. Rev. Plant Biol.* 51, 401–432. doi: 10.1146/annurev.arplant.51.1.401

Turakainen, M., Hartikainen, H., and Seppänen, M. M. (2004). Effects of selenium treatments on potato (*Solanum tuberosum* L.) growth and concentrations of soluble sugars and starch. *J. Agr. Food Chem.* 52, 5378–5382. doi: 10.1021/jf040077x

Vallverdú-Queralt, A., Oms-Oliu, G., Odriozola-Serrano, I., Lamuela-Raventós, R. M., Martín-Belloso, O., and Elez-Martínez, P. (2013). Metabolite profiling of phenolic and carotenoid contents in tomatoes after moderate-intensity pulsed electric field treatments. *Food Chem.* 136, 199–205. doi: 10.1016/j.foodchem.2012.07.108

Van Hoewyk, D. (2013). A tale of two toxicities: malformed selenoproteins and oxidative stress both contribute to selenium stress in plants. *Ann. Bot.* 112, 965–972. doi: 10.1093/aob/mct163

Vatansever, R., Ozyigit, I. I., and Filiz, E. (2017). Essential and beneficial trace elements in plants, and their transport in roots: a review. *Appl. Biochem. Biotechnol.* 181, 464–482. doi: 10.1007/s12010-016-2224-3

Voogt, W., Holwerda, H. T., and Khodabaks, R. (2010). Biofortification of lettuce (*Lactuca sativa* L.) with iodine: the effect of iodine form and concentration in the nutrient solution on growth, development and iodine uptake of lettuce grown in water culture. *J. Sci. Food Agric.* 90, 906–913. doi: 10.1002/jsfa.3902

Wang, G., Wu, L., Zhang, H., Wu, W., Zhang, M., and Li, X. (2016). Regulation of the phenylpropanoid pathway: a mechanism of selenium tolerance in peanut (*Arachis hypogaea* L.) seedlings. *J. Agr. Food Chem.* 64, 3626–3635. doi: 10.1021/acs.jafc.6b01054

White, P. J., Bowen, H. C., Parmaguru, P., Fritz, M., Spracklen, W. P., and Spiby, R. E. (2004). Interactions between selenium and sulphur nutrition in *Arabidopsis thaliana. J. Exp. Bot.* 55, 1927–1937. doi: 10.1093/jxb/erh192

White, P. J., and Broadley, M. R. (2005). Biofortifying crops with essential mineral elements. *Trends Plant Sci.* 10, 586–593. doi: 10.1016/j.tplants.2005.10.001

Wiesner-Reinhold, M., Schreiner, M., Baldermann, S., Schwarz, D., Hanschen, F. S., and Kipp, A. P. (2017). Mechanisms of selenium enrichment and measurement in brassicaceous vegetables, and their application to human health. *Front. Plant Sci.* 8, 1365. doi: 10.3389/fpls.2017.01365

Wu, L., and Huang, Z. Z. (1992). Selenium assimilation and nutrient element uptake in white clover and tall fescue under the influence of sulphate concentration and selenium tolerance of the plants. *J. Exp. Bot.* 43, 549–555. doi: 10.1093/jxb/43.4.549

Xu, J., Yang, F., Chen, L., Hu, Y., and Hu, Q. (2003). Effect of selenium on increasing the antioxidant activity of tea leaves harvested during the early spring tea producing season. *J. Agr. Food Chem.* 51, 1081–1084. doi: 10.1021/jf020940y

Yang, F., Chen, L., Hu, Q., and Pan, G. (2003). Effect of the application of selenium on selenium content of soybean and its products. *Bio. Trace Elem. Res.* 93, 249–256. doi: 10.1385/bter:93:1-3:249

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.

Metabolic Engineering a Model Oilseed *Camelina sativa* for the Sustainable Production of High-Value Designed Oils

Lixia Yuan[1] and Runzhi Li[2*]

[1] College of Biological Science and Technology, Jinzhong University, Jinzhong, China, [2] Institute of Molecular Agriculture and Bioenergy, Shanxi Agricultural University, Taigu, China

Camelina sativa (L.) Crantz is an important Brassicaceae oil crop with a number of excellent agronomic traits including low water and fertilizer input, strong adaptation and resistance. Furthermore, its short life cycle and easy genetic transformation, combined with available data of genome and other "-omics" have enabled camelina as a model oil plant to study lipid metabolism regulation and genetic improvement. Particularly, camelina is capable of rapid metabolic engineering to synthesize and accumulate high levels of unusual fatty acids and modified oils in seeds, which are more stable and environmentally friendly. Such engineered camelina oils have been increasingly used as the super resource for edible oil, health-promoting food and medicine, biofuel oil and high-valued chemical production. In this review, we mainly highlight the latest advance in metabolic engineering towards the predictive manipulation of metabolism for commercial production of desirable bio-based products using camelina as an ideal platform. Moreover, we deeply analysis camelina seed metabolic engineering strategy and its promising achievements by describing the metabolic assembly of biosynthesis pathways for acetyl glycerides, hydroxylated fatty acids, medium-chain fatty acids, ω-3 long-chain polyunsaturated fatty acids, palmitoleic acid (ω-7) and other high-value oils. Future prospects are discussed, with a focus on the cutting-edge techniques in camelina such as genome editing application, fine directed manipulation of metabolism and future outlook for camelina industry development.

Keywords: *Camelina sativa* (L.) Crantz, model oilseed, metabolic engineering, fatty acids, designed oil

INTRODUCTION

Plant seed oils enriched in triacylglycerols (TAGs) consisting of three fatty acids esterified to a glycerol backbone are energy-dense molecules that are utilized for energy production in the life cycle of plants (Athenstaedt and Daum, 2006). More importantly, they not only provide the nutritional requirements of humans and animals but also serve as a renewable chemical feedstock for biofuels and various industrial applications (van Erp et al., 2011). Although global production of vegetable oils has increased in recent decades, a wider gap between the production and consumption still

exists. To meet the ever-growing market demands for vegetable oils, it is much needed to genetically improve seed oil yield and quality from oil crops.

The fatty acid profile and their distribution in TAGs of plant oils determines oil quality, physicochemical properties, and uses. TAGs from commercially grown oilseed crops typically contain mainly five fatty acids including palmitic (16:0), stearic (18:0), oleic (18:1Δ9), linoleic (18:1Δ9,12), and a-linolenic (18:1Δ9,12,15) acids. In contrast, a wide variety of fatty acids with different chain lengths and functional groups were found to be highly accumulated in seeds of many uncultivated plant species (Badami and Patil, 1980). TAG containing modified fatty acids with functionality beyond those found in commercially-grown oil seed crops represents a valued resource for bio-based materials and other diverse uses (Dyer et al., 2008).

Over the years, an increased understanding has been made on plant lipid metabolism and its regulation, coupled by well characterization of fatty acid biosynthesis, modification and assembly into TAGs (Cahoon et al., 2002; Kagale et al., 2014). With this information and the wealth of genetic diversity for synthesis of novel fatty acids and storage oils, plant seeds have been developing as platform for the design and tailoring of biochemical pathways to synthesize diverse nutritional and industrial oils not currently found in oilseed crops (Haslam et al., 2016). Various genetic modification tools have been developed including gene editing and synthetic biology techniques, which allow to rapidly assembly novel pathways in oilseed crops for commercially producing high levels of designed lipids/oils and high-valued compounds.

Until recently, much of this work was made in metabolically engineering the model plant *Arabidopsis thaliana* for the production of various modified fatty acids (Arondel et al., 1992; Farmer et al., 1998; Mu et al., 2008; Bates and Browse, 2011). However, this model plant has poor agronomic traits such as small seed yield and unable large-scale field cultivation, which has limited the functional testing of the modified oil. In contrast, *Camelina sativa* (L.) Crantz, an important oilseed crop in the family Brassicaceae, possesses a number of valuable agronomic traits that recommend it as both a new model system and an ideal crop platform for lipid metabolic engineering (Zubr, 1997; Kagale et al., 2014; Ruiz-Lopez et al., 2015; Bansal and Durrett, 2016; Malik et al., 2018). Camelina has a relatively short life cycle, low water and fertilizer requirements. Camelina seed yield is comparable to other oil seed crops, particularly under stress conditions. It's simple, effective transformation system, combined with the availability of abundant transcriptomic and genomic data, has allowed the generation of engineered camelina lines capable of synthesizing high levels of novel oils or UFAs, further enabling subsequent field testing of such traits at a large scale (Lu et al., 2011; Nguyen et al., 2013; Ruiz-Lopez et al., 2014; Malik et al., 2018).

This review was conducted to investigate why camelina is particularly attractive as an ideal model oilseed for metabolic engineering and a platform for commercial production of high-valued bioproducts. We will briefly overview advances in the metabolic engineering of unusual lipids or novel oils in this oil seed crop, combined with author group's work on camelina functional genomics and genetic improvement (Li et al., 2010; Wu et al., 2012; Yuan et al., 2017a; Yuan et al., 2017b). Main description in the selected examples focus on the pathway reconstruction for high synthesis and accumulation of acetyl triacylglycerols, hydroxylated fatty acids, medium-chain fatty acids, ω-3 long-chain polyunsaturated fatty acids, ω-7 monounsaturated fatty acids, and other novel lipids having beneficial functional groups or properties. Moreover, we discuss the cutting-edge research directions in camelina such as genome editing application, a flexible and useful substrate for applied synthetic biology, and future outlook for camelina industry development.

AN IDEAL MODEL OILSEED FOR LIPID METABOLIC ENGINEERING

Camelina has been identified as a promising new crop for oil production due to its several excellent characteristics of low requirements, a short crop cycle (80–100 days), high disease-pest resistance and stress tolerance (Zubr, 1997; Zanetti et al., 2013). In terms of performance, camelina has high yield in favorable environments, and camelina seeds accumulate high levels of oil (40%) and protein (30%) compared to other Brassicaceae crops (Vollmann and Eynck, 2015). Particularly, in camelina oil, UFAs make up 90%, including 40% of a-linolenic acid (ω-3), 25% of linoleic acid, 15% of oleic acid, and 15% of eicosenoic acid. This desirable fatty acid composition enables camelina to be developed as nutritionally enhanced oils. With several agronomic advantages, camelina could be easily developed for commercial production of vegetable oil as much healthy food and a renewable resource for green manufacture of high-quality biofuels (Lu and Kang, 2008).

Meanwhile, camelina has been considered as a platform for the production of specific oils (Collins-Silva et al., 2011; Bansal and Durrett, 2016; Haslam et al., 2016). Camelina shares many (> 90%) of the genes involved in lipid metabolism with the genetic model plant Arabidopsis (Nguyen et al., 2013; Kagale et al., 2014). The rational design of lipid pathways in camelina had made more reliable with the reference genome in 2014 (Kagale et al., 2014). Moreover, the expanded lipid gene family in *C. sativa* provides greater diversity in enzyme expression and substrate specificity. Many of the shortcomings associated with model species, discussed later, can be overcome with camelina, as it has the ability to be both an experimental model system and recognized oilseed crop. Currently, camelina is getting the rising interest and increasing expanding of cultivation across the world (Gugel and Falk, 2006; Mcvay and Khan, 2011; Guy et al., 2014; Malik et al., 2018). A number of lipid metabolic engineering in seeds of *C. sativa* in recent years were summarized in **Table 1**.

TABLE 1 | Lipid metabolic engineering in seeds of *Camelina sativa* in recent years.

Target lipid	Target gene	Manipulation	Phenotype	Reference
Cyclopropane fatty acids (CPAs)	Lychee phosphatidylcholine: diacylglycerol cholinephosphotransferase (LcPDCT) and *Escherichia coli* cyclopropane synthase (EcCPS)	Co-overexpression	50% increase of CPAs	Yu et al. (2019)
Medium-chain FA (MCFA)-containing acetyl-TAGs (MCFA-AcTAGs)	EaDAcT, ChFatB2, cpFatB2, UcFatB1, CnLPAAT, CsDGAT1, and CsPDAT1	Co-expression of EaDAcT with one or two of ChFatB2, cpFatB2, UcFatB1, CnLPAAT, plus combination with CsDGAT1-RNAi and CsPDAT1-RNAi	Significant increased levels of MCFA-AcTAGs, MCFA-AcTAGs was up to 77% more in the best lines	Bansal et al. (2018)
Acetyl-TAGs (AcTAGs)	EaDAcT, and CsDGAT1	Co-expression of EaDAcT together with RNAi suppression of CsDGAT1	AcTAGs with a 2-fold reduction in very long chain fatty acids was up to 85 mol % in the field-grown transgenic line	Liu et al. (2015a)
Saturated FAs, Oleic acid	Fatty acyl-ACP thioesterases (FATB)	Artificial microRNA mediated CsFATB gene suppression (amiFATB)	35% reduction of total saturated FAs and an increase of oleic acid in seed oil	Ozseyhan et al. (2018)
Hydroxy fatty acids (HFAs)	Lesquerella (*Physaria fendleri*) fatty acid elongase (LfKCS3) and castor fatty acid hydroxylase (RcFAH12)	Co-expression of LfKCS3 and RcFAH12	Increased HFAs higher than that in the transgenics expressing RcFAH12 alone	Snapp et al. (2014)
C(20-C24 very long chain FAs (VLCFAs,), C18 unsaturated FAs	Fatty acid elongase1 (FAE1)	Knocking out three CsFAE1 alleles by CRISPR technology with an egg cell-specific Cas9 expression	Reduction of VLCFAs from over 22% to less than 2% of total FAs and concomitant increase of C18 unsaturated FAs	Ozseyhan et al. (2018)
Hydroxy fatty acids (HFAs)	Phospholipase C-like protein (RcPLCL1) and fatty acid hydroxylase (RcFAH12) from *Ricinus communis*	Seed-specific coexpression of RcPLCL1 and RcFAH12	HFAs was increased to 24% of total FAs	Aryal and Lu (2018)
Oleic acid	Fatty acid desaturase 2 (FAD2)	Knockout of all three CsFAD2s by CRISPR/Cas9	Oleic acid level was increased from 16% to >50% and total MUSFA (18:1, (20:1, 22:1) increased from 32% to >70% with concurrent decreases in linoleic and linolenic acids	Jiang et al. (2017)
Oleic acid	Fatty acid desaturase2 (FAD2)	Knockout of three, two, and an isologous CsFAD2 by CRISPR/Cas9	Oleic acid level was increased from 10% to 62% of total FAs in different allelic combinations	Morineau et al. (2017)
Linolenic(18:3) and eicosenoic acids (20:1)	Two Arabidopsis phospholipase Dζ genes (AtPLDζ1 and AtPLDζ2)	Co-expression of AtPLDζ1 and AtPLDζ2	TAG was increased by 2% to 3% more compared to wild type. Increase of 18:3 and (20:1 FAs was concurrent with decrease of other FAs	Yang et al. (2017)
Oil and seed yields	Arabidopsis diacylglycerol acyltransferase1 (AtDGAT1) and a yeast cytosolic glycerol-3-phosphate dehydrogenase (ScGPD1)	Seed-specific coexpression of AtDGAT and ScGPD1	The transgenic seeds showed up to 13% higher seed oil content and up to 52% increase in seed mass, with decreased 18:1 level	Chhikara et al. (2018)
Linolenic(18:3) and linoleic acid (18:2)	CsPDAT1 and CsDAT1	Overexpression of CsPDAT; silencing of CsDGAT1 by amiRNA	Levels of 18:3 was up to 56 mol% of total FAs in CsDGAT1-silenced lines; Levels of 18:2 was increased with concurrent decrease of 18:3 in the CsPDAT-overexpressed lines.	Marmon et al. (2017)
α-linolenic acid, linoleic acid	microRNA167A (miR167A), camelina fatty acid desaturase3 (CsFAD3) and, auxin response factor 8 (CsARF8)	Seed-specific expression of miR167 which suppresses CsARF8 and then mediates transcriptional cascade for CsFAD3 suppression via the ABI3-bZIP67 pathway	Decrease of α-linolenic acid and concomitant increase of linoleic acid , and also increased seed size	Na et al. (2018)
ω-3 long chain PUFAs, eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA)	Δ6-desaturase from *Ostreococcus. tauri* (OtΔ6), Δ6 fatty acid elongase from *Physcomitrella patens* (PSE1), Δ5-desaturase from *Thraustochytrium sp*. (TcΔ5), Δ12-desaturase from *Phytophthora sojae* (PsΔ12), ω3-desaturase from *Phytophthora infestans* (Piω3) and *Hyaloperonospora parasitica* (Hpω3), *O. tauri* Δ5 fatty acid elongase (OtElo5) and Δ4-desaturase gene from *Emiliania huxleyi* (EhΔ4)	Assembly of EPA and DHA pathway by seed-specific expression 5 genes (PSE1, TcΔ5, OtΔ6, Hpω3, and PsΔ12) and 7 genes (PSE1, TcΔ5, OtΔ6, Piω3, PsΔ12, OtElo5 and EhΔ4), respectively	EPA content was 16.2-17.2 % of total FAs in the 5-gene transgenics; EPA and DHA levels were 4.3-4.9% and 4.0% of total FAs in the 7-gene transgenics, respectively	Ruiz-Lopez et al. (2014); Usher et al. (2017)

(Continued)

Target lipid	Target gene	Manipulation	Phenotype	Reference
ω-3 long chain PUFAs, EPA and eicosatetraenoic acid (ETA)	Δ9-elongase from *E. huxleyi* (EhElo9), Δ9-elongase from *I. galbana* (IsoElo9), Δ8-desaturase from *P. Salina* (PsΔ8), Δ5-desaturase from *E. huxleyi* (EhΔ5), ω3-desaturase from *P. infestans* (Piω3), and Δ12-desaturase from *P. sojae* (PsΔ12)	Assembly of Δ9-alternative pathway for ω-3 long chain PUFAs biosynthesis by seed-specific expression 6 genes (EhElo9, IsoElo9, PsΔ8, EhΔ5, Piω3 and PsΔ12)	Mean EPA levels in T3 seeds increased on average from 4.7 to 8.8%, with 16.9% of total FAs in the best line. Mean ETA content was 5.8-7.1%.	Ruiz-Lopez et al. (2015)
Medium-chain, saturated fatty acids (MCSFAs)	12:0-acyl-carrier thioesterase (UcFATB1) from *Umbellularia californica* and CsKASII	Seed-specific expression of UcFATB1 alone or together with CsKASII RNAi	Level of laurate (12:0) and myristate (14:0) were up to 40% of the seed oil in UcFATB1 transgenics; Level of 16:0 was increased from 7.5% up to 28.5% of total FAs in KASII-RNAi lines. Level of MCSFAs (12:0, 14:0 and 16:0) was up to 30% of the co-transformed seeds	Hu et al. (2017)
Medium-chain fatty acids (MCFAs; 8:0-14:0)	CpuFatB3 and CpuFatB4 from *C. pulcherrima*, CvFatB1 and CvFatB3 from *C. viscosissima*, CpFatB2 from *C. palustris*, UcFatB1 from California bay, and ChFatB2 from *C. hookeriana;* a coconut lysophosphatidic acid acyltransferase (CnLPAT)	Seed-specific expression of each of those FatBs; Co-expression of two or three of those FatBs; Co-expression of FatB and CnLPAT	MCFAs were greatly accumulated in the individual FatB transgenics and much high levels occurred in coexpression lines	Kim et al. (2015)
Novel wax esters (WEs)	Acyl-ACP thioesterases (Thio), a fatty acid hydroxylase (FAH), *Marinobacter hydrocarbonoclasticus* WS (wax synthase), *Marinobacter aquaeolei* FAR (fatty acid reductase)	Co-expression of WS and FAR; Co-expression of WS, FAR and Thio or FAH	WEs with different compositions were produced in the transgenic seeds of FAR, WS and Thio; WEs with reduced chain lengths were produced in coexpression lines of MaFAR, MhWS, and Thio	Ruiz-Lopez et al. (2017)
Oleyl oleate wax esters (Oleyl oleate WEs)	*Marinobacter aquaeolei* FAR (Ma FAR) and *Simmondsia chinensis* WS (ScWS)	Co-expression of ScWS and MaFAR	Levels of oleyl oleate WEs were 21% of the seed oil TAGs in the coexpression lines of MaFAR and ScWS	Iven et al. (2016)
Nervonic acid (24:1Δ15)	β-ketoacyl-CoA synthase (LaKCS) from *Lunaria annua*, Arabidopsis β-ketoacyl-CoA reductase (AtKCR) and β-hydroxyacyl-CoA dehydratase (AtHCD)	Seed-specific expressing of LaKCS (*Vector1*); Co-expression of LaKCS, AtKCR and AtHCD (*Vector2*)	Nervonic acid in seed oil increased from null to 6-12% in *Vector1* transgenics; Nervonic acid level from *Vector2*-expressing line was significantly high in early seed development	Huai et al. (2015)
Seed oil	A patatin-related phospholipase AIIIδ (pPLAIIIδ) from Arabidopsis	Constitutive or seed-specific expression of pPLAIIIδ (*Vector1* or *Vector2*)	Seed oil increased greatly and cellulose reduced in *Vector1* transgenics; A increased seed oil content without negative effects in *Vector2* transgenics	Li et al. (2015)
ω-7 FAs (16:1Δ9, 18:1Δ11, 20:1Δ13)	A mutant Δ9-acyl-ACP (Δ9DES1), a 16:0-specific acyl-CoA desaturase (Δ9DES2), 3-keto-acyl-ACP synthase II (KASII) and FatB 16:0-ACP thioesterase (FatB)	Seed-specific co-expression of Δ9DES1 and Δ9DES2 (*Vector1*); seed-specific suppression of CsKASII and CsFatB (*Vector2*)	ω-7 FAs was increased to 17% in *Vector1*-expression lines; ω-7 FAs was increased to 60-65% of total FAs in *Vector1* and *2* coexpression lines	Nguyen et al. (2015)
DHA	Δ;5-desaturase (Δ5D), Δ6-elongase (Δ5E), Δ6D, Δ6E, Δ4D, Δ12D , and ω3-desaturase (ω3D)	Overexpression of the 7 gens(Δ4D, Δ5D, Δ5E, Δ6D, Δ6E, Δ12D and ω3D)	levels of DHA was up to > 12% with very high ω3: ω6 ratios in transgenic seeds	Petrie et al. (2014)
Cuticular wax	Arabidopsis MYB96 (AtMYB96)	Overexpression of AtMYB96 driven by CaMV35S promoter	Significant increase of epicuticular wax crystals and total wax loads on the surfaces of transgenic leaves	Lee et al. (2014)

SELECTED EXAMPLES OF METABOLIC ENGINEERING FOR PRODUCTION OF THE DESIGNED OILS IN *C. SATIVA*

Redesigning Acetyl Triacylglycerol (acetyl-TAG, acTAG) Synthesis for the Production of Superior Biodiesel and Lubricant Oils

AcTAGs (3-acetyl-1,2-diacyl-*sn*-glycerols) are unique and valuable triacylglycerols. Their molecular characteristics include two long-chain fatty acid acyl groups bound to the *sn*-1 and *sn*-2 positions of the glycerol molecule, respectively, and one acetyl group linked to the *sn*-3 position of the glycerol molecule (Durrett et al., 2010). Unlike ordinary TAGs (lcTAG) (three long-chain fatty acid acyl groups were bond on the three carbon atoms of the glycerol molecule, respectively), acTAGs have unique physical and chemical properties, showing their utility in a variety of applications. For example, their kinematic viscosity is 40% lower than that of ordinary TAGs. Therefore, acTAGs are excellent oils for use in the production of low-viscosity biofuels. The biodiesel produced can be used as a premium fuel for ships, trains, and generators. Another advantage of acTAGs is their low temperature tolerance. The biodiesel made from them is less prone to agglomeration and

burning defects in low-temperature climates. AcTAGs can also be used to produce high-quality biodegradable lubricants (Liu et al., 2015a).

These high-value industrial oils cannot be synthesized in ordinary field oilseed crops, but in some wild plants, they are synthesized at high levels. For example, 98% of the seed oil of *Euonymus alatus* (burning bush) is acTAGs (Milcamps et al., 2005). Studies have shown that EaDAcT (*E. alatus* diacylglycerol acetyltransferase) catalyses the formation of acTAGs by the binding of a diacylglycerol (DAG) to an acetyl group at the *sn*-3 position. In the developing seeds of common oil crops, DGAT (diacylglycerol acyltransferase) or PDAT (phosphatidylcholine: diacylglycerol acyltransferase) catalyses DAG binding to a long-chain fatty acid acyl group to generate "regular" long-chain TAG (lcTAG) but does not catalyse the formation of acTAGs. The gene encoding EaDAcT was overexpressed in Camellia seeds, resulting in the seed oil containing up to 55% acTAGs. EaDAcT and DGAT or PDAT share the same substrate, DAG. If the activity of endogenous DGAT or PDAT can be silenced, more DAG can be used to generate acTAGs for EaDAcT. Silencing the expression of three *DGAT1* genes by RNAi suppression in *C. sativa* and overexpressing *EaDAcT* resulted in increasing acTAGs in the seed oil of transgenic *C. sativa* to 85%, and it was stably inherited (Liu et al., 2015a). Continuous field trials showed that compared with the wild-type, the seed weight and oil and protein content of this transgenic *C. sativa* were not significantly different, and the seed germination was normal. During seed germination, AcTAGs can also be degraded like regular lcTAGs for seedling growth (Liu et al., 2015b).

In order to further improve the physicochemical properties of acTAGs and expand their industrial applications, some studies have attempted to replace the long-chain polyunsaturated fatty acid acyl groups at the *sn*-1 and *sn*-2 positions of the acTAG molecule with other long-chain fatty acid acyl groups. The combination of monounsaturated oleic acid (18:1) acyl groups at the *sn*-1 and *sn*-2 positions of acTAGs improves the oxidation resistance of acTAGs. The oleic acid-rich strains of Camelina obtained by RNAi silencing of *GsFAD2* and were used as recipients, and vectors to over-express *EaDAcT* and RNAi-silence *CsDGAT* and *CsPDAT* were transferred to the receptor, resulting in an acTAG content in the seed oil of transgenic *C. sativa* as high as 70%, particularly, the content of 3-acetyl-1,2-dioleoyl-*sn*-glycerol was 47% (Liu et al., 2015a). Such oil showed a significant increase in the oxidation resistance. In addition, the middle chain fatty acid acyl group is bonded to the *sn*-1 and *sn*-2 positions of the acTAG molecule to further reduce the viscosity of the acTAG lipids (Durrett et al., 2010). This shows that changing the fatty acid acyl species at the *sn*-1 and *sn*-2 positions gives the acTAGs more desirable properties and broader industrial applications. Future efforts are needed to design a pathway to incorporate acetate into *sn*-1 and *sn*-2 positions of glycerol backbone (**Figure 1**).

Additionally, the introduction of acTAG into edible oilseed crops may provide an opportunity to develop reduced-calorie fats and oils with a molecular structure similar to that of existing commercial products such as SALATRIM (short and long acyl triglyceride molecule).

Redesigning Hydroxylated Fatty Acid Biosynthesis for the Production of Oxidation Resistance Oils

The oxidation level of vegetable oils depends on the fatty acid composition. If the content of polyunsaturated fatty acid is high, the oxidizability of the vegetable oils is high. But high contents of saturated fatty acids, although more resistance to oxidation, reduce the fluidity of the oil, making it easily solidified. Vegetable oils containing high level of monounsaturated fatty
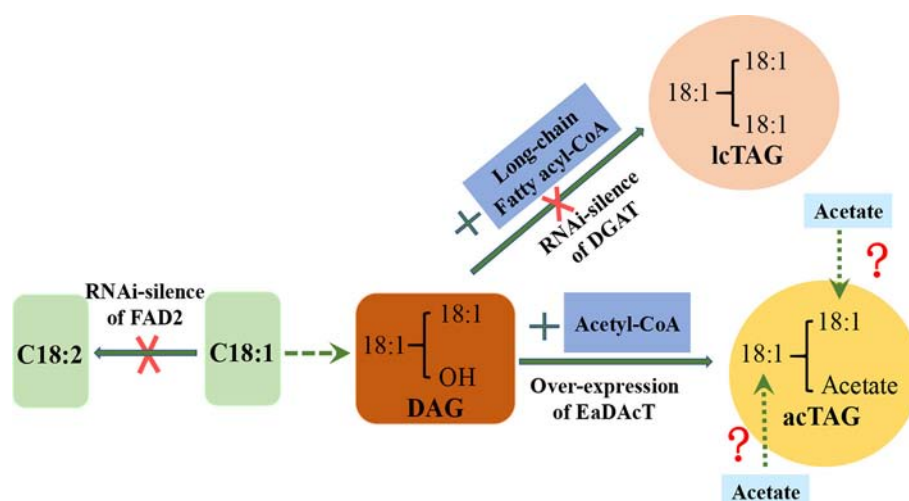

FIGURE 1 | Strategy for redesigning biosynthesis of acetyl-TAG in *C. sativa* seeds. DAG, Diacylglycerol; lcTAG, Long-chain Triacylglycerol; acTAG, Acetyl Triacylglycerol; DGAT, Diacylglycerol acyltransferases; EaDAcT, *Euonymus alatus* diacylglycerol acetyl-CoA transferase.

acids (e.g. oleic acid,18:1Δ9) have high oxidation resistance and other good properties.

Camelina seed oil consists of about 45% polyunsaturated fatty acids, namely linoleic acid (18:2Δ9,12) and linolenic acid (18:1Δ9,12,15), while only 17% of seed oil is monounsaturated oleic acid (18:1Δ9). This seed oil containing high levels of polyunsaturated fatty acids is easily oxidized. RNAi technology was used to specifically silence the *FAD2* (Fatty acid desaturase 2), *FAD3* (Fatty acid desaturase 3), and *FAE1* (Fatty acid elongase 1) genes in camelina, obtaining a transgenic camelina seed oil with significantly reduced content of polyunsaturated fatty acids and a high level of oleic acid (from 18% in wild-type to 65% in the transgenic seeds). Such camelina oil exhibited significantly improved oxidation resistance (Cahoon et al., 2007). In addition, another strategy is to express an enzyme that catalyses the formation of monounsaturated hydroxylated fatty acids in the seeds of camelina, promoting the biosynthesis and accumulation of high levels of hydroxylated fatty acids. Because this fatty acid is extremely resistant to oxidation, the inoxidizability of the seed oil can be increased. Hydroxylated fatty acids have been widely used in the industrial production of resins, waxes, nylons, plastics, lubricants, and cosmetics.

Castor bean (*Ricinus communis*) seed oil consist of up to 90% ricinoleic acid (18:1Δ9,12OH) (a type of monounsaturated hydroxylated fatty acid, HFA). A fatty acid hydroxylase FAH12 was found to catalyze the oleic acid molecule (18:1Δ9) bound to PC to generate a hydroxyl group at the Δ12 carbon atom. Seed-specific expression of the *RcFAH12* gene from *R. communis* resulted in ricinoleic acid in camelina seeds reaching up to 6% (Lu and Kang, 2008). Correspondingly, the oxidation resistance is significantly higher than that of non-transgenic camelina oil. Lesquerelic acid is another hydroxylated fatty acid similar to ricinoleic acid and accumulates in high levels in plant seeds of the *Physaria* genus and *Cruciferae* family. A fatty acid condensing enzyme, LfKCS derived from *Physaria fendleri,* specifically catalyses the elongation of ricinoleic acid to hydroxyalkanoic acid. The synergistic expression of *RcFAH12* and *LfKCS* in camelina seeds not only increased ricinoleic acid from 14% to 19% but also resulted in hydroxyarsenoic acid reaching 8%. The total amount of hydroxylated fatty acids reached 27%. The oxidation resistance of this camelina oil was greatly improved, and seed vigor was not affected (Snapp et al., 2014). Clearly, LfKCS expression accelerates the removal of hydroxylated fatty acids from the phosphatidylcholine (PC) synthesis pool into the acyl-CoA pool to finally form TAGs. A phospholipase C-like protein (RcPLCL1) from castor bean was identified to have hydrolyzing activities on both PC and phosphatidylinositol (PI) substrates (Aryal and Lu, 2018). Co-expression of *RcPLCL1* and *RcFAH12* resulted in accumulation of HFAs up to 24% of total FAs in *C. sativa* seeds (Aryal and Lu, 2018) with less detrimental effect on seed germination, showing that RcPLCL1 can promote the transfer of RcFAH12-formed HFAs on PC into DAG to generate TAGs containing HFAs.

With the Arabidopsis model plant as a receptor, the co-expression of *R. communis RcFAH12* with *RcDGAT2* and *RcPDAT* which controls the final acylation reaction of TAG synthesis can further increase the hydroxylated fatty acid content in seed oil up to 29% (Burgal et al., 2008; van Erp et al., 2011). Co-expression of *RcFAH12* and *RcPDCT* (phosphatidylcholine: diacylglycerol choline phosphotransferase) increased the HFA accumulation in Arabidopsis seeds from approximately 10% to 20% (Hu et al., 2012). Unlike conventional DGAT, PDAT, and PDCT, R. communis homologs are specific for hydroxylated fatty acid substrates. These castor enzymes can accelerate the transfer of HFAs from the PC pool and CoA pool into DAG to form hydroxylated TAGs. It is hypothesized that the co-expression of these three enzyme genes with *RcFAH12* in camelina seed will allow the amount of hydroxylated TAG to accumulate to levels appropriate for commercial use. Since fatty acid thioesterase A (FatA) and fatty acid thioesterase B (FatB) in camelina plastid were identified to be specific for oleoyl-ACP and palmitoleic acid-ACP, respectively (Rodriguez-Rodriguez et al., 2014), in the future, the molecular manipulation of these two enzymes as targets will allow cells to selectively accumulate oleic acid (18:1Δ9) or palmitoleic acid (16:1Δ9) in TAGs and increase the resistance to oxidation of camelina seed oil.

Assembling Medium-Chain Fatty Acid Biosynthesis for the Production of High-Quality Jet Oils

Jet fuels (Jet A and Jet-A1 fuels) consist mainly of C8-C16 alkanes and aromatic hydrocarbons (Kallio et al., 2014). The main fatty acids of most common oilseeds, such as camelina, are 18C fatty acids, which are directly used to process aviation fuels with poor quality and lengthy processes. Vegetable oils rich in caprylic acid (8:0), capric acid (10:0), lauric acid (12:0), myristic acid (14:0) are excellent resources for the production of aviation biofuels (Dyer et al., 2008). Medium-chain fatty acids (MCFAs) are also widely used in the production of detergents, soaps, cosmetics, surfactants, and lubricants.

The palm kernel of the tropical palm plant (*Elaeis guineensis* Jacq.) and the coconut meat of coconut (*Cocos nucifera* L.) are rich in lauric acid (46% to 52%) and decanoic acid (16% to 19%), serving as the main sources of commercial MCFAs. Some of the Lythraceae Cuphea plants can be enriched with >90% of MCFAs. For example, the seed of *Cuphea viscosissima* contains about 25% caprylic acid and about 64% capric acid. The seed of *Cuphea pulcherrima* is enriched with about 95% caprylic acid (Knothe, 2014). These plants have indeterminate agronomic traits such as infinity of inflorescence, shattering, and seed dormancy, being difficult to use in the commercial production of seed oil. However, they can be used as an excellent gene source for MCFA biosynthesis (Filichkin et al., 2006).

The *de novo* synthesis of plant fatty acids occurs in the plastids. Acetyl-CoA and malonyl-ACP are catalyzed by β-ketoacyl-ACP synthase III (KASIII) to form 4C β-ketoacyl-ACP. Then, under the action of the fatty acid synthase (FAS) complex, each cycle adds 2 carbon atoms until 16-carbon palmitate-ACP (16:0-ACP) is formed. 16:0-ACP can be further extended to stearic acid-ACP (18:0-ACP) under the action of KASII. It is also possible to dissociate palmitate from ACP and terminate the elongation of the fatty acid carbon chain under the

catalysis of acyl-ACP thioesterase FatB. The 18C fatty acids synthesized in the plastids, namely stearic acid (18:0) and oleic acid (18:1), are catalyzed by FatA thioesterase to dissociate from ACP (**Figure 2**). Therefore, acyl-ACP thioesterases are the major determinants of the synthesis of fatty acid carbon chain lengths in the plastids (Li-Beisson et al., 2013).

After being transferred from the plastids, MCFAs and palmitic acid (16:0) bind to CoA and then enter the endoplasmic reticulum of the cytoplasm. After a series of reactions, they eventually bind to the *sn*-1, 2, and 3 positions of the glycerol carbon skeleton and turn into TAG. Glycerol-3-phosphate acyltransferase (GPAT), lysophosphatidic acid acyltransferase (LPAT), and DGAT in turn catalyze the esterification of MCFAs-CoA and 16:0-CoA into TAG molecules, and DGAT is thus the key enzyme to accumulate high levels of MCFAs and palmitic acid (Lu et al., 2011). It has been found that LPAT in most oil crops has substrate selectivity for unsaturated fatty acid-CoA (such as oleic acid-CoA) and no selectivity for MCFAs-CoA (Nlandu Mputu et al., 2009). CnLPAT derived from *Cocos nucifera* has a strong substrate specificity for lauric acid (12:0)-CoA (Kim et al., 2015).

To date, genes encoding FatB enzymes with higher substrate specificity for MCFAs than for palmitoyl-ACP has been isolated from *Cuphea* and *Umbellularia californica* seeds containing high levels of MCFAs. Heterologous overexpression of these *FatBs* resulted in the synthetic accumulation of MCFAs in *Brassica*

napus and *Arabidopsis* seeds (Tjellstrom et al., 2013). RNAi was used to silence *KAS*II to block the production of 18:0-ACP from 16:0-ACP, obtaining transgenic seeds that accumulated high levels of 16:0-ACP (Pidkowich et al., 2007). The transgenic camelina seeds specifically expresses the *CpFatB2* gene derived from *C. palustris* accumulated 25% myristic acid (14:0). The *C. sativa* seed respectively accumulated lauric acid (12:0) to 18%, and capric acid (10:0) to 10% following the overexpression of *UcFatB1*, and *ChFatB2* (*C. hookeriana FatB2*), respectively. The *UcFatB1* gene from *U. californica* and the *CnLPAT* gene from coconut were co-expressed in camelina seeds, resulting in up to 30% accumulation of lauric acid (12:0) in camelina seeds (Collins-Silva et al., 2011).

Kim et al. (2015) performed the functional identification of FatBs from the *C. viscosissima* and *C. pulcherrima,* and subsequently used these FatBs for assembly of MCFA synthesis pathway in camelina seeds. Transcriptome analysis revealed that three FatB cDNAs, namely *CpuFatB3*, *CvFatB1*, and *CpuFatB4*, were abundantly expressed in developing seeds, showing positive association with the accumulation of MCFAs.

CpuFatB4 is selective for 12:0-ACP, 14:0-ACP, and 16:0-ACP. The content of palmitic acid (16:0) in camelina seeds overexpressing *CpuFatB4* rose to 43.5% (5 times higher than in the wild-type), and the content of myristic acid (14:0) reached 8%. Similar to CpuFatB4, CpuFatB3 has broad-spectrum substrate specificity for various MCFA-ACPs. The
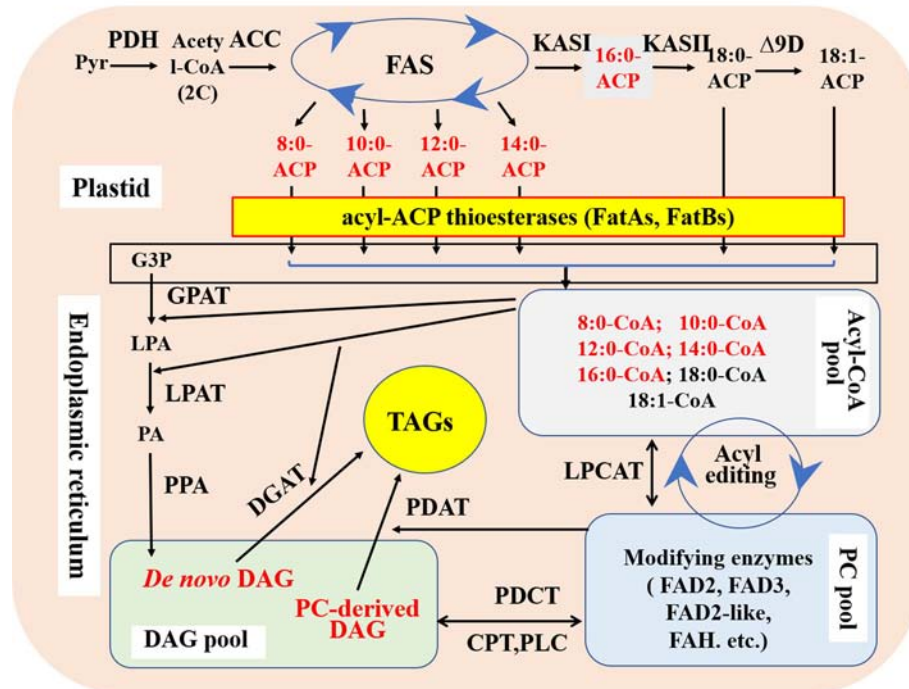
FIGURE 2 | Schematic diagram showing the key steps of *de novo* synthesis of fatty acids and triacylglitcerols with main modified targets for medium-chain fatty acid production in plant seeds. FAS; Fatty acid synthase; FatA, Fatty acid thioesterase A; FatB, Fatty acid thioesterase B; KASII, β-ketoacyl-ACP synthase II; DAG, Diacylglycerol; DGAT, Diacylglycerol acyltransferases; TAG, Triacylglycerol; PDAT, Phospholipid diacylglycerol acyltransferase; GPAT, glycerol-3-phosphate acyltransferase; LPAT, Lysophosphatidic acid acyltransferase; PDH, Pyruvate dehydrogenase; ACC, Acetyl-CoA carboxylase; Pyr, Pyruvic acid.

accumulation of capric acid (10:0) in the seeds of *C. sativa* overexpressing *CpuFatB3* reached 1.2%, while only small amounts of other MCFAs (8:0, 12:0, 14:0) were synthesized. The accumulation of capric acid and palmitic acid in camelina seeds overexpressing *CvFatB1* reached 9% and 16%, respectively, and the contents of other MCFAs (8:0, 12:0, 14:0) were also low. Two or three genes encoding FatB enzymes were co-expressed, resulting in the accumulation of various MCFAs such as C8-C16 in transgenic *C. sativa* seeds, but each fatty acid content was lower than the corresponding fatty acid content in the seeds with only one gene for FatB enzyme. Further over-expression of *CpFatB2* or *UcFatB1* and coconut *CnLPAT* in camelina seeds resulted in the synthesis of more MCFAs. More importantly, co-expressing MCFA-specific FatB and CnLPAT not only increased the MCFA content but also had no negative effect on the total oil content. The above-mentioned metabolically modified engineered camelina strains enriched with MCFAs can be directly used for the production of high-quality jet fuel that is highly resistant to low temperatures.

Assembling ω-3 Long-Chain Polyunsaturated Fatty Acid Biosynthesis for the Production of Oils With Healthcare Applications

Omega-3 fatty acids (ω-3 FAs) are fatty acids that have one double bond at the 3rd carbon atom of the methyl terminal of the carbon chain, including a-linolenic acid (ALA,18:3Δ9,12,15), eicosapentaenoic acid (EPA,20:5Δ5,8,11,14,17), and docosahexaenoic acid (DHA, 22:6Δ4,7,10,13,16,19). In particular, long-chain polyunsaturated fatty acids (ω-3-LC-PUFAs), such as EPA and DHA, derived from fish oil are extremely important for human health, dietary nutrition, and brain development. In order to establish a renewable resource that can replace fish oil for the production of EPA and DHA, many studies have been devoted to the assembly of the EPA and DHA biosynthetic pathways in the developing seeds of common oil crops in order to achieve the "factory" production of EPA and DHA to meet growing market demand (Ruiz-Lopez et al., 2014; Betancor et al., 2015; Malik et al., 2015; Ruiz-Lopez et al., 2015). The EPA and DHA pathways have been assembled in the seeds of oil crops such as soybeans. However, the accumulation of EPA and DHA is low, and it is difficult to commercialize. Camelina seed contains >30% ALA, which is the starting substrate required for the synthesis of EPA and DHA, making it a good platform for assembling the ω-3-LC-PUFA synthesis pathway (**Figure 3**).

There are two ways for biosynthesis of ω-3-LC-PUFA: the conventional Δ6 pathway and the unconventional Δ8 pathway (Ruiz-Lopez et al., 2015). The Δ6 pathway begins with ALA. The synthesis of EPA requires one Δ6 and Δ5 desaturase (Δ6Des and Δ5Des) and one ELO-type Δ6 carbon chain elongase (Δ6Elo); the further synthesis of DHA requires catalysis by one ELO-type Δ5 carbon chain elongase (Δ5Elo) and one Δ4 desaturase (Δ4Des).

The seed-specific coexpression vectors for five genes were constructed and introduced into camelina to assemble the conventional Δ6 pathway of EPA synthesis in developing seeds. The genes used include OtΔ6 desaturase from the eukaryotic

microalgae *Ostreococcus tauri*, TcΔ5 desaturase from marine fungus *Thraustochytrium* sp, PiΩ-3 desaturase from *Phytophthora infestans*, PsΔ12 desaturase from *Phytophthora sojae*, and Δ6 fatty acid carbon chain elongation enzyme PSE1 from *Physcomitrella patens*. The co-expression of these five genes led to 31% EPA accumulation in camelina oil (average EPA 24%) (Ruiz-Lopez et al., 2014). EhΔ4 desaturase from *Emiliania huxleyi* and Δ5 elongase OtElo5 from eukaryotic microalgae were then inserted with these five gene vectors to construct seven gene coexpression vectors. The co-expression of these seven genes in the seed of *C. sativa* resulted in the accumulation of up to 14% DHA (average 8%) (Ruiz-Lopez et al., 2014). The contents of EPA and DHA in this transgenic camelina reached the levels in fish oil, representing the highest EPA and DHA accumulation levels to date have been successfully obtained by assembling biosynthetic pathways in commercial oil crop seeds. More importantly, there is no accumulation of other harmful intermediate metabolites in the seed oil of camelina with these complete EPA and DHA synthetic pathways. The accumulation of EPA and DHA does not negatively affect other agronomic traits, and thus the transgenic plants exhibit normal growth and development. This new camelina germplasm obtained by such metabolic engineering can serve as a high-quality renewable resource for EPA and DHA, and it will be further developed in the future for the commercial production of a series of medicines and nutraceuticals rich in the long-chain ω-3 fatty acids EPA and DHA.

The Δ8 pathway of the unconventional synthesis of EPA is essentially a by-pass pathway, found in some microalgae (e.g. *Pavlova salina* and *Isochrysis galbana*) cells (Ruiz-Lopez et al., 2015). The starting substrate for the Δ8 pathway is also a-linolenic acid (ALA). ALA is elongated two carbon atoms by the ELO-typeΔ9 carbon chain elongase (Δ9Elo) to produce eicosatrienoic acid (20:3Δ11,14,17; ω-3, ERA). Next, Δ8 desaturase catalyses ERA to form arachidonic acid (20:4Δ8,11,14,17; ω-3, ETA). Finally, ETA is catalytically converted to EPA by Δ5 desaturase (20:5Δ5,8,11,14,17; ω-3). The Δ8 pathway has now been successfully assembled and expressed in the developing seeds of camelina, and the accumulation of long-chain ω-3 fatty acids (EPA and ETA) has reached levels up to 26.4% in transgenic *C. sativa* seed oil (Ruiz-Lopez et al., 2015). In addition, linoleic acid (18:2Δ9, 12, LA) is also the substrate for the elongase Δ9Elo. Δ9Elo catalyses the addition of two carbon atoms to LA to extend it to eicosenoic acid (20:2Δ11,14; ω-6, EDA). Under the action of Δ8 desaturase, EDA is converted to eicosatrienoic acid (20:3Δ8,11,14; ω-6, DGLA). Finally, the Δ5 desaturase catalyzes the conversion of DGLA into arachidonic acid (20:4Δ5,8,11,14; ω-6, ARA), which in turn is acted on by Δ5Elo, Δ4Des, and ω-3Des to form DHA. Camelina expressing the Δ8 pathway also requires a precise metabolic modification to reduce the synthesis of the two ω-6 fatty acids DGLA and ARA and promote synthesis to accumulate more EPA and DHA.

For the plant-based production of ω-3 LC-PUFAs, two reviews (Napier et al., 2015; Haslam et al., 2016) described several examples of "proof-of-concept" and progresses in developing transgenic plants enriched ω-3 fish oils. Although promising achievements were obtained in recent years, the metabolic bottleneck remains to be resolved, such as low levels of these non-native fatty acids
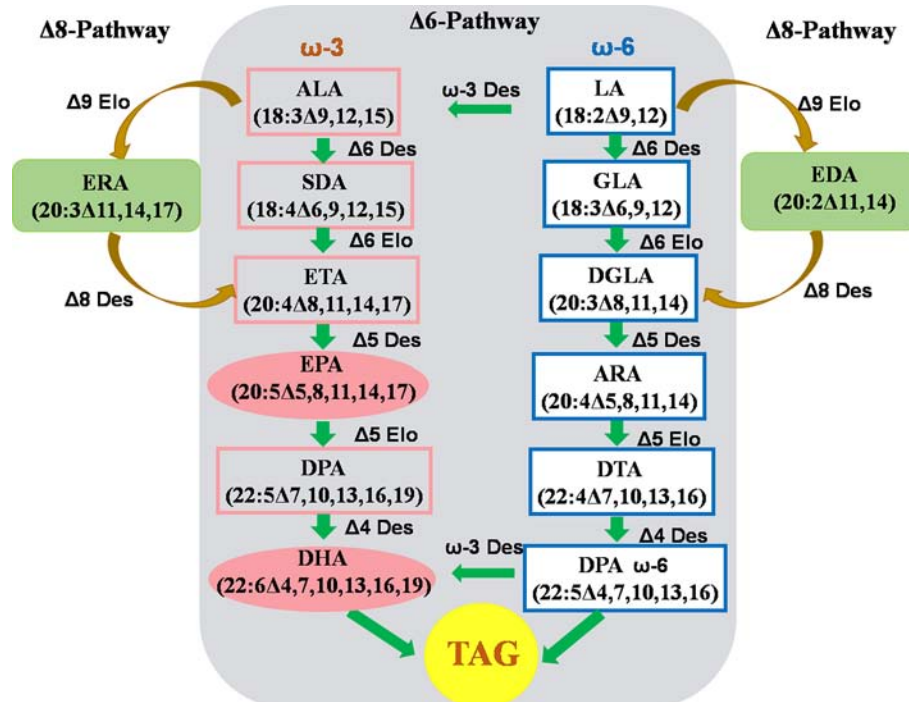
FIGURE 3 | Strategy to assembly biosynthesis of ω-3 LC-PUFAs in *C. sativa* seeds. Δ3 Des, Δ3-desaturase; Δ4 Des, Δ4-desaturase; Δ5 Des, Δ5-desaturase; Δ6 Des, Δ6 -desaturase; Δ8 Des, Δ8-desaturase; Δ5 Elo, Δ5-elongase; Δ6 Elo, Δ6-elongase; Δ9 Elo, Δ9-elongase; ALA,α-Linolenic acid; ARA, Arachidonic acid; DGLA, Dihomo-γ-linolenic acid; DHA, Docosahexaenoic acid; DPA, Docosapentaenoic acid; DTA, Adrenic acid; EDA, Eicosadienoic acid; EPA, Eicosapentaenoic acid; ERA, Eicosatrienoic acid; ETA, Eicosatetraenoic acid; GLA, γ-Linolenic acid; LA, Linoleic acid; SDA, Stearidonic acid; TAG, Triacylglycerol. (The figure is modified according to Ruiz-Lopez et al., 2015.)

accumulated in the transgenic seeds and less knowledge on interaction between the introduced pathway and the endogenous metabolic network. With the success of the field testing of ω-3 LC-PUFAs-enriched camelina, large-scale production of the novel oil by plants will provide an alternative, sustainable source of ω-3 fish oils for increasing end-users.

Redesigning ω-7 Monounsaturated Fatty Acid Biosynthesis for the Production of High-Valued-Added Oils

Omega-7 fatty acids (ω-7 FAs), including palmitoleic acid (16:1Δ9), 11-octadecenoic acid (18:1Δ11) and 13-eicosenoic acid (20:1Δ13), have important industrial, nutritional, and pharmaceutical values. ω-7 FAs are also the best fatty acids for producing high-quality biodiesel (Durrett et al., 2010). These rare fatty acids are mostly synthesized in the seeds of some wild plants such as cat's claw (*Dolichandra unguis-cati*, 64% 16:1Δ9 and 15% 18:1Δ11) and sea buckthorn (*Hippophae rhamnoides*, 32% 16:1Δ9). Because of the poor agronomic traits of these nonergonomic plants, they have not been commercialized to date. Seeds of common oil crops accumulate very small amounts of ω-7 FAs. Metabolic engineering techniques could be employed to assemble ω-7 FA synthesis pathways in common oilseed crops in order to achieve the commercial production of ω-7 FA-rich vegetable oils (Wu et al., 2012; Nguyen et al., 2015). To assemble

this pathway in soybean, our laboratory constructed a seed-specific expression vector for co-expression of the Δ9 desaturase gene that catalyzes palmitic acid (16:0) to palmitoleic acid (16:1Δ9) and the gene encoding a DGAT enzyme. The content of ω-7 FAs in the transgenic seeds with the co-expression of the two genes was greater than 29%. With this soybean oil, a series of functional foods rich in ω-7 FAs can be produced. In addition, in order to develop a new-type of tobacco specifically used for biofuel production, our laboratory has also assembled ω-7 FA biosynthetic pathways in tobacco vegetative organs. The transgenic tobacco leaves accumulated high levels of ω-7 FAs (Xue et al., 2013).

The starting substrate for the synthesis of ω-7 FAs like palmitoleic acid is palmitoyl (16:0)-ACP, which is synthesized in the plastid. There are two downstream enzymatic reaction pathways following the 16:0-ACP generated in the plastid of common oil crops. In the first pathway, under the action of KASII, 16:0-ACP plus two carbon atoms are extended to produce stearoyl (18:0)-ACP. 18:0-ACP can be directly transferred from the plastid into the cytoplasmic endoplasmic reticulum as 18:0-CoA. 18:0-ACP can also be catalyzed by the acyl-ACP-Δ9 desaturase (acyl-ACP-Δ9Des) in the plastid to generate monounsaturated oleoyl (18:1Δ9)-ACP, which is then transferred from the plastid and bound to CoA (18:1Δ9-CoA) to enter the cytoplasmic endoplasmic reticulum. In the second pathway, 16:0-ACP is dissociated from ACP by the action of FatB,

and palmitic acid (16:0) is transferred from the plastid and bound to CoA to enter the cytoplasmic endoplasmic reticulum as 16:0-CoA. In some wild plant seeds rich in ω-7 fatty acids, 16:0-ACP is first converted to palmitoleoyl (16:1Δ9)-ACP (ω-7) in the plastid under the action of acyl-ACP-Δ9Des with substrate specificity for 16:0-ACP, followed by the action of the fatty acid elongate (FAE) to produce 11-octadecenoyl (18:1Δ11)-ACP (ω-7). Finally, under the influence of FatA, 16:1Δ9 and 18:1Δ11 are dissociated from ACPs and transferred out of the plastid and bound to CoA to enter the cytoplasmic endoplasmic reticulum as 16:1Δ9-CoA and 18:1Δ11-CoA, respectively (**Figure 4**).

Currently, a successful strategy for assembling ω-7 FA biosynthesis in *C. sativa* seeds involves expressing an acyl-ACP-Δ9Des that specifically catalyzes 16:0-ACP to produce 16:1Δ9-ACP in the plastids to increase plastid 16:1 Δ9-ACP synthesis. The simultaneous expression of one or two acyl-CoA-Δ9 desaturase (acyl-CoA-Δ9Des) capable of specifically catalyzing 16:0-CoA produced in the cytoplasmic endoplasmic reticulum to form 16:1Δ9-CoA will further increase the accumulation of 16:1Δ9 and its elongation products (18:1Δ11 and 20:1Δ13) in the transgenic camelina seeds.

Nguyen et al. (2015) applied this strategy to assemble the ω-7 FA biosynthetic pathway in camelina and obtained an engineered strain that accumulated high levels of ω-7 FAs. The seed-specific co-expression of one acyl-ACP-Δ9Des with strong specificity for 16:0-ACP and two acyl-CoA-Δ9Des with strong specificity for 16:0-CoA resulted in an increase in the accumulation of ω-7 FAs in the seeds of transgenic camelina from 0.6% in the wild-type to 17%. In this transgenic camelina, resynchronizing the RNAi

silencing of endogenous *CsKASII* and *CsFatB* further increased the ω-7 FA content to an average of 60% of total lipids, and the saturated fatty acid content was correspondingly reduced from 12% in wild-type to 5% in transgenic seeds. This engineered camelina line with high accumulation of ω-7 FAs did not exhibit poor agronomic traits. Seed weight, protein content, and oil content did not differ from those in the control, and seed germination and seedling development were normal. Camelina oil rich in ω-7 fatty acid can be used to more efficiently produce various medicines and functional health products and foods. Physicochemical tests have shown that the thermodynamic properties of these camelina oils rich in ω-7 FAs are significantly improved compared to those of wild type camelina oils (Nguyen et al., 2015). The ω-7 FA-enriched oils can be used to produce high-value-added industrial products such as high-value biodiesel and polyurethanes.

CRISPR/Case9 Gene Editing Technology in Designer Oil Production

The CRISPR/Cas9 gene editing has evolved as the most powerful tool for efficient and specific genome engineering to create genetic models for both fundamental research and crop genetic improvements although this new technology is being developed rapidly (Shan et al., 2013; Van Campenhout et al., 2019; Wolter et al., 2019). This gene editing system has been successfully used to generate a broad range of stable mutagenesis at specific locus in several important crops such as rice (Zhang et al., 2014), wheat (Wang et al., 2019), maize (Svitashev et al., 2015), soybean (Li et al., 2015), tomato (Ito and Nakano, 2015), potato (Butler et al., 2015),
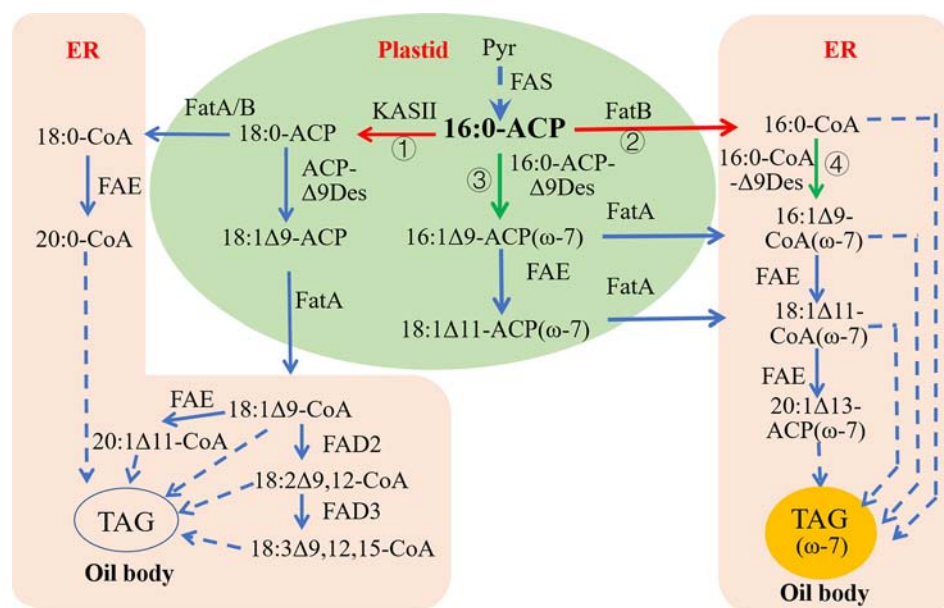
FIGURE 4 | Strategy redesigning biosynthetic pathways for ω-7 FAs production in common oil seeds. ER, Endoplasmic reticulum; FAS, Fatty acid synthase; FatA, Fatty acid thioesterase A; FatB, Fatty acid thioesterase B; FAE, Fatty acid elongase; KASII, β-ketoacyl-ACP synthase II; TAG, Triacylglycerol. Green arrows indicate up-regulation, red arrows indicate down-regulation; ① and ② are to silence KASII and FatB, respectively, to increase pool sizes of 16:0-ACP in plastid to provide more substrates fo16:0-ACP Δ9-desaturase; ③ is to introduce a plastid-localized acyl-ACP Δ9-desaturase with high 16:0-ACP activity; ④ is to introduce an ER-localized 16:0-CoA-specific Δ9-desaturase to catalyze 16:0-CoA into 16:1-CoA in cytoplasm.

barley and *Brassica oleracea* (Lawrenson et al., 2015). The CRISPR-Cas9 system can also be employed to achieve simultaneous editing multiple loci by expressing different sgRNAs targeting the genes (Lowder et al., 2015; Ma et al., 2015). Many agronomic traits of crops were modified by CRISPR-Cas9, including yield and quality-related traits, plant nutrition and development, resistances to various stresses, and domestication (Chen et al., 2019). Notably, simultaneous targeting genes related to meiosis and fertilization by this technology produced a system enabling the clonal reproduction from F1 hybrid rice, thus stably preserving the favorable high degree of heterozygosity (Wang et al., 2019).

As highlighted by recent studies, CRISPR-Cas9 system was successfully used in *C. sativa* for seed oil quality improvement (Jiang et al., 2017; Morineau et al., 2017; Ozseyhan et al., 2018). The allohexaploid camelina contains three closely-related sub-genomes with each gene having three pair of alleles. Camelina seed oil is dominated by polyunsaturated fatty acids (linoleic and linolenic acids) and the development of new varieties rich in monounsaturated fatty acids (particularly oleic acid) is desirable. By targeting all three homeologs of the *CsFAD2* gene, Jiang et al. (2017) obtained a diverse set of genetic combinations with single, double and triple knockouts. In these mutant lines, oleic acid content was increased from 16% to over 50% of total fatty acids, whereas polyunsaturated fatty acids were significantly reduced. Similarly, Morineau et al. (2017) also obtained selectively targeted mutagenesis of *CsFAD2* genes by CRISPR-Cas9, showing the mutants with decreased polyunsaturated fatty acids and concomitant increase of oleic acid in the oil. Another lipid-related gene, *FAE1*(fatty acid elongase1), was effectively deactivated in camelina by CRISPR-Cas9 targeting three *FAE1* alleles simultaneously (Ozseyhan et al., 2018). Homozygous knockout mutants without growth defect were generated in a single generation by an egg cell-specific Cas9 expression, exhibiting that content of C20-C24 very long-chain fatty acids are reduced to less than 2% of total fatty acid compared to over 22% in the wild type. However, knocking out FAE1 increased levels of oleic acid or α-linolenic acid in camelina oils, which are desirable for industrial or food/feed utilizations. The different allelic combinations generated by the gene editing above allowed an unbiased characterization of gene dosage and function in this allohexaploid species, providing a valuable resource of genetic variability for precise plant breeding. It is promising that CRISPR-Cas9 technology will be used to effectively mutate other oil/fatty acid-related genes in camelina for genetic fine-tunning of oil trait.

FUTURE PROSPECTS

Arabidopsis has long been used as a model plant although it is in fact a "wild grass", not a cultivated agricultural plant. For comprehensive understanding of the complexity of many commercial crops, it is necessary to develop a model crop. In this regard, rice (*Oryza sativa*) is recognized as the model plant for cereal crops and species in Gramineae. Similarly, *Camelina*

sativa has emerged as a new model plant for agricultural plants, especially for oil crops. Due to its short life cycle and an efficient genetic transformation like Arabidopsis, camelina is also developed as a "platform crop" with its seeds as "bioreactors" for commercial production of many high-valued products, followed by a new round of camelina commercial planting and a research boom on camelina-based industry worldwide.

As described above, several strategies are available to design high-valued fatty acids/oils by metabolic engineering in camelina seeds. These tools include optimizing enzymatic activity of the lipid-related enzymes, modifying a key gene, and/or coordinately expressing multiple genes associated with fatty acid synthesis and oil accumulation. Future achievements in this field require to obtain more knowledge on the key genes in lipid synthesis and accumulation, the networks responsible for regulation of metabolic carbon partitioning and target lipid synthesis, the rational redesign of novel enzymatic activities to synthesize the target FAs, and the effective assembly of lipid pathways consisting of multiple genes in the host plant.

Genome-wide association studies (GWAS) and multiple "omics" tools including genome resequencing, transcriptomics, and metabolomics are increasingly used to identify new genes involved in FA synthesis and modification and their incorporation into TAGs. For example, a GWAS using 391 wild Arabidopsis accessions revealed four to nineteen genomic regions responsible for FA composition in seed oil, showing that the most regions identified contained candidate genes that had never before been implicated in lipid metabolism (Branham et al., 2015). This tool was also used to examine a number of loci containing candidate genes involved in wax biosynthesis in camelina (Luo et al., 2019). By mining sequence data, the candidate enzymes for target FAs were identified, and subsequently used to manipulate *C. sativa* oil composition towards a superior biofuel and bio-based lubricant oil (Kim et al., 2015; Nguyen et al., 2015).

To enrich the designed FAs/oils in the host, systems biology approaches are needed to obtain a greater understanding of the networks responsible for metabolic carbon partitioning and the transcription factors regulating the related metabolism. More promising, synthetic biology tools like GoldenBraid 2.0 that enable interchangeable, modular assembly of transgene cassettes (Sarrion-Perdigones et al., 2013) can also be employed to effectively introduce the complex pathways involved in multiple genes to generate high-value oil products in this easily transformable camelina host. Synthetic scaffolds could provide a modular and highly flexible tool for rationally organizing multiple function-related enzymes in a controllable manner (Pröschel et al., 2015). It is envisioned that *C. sativa* can be developed as the adaptable chassis plant for commercial production of high-valued compounds by synthetic biology tools.

As CRISPR/Cas9 advance for targeted genome editing in recent years (Wolter et al., 2019), this tool makes it possible to introduce the assembled transgene modules into specific loci in the host crop, leading to an enhancement in the predictability of transgenesis and metabolic outcomes. Like the cases discussed above, CRISPR/Cas9 was used to "knock out" or disable the target gene in the competing

pathways so as to direct metabolic flux toward the desired route. Our laboratory used this technology to construct a library of gene-related mutants associated with the oil metabolism in camelina seeds and obtained various mutants with a single copy of the gene. This shows that the CRISP/Case9 technology has multi-faceted application values in *C. sativa* to overcome the tripled orthologous genes. Undoubtedly, advances in synthetic biology could make it more effectively to perform modular transgene assembly and its targeting insertion into the host genome. Improvements in genome editing could allow the accurate optimization of the host metabolism for the desired pathway. Combination of these two technologies could significantly increase the efficiency of plant lipid redesign and pathway assembly.

In order to obtain various high-valued lipid products, it is also required to redesign the enzyme with improved activities or specific properties. Several approaches can be used to modify or generate novel enzymes necessary for the target lipid synthesis and accumulation, including directed protein mutation, codon-optimization and post-translation modifications. Finally, re-design of metabolic pathways for enriching novel and high-valued lipids should be combined with other breeding programs in camelina such as increasing yield and seed oil content, and improving tolerance to various stress conditions. By doing this, new types of engineered camelina stacking the target FAs and other improved agronomic traits will be created, greatly promoting sustainable production of the designed oil and its commercialization.

CONCLUSIONS

Camelina, an important oil crop in the world, is being developed as an ideal "platform crop" for production of high-valued lipids/oils by metabolic engineering. This review described genetic improvements of the seed oil quality in camelina, particularly focusing on the metabolic pathway assembly of novel fatty acids/oils with multiple nutritional and industrial applications. Several successful examples were provided here to show how lipid biosynthesis in camelina seeds can be redesigned to enable the high accumulation of the target oils that have beneficial functional groups or properties. These target products include acTAGs, hydroxylated FAs, medium chain FAs, ω-3-LC-PUFAs and ω-7 FAs. Technological strategies used in this regard include introduction of a new pathway by synergistical overexpression of multiple genes, blocking the competitive pathway by RNAi-mediated suppression or genomic editing, and optimizing the related metabolic pathways using mutant enzymes. New knowledge and technology, particularly synthetic biology and targeting genome editing will be collectively employed in this field to achieve a more complete orchestration of fatty acid synthesis and metabolism as well as sustainable yields of target lipids in camelina and other oil crop seeds although some significant challenges remain to address.

AUTHOR CONTRIBUTIONS

LY was involved in the review writing. RL was involved in manuscript refinement. All authors read and approved the final manuscript.

FUNDING

REFERENCES

Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H. M., and Somerville, C. R. (1992). Map-based cloning of a gene controlling omega-3 fatty acid desaturation in Arabidopsis. *Science* 258, 1353–1355. doi: 10.1126/science.1455229

Aryal, N., and Lu, C. (2018). A Phospholipase C-Like protein from Ricinus communis increases Hydroxy fatty acids accumulation in transgenic seeds of Camelina sativa. *Front. Plant Sci.* 9, 1576. doi: 10.3389/fpls.2018.01576

Athenstaedt, K., and Daum, G. (2006). The life cycle of neutral lipids: synthesis, storage and degradation. *Cell Mol. Life Sci.* 63, 1355–1369. doi: 10.1007/s00018-006-6016-8

Badami, R., and Patil, K. B. (1980). Structure and occurrence of unusual fatty acids in minor seed oils. *Prog. Lipid Res.* 19, 119–153. doi: 10.1016/0163-7827(80)90002-8

Bansal, S., and Durrett, T. P. (2016). Camelina sativa: an ideal platform for the metabolic engineering and field production of industrial lipids. *Biochimie* 120, 9–16. doi: 10.1016/j.biochi.2015.06.009

Bansal, S., Kim, H. J., Na, G., Hamilton, M. E., Cahoon, E. B., Lu, C., et al. (2018). Towards the synthetic design of camelina oil enriched in tailored acetyl-triacylglycerols with medium-chain fatty acids. *J. Exp. Bot.* 69 (18), 4395–4402. doi: 10.1093/jxb/ery225

Bates, P. D., and Browse, J. (2011). The pathway of triacylglycerol synthesis through phosphatidylcholine in Arabidopsis produces a bottleneck for the accumulation of unusual fatty acids in transgenic seeds. *Plant J.* 68, 387–399. doi: 10.1111/j.1365-313X.2011.04693.x

Betancor, M. B., Sprague, M., Usher, S., Sayanova, O., Campbell, P. J., Napier, J. A., et al. (2015). A nutritionally-enhanced oil from transgenic Camelina sativa effectively replaces fish oil as a source of eicosapentaenoic acid for fish. *Sci. Rep.* 5, 8104. doi: 10.1038/srep08104

Branham, S. E., Wright, S. J., Reba, A., and Linder, C. R. (2015). Genome-wide association study of Arabidopsis thaliana identifies determinants of natural variation in seed oil composition. *J. Hered.* 107, 248–256. doi: 10.1093/jhered/esv100

Burgal, J., Shockey, J., Lu, C., Dyer, J., Larson, T., Graham, I., et al. (2008). Metabolic engineering of hydroxy fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil. *Plant Biotechnol. J.* 6, 819–831. doi: 10.1111/j.1467-7652.2008.00361.x

Butler, N. M., Atkins, P. A., Voytas, D. F., and Douches, D. S. (2015). Generation and inheritance of targeted mutations in potato (Solanum tuberosum L.) using the CRISPR/Cas system. *PloS One* 10, e0144591. doi: 10.1371/journal.pone.0144591

Cahoon, E. B., Ripp, K. G., Hall, S. E., and McGonigle, B. (2002). Transgenic production of epoxy fatty acids by expression of a cytochrome P450 enzyme

from Euphorbia lagascae seed. *Plant Physiol.* 128, 615–624. doi: 10.1104/pp.010768

Cahoon, E. B., Shockey, J. M., Dietrich, C. R., Gidda, S. K., Mullen, R. T., and Dyer, J. M. (2007). Engineering oilseeds for sustainable production of industrial and nutritional feedstocks: solving bottlenecks in fatty acid flux. *Curr. Opin. Plant Biol.* 10, 236–244. doi: 10.1016/j.pbi.2007.04.005

Chen, K. L., Wang, Y. P., Zhang, R., Zhang, H. W., and Gao, C. X. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* 70, 667–697. doi: 10.1146/annurev-arplant-050718-100049

Chhikara, S., Abdullah, H. M., Akbari, P., Schnell, D., and Dhankher, O. P. (2018). Engineering Camelina sativa (L.) Crantz for enhanced oil and seed yields by combining diacylglycerol acyltransferase1 and glycerol-3-phosphate dehydrogenase expression. *Plant Biotechnol. J.* 16 (5), 1034–1045. doi: 10.1111/pbi.12847

Collins-Silva, J. E., Lu, C., and Cahoon, E. B. (2011). Camelina: a designer biotech oilseed crop.

Durrett, T. P., McClosky, D. D., Tumaney, A. W., Elzinga, D. A., Ohlrogge, J., and Pollard, M. (2010). A distinct DGAT with sn-3 acetyltransferase activity that synthesizes unusual, reduced-viscosity oils in Euonymus and transgenic seeds. *Proc. Natl. Acad. Sci. U. States America* 107, 9464–9469. doi: 10.1073/pnas.1001707107

Dyer, J. M., Stymne, S., Green, A. G., and Carlsson, A. S. (2008). High-value oils from plants. *Plant J.: Cell Mol. Biol.* 54, 640–655. doi: 10.1111/j.1365-313X.2008.03430.x

Farmer, E. E., Weber, H., and Vollenweider, S. (1998). Fatty acid signaling in Arabidopsis. *Planta* 206, 167–174. doi: 10.1007/s004250050388

Filichkin, S. A., Slabaugh, M. B., and Knapp, S. J. (2006). New FATB thioesterases from a high-laurate Cuphea species: functional and complementation analyses. *Eur. J. Lipid Sci. Tech.* 108, 979–990. doi: 10.1002/ejlt.200600158

Gugel, R. K., and Falk, K. C. (2006). Agronomic and seed quality evaluation of Camelina sativa in western Canada. *Can. J. Plant Sci.* 86, 1047–1058. doi: 10.4141/P04-081

Guy, S. O., Wysocki, D. J., Schillinger, W. F., Chastain, T. G., Karow, R. S., Garland-Campbell, K., et al. (2014). Camelina: adaptation and performance of genotypes. *Field Crops Res.* 155, 224–232. doi: 10.1016/j.fcr.2013.09.002

Haslam, R. P., Sayanova, O., Kim, H. J., Cahoon, E. B., and Napier, J. A. (2016). Synthetic redesign of plant lipid metabolism. *Plant J.* 87 (1), 76–86. doi: 10.1111/tpj.13172

Hu, Z., Ren, Z., and Lu, C. (2012). The phosphatidylcholine diacylglycerol cholinephosphotransferase is required for efficient hydroxy fatty acid accumulation in transgenic Arabidopsis. *Plant Physiol.* 158, 1944–1954. doi: 10.1104/pp.111.192153

Hu, Z., Wu, Q., Dalal, J., Vasani, N., Lopez, H. O., Sederoff, H. W., et al. (2017). Accumulation of medium-chain, saturated fatty acyl moieties in seed oils of transgenic Camelina sativa. *PloS One* 12 (2), e0172296. doi: 10.1371/journal.pone.0172296

Huai, D., Zhang, Y., Zhang, C., Cahoon, E. B., and Zhou, Y. (2015). Combinatorial effects of fatty acid Elongase enzymes on nervonic acid production in Camelina sativa. *PloS One* 10 (6), e0131755. doi: 10.1371/journal.pone.0131755

Ito, Y., and Nakano, T. (2015). Development and regulation of pedicel abscission in tomato. *Front. Plant Sci.* 6, 442. doi: 10.3389/fpls.2015.00442

Iven, T., Hornung, E., Heilmann, M., and Feussner, I. (2016). Synthesis of oleyl oleate wax esters in Arabidopsis thaliana and Camelina sativa seed oil. *Plant Biotechnol. J.* 14 (1), 252–259. doi: 10.1111/pbi.12379

Jiang, W. Z., Henry, I. M., Lynagh, P. G., Comai, L., Cahoon, E. B., and Weeks, D. P. (2017). Significant enhancement of fatty acid composition in seeds of the allohexaploid, Camelina sativa, using CRISPR/Cas9 gene editing. *Plant Biotechnol. J.* 15, 648–657. doi: 10.1111/pbi.12663

Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W. E., Tuteja, R., et al. (2014). The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure. *Nat. Commun.* 5, 3706. doi: 10.1038/ncomms4706

Kallio, P., Pasztor, A., Akhtar, M. K., and Jones, P. R. (2014). Renewable jet fuel. *Curr. Opin. Biotechnol.* 26, 50–55. doi: 10.1016/j.copbio.2013.09.006

Kim, H. J., Silva, J. E., Vu, H. S., Mockaitis, K., Nam, J. W., and Cahoon, E. B. (2015). Toward production of jet fuel functionality in oilseeds: identification of FatB acyl-acyl carrier protein thioesterases and evaluation of combinatorial expression strategies in Camelina seeds. *J. Exp. Bot.* 66, 4251–4265. doi: 10.1093/jxb/erv225

Knothe, G. (2014). Cuphea oil as a potential biodiesel feedstock to improve fuel properties. *J. Energy Eng.* 140 (26), 50–55. doi: 10.1111/pbi.12663

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N., et al. (2015). Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease. *Genome Biol.* 16, 258. doi: 10.1186/s13059-015-0826-7

Lee, S. B., Kim, H., Kim, R. J., and Suh, M. C. (2014). Overexpression of Arabidopsis MYB96 confers drought resistance in Camelina sativa *via* cuticular wax accumulation. *Plant Cell Rep.* 33 (9), 1535–1546. doi: 10.1007/s00299-014-1636-1

Li, R., Yu, K., Hatanaka, T., and Hildebrand, D. F. (2010). Vernonia DGATs increase accumulation of epoxy fatty acids in oil. *Plant Biotechnol. J.* 8, 184–195. doi: 10.1111/j.1467-7652.2009.00476.x

Li, Z., Liu, Z.-B., Xing, A., Moon, B. P., Koellhoffer, J. P., Huang, L., et al. (2015). Cas9-guide RNA directed genome editing in soybean. *Plant Physiol.* 169, 960–970. doi: 10.1104/pp.15.00783

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., et al. (2013). Acyl-lipid metabolism. *Am. Soc. Plant Biolog.* 11, e0161. doi: 10.1199/tab.0161

Liu, J., Rice, A., Mcglew, K., Shaw, V., Park, H., Clemente, T., et al. (2015a). Metabolic engineering of oilseed crops to produce high levels of novel acetyl glyceride oils with reduced viscosity, freezing point and calorific value. *Plant Biotechnol. J.* 13, 858–865. doi: 10.1111/pbi.12325

Liu, J., Tjellström, H., Mcglew, K., Shaw, V., Rice, A., Simpson, J., et al. (2015b). Field production, purification and analysis of high-oleic acetyl-triacylglycerols from transgenic Camelina sativa. *Ind. Crops Prod.* 65, 259–268. doi: 10.1016/j.indcrop.2014.11.019

Lowder, L. G., Zhang, D., Baltes, N. J., Paul, J. W., Tang, X., Zheng, X., et al. (2015). A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant Physiol.* 169, 971–985. doi: 10.1104/pp.15.00636

Lu, C., and Kang, J. (2008). Generation of transgenic plants of a potential oilseed crop Camelina sativa by Agrobacterium-mediated transformation. *Plant Cell Rep.* 27, 273–278. doi: 10.1007/s00299-007-0454-0

Lu, C., Napier, J. A., Clemente, T. E., and Cahoon, E. B. (2011). New frontiers in oilseed biotechnology: meeting the global demand for vegetable oils for food, feed, biofuel, and industrial applications. *Curr. Opin. Biotechnol.* 22, 252–259. doi: 10.1016/j.copbio.2010.11.006

Luo, Z., Tomasi, P., Fahlgren, N., and Abdel-Haleem, H. (2019). Genome-wide association study (GWAS) of leaf cuticular wax components in Camelina sativa identifies genetic loci related to intracellular wax transport. *BMC Plant Biol.* 19, 187. doi: 10.1186/s12870-019-1776-0

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., et al. (2015). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol. Plant* 8, 1274–1284. doi: 10.1016/j.molp.2015.04.007

Malik, M. R., Yang, W., Patterson, N., Tang, J., Wellinghoff, R. L., Preuss, M. L., et al. (2015). Production of high levels of poly-3-hydroxybutyrate in plastids of Camelina sativa seeds. *Plant Biotechnol. J.* 13, 675–688. doi: 10.1111/pbi.12290

Malik, M. R., Jihong, T., Nirmala, S., Claire, B., Yuanyuan, J., Marie, M., et al. (2018). Camelina sativa, an oilseed at the nexus between model system and commercial crop. *Plant Cell Rep.* 37 (10), 1367–1381. doi: 10.1007/s00299-018-2308-3

Marmon, S., Sturtevant, D., Herrfurth, C., Chapman, K., Stymne, S., and Feussner, I. (2017). Two Acyltransferases contribute differently to linolenic acid levels in seed oil. *Plant Physiol.* 173 (4), 2081–2095. doi: 10.1104/pp.16.01865

Mcvay, K. A., and Khan, Q. A. (2011). Camelina yield response to different plant populations under dryland conditions. *Agron. J.* 103, 1265–1269. doi: 10.2134/agronj2011.0057

Milcamps, A., Tumaney, A. W., Paddock, T., Pan, D. A., Ohlrogge, J., and Pollard, M. (2005). Isolation of a gene encoding a 1,2-diacylglycerol-sn-acetyl-CoA acetyltransferase from developing seeds of Euonymus alatus. *J. Biol. Chem.* 280, 5370–5377. doi: 10.1074/jbc.M410276200

Morineau, C., Bellec, Y., Tellier, F., Gissot, L., Kelemen, Z., Nogué, F., et al. (2017). Selective gene dosage by CRISPR-Cas9 genome editing in hexaploid Camelina sativa. *Plant Biotechnol. J.* 15, 729–739. doi: 10.1111/pbi.12671

Mu, J., Tan, H., Zheng, Q., Fu, F., Liang, Y., Zhang, J., et al. (2008). LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. *Plant Physiol.* 148, 1042–1054. doi: 10.1104/pp.108.126342

Na, G., Aryal, N., Fatihi, A., Kang, J., and Lu, C. (2018). Seed-specific suppression of ADP-glucose pyrophosphorylase in Camelina sativa increases seed size and weight. *Biotechnol. Biofuels* 11, 330. doi: 10.1186/s13068-018-1334-2

Napier, J. A., Usher, S., Haslam, R. P., Ruiz-Lopez, N., and Sayanova, O. (2015). Transgenic plants as a sustainable, terrestrial source of fish oils. *Eur. J. Lipid Sci.* 117, 1317–1324. doi: 10.1002/ejlt.201400452

Nguyen, H. T., Silva, J. E., Podicheti, R., Macrander, J., Yang, W., Nazarenus, T. J., et al. (2013). Camelina seed transcriptome: a tool for meal and oil improvement and translational research. *Plant Biotechnol. J.* 11, 759–769. doi: 10.1111/pbi.12068

Nguyen, H. T., Park, H., Koster, K. L., Cahoon, R. E., Nguyen, H. T., Shanklin, J., et al. (2015). Redirection of metabolic flux for high levels of omega-7 monounsaturated fatty acid accumulation in camelina seeds. *Plant Biotechnol. J.* 13 (1), 38–50. doi: 10.1111/pbi.12233

Nlandu Mputu, M., Rhazi, L., Vasseur, G., Vu, T. D., Gontier, E., and Thomasset, B. (2009). Identification of a potential bottleneck in branched chain fatty acid incorporation into triacylglycerol for lipid biosynthesis in agronomic plants. *Biochimie* 91, 703–710. doi: 10.1016/j.biochi.2009.03.011

Ozseyhan, M. E., Kang, J., Mu, X., and Lu, C. (2018). Mutagenesis of the FAE1 genes significantly changes fatty acid composition in seeds of Camelina sativa. *Plant Physiol. Biochem.* 123, 1–7. doi: 10.1016/j.plaphy.2017.11.021

Petrie, J. R., Shrestha, P., Belide, S., Kennedy, Y., Lester, G., Liu, Q., et al. (2014). Metabolic engineering Camelina sativa with fish oil-like levels of DHA. *PloS One* 9 (1), e85061. doi: 10.1371/journal.pone.0085061

Pidkowich, M. S., Nguyen, H. T., Heilmann, I., Ischebeck, T., and Shanklin, J. (2007). Modulating seed beta-ketoacyl-acyl carrier protein synthase II level converts the composition of a temperate seed oil to that of a palm-like tropical oil. *Proc. Natl. Acad. Sci. U. S. A.* 104, 4742–4747. doi: 10.1073/pnas.0611141104

Pröschel, M., Detsch, R., Boccaccini, A. R., and Sonnewald, U. (2015). Engineering of metabolic pathways by artificial enzyme channels. *Front. Bioeng. Biotechnol.* 3, 168. doi: 10.3389/fbioe.2015.00168

Rodriguez-Rodriguez, M. F., Salas, J. J., Garces, R., and Martinez-Force, E. (2014). Acyl-ACP thioesterases from Camelina sativa: cloning, enzymatic characterization and implication in seed oil fatty acid composition. *Phytochemistry* 107, 7–15. doi: 10.1016/j.phytochem.2014.08.014

Ruiz-Lopez, N., Haslam, R. P., Napier, J. A., and Sayanova, O. (2014). Successful high-level accumulation of fish oil omega-3 long-chain polyunsaturated fatty acids in a transgenic oilseed crop. *Plant J. Cell Mol. Biol.* 77, 198–208. doi: 10.1111/tpj.12378

Ruiz-Lopez, N., Haslam, R. P., Usher, S., Napier, J. A., and Sayanova, O. (2015). An alternative pathway for the effective production of the omega-3 long-chain polyunsaturates EPA and ETA in transgenic oilseeds. *Plant Biotechnol. J.* 13 (9), 1264–1275. doi: 10.1111/pbi.12328

Ruiz-Lopez, N., Broughton, R., Usher, S., Salas, J. J., Haslam, R. P., Napier, J. A., et al. (2017). Tailoring the composition of novel wax esters in the seeds of transgenic Camelina sativa through systematic metabolic engineering. *Plant Biotechnol. J.* 15 (7), 837–849. doi: 10.1111/pbi.12679

Sarrion-Perdigones, A., Vazquez-Vilar, M., Palací, J., Castelijns, B., Forment, J., Ziarsolo, P., et al. (2013). GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol.* 162, 1618–1631. doi: 10.1104/pp.113.217661

Shan, Q. W., Wang, Y. P., Li, J., Zhang, Y., Chen, K. L., Liang, Z., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 686–688. doi: 10.1038/nbt.2650

Snapp, A. R., Kang, J., Qi, X., and Lu, C. (2014). A fatty acid condensing enzyme from Physaria fendleri increases hydroxy fatty acid accumulation in transgenic oilseeds of Camelina sativa. *Planta* 240, 599–610. doi: 10.1007/s00425-014-2122-2

Svitashev, S., Young, J. K., Schwartz, C., Gao, H., Falco, S. C., and Cigan, A. M. (2015). Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and Guide RNA. *Plant Physiol.* 169, 931–945. doi: 10.1104/pp.15.00793

Tjellstrom, H., Strawsine, M., Silva, J., Cahoon, E. B., and Ohlrogge, J. B. (2013). Disruption of plastid acyl:acyl carrier protein synthetases increases medium chain fatty acid accumulation in seeds of transgenic Arabidopsis. *FEBS Lett.* 587, 936–942. doi: 10.1016/j.febslet.2013.02.021

Usher, S., Han, L., Haslam, R. P., Michaelson, L. V., Sturtevant, D., Aziz, M., et al. (2017). Tailoring seed oil composition in the real world: optimising omega-3 long chain polyunsaturated fatty acid accumulation in transgenic Camelina sativa. *Sci. Rep.* 7 (1), 6570. doi: 10.1038/s41598-017-06838-0

Van Campenhout, C., Cabochette, P., Veillard, A. C., Laczik, M., Zelisko-Schmidt, A., Sabatel, C., et al. (2019). Guidelines for optimized gene knockout using CRISPR/Cas9. *Biotechniques* 66, 295–302. doi: 10.2144/btn-2018-0187

van Erp, H., Bates, P. D., Burgal, J., Shockey, J., and Browse, J. (2011). Castor phospholipid:diacylglycerol acyltransferase facilitates efficient metabolism of hydroxy fatty acids in transgenic Arabidopsis. *Plant Physiol.* 155, 683–693. doi: 10.1104/pp.110.167239

Vollmann, J., and Eynck, C. (2015). Camelina as a sustainable oilseed crop: contributions of plant breeding and genetic engineering. *Biotechnol. J.* 10 (4), 525–535. doi: 10.1002/biot.201400200

Wang, C., Liu, Q., Shen, Y., Hua, Y. F., Wang, J. J., Lin, J. R., et al. (2019). Clonal seeds from hybrid rice by simultaneous genome engineering of meiosis and fertilization genes. *Nat. Biotechnol.* 37, 283–28+. doi: 10.1038/s41587-018-0003-0

Wolter, F., Schindele, P., and Puchta, H. (2019). Plant breeding at the speed of light: the power of CRISPR/Cas to generate directed genetic diversity at multiple sites. *BMC Plant Biol.* 19, 176. doi: 10.1186/s12870-019-1775-1

Wu, Y., Li, R., and Hildebrand, D. F. (2012). Biosynthesis and metabolic engineering of palmitoleate production, an important contributor to human health and sustainable industry. *Prog. Lipid Res.* 51, 340–349. doi: 10.1016/j.plipres.2012.05.001

Xue, J. A., Mao, X., Yang, Z. R., Wu, Y. M., Jia, X. Y., Zhang, L., et al. (2013). Expression of yeast acyl-CoA-9 desaturase leads to accumulation of unusual monounsaturated fatty acids in soybean seeds. *Biotechnol. Lett.* 35, 951–959. doi: 10.1007/s10529-013-1149-y

Yang, W., Wang, G., Li, J., Bates, P. D., Wang, X., and Allen, D. K. (2017). Phospholipase Dζ Enhances Diacylglycerol Flux into Triacylglycerol. *Plant Physiol.* 174 (1), 110–123. doi: 10.1104/pp.17.00026

Yu, X. H., Cai, Y., Chai, J., Schwender, J., and Shanklin, J. (2019). Expression of a Lychee PHOSPHATIDYLCHOLINE: DIACYLGLYCEROL CHOLINEPHOSPHOTRANSFERASE with an Escherichia coli CYCLOPROPANE SYNTHASE enhances cyclopropane fatty acid accumulation in Camelina Seeds. *Plant Physiol.* 180 (3), 1351–1361. doi: 10.1104/pp.19.00396

Yuan, L., Mao, X., Zhao, K., Ji, X., Ji, C., Xue, J., et al. (2017a). Characterization of Phospholipid: Diacylglycerol Acyltransferases (PDATs) from Camelina sativa and their roles in stress responses. *Biol. Open* 6, 1024–1034. doi: 10.1242/bio.026534

Yuan, L., Sun, Y., Xue, J., and Li, R. (2017b). Spatio-temporal expression and stress responses of DGAT1, DGAT2 and PDAT responsible for TAG biosynthesis in Camelina sativa. *Emir. J. Food Agr.* 29, 274–284. doi: 10.9755/ejfa.2016-10-1420

Zanetti, F., Monti, A., and Berti, M. T. (2013). Challenges and opportunities for new industrial oilseed crops in EU-27: a review. *Ind. Crops Prod.* 50, 580–595. doi: 10.1016/j.indcrop.2013.08.030

Zhang, H., Zhang, J. S., Wei, P. L., Zhang, B. T., Gou, F., Feng, Z. Y., et al. (2014). The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.* 12, 797–807. doi: 10.1111/pbi.12200

Zubr, J. (1997). Oil-seed crop: Camelina sativa. *Ind. Crops Prod.* 6, 113–119. doi: 10.1016/S0926-6690(96)00203-8

Methylobacterium sp. 2A Is a Plant Growth-Promoting Rhizobacteria That Has the Potential to Improve Potato Crop Yield Under Adverse Conditions

Cecilia Eugenia María Grossi[1], Elisa Fantino[1†], Federico Serral[2], Myriam Sara Zawoznik[3], Darío Augusto Fernandez Do Porto[2] and Rita María Ulloa[1,4*]

[1] Laboratorio de Transducción de Señales en Plantas, Instituto de Investigaciones en Ingeniería Genética y Biología Molecular (INGEBI), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ciudad Autónoma de Buenos Aires, Argentina, [2] Plataforma de Bioinformática Argentina, Instituto de Cálculo, Ciudad Universitaria, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires (UBA), Ciudad Autónoma de Buenos Aires, Argentina, [3] Cátedra de Química Biológica Vegetal, Departamento de Química Biológica, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires (UBA), Ciudad Autónoma de Buenos Aires, Argentina, [4] Departamento de Química Biológica, Universidad de Buenos Aires (UBA), Ciudad Autónoma de Buenos Aires, Argentina

A Gram-negative pink-pigmented bacillus (named 2A) was isolated from *Solanum tuberosum* L. cv. Desirée plants that were strikingly more developed, presented increased root hair density, and higher biomass than other potato lines of the same age. The 16S ribosomal DNA sequence, used for comparative gene sequence analysis, indicated that strain 2A belongs to the genus *Methylobacterium*. Nucleotide identity between *Methylobacterium* sp. 2A sequenced genome and the rest of the species that belong to the genus suggested that this species has not been described so far. *In vitro,* potato plants inoculated with *Methylobacterium* sp. 2A had a better performance when grown under 50 mM NaCl or when infected with *Phytophthora infestans*. We inoculated *Methylobacterium* sp. 2A in *Arabidopsis thaliana* roots and exposed these plants to salt stress (75 mM NaCl). *Methylobacterium* sp. 2A-inoculated plants, grown in control or salt stress conditions, displayed a higher density of lateral roots ($p < 0.05$) compared to noninoculated plants. Moreover, under salt stress, they presented a higher number of leaves and larger rosette diameter. In dual confrontation assays, *Methylobacterium* sp. 2A displayed biocontrol activity against *P. infestans*, *Botrytis cinerea*, and *Fusarium graminearum*, but not against *Rhizoctonia solani*, and *Pythium dissotocum*. In addition, we observed that *Methylobacterium* sp. 2A diminished the size of necrotic lesions and reduced chlorosis when greenhouse potato plants were infected with *P. infestans*. *Methylobacterium* sp. 2A produces indole acetic acid, solubilizes mineral phosphate and is able to grow in a N_2 free medium. Whole-genome sequencing revealed metabolic pathways associated with its plant growth promoter capacity. Our results suggest that *Methylobacterium* sp. 2A is a plant growth-promoting rhizobacteria (PGPR) that can

alleviate salt stress, and restricts *P. infestans* infection in potato plants, emerging as a potential strategy to improve crop management.

INTRODUCTION

In 30 years, the world population will be close to 10 billion people, there will be a further 2–3 billion people to feed. According to FAO Statistical Database[1], 50% of the habitable land is nowadays used for agriculture, most of which is used for the rearing of livestock and only 23% (11 million km^2) is for food crop production. These 11 million km^2 supply more calories and proteins for the global population than the almost four-time larger area devoted to livestock (Our World in Data[2]). Today, our challenge is to increase crop productivity at a faster rate than population growth. Most countries have managed to achieve this goal in recent decades; a combination of agricultural technologies, irrigation, improved crop varieties, fertilizers, and pesticides, were used to obtain higher yields (Roser and Ritchie, 2017). However, it is possible that yield gains in the decades to follow will be offset by a growing population. In the near future, especially in the developing world, there will be an increasing prevalence of farming on marginal, arid, and semiarid lands (Coleman-Derr and Tringe, 2014). Therefore, crop yield has to be further improved and crop varieties should be adapted to hostile environments.

Potato is produced in over 100 countries and is the third most important food crop in the world after rice and wheat in terms of human consumption. One hectare (ha) of potato can yield two to four times the food quantity of grain crops; potatoes produce more food per unit of water than any other major crop and are up to seven times more efficient in using water than cereals (CIP International Potato Center[3]). From 1994 to 2014, productivity gains in potato increased 15 points above the population growth rate (Our World in Data[2]). In Argentina, 70 to 80 thousand ha are allocated to potato production, yield values are around 30 to 35 TN/ha, and total production is 2.1 to 2.5 million TNs per year (Garzón and Young, 2016). However, it was estimated that potato productive yield is only 40% to 76% or 47% to 81% of the potential yield, depending on the method used to estimate potential yield, the year and the region (Cantos de Ruiz et al., 1989; de la Casa et al., 2014).

Potential yields can be calculated assuming that there are no abiotic limitations to the growth and there are no biotic factors that reduce growth. The potato plant is well adapted to a number of environmental conditions but certain biotic and abiotic stresses cause significant reductions in growth and yield. Among abiotic stresses, salinity is one of the main constraints that limit plant productivity and cause loss of arable land (Isayenkov and Maathuis, 2019). In particular, potato plants are glycophytes moderately sensitive to salt stress. Although

there are differences between cultivars, all cultivars showed a reduction in shoot length, reduced root system development, and reduced tuber yield due to salinity (Dahal et al., 2019). Among biotic stresses, late blight of potato caused by the hemibiotrophic oomycete *Phytophthora infestans* is still a devastating disease worldwide. Chemical management is a popular strategy to control late blight but it may have a negative impact on the environment (Lal et al., 2018).

Numerous techniques have been used to understand the mechanisms and provide tools to enhance plant tolerance under environmental stresses. To ensure long-term food production, we must develop sustainable agricultural practices with minimal adverse impact on the environment. In this context, the use of microbial inoculants plays a key role. When introduced to seeds, roots or into the soil, plant growth-promoting rhizobacteria (PGPR) can solubilize insoluble phosphates, produce plant growth hormones, convert atmospheric nitrogen to ammonia or suppress the growth of plant pathogens (Pérez-Montaño et al., 2014). Naturally occurring PGPR were shown to be effective in enhancing plant growth and development, and in promoting crop productivity and disease management under stress conditions (García et al., 2017; Singh and Jha, 2017; Cordero et al., 2018). This environmentally friendly approach could be among the most efficient methods for minimizing the use of chemicals.

In this work, we have identified and characterized *Methylobacterium* sp. 2A, a PGPR that promotes plant growth and is able to mitigate the harsh effect of salinity on *in vitro* potato and Arabidopsis plants. Moreover, it can reduce *P. infestans* infection on potato greenhouse plants. Genome sequencing allowed us to identify genes that could be involved in its plant growth-promoting (PGP) capacity. We confirmed that this rhizobial can produce indole acetic acid (IAA), is able to grow in N_2 depleted media, and can solubilize inorganic phosphate.

MATERIALS AND METHODS

Plant and Pathogen Material

Solanum tuberosum L. cv. Desirée wild-type (WT) and *Arabidopsis thaliana* Columbia (Col-0) plants were used. Fresh internodes (the first three from the top) from 1-month-old virus-free potato plants were micropropagated in MS medium (Murashige and Skoog, 1962) with the addition of 2% (w/v) sucrose and 0.7% (w/v) agar. Surface disinfected Arabidopsis seeds were vernalized for two days at 4°C and, were then allowed to grow in MS medium (0.5X, 0.8% (w/v) agar). Plants were grown for the indicated times in a growth chamber under a 16 h light photoperiod at 21°C–23°C.

Certified pathogen tested *Solanum tuberosum*, L. var. Spunta (Diagnósticos Vegetales S.A., Mar del Plata, Argentina) tubers were planted in 1-L pots filled with MultiProTM substrate (GrowMix®, Terrafertil S.A., Bs. As., Argentina). Soil-grown plants were cultivated in a greenhouse under a 16-h light photoperiod; natural light was supplemented by sodium lamps providing 100–300 μmols^{-1}m^{-2}; the temperature was set at 25°C during the day and 20°C in the night.

P. infestans isolate Pi-60 was kindly provided by Dr. Natalia Norero (Laboratorio de Agrobiotecnología, INTA EEA-Balcarce), maintained on rye sucrose agar (RSA) medium at 19°C ± 1°C in the dark. Suspensions of *Phytophthora* zoospores were performed as described in Fantino et al. (2017). The concentration was adjusted to 25 sporangia μl^{-1} to be used as an inoculum. The fungi *Fusarium graminearum, Botrytis cinerea, Phytium dissotocum,* and *Rhizoctonia solani* were maintained in potato dextrose agar (PDA) medium at room temperature in the dark, or at 19°C ± 1°C in the case of *P. dissotocum*.

Isolation and Identification of *Methylobacterium* sp. 2A

Bacterial isolation was conducted from roots of potato plants established *in vitro* that were previously grown in the greenhouse at Instituto de Investigaciones en Ingeniería Genética y Biología Molecular (INGEBI-CONICET, Argentina; 34° 33′ 28.0″ S 58° 27′ 32.0″ W). Fresh root samples were collected, surface sterilized, and placed in LB agar (Luria-Bertani) at 28°C for 4 days. Pink-pigmented colonies were selected and the isolate was maintained on LB medium without the addition of NaCl (LBNS). To determine the optimal growth conditions, different temperatures (18°C, 24°C, 28°C, 30°C, and 37°C) and culture media (LBNS, PDA, pea sucrose agar (PSA), and RSA) were tested. pH tolerance was determined in LB adjusted to different pH values (4.5 to 7.5). A catalase slide test was performed and oxidase activity was determined by the oxidation of discs impregnated with N, N-dimethyl-p-phenylenediamine oxalate (Britania S.A., Argentina). Antibiotic resistance to streptomycin, hygromycin, rifampicin, ampicillin, kanamycin, chloramphenicol, and cefotaxime were tested in liquid LBNS.

16s Ribosomal DNA Gene-Based Analyses

Genomic DNA (gDNA) was extracted manually according to Chen and Kuo (1993) from a 48-h culture. The 16S ribosomal DNA (*16S rDNA*) gene was amplified by Polymerase Chain Reaction (PCR) with universal primers fD1 and rP2 (**Table S1**) (Genbiotech, Buenos Aires, Argentina) (Weisburg et al., 1991) and sequenced at Macrogen Sequencing facility (Seoul, South Korea). Phylogenetic analysis was performed comparing *Methylobacterium* sp. 2A *16S rDNA* against representative type species of the *Methylobacterium* genus, and against eleven species that have been recently reclassified into the new *Methylorubrum* genus (Green and Ardley, 2018). Multiple sequence alignment and phylogenetic tree reconstruction were performed with MEGA7 (Molecular Evolutionary Genetics Analysis 7, Kumar et al., 2016) software using the Maximum Likelihood method based on the Kimura two-parameter model with 1,000 bootstrap values. Members of the Methylocystaceae family were included as outgroups. Sequence similarity was also compared to those on the EzBioCloud[3] database (Yoon et al., 2017).

Plant Inoculation With *Methylobacterium* sp. 2A and Salt Treatments

In vitro potato plants were inoculated with *Methylobacterium* sp. 2A by root contact; plant internodes of potato plants showing root pink-pigmentation were equidistantly placed in flasks containing solid MS media with noninoculated potato internodes. Bacterial colonization was evident to the naked-eye two weeks later.

Internodes from potato plants were grown in solid MS media with the addition or not (controls) of increasing NaCl concentrations for 21 days; root and shoot length (cm), and total chlorophyll content (Hiscox and Israelstam, 1979) were assessed. Significant stress was evidenced at 50 mM, but it was still possible to determine chlorophyll content and root growth (**Figure S1**). Therefore, this concentration was chosen to perform further studies. Internodes from potato plants, inoculated or not with *Methylobacterium* sp. 2A, were grown for 21 days in solid MS media with the addition or not of 50 mM NaCl. The experiment was performed three times, using two flasks with four plants each, for each condition (control, C; inoculated, I; salt stress, S; inoculated plants with salt stress, I+S). Root and shoot length, the number of leaves and chlorophyll content were determined at the time of harvest in both control and salt treatments. When measuring root and shoot length, the original internode was not considered. Chlorophyll content was measured on fully expanded leaves as SPAD units using a portable chlorophyll spectrophotometer (Clorofilio®, Cavadevices, Argentina).

One-week-old Arabidopsis plants with similar growth were selected and established in Petri dishes containing MS 0.5X, or MS 0.5X with 75 mM NaCl (salt stress conditions; Ruiz Carrasco et al., 2007). After 48 h, plants were inoculated with 2 μl of a sterile saline solution (0.85% NaCl) or with a suspension of *Methylobacterium* sp. 2A cells (0.05 OD$_{600nm}$ units in 0.85% NaCl). Ten days later, the number of leaves and rosette diameter were determined and lateral root density was counted in each plant as the number of lateral roots/primary root length. For this, Petri dishes were photographed; the primary root length was measured and lateral roots were counted using Fiji software (Schindelin et al., 2012). Samples were stored in liquid nitrogen for further studies. Protein extracts were obtained and catalase activity was determined as described in Aebi (1984). Hydrogen peroxide (H$_2$O$_2$) quantification was performed according to Junglee et al. (2014). Protein content was estimated by Bradford using bovine serum albumin as standard. The experiment was conducted three times using two plates with five plants each for each condition (control, C; *Methylobacterium* sp. 2A-inoculated, I; salt stress, S; *Methylobacterium* sp. 2A-inoculated plants with salt stress, I+S).

Plant Inoculation With *Methylobacterium* sp. 2A and Infection With *P. infestans*

In vitro four-weeks-old potato plants inoculated or not with *Methylobacterium* sp. 2A, were infected with *P. infestans* isolate Pi-60; 10-μl droplets of zoospore suspension were pipetted on

three leaves per plant. Two flasks containing five plants each were used for each condition (Pi-60 and I+Pi-60). Five days later, *P. infestans* aggressiveness was observed (**Figure S2**). The result obtained encouraged us to conduct experiments in greenhouse plants. To this end, three-weeks-old potato plants were sprayed with *Methylobacterium* sp. 2A bacterial suspension (0.05 OD_{600nm} units in 0.85% NaCl), or with saline solution (control). Infection with Pi-60 was performed two days later; 10-μl droplets of sporangia suspension or water were pipetted on the abaxial side of apical leaflets of three leaves per plant (two equidistant spots per leaf). Five days post-infection (dpi), leaves were observed and photographed with a phone camera. Necrotic lesions were counted and Fiji software was used to measure the lesion area. The experiment was conducted twice using three plants for each condition (control, C; *Methylobacterium* sp. 2A-inoculated, I; infected, Pi-60; *Methylobacterium* sp. 2A-inoculated and infected, I+Pi-60).

Leaf samples were collected and total RNA was extracted using the TRIZOL Reagent (Invitrogen) following the manufacturer's instructions. Total RNAs (1 μg) were pre-treated with DNAse (RQ1 RNAse-free DNAse, Promega) and reverse transcribed with M-MLV-Reverse Transcriptase (Promega) using an oligo-dT primer and random hexamers. Expression levels of *StPR-1b* and *StPAL* genes were analyzed by RT-qPCR on an Applied Biosystems 7500 Real-Time PCR System, with the indicated primers (**Table S1**) (Genbiotech, Buenos Aires, Argentina) and FastStart Universal SYBR Green Master Rox (Roche). Elongation factor 1 alpha (*EF-1α*) was used as a reference gene. PCR reactions were incubated at 95°C for 10 min followed by 40 cycles of 95°C for 10s; and 60°C for 1 min. PCR specificity was checked by melting curve analysis. Expression data was analyzed using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

In Plate Confrontation Assay

Dual culture assays were performed in PDA or in RSA. *Methylobacterium* sp. 2A was striked on one half of the plate and, after 3 days of incubation at 25°C, a 1 cm² plug from an actively growing culture of the oomycete *P. infestans* or the fungi *F. graminearum, B. cinerea, P. dissotocum,* and *R. solani,* was placed at the other half of the plate. Petri dishes with PDA or RSA containing only the corresponding plugs served as controls. The radial mycelial growth of the pathogens toward *Methylobacterium* sp. 2A (Ri) and that on a control plate (Rc) were measured and mycelial growth inhibition was calculated according to the formula: (Rc-R)/Rc x 100 (Lahlali and Hijri, 2010).

De Novo Genome Assembly and Annotation

Whole Genome Sequencing (WGS) of *Methylobacterium* sp. 2A was carried out by an Illumina TruSeq Nano platform at Macrogen Laboratories. *De novo* assembly was done using the standard procedures from our own prokaryotic assembly pipeline (Sosa et al., 2018), based on SPAdes version 3.9.0 (Bankevich et al., 2012) and SSPACE version 3.0 (Boetzer et al., 2011). Genome annotation was done using the Rapid

Annotations Subsystems Technology (RAST) server (Aziz et al., 2008).

Whole-Genome Computational Analysis

The ANI (Average Nucleotide Identity) values between *Methylobacterium* sp. 2A and the other reference strains were calculated with the JspeciesWS web service (Richter et al., 2015). *In silico* DNA-DNA hybridization (DDH) was conducted between *Methylobacterium* sp. 2A and the reference strains with the Genome-to-Genome Distance Calculator web service (GGDC 2.1; Meier-Kolthoff et al., 2013).

Search for Metabolic Pathways Associated With PGP Traits

On the basis of the annotated genome, computational prediction of metabolic pathways was made through automatic reconstruction by Pathway Tools v23.0. The Pathologic software of Pathway Tools (Karp et al., 2002) and the MetaCyc database (Caspi, 2005) were used to automatically generate a pathway-genome database (PGDB) from the GenBank file of the *Methylobacterium* sp. 2A annotated genome. The PGDB links the coding sequences and potential genes to enzymatic reactions and biochemical pathways. In addition, gene clusters were identified with SnapGene[4] software (GSL Biotech) and with antiSMASH 5.0 (Blin et al., 2019) and manually curated.

Assays for Detection of PGP Abilities

Dinitrogen fixation, phosphate solubilization, and IAA production were analyzed. The amount of IAA produced by *Methylobacterium* sp. 2A was estimated using Salkowski's method (Ehmann, 1977). The IAA-producing strains *Azospirillum brasilense* Az39 and *Pantoea* sp. were used as positive controls (Díaz Herrera et al., 2016). The different strains were grown in LB broth plus Trp (0.1 mg ml⁻¹) and incubated at 28°C for 3, 4, or 5 days. After incubation, 2-ml aliquots were centrifuged and 1-ml supernatant samples were mixed with 1 ml of Salkowski's reagent (2% 0.5 $FeCl_3$ in 35% H_2SO_4 solution) and kept in the dark. OD was recorded at 530 nm after 30 min. Nitrogen fixation was qualitatively determined by culturing single colonies in semisolid Nfb media, as described by Hartmann and Baldani (2006). An assay to evaluate phosphate solubilization was performed using tricalcium phosphate (TCP) in NBRIP solid and liquid media. Phosphate determination was made by the vanado-molybdate colorimetric method according to Pearson (1976). The strain *Pseudomonas fluorescens* BNM 233 currently used as a biofertilizer (Okon et al., 2015) was used as a positive control for phosphate solubilization.

Statistical Analysis

Statistical analysis was performed by one-way or two-way ANOVA followed by Tukey's HSD test (p < 0.05) or by T-test as indicated in the figures, using GraphPad Prism[5] version 5.03 (GraphPad Software, La Jolla, California, USA).

[4] http://snapgene.com
[5] http://graphpad.com

RESULTS

General Features of *Methylobacterium* sp. 2A

One-week-old potato plants micropropagated *in vitro* attracted our attention due to the pink-pigmentation of its roots (**Figure 1A**). MS media was not contaminated but a strong association with a microorganism was evident. One month later, these plants were more developed than other potato lines of the same age (**Figure 1B**) and the density of root hairs was increased (**Figure 1C**). A gram-negative bacillus, named *Methylobacterium* sp. 2A, was isolated from the roots and characterized. Colonies were pink-pigmented and circular, reaching a diameter of 0.2 mm after 3 days of incubation, denoting slow growth. It had a better performance in solid media containing plant extracts (PDA, PSA, and RSA media) while growth on LB agar was lower. Optimal growth was observed at 28°C, and the optimal pH growth range was between 5 and 7, being 6 the optimal pH value (**Table S2**). Oxidase and catalase reactions were positive. Resistance to hygromycin (10 μg/μl), ampicillin and chloramphenicol (20 μg/μl) was evidenced; however, *Methylobacterium* sp. 2A was sensitive to the other antibiotics tested.

Phylogenetic Analysis of *Methylobacterium* sp. 2A

The nearly full-length *16S rDNA* sequence (1,218 nt; GenBank accession number: MG818293.1) used for comparative gene sequence analysis, indicated that strain 2A belongs to the genus *Methylobacterium*, family Methylobacteriaceae, order Rhizobiales. EzBioCloud comparison revealed that *16S rDNA* from *Methylobacterium* sp. 2A shared high sequence similarity (>97%) with fourteen validated type species of the genus (**Table S3**). The phylogenetic tree (**Figure 2**) indicated that this isolate is most closely related to *M. fujisawaense*, *M. phyllosphaerae*, *M. oryzae*, *M. radiotolerans*, *M. tardum*, *M. longum*, and *M. phyllostachyos* (sequence similarity between 98.52% and 99.10%). Characteristics of *Methylobacterium* sp. 2A, such as colony pigmentation, cell size, growth conditions, and catalase and oxidase reactions, were compared (**Table S2**) with those of the most closely related species (Ito and Iizuka, 1971; Kato et al., 2008; Knief et al., 2012; Madhaiyan and Poonguzhali, 2014; Madhaiyan et al., 2007; Madhaiyan et al., 2009). As observed, though isolated from different sources, all strains were positive for catalase and oxidase reactions, had a similar pigmentation, and shared several growth conditions.

Inoculation of *Methylobacterium* sp. 2A Conferred Stress Tolerance Under Salt Conditions

Salinity impairs plant growth and development *via* water stress, cytotoxicity due to excessive uptake of ions (Na^+ and Cl^-), and nutritional imbalance; it is accompanied by oxidative stress due to the generation of reactive oxygen species (ROS) (reviewed in Isayenkov and Maathuis, 2019). *In vitro* potato plants were grown in MS media with or without 50 mM NaCl for 21 days. Strong inhibition of shoot and root growth, and total biomass, together with a significant decrease in chlorophyll content (SPAD units) and leaf number (**Figures 3A, B**) was observed compared to control conditions (C vs. S). On the other hand, when the *Methylobacterium* sp. 2A-inoculated potato plants were grown in control (I) or salt stress conditions (I+S), the inhibition observed in shoot and root length and the decrease in the number of leaves and total biomass was less severe (I vs. I+S). Moreover, chlorophyll content was not reduced. When comparing
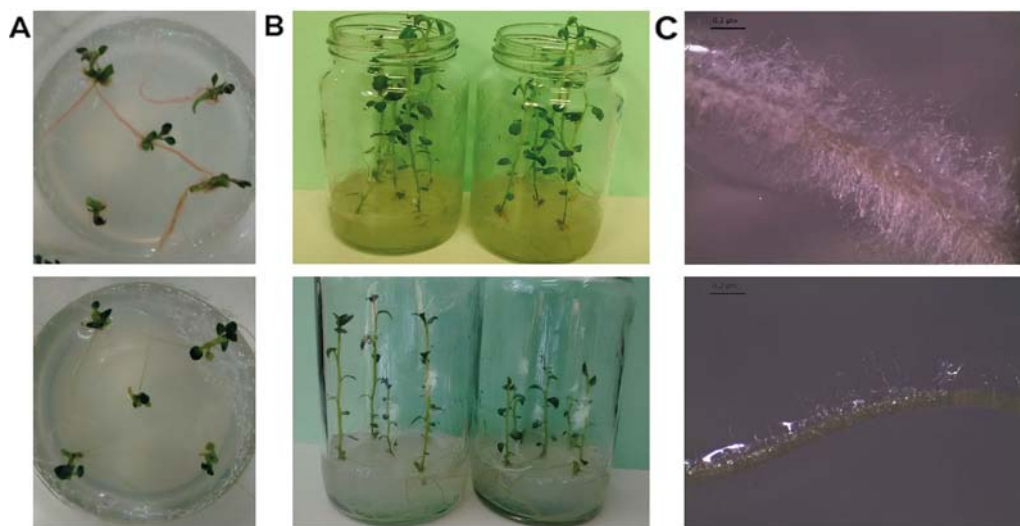
FIGURE 1 | *Methylobacterium* sp. 2A was found associated with roots of potato plants established *in vitro*. **(A)** Bacterial pink-pigmentation was evident in the roots of one-week-old potato plants. Five-weeks-old plants associated with *Methylobacterium* sp. 2A exhibited increased growth **(B)** and increased root hair density **(C)**. Plants not associated with *Methylobacterium* sp. 2A that were micropropagated simultaneously are shown in lower panels as reference. Bar: 0.2 μm.
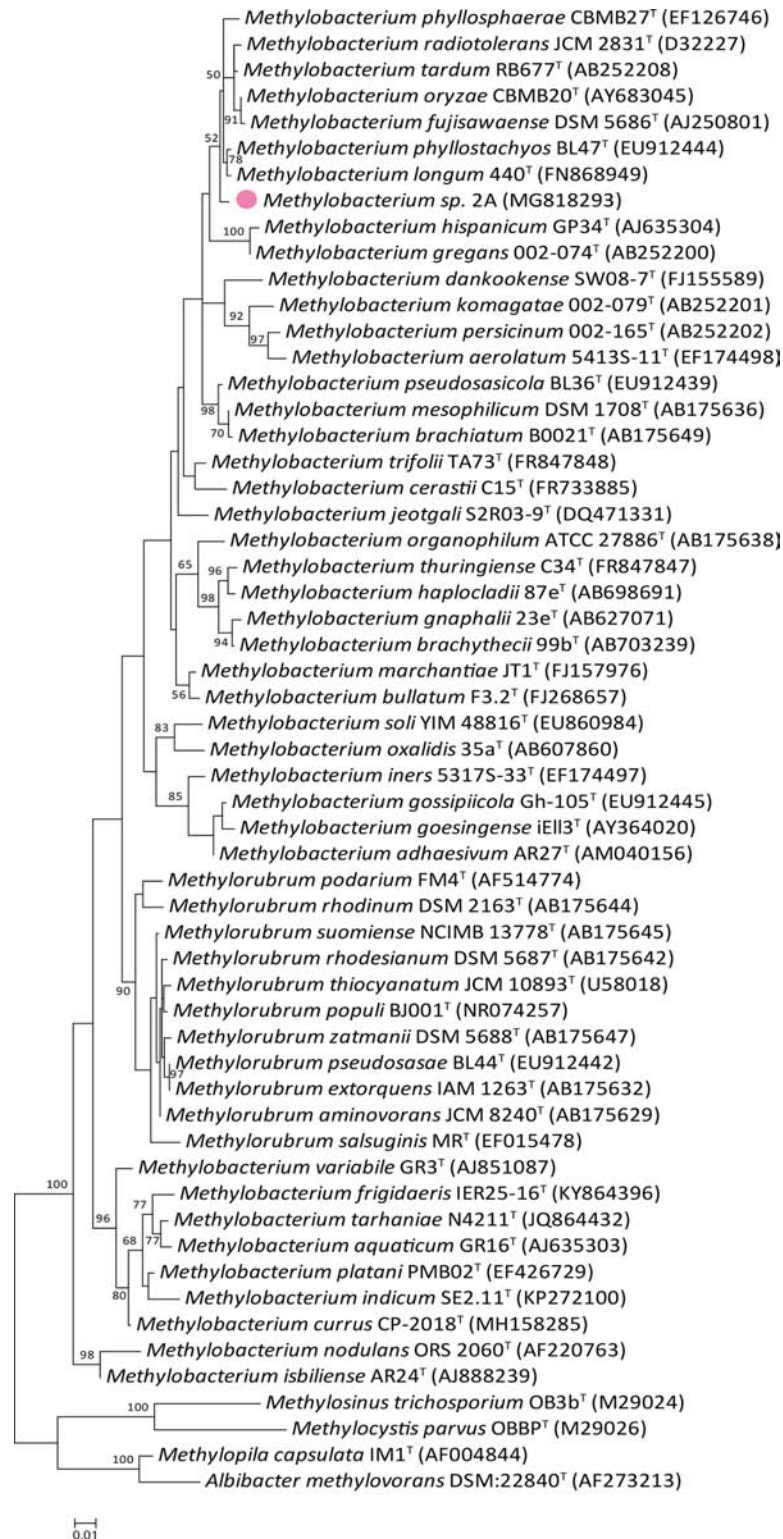
FIGURE 2 | Phylogenetic tree of *Methylobacterium* sp. 2A and related members of the *Methylobacterium* genus based on *16S rDNA* gene sequence comparison. The maximum likelihood tree was reconstructed using the Kimura two-parameter model. The frequencies with which a given branch appeared in 1,000 bootstrap replications are indicated at nodes. Values below 50% are not displayed. Accession numbers are indicated in brackets. *Methylobacterium* sp. 2A position is indicated with a pink colored filled dot. Eleven species recently housed in the new *Methylorubrum* genus are included and four members of the Methylocystaceae family were used as outgroups.
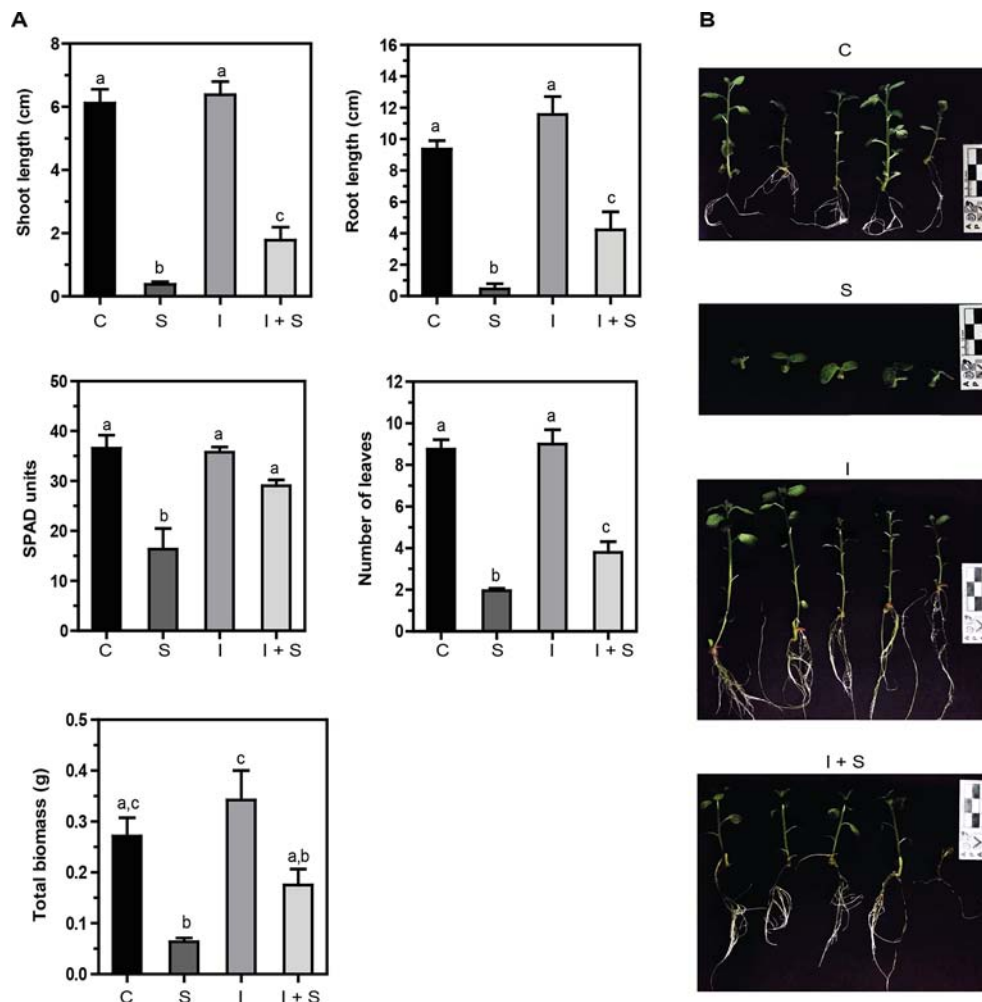
FIGURE 3 | *Methylobacterium* sp. 2A mitigates salt stress in inoculated potato plants. Internodes from *in vitro* plants, inoculated or not with *Methylobacterium* sp. 2A, were grown in solid MS media (C or I) or in MS media with the addition of 50 mM NaCl (S or I+S). **(A)** Shoot and root length (cm), SPAD units, the number of leaves, and total biomass (g) were determined. Mean ± SEM of three biological replicates, with three technical replicates each, were plotted. Two-way ANOVA analysis was performed and Tukey´s HSD test was applied. Different letters above the bars indicate significant differences (p < 0.05). **(B)** Representative pictures of each condition are shown.

noninoculated versus *Methylobacterium* sp. 2A-inoculated plants under salt conditions (S vs. I+S), a significant difference was observed in most parameters, indicating that this isolate is able to mitigate the negative effect of salinity.

Arabidopsis plantlets inoculated or not with *Methylobacterium* sp. 2A were grown under control or salt stress conditions. After a week of conducting the experiment, a significant increase in lateral root density (p < 0.05) and in the number of leaves (p < 0.01) was observed in *Methylobacterium* sp. 2A-inoculated plants, both under control and stress conditions, compared to noninoculated ones (C vs. I, and S vs. I+S; **Figures 4A, B**). The reduction in rosette diameter observed in control plants under salt stress (C vs. S; **Figure 4A**) was not perceived when *Methylobacterium* sp. 2A was present (I vs. I+S). As depicted in **Figure 4A**, a sixfold increase in

catalase activity was observed upon salt stress in noninoculated plants (C vs. S), but not in *Methylobacterium* sp. 2A-inoculated ones (C vs. I+S). However, upon inoculation, catalase activity increased three-fold under control conditions (C vs. I). The increase observed in catalase activity in stressed plants (S) was not sufficient to reduce H_2O_2 content, in fact, these plants present fivefold more peroxide than control ones (p = 0.01). However, no significant difference in peroxide content was observed when comparing control with *Methylobacterium* sp. 2A-inoculated plants under control (p = 0.656) or under salt stress conditions (p = 0.651). Furthermore, a similar peroxide content was observed in *Methylobacterium* sp. 2A-inoculated plants grown under control or saline conditions (p = 0.796). Our results indicate that this isolate can exert a salt-protective effect on different plant species.
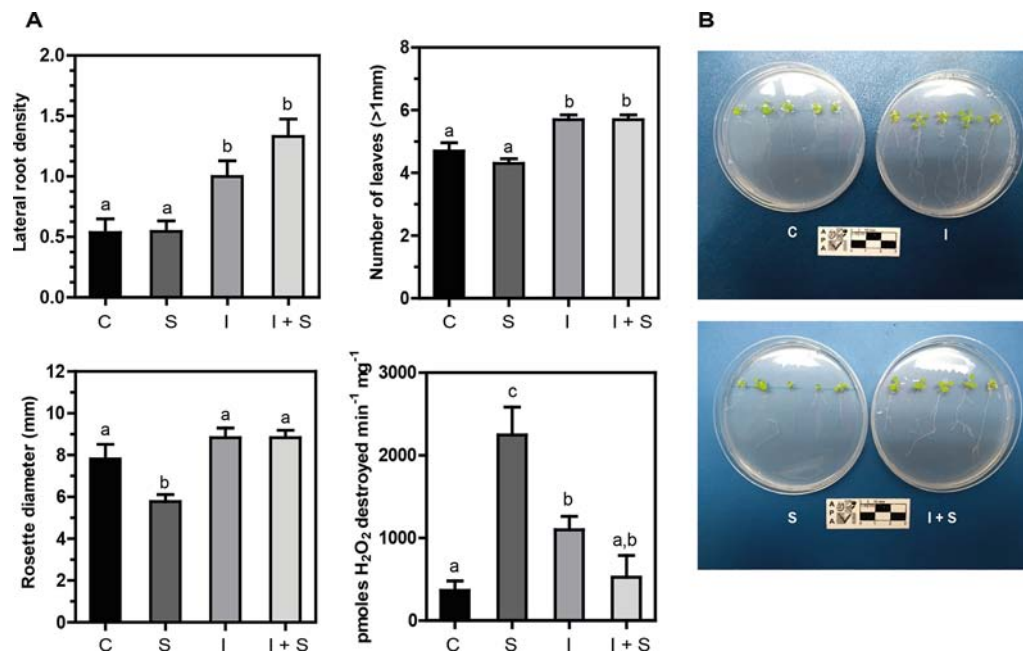
FIGURE 4 | *Methylobacterium* sp. 2A mitigates salt stress in inoculated Arabidopsis plants. Col-0 seeds were inoculated or not with *Methylobacterium* sp. 2A and grown in solid MS media (C or I) or in MS media with the addition of 75 mM NaCl (S or I+S). **(A)** Lateral root density, the number of leaves, rosette diameter (mm), and catalase activity were determined. Catalase activity is expressed as pmoles H_2O_2 destroyed min^{-1} mg^{-1}. Mean ± SEM of three biological replicates, with three technical replicates each, were plotted. Two-way ANOVA analysis was performed and Tukey´s HSD test was applied. Different letters above the bars indicate significant differences ($p < 0.05$). **(B)** Representative pictures of each condition are shown.

Methylobacterium sp. 2A Displays Biocontrol Activity Against *P. infestans*

PGPR can promote plant growth indirectly by preventing the deleterious effects of plant pathogens. Our result with *in vitro* plants suggested that *Methylobacterium* sp. 2A protected potato plants against *P. infestans* (**Figure S2**). In order to evaluate the biocontrol effect of this strain, an in plate confrontation assay against *P. infestans, B. cinerea, Fusarium* sp*., R. solani,* and *P. dissotocum* was performed. As shown in **Figures 5A, B** and **S3**, *Methylobacterium* sp. 2A inhibited mycelial growth of *P. infestans* (24.8%), *B. cinerea* (42.1%), and *Fusarium* sp. (34.7%). On the contrary, no notorious effect was observed against *R. solani* (7.2%) and *P. dissotocum* (1.2%).

To further analyze the antagonistic effect of *Methylobacterium* sp. 2A, *P.infestans* infection assays were conducted on greenhouse plants that were previously sprayed with *Methylobacterium* sp. 2A (I) or with water (controls, C). At 5 dpi we evaluated the presence and area of necrotic lesions (**Figures 5C, D**). No lesions were observed in leaves sprayed with *Methylobacterium* sp. 2A (I), while lesions were evident in *P. infestans* treated ones (Pi-60 and I+Pi-60). However, lesion size was significantly bigger in the absence of *Methylobacterium* sp. 2A and 87.5% of infected leaves presented chlorosis. On the other hand, in *Methylobacterium* sp. 2A-sprayed plants, only 28.5% of the leaves were chlorotic. This result suggests that *Methylobacterium* sp. 2A is able to restrict *P. infestans* growth.

Induced systemic resistance (ISR) is a physiological "state of enhanced defensive capacity" elicited by PGPR where the plant's innate defenses are potentiated against subsequent biotic challenges (reviewed in Pieterse et al., 2014). We previously reported that the pathogenesis-related protein 1b (*StPR-1b*) and phenylalanine ammonia-lyase (*StPAL*) genes were induced in infected and in distal potato leaves upon *P. infestans* infection (Fantino et al., 2017). Therefore, we decided to analyze their expression in leaves of plants sprayed with *Methylobacterium* sp. 2A that were then infected or not with Pi-60 and compare them with leaves infected with Pi-60 that were not previously inoculated (**Figures 5E, F**). At the time point analyzed, no induction of *StPR-1b* or *StPAL* expression was observed in inoculated leaves (I). *StPR-1b* expression was strongly upregulated (*ca.* 35-fold, $p < 0.001$) in the presence of the oomycete (Pi-60), however, no upregulation was observed when *P. infestans* zoospores were spotted in leaves previously sprayed with *Methylobacterium* sp. 2A (I+Pi-60). On the other hand, *StPAL* was induced tenfold in infected leaves (Pi-60), and fourfold in I+Pi-60 leaves. The reduction observed in the expression of *PR-1b* upon *P. infestans* infection in *Methylobacterium* sp. 2A-inoculated plants suggests that the defense mechanisms triggered by this isolate do not involve this gene.

General Genome Features of *Methylobacterium* sp. 2A

Our data indicated that this isolate has the potential to be a PGPR, so we decided to sequence its genome in order to identify the genes
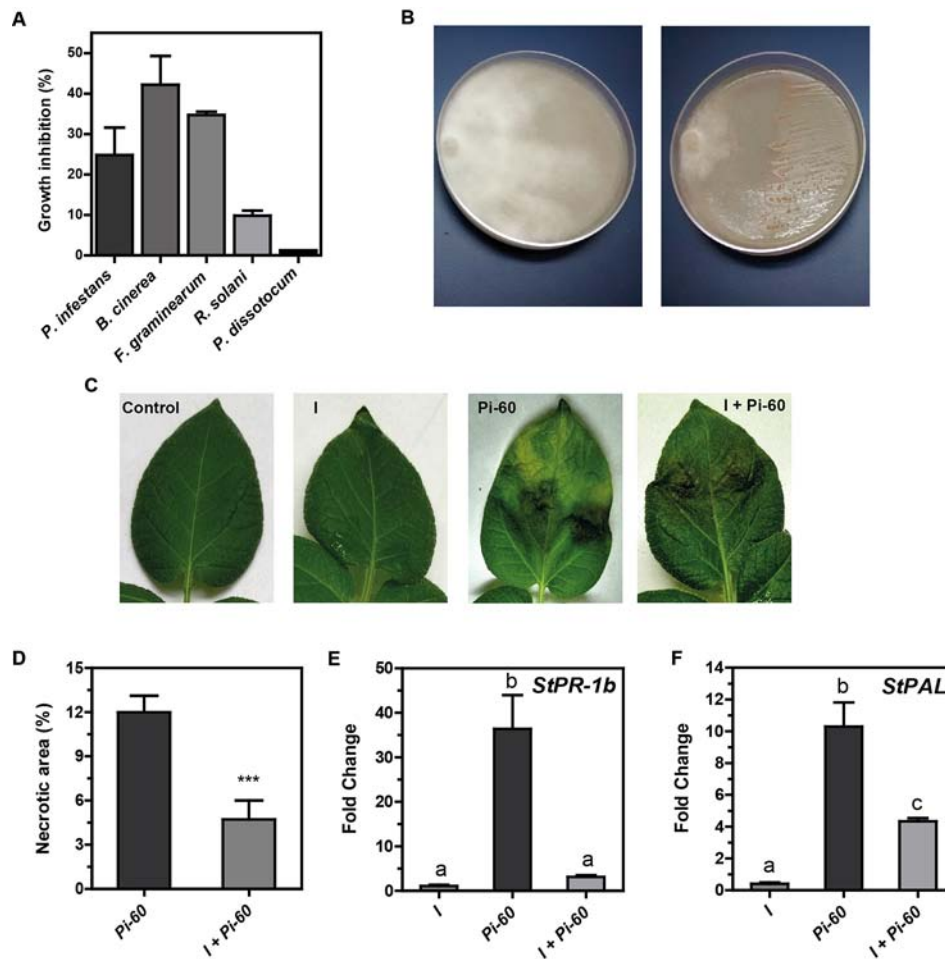
FIGURE 5 | *Methylobacterium* sp. 2A is effective in controlling *P. infestans*. **(A)** Biocontrol activity of *Methylobacterium* sp. 2A against different phytopathogens in a dual confrontation assay. The mycelial growth inhibition (%) was calculated comparing the radial size of the pathogen colony in the dual culture and in control plates. **(B)** Illustrative image showing *Methylobacterium* sp. 2A antagonistic effect against *P. infestans*; left plate: pathogen; right plate: dual culture (pathogen + *Methylobacterium* sp. 2A). Images illustrating the antagonistic effect of *Methylobacterium* sp. 2A against the other pathogens are shown in **Figure S3**. **(C)** Greenhouse potato plants were inoculated or not with *Methylobacterium* sp. 2A 48 h prior to infection with *P. infestans*. Illustrative images show control, inoculated (I), infected (Pi-60) and inoculated and infected (I+Pi-60) apical leaflets. **(D)** Histogram depicting the percentage of necrotic area (%) in leaves infected with *P. infestans* that were previously inoculated (I+Pi-60) or not (Pi-60) with *Methylobacterium* sp. 2A. RT-qPCR analysis of *StPR-1b* **(E)** and *StPAL* **(F)** in inoculated leaves (I), in leaves infected with *P. infestans* (Pi-60) and in leaves inoculated with *Methylobacterium* sp. 2A and infected with *P. infestans* (I+Pi-60). EF-1α was used as a reference gene. Mean ± SEM of three biological replicates, with three technical replicates each, were plotted. Two-way ANOVA analysis was performed and Tukey´s HSD test was applied. Different letters above the bars indicate significant differences in transcript levels between treatments (p < 0.001).

that could be responsible for this effect. *Methylobacterium* sp. 2A genomic sequence is 6,395,352 bp in length, and G+C content is 69.34 mol%. The ANI values between *Methylobacterium* sp. 2A and *M. phyllosphaerae* CBMB27 (GenBank accession no. NZ_CP015367.1), *M. oryzae* CBMB20 (GenBank accession no. NZ_CP003811.1), *M. radiotolerans* JCM 2831 (GenBank accession no. NC_010505.1), and *M. phyllostachyos* BL47 (GenBank accession no. NZ_FNHS00000000.1) were 84.53%, 84.37%, 85.21%, and 87.79%, respectively. These results were lower than the established threshold of 95%–96% ANI for prokaryotic species boundary (Chun et al., 2018). Furthermore, *in silico* DDH values between *M. phyllosphaerae, M. oryzae, M. radiotolerans,* and *M. phyllostachyos* were 30.50, 30.50, 31.00, and

37.00, respectively. The abovementioned genomic results suggest that this species has not been described so far.

A total of 6,142 coding DNA sequences (CDSs) and 55 structural RNAs (50 tRNAs) were predicted; 2,022 (33%) were classified as hypothetical and 2,419 CDSs (40%) were assigned to RAST subsystems (**Figure S4**). Using the GenBank file of the annotated genome as input, the Pathologic software automatically assigned 306 pathways and 2,116 enzymatic reactions.

Plant Growth Promotion Traits in *Methylobacterium* sp. 2A Genome

Genes involved in chemotaxis and motility were found in *Methylobacterium* sp. 2A genome; there are 52 genes encoding

methyl-accepting chemotaxis proteins (MCP), a *cheV*, the *cheAWRB* gene cluster, and *cheY*. The chemotaxis response regulator *cheA* gene is essential for motility toward root exudates (De Weert et al., 2002). Moreover, it has 31 flagellar-related genes grouped in at least five clusters (**Table S4**). As other methylotrophic bacteria, *Methylobacterium* sp. 2A genome contains the *mxa* cluster (12,475 bp) responsible for methanol oxidation.

In addition, several genes were found in the *Methylobacterium* sp. 2A genome attributable to the PGP traits observed in potato and Arabidopsis plants (**Figure 6** and **Table S4**). We identified the five enzymes involved in L-tryptophan biosynthesis that are encoded by seven genes (*trp A-G)* in this isolate as in all microbial genomes. Though the chemical reaction steps are highly conserved, the genes of the pathway enzymes show considerable variations in arrangements, operon structure, and regulation in diverse microbial genomes (Priya et al., 2014). In *Methylobacterium* sp. 2A, *TrpEGDC* genes are clustered in an operon while *trp A, B,* and *F* are unconnected. L-tryptophan is the auxin (IAA) precursor.

Two proposed IAA biosynthesis pathways were identified: the indole-3-acetamide (IAM) and the indole-3-acetonitrile (IAN) pathways. The IAM pathway consists of two-steps were L-tryptophan is first converted to IAM by the enzyme tryptophan-2-monooxygenase *(IaaM),* and then IAM is converted to IAA by an aliphatic amidase *(AmiE)*. In the IAN pathway, IAN can first be converted to indole-3-acetamide (IAM) by nitrile hydratase and then IAM is converted to IAA by an aliphatic amidase. Also, a nitrilase enzyme has been suggested to convert IAN to IAA directly without the IAM intermediate step. The production of IAA was confirmed using the Salkowski reagent. In fact, *Methylobacterium* sp. 2A produced higher levels of IAA than the *A. brasilense* Az39 and *Pantoea* sp. strains (**Figure 7A**). This tightly correlates with the increase observed in root hair development in potato and in lateral root density in Arabidopsis. As many PGPR associated with plant roots, *Methylobacterium* sp. 2A contains the gene encoding 1-aminocyclopropane-1-carboxylate (ACC) deaminase that can metabolize ACC, the immediate precursor of ethylene, thereby decreasing plant ethylene production and increasing plant growth.

Cobalamin (Vitamin B_{12}) has been suggested to stimulate plant development and could be synthesized either *via de novo* or salvage pathways. For *de novo* biosynthesis more than thirty enzymatic steps are required, while the salvage pathway is a cost-effective way for bacteria (Palacios et al., 2014). In the *Methylobacterium* sp. 2A genome, all the genes involved in the aerobic *de novo* biosynthesis were identified (**Table S4**); some of them are grouped in clusters of genes: *cobP/UA/OQ, cobST, cobSC,* and *cobWNGHIJKLMB* (**Figure 6**). *Methylobacterium* sp.
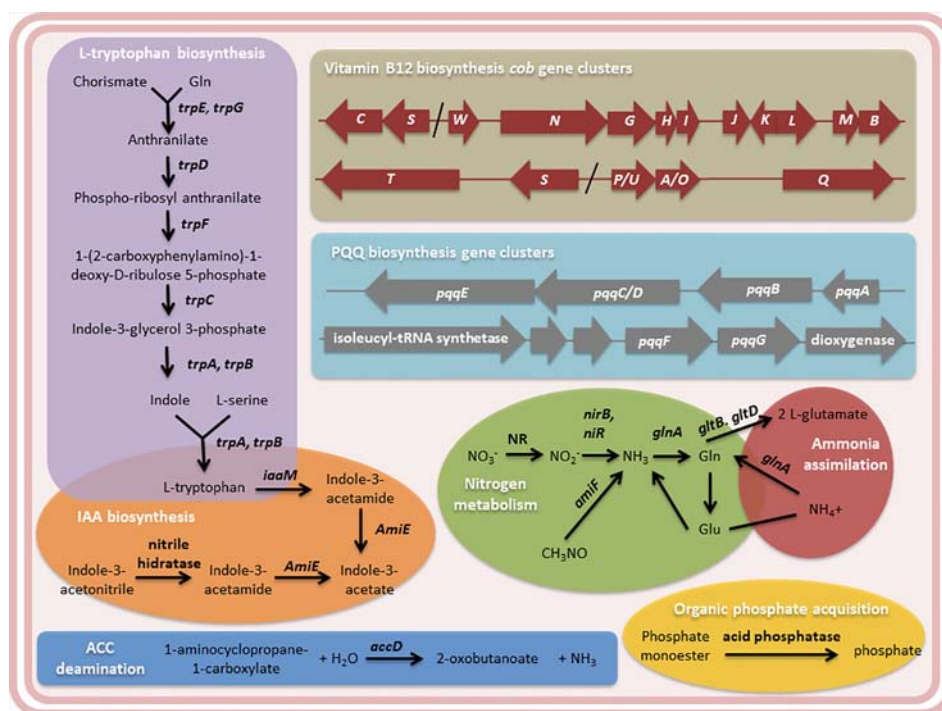
FIGURE 6 | The genome of *Methylobacterium* sp. 2A contains genes that may contribute to plant growth stimulation and fertilization. Metabolic pathways related to L-tryptophan and indole acetic acid (IAA) biosynthesis, and N_2 metabolism and ammonia assimilation pathways are shown. The *cob* gene clusters involved in vitamin B12 biosynthesis and the pyrroloquinoline quinone (*PQQ*) operons are indicated. PQQ is a cofactor of GDH, involved in phosphate solubilization. 1-aminocyclopropane-1-carboxylate (ACC)-deamination and organic phosphate acquisition reactions are included. Gene annotation, EC number, and KEGG gene names of the enzymes mentioned in the diagram are listed in **Table S4**.
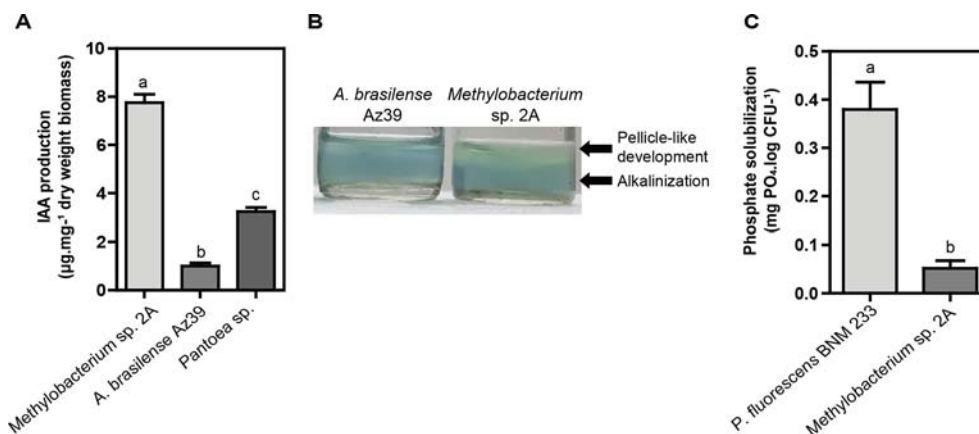
FIGURE 7 | Plant growth-promoting (PGP) abilities of *Methylobacterium* sp. 2A. **(A)** *Methylobacterium* sp. 2A produced higher indole acetic acid (IAA) amounts than *A. brasilense* Az39 and *Pantoea* sp. strains used as positive controls. **(B)** The nitrogen-fixing ability of *Methylobacterium* sp. 2A was inferred from its growth in a nitrogen-free culture medium (Nfb); the pH change observed was similar to that of the well-known nitrogen-fixing bacterium *A. brasilense*. **(C)** Inorganic phosphate solubilization was confirmed in liquid NBRIP medium. *P. fluorescens* strain BNM 233 was used as positive control. One-way ANOVA analysis was performed and Tukey's HSD test was applied. Different letters above the bars indicate significant differences (p < 0.05).

2A harbors two copies of *cobS*. In gram-negative bacteria, exogenous corrinoids are transported into the cell through an ATP-binding cassette (ABC) transport system consisting of *Btu* genes. *BtuB*, *BtuC,* and *BtuN* genes coding for receptors were identified suggesting that this strain can also synthesize cobalamin by absorbing corrinoids *via* the salvage pathway (**Table S4**).

On the other hand, *Methylobacterium* sp. 2A presents the two-component regulatory genes involved in nitrogen fixation, *NtrBC*, and *NtrXY* (**Table S4**). This system allows the survival of the bacteria in a nitrogen-depleted medium. In fact, we confirmed that *Methylobacterium* sp. 2A is able to grow in NFb medium (**Figure 7B**). A two-component nitrogen fixation transcriptional regulator, *FixJ* and a *FixL* gene were identified in its genome, together with *nifJ*, *nifP*, *nifS*, and *nifU* genes (**Table S4**); however, other *nif* genes were not assigned by the RAST annotation. In addition, it harbors the enzymes involved in nitrogen metabolism and ammonia assimilation (**Figure 6**, **Table S4**).

Moreover, there are four glucose dehydrogenase (*gdh*) genes present in *Methylobacterium* sp. 2A genome (**Table S4**). GDH requires pyrroloquinoline-quinone (PQQ) as a redox cofactor for direct oxidation of glucose to gluconic acid, which diffuses and helps in acidic solubilization of mineral phosphates in soil. Synthesis of PQQ requires the expression of six genes, *pqqA-G* (Puehringer et al., 2008). *Methylobacterium* sp. 2A harbors the complete *pqqABC/DE* operon where *pqqC* and *pqqD* are fused into a single polypeptide designated *pqqC/D* (**Figure 6**, **Table S4**) as in *Methylorubrum extorquens* AM1 (Toyama et al., 1997). In *M. extorquens* the cluster containing *pqqF/G* is a group of six genes transcribed in the same direction, the first is predicted to encode an isoleucyl tRNA synthetase and the sixth a dioxygenase (Zhang and Lidstrom, 2003). In *Methylobacterium* sp. 2A we identified this same region (**Figure 6**). The DNA sequence

encoding the last five genes shared 84.65% identity with the *M. extorquens* cluster. In addition, there are three kinds of microbial enzymes that can solubilize organic phosphate: a nonspecific acid phosphatase, a phytase, and C-P lyase or phosphonatase. *Methylobacterium* sp. 2A genome has three genes encoding acid phosphatases that can release phosphate from phosphoric ester or phosphoric anhydride (**Figure 6**).

Phosphate solubilization is an important trait to increase the availability and uptake of mineral nutrients for plants. From the genomic data, we can infer that *Methylobacterium* sp. 2A is able to solubilize phosphate, so we grew it in NBRIP solid and liquid media containing TCP. In the solid media, a yellow halo indicative of phosphate solubilization was observed around the positive control, *P. fluorescens* BNM 233 colony, but not around *Methylobacterium* sp. 2A. On the other hand, when grown in liquid media, *Methylobacterium* sp. 2A was able to solubilize mineral phosphate but to a lower extent than the positive control (**Figure 7C**).

In addition, antiSMASH analysis identified 10 genes associated with terpene biosynthesis (four phytoene synthases, four phytoene desaturases, one phytoene cyclase, and one phytoene dehydrogenase) in three genomic regions. Recent reports highlight the ecological importance of microbial terpenes emitted by beneficial microorganisms that function as volatile organic compounds (mVOCs) altering plant development (reviewed in Schulz-Bohm et al., 2017).

The dual confrontation assays and the *in planta* infection assays suggest that this isolate may produce compounds to reduce the growth of phytopathogens. In fact, its genome carries a chitinase, a phenazine biosynthesis *PhzF* gene, two genes involved in 4-hydroxybenzoate synthesis (the transferase, *ubiA*, and a transporter), a penicillin acylase (PAH), and a *CvpA* gene encoding a Colicin V production protein. Several copies of the genes responsible for translocating colicins across the

membrane (_TolA, TolB, TolC, TonB_) were also present (**Table S4**). In addition, genes responsible for aerobactin and enterobactin biosynthesis are present in its genome (**Table S4**). The presence of a siderophore biosynthetic gene cluster (_IucABCD_) and a siderophore transporter protein was confirmed with antiSMASH analysis. Concerning antibiotic resistance, _Methylobacterium_ sp. 2A has two chloramphenicol acetyltransferases consistent with the observed chloramphenicol resistance, and a B-lactamase gene that could be responsible for the ampicillin resistance.

Furthermore, its genome encodes ROS scavenging enzymes that could alleviate oxidative damage: eleven catalases, two glutathione reductases, two superoxide dismutases, and five peroxidases, among them, are two glutathione, a cytochrome c, and a thiol peroxidase (**Table S4**).

DISCUSSION

PGPR stimulate plant growth by either providing plant hormones (auxin or cytokinin), lowering plant ethylene levels through the action of the enzyme 1-aminocyclopropane-1-carboxylate (ACC) deaminase, helping in the acquisition of nutritional resources (nitrogen, iron, and phosphorus), or antagonizing phytopathogens (Glick, 1995; Glick et al., 1999). _In vitro_ potato and Arabidopsis plants inoculated with _Methylobacterium_ sp. 2A were more developed than noninoculated plants of the same age, suggesting that this isolate could stimulate plant growth. Genome sequencing indicates that _Methylobacterium_ sp. 2A presents several traits related to phyto-stimulation and phyto-fertilization, such as L-tryptophan and IAA biosynthesis, phosphate acquisition and solubilization, siderophore and vitamin B12 biosynthesis, N_2 fixation and ammonia assimilation. In fact, the increase observed in root hair development and in lateral root density are auxin-associated phenotypes (Overvoorde et al., 2010) that are in agreement with the IAA produced by this isolate.

The experiments presented in this work were conducted using _in vitro_ plants; however, preliminary experiments were performed with tubers inoculated with _Methylobacterium_ sp. 2A planted in the greenhouse. These plants were more developed than noninoculated ones (data not shown). Apart from modifying IAA levels, this isolate could stimulate plant growth either by fixing N_2 or by improving phosphate acquisition. We recognized several genes involved in fixing N_2 in _Methylobacterium_ sp. 2A; however, we could not identify the _NifH_ gene of nitrogenase. Multiple independent horizontal gene transfer events took place in the evolution of nitrogen-fixing bacteria. Leaf isolates from the _Methylobacterium_ genus present sequences highly divergent from _NifH_ consensus sequence, more related to the Pfam NifH/frxC-family protein, such as BchX (Madhaiyan et al., 2015). _Methylobacterium_ sp. 2A harbors a chlorophyllide reductase iron protein subunit X (BchX); this fact, together with its ability to grow in Nfb medium, suggests that it is a diazotrophic bacterium. Nevertheless, the

nitrogenase activity of _Methylobacterium_ sp. 2A should be assessed by the acetylene reduction assay to conclude this.

Potato is a high fertilizer-demanding crop, which requires 250 kg ha^{-1} of nitrogen and 150 kg ha^{-1} of phosphorus to get an optimum yield (Khan et al., 2012). Phosphorus availability is often limited due to the formation of insoluble inorganic and organic phosphate complexes (Adesemoye and Kloepper, 2009). The genomic data indicate that _Methylobacterium_ sp. 2A is able to mobilize organic sources of phosphorus _via_ acid phosphatases, and solubilize mineral phosphate, thus contributing to plant phosphate acquisition. The solubilization assay confirmed that this isolate is able to solubilize inorganic phosphate. Therefore, _Methylobacterium_ sp. 2A emerges as a potential biofertilizer for potato plants to improve phosphorus and nitrogen uptake as reported for other PGPR (Hanif et al., 2015; Naqqash et al., 2016).

Furthermore, our results indicate that _Methylobacterium_ sp. 2A mitigates the deleterious effects of salt stress in potato and Arabidopsis plants. Under stress conditions, ethylene regulates plant homeostasis and reduces root and shoot growth. However, under saline conditions, _Methylobacterium_ sp. 2A-inoculated plants presented higher biomass, increased root and shoot growth, and more chlorophyll content than noninoculated plants. The modification in root-architecture promoted by auxin possibly increases the uptake of water and nutrients, thus explaining the improved fitness of the plants. In addition, like many other PGPRs, this isolate encodes an ACC deaminase; it was reported that degradation of ACC by bacterial ACC deaminase releases plant stress and rescues normal plant growth (Glick et al., 2007). In particular, the ACC deaminase-producing bacteria _Achromobacter piechaudii_ ARV8, lowered the level of ethylene and prevented inhibition of plant growth when inoculated in tomato plants grown in the presence of high salt (Mayak et al., 2004a) and drought stress (Mayak et al., 2004b). Therefore _Methylobacterium_ sp. 2A could alleviate salt stress by increasing auxins and decreasing ethylene content in the rhizosphere.

Another important feature observed was that when grown under salt stress, _Methylobacterium_ sp. 2A-inoculated Arabidopsis plants presented lower levels of peroxide than noninoculated ones. Its genome contains several ROS scavenging enzymes and was positive for catalase reaction. It was reported that ROS degradation could be the result of enhancing plant antioxidant activities by PGPR thus protecting plants from salt toxicity (Chu et al., 2019). Catalase activity increased upon inoculation with _Methylobacterium_ sp. 2A in control conditions, but its activity was lower than in noninoculated plants when exposed to salt stress. A meta-analysis performed by Pan et al. (2019) suggests that PGPR helps host plants to alleviate oxidative stress mainly through reducing the generation of ROS formed on the onset of ionic stress, not _via_ scavenging ROS by accumulating antioxidant enzymes in host plants. This might be the case for _Methylobacterium_ sp. 2A.

To successfully thrive within bacterial populations, _Methylobacterium_ sp. 2A has to be highly competitive. Its genome harbors bacterial chemotaxis and motility genes, and several genes involved in the production of antimicrobial

compounds (listed in **Table S4**). The first could be responsible for root association, and the later could function as antibiotics conferring an advantage against other microorganisms while protecting the plant against phytopathogens. In this regard, chitinase has been associated with protection against plant fungal pathogens (Kumar et al., 2018); colicins are the most representative bacteriocins produced by gram-negative bacteria (Beneduzi et al., 2012), while chemicals such as phenazine and 4-hydroxybenzoate act as antibiotics and suppress plant pathogenic microbes (Gupta et al., 2015). As an example, phenazine is known to suppress the plant pathogen *F. oxysporum* (Chin-A-Woeng et al., 2003). It was reported that PAH participates in acyl-homoserine lactones degradation thus interfering with quorum sensing of competing bacteria (Mukherji et al., 2014). Genes related to aerobactin and enterobactin biosynthesis were also identified in the *Methylobacterium* sp. 2A genome. Siderophore producing bacteria are of significant importance in the field of agriculture; in addition to supplementing iron to the plant, siderophores prevent the growth of the soil-borne phytopathogens (Fernández Scavino and Pedraza, 2013).

The biocontrol capacity of *Methylobacterium* sp. 2A was evidenced in the dual confrontation assays against *B. cinerea*, *F. graminearum* and *P. infestans*, and in greenhouse plants infected with *P. infestans*. Lesion size was smaller and chlorosis was seldom observed in *Methylobacterium* sp. 2A-inoculated plants infected with Pi-60 compared to noninoculated plants. In addition to producing antimicrobial compounds, PGPR are capable of potentiating plant defense mechanisms against pathogens. ISR has been reported in potato after inoculation with a Rhizobium strain (Reitz et al., 2001). This response differs from systemic acquired resistance (SAR) and is not salicylic acid (SA) dependent (Pieterse et al., 2014). In fact, at the time points analyzed we observed that *St-PR1b* was not induced in *Methylobacterium* sp. 2A-inoculated leaves after infection with Pi-60. On the contrary, a very strong induction of *StPR-1b* was observed when noninoculated plants were infected with *P. infestans*. This induction, however, did not restrain the oomycete growth. On the other hand, induction of *StPAL* was observed in *Methylobacterium* sp. 2A-inoculated plants upon infection with *P. infestans*. While its antagonizing effect is clear, at the moment, we cannot infer which plant defense mechanisms are primed by *Methylobacterium* sp. 2A.

In this work, we present a new PGPR isolate capable of promoting plant growth under control and salt stress conditions in two dicots. Probably this beneficial effect is not limited to potato and Arabidopsis plants. In addition, in dual confrontation assays, *Methylobacterium* sp. 2A restricted the growth of two necrotrophic fungi and of the hemibiotrophic oomycete *P. infestans*, denoting broad spectrum antagonism. *In planta* assays confirmed that it was able to reduce *P. infestans* deleterious effect in potato plants. These promissory results allow us to envisage that *Methylobacterium* sp. 2A has the potential to be used as a substitute for chemical fertilizers and fungicides, preventing pollution in farmlands. This research is the first approach to understand the PGP capacities of *Methylobacterium* sp. 2A; we

have to test this isolate in the field in order to assess the productivity, efficacy, and viability of this inoculum. Certainly, this eco-friendly microbial technology will contribute to sustainable agricultural practices and is an alternative strategy to improve crop production for an increasing world population.

DATA AVAILABILITY STATEMENT

This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the BioProject PRJNA560067. The version described in this paper is version VUOK00000000.1.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00071/full#supplementary-material

REFERENCES

Adesemoye, A. O., and Kloepper, J. W. (2009). Plant-microbes interactions in enhanced fertilizer-use efficiency. *Appl. Microbiol. Biotechnol.* 85, 1–12. doi: 10.1007/s00253-009-2196-0

Aebi, H. (1984). Catalase *in vitro*. *Meth. Enzymol.* 105, 121–126. doi: 10.1016/S0076-6879(84)05016-3

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75. doi: 10.1186/1471-2164-9-75

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Beneduzi, A., Ambrosini, A., and Passaglia, L. M. P. (2012). Plant growth-promoting rhizobacteria (PGPR): their potential as antagonists and biocontrol agents. *Genet. Mol. Biol.* 35, 1044–1051. doi: 10.1590/S1415-47572012000600020

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., et al. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47, W81–W87. doi: 10.1093/nar/gkz310

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27, 578–579. doi: 10.1093/bioinformatics/btq683

Cantos de Ruiz, S. T., Andrade, F. H., and Mendiburu, A. (1989). Rendimiento potencial del cultivo de papa en Balcarce, causas que limitan la productividad real. *Rev. Latinoam. La Papa.* 2, 29–45.

Caspi, R. (2005). MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* 34, D511–D516. doi: 10.1093/nar/gkj128

Chen, W., and Kuo, T. T. (1993). A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. *Nucleic Acids Res.* 21, 2260. doi: 10.1093/nar/21.9.2260

Chin-A-Woeng, T. F. C., Bloemberg, G. V., and Lugtenberg, B. J. J. (2003). Phenazines and their role in biocontrol by Pseudomonas bacteria. *New Phytol.* 157, 503–523. doi: 10.1046/j.1469-8137.2003.00686.x

Chu, T. N., Tran, B. T. H., Van Bui, L., and Hoang, M. T. T. (2019). Plant growth-promoting rhizobacterium Pseudomonas PS01 induces salt tolerance in Arabidopsis thaliana. *BMC Res. Notes* 12, 11. doi: 10.1186/s13104-019-4046-1

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., Costa, M. S., et al. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68, 461–466. doi: 10.1099/ijsem.0.002516

Coleman-Derr, D., and Tringe, S. G. (2014). Building the crops of tomorrow: advantages of symbiont-based approaches to improving abiotic stress tolerance. *Front. Microbiol.* 5, 1–6. doi: 10.3389/fmicb.2014.00283

Cordero, I., Balaguer, L., Rinco, A., and Pueyo, J. J. (2018). Inoculation of tomato plants with selected PGPR represents a feasible alternative to chemical fertilization under salt stress. *J. Plant Nutr. Soil Sci.* 181, 694–703. doi: 10.1002/jpln.201700480

Dahal, K., Li, X.-Q., Tai, H., Creelman, A., and Bizimungu, B. (2019). Improving potato stress tolerance and tuber yield under a climate change scenario – a current overview. *Front. Plant Sci.* 10, 563. doi: 10.3389/fpls.2019.00563

de la Casa, A., Ovando, G., Bressanini, L., and Martínez, J. (2014). Estimación del rendimiento potencial de papa en Córdoba, Argentina, a partir de la eficiencia en el uso de la radiación y de la productividad del agua. *Agriscientia* 31, 75–91. doi: 10.31047/1668.298x.v31.n2.16535

De Weert, S., Vermeiren, H., Mulders, I. H. M., Kuiper, I., Hendrickx, N., Bloemberg, G. V., et al. (2002). Flagella-driven chemotaxis towards exudate components is an important trait for tomato root colonization by Pseudomonas fluorescens. *Mol. Plant Microbe Interact.* 15, 1173–1180. doi: 10.1094/MPMI.2002.15.11.1173

Díaz Herrera, S., Grossi, C., Zawoznik, M., and Groppa, M. D. (2016). Wheat seeds harbour bacterial endophytes with potential as plant growth promoters and biocontrol agents of Fusarium graminearum. *Microbiol. Res.* 186, 37–43. doi: 10.1016/j.micres.2016.03.002

Ehmann, A. (1977). The van urk-Salkowski reagent—a sensitive and specific chromogenic reagent for silica gel thin-layer chromatographic detection and identification of indole derivatives. *J. Chromatogr.* 132, 267–276. doi: 10.1016/S0021-9673(00)89300-0

Fantino, E., Segretin, M. E., Santin, F., Mirkin, F. G., and Ulloa, R. M. (2017). Analysis of the potato calcium-dependent protein kinase family and characterization of StCDPK7, a member induced upon infection with Phytophthora infestans. *Plant Cell Rep.* 36, 1137–1157. doi: 10.1007/s00299-017-2144-x

Fernández Scavino, A., and Pedraza, R. (2013). "The Role of Siderophores in Plant Growth-Promoting Bacteria," in *Bacteria in Agrobiology: Crop Productivity* (Heidelberg: Springer). doi: 10.1007/978-3-642-37241-4_11

García, J. E., Maroniche, G., Creus, C., Suárez-Rodríguez, R., Ramirez-Trujillo, J. A., and Groppa, M. D. (2017). In vitro PGPR properties and osmotic tolerance of different Azospirillum native strains and their effects on growth of maize under drought stress. *Microbiol. Res.* 202, 21–29. doi: 10.1016/j.micres.2017.04.007

Garzón, J. M., and Young, M. (2016). La producción de papa en Córdoba: Aspectos básicos y potencial productivo. Ministerio de Agricultura y Ganadería de Córdoba. Córdoba. http://magya.cba.gov.ar/upload/Informe_IERAL_Papa_Nov_2016.pdf.

Glick, B. R., Patten, C. L., Holguin, G., and Penrose, D. M. (1999). *Biochemical and Genetic Mechanisms Used by Plant Growth Promoting Bacteria*, (Ontario: Imperial College Press). doi: 10.1142/p130

Glick, B. R., Todorovic, B., Czarny, J., Cheng, Z., Duan, J., and McConkey, B. (2007). Promotion of plant growth by bacterial ACC deaminase. *CRC Crit. Rev. Plant Sci.* 26, 227–242. doi: 10.1080/07352680701572966

Glick, B. R. (1995). The enhancement of plant growth by free-living bacteria. *Can. J. Microbiol.* 41, 109–117. doi: 10.1139/m95-015

Green, P. N., and Ardley, J. K. (2018). Review of the genus Methylobacterium and closely related organisms: a proposal that some Methylobacterium species be reclassified into a new genus. Methylorubrum Gen. Nov. *Int. J. Syst. Evol. Microbiol.* 68, 2727–2748. doi: 10.1099/ijsem.0.002856

Gupta, G., Parihar, S., Ahirwar, N., Snehi, D. S. K., and Singh, V. (2015). Plant growth promoting rhizobacteria (PGPR): current and future prospects for development of sustainable agriculture. *J. Microb. Biochem. Technol.* 7, 96–102. doi: 10.4172/1948-5948.1000188

Hanif, M. K., Hameed, S., Imran, A., Naqqash, T., Shahid, M., and Van Elsas, J. D. (2015). Isolation and characterization of a β-propeller gene containing phosphobacterium Bacillus subtilis strain KPS-11 for growth promotion of potato (Solanum tuberosum L.). *Front. Microbiol.* 6, 1–12. doi: 10.3389/fmicb.2015.00583

Hartmann, A., and Baldani, I. (2006). *The Genus Azospirillum. In: The Prokaryotes* (New York: Springer), 115–140. doi: 10.1007/0-387-30745-1_6

Hiscox, J. D., and Israelstam, G. F. (1979). A method for the extraction of chlorophyll from leaf tissue without maceration. *Can. J. Bot.* 57 (12), 1332–1334. doi: 10.1139/b79-163

Isayenkov, S. V., and Maathuis, F. J. M. (2019). Plant salinity stress: many unanswered questions remain. *Front. Plant Sci.* 10, 80. doi: 10.3389/fpls.2019.00080

Ito, H., and Iizuka, H. (1971). Taxonomic studies on a radio-resistant pseudomonas. *Agric. Biol. Chem.* 35, 1566–1571. doi: 10.1080/00021369.1971.10860119

Junglee, S., Urban, L., Sallanon, H., and Lopez-Lauri, F. (2014). Optimized assay for hydrogen peroxide determination in plant tissue using potassium iodide. *Am. J. Anal. Chem.* 05, 730–736. doi: 10.4236/ajac.2014.511081

Karp, P. D., Paley, S., and Romero, P. (2002). The pathway tools software. *Bioinformatics* 18, 225–232. doi: 10.1093/bioinformatics/18.suppl_1.S225

Kato, Y., Asahara, M., Goto, K., Kasai, H., and Yokota, A. (2008). Methylobacterium persicinum sp. Nov., Methylobacterium komagatae sp. Nov., Methylobacterium brachiatum sp. Nov., Methylobacterium tardum sp. Nov. and Methylobacterium gregans sp. Nov., isolated from freshwater. *Int. J. Syst. Evol. Microbiol.* 58, 1134–1141. doi: 10.1099/ijs.0.65583-0

Khan, M. Z., Akhtar, M. E., Mahmood-ul-Hassan, M., Mahmood, M. M., and Safdar, M. N. (2012). Potato tuber yield and quality as affected by rates and sources of potassium fertilizer. *J. Plant Nutr.* 35, 664–677. doi: 10.1080/01904167.2012.653072

Knief, C., Dengler, V., Bodelier, P. L. E., and Vorholt, J. A. (2012). Characterization of Methylobacterium strains isolated from the phyllosphere and description of

Methylobacterium longum sp. nov. *Antonie van Leeuwenhoek. Int. J. Gen. Mol. Microbiol.* 101, 169–183. doi: 10.1007/s10482-011-9650-6

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Kumar, M., Brar, A., Yadav, M., Chawade, A., and Pareek, N. (2018). Chitinases — potential candidates for enhanced plant resistance towards fungal pathogens. *Agriculture* 8, 88. doi: 10.3390/agriculture8070088

Lahlali, R., and Hijri, M. (2010). Screening, identification and evaluation of potential biocontrol fungal endophytes against Rhizoctonia solani AG3 on potato plants. *FEMS Microbiol. Lett.* 311, 152–159. doi: 10.1111/j.1574-6968.2010.02084.x

Lal, M., Sharma, S., Yadav, S., and Kumar, S. (2018). *Management of Late Blight of Potato. In: Potato - From Incas to All Over the World.* (London: IntechOpen). doi: 10.5772/intechopen.72472

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Madhaiyan, M., and Poonguzhali, S. (2014). Methylobacterium pseudosasicola sp. Nov. and Methylobacterium phyllostachyos sp. Nov., isolated from bamboo leaf surfaces. *Int. J. Syst. Evol. Microbiol.* 64, 2376–2384. doi: 10.1099/ijs.0.057232-0

Madhaiyan, M., Kim, B. Y., Poonguzhali, S., Kwon, S. W., Song, M. H., Ryu, J. H., et al. (2007). Methylobacterium oryzae sp. Nov., an aerobic, pink-pigmented, facultatively methylotrophic, 1-aminocyclopropane-1-carboxylate deaminase-producing bacterium isolated from rice. *Int. J. Syst. Evol. Microbiol.* 57, 326–331. doi: 10.1099/ijs.0.64603-0

Madhaiyan, M., Poonguzhali, S., Kwon, S. W., and Sa, T. M. (2009). Methylobacterium phyllosphaerae sp. Nov., a pink-pigmented, facultative methylotroph from the phyllosphere of rice. *Int. J. Syst. Evol. Microbiol.* 59, 22–27. doi: 10.1099/ijs.0.001693-0

Madhaiyan, M., Alex, T. H. H., Ngoh, S., Te,, Prithiviraj, B., and Ji, L. (2015). Leaf-residing Methylobacterium species fix nitrogen and promote biomass and seed production in Jatropha curcas. *Biotechnol. Biofuels* 8, 222. doi: 10.1186/s13068-015-0404-y

Mayak, S., Tirosh, T., and Glick, B. R. (2004a). Plant growth-promoting bacteria confer resistance in tomato plants to salt stress. *Plant Physiol. Biochem.* 42, 565–572. doi: 10.1016/j.plaphy.2004.05.009

Mayak, S., Tirosh, T., and Glick, B. R. (2004b). Plant growth-promoting bacteria that confer resistance to water stress in tomatoes and peppers. *Plant Sci.* 166, 525–530. doi: 10.1016/j.plantsci.2003.10.025

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinf.* 14, 60. doi: 10.1186/1471-2105-14-60

Mukherji, R., Varshney, N. K., Panigrahi, P., Suresh, C. G., and Prabhune, A. (2014). A new role for penicillin acylases: degradation of acyl homoserine lactone quorum sensing signals by Kluyvera citrophila penicillin G acylase. *Enzyme Microb. Technol.* 56, 1–7. doi: 10.1016/j.enzmictec.2013.12.010

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant* 15, 473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x

Naqqash, T., Hameed, S., Imran, A., Hanif, M. K., Majeed, A., and van Elsas, J. D. (2016). Differential response of potato toward inoculation with taxonomically diverse plant growth promoting rhizobacteria. *Front. Plant Sci.* 7, 1–12. doi: 10.3389/fpls.2016.00144

Okon, Y., Labandera-Gonzales, C., Lage, M., and Lage, P. (2015). Agronomic Applications of Azospirillum and Other PGPR. *Biol. Nitrogen Fixat.* 2–2, 925–936. doi: 10.1002/9781119053095.ch90

Overvoorde, P., Fukaki, H., and Beeckman, T. (2010). Auxin control of root development. *Cold Spring Harb. Perspect. Biol.* 2, 1–17. doi: 10.1101/cshperspect.a001537

Pérez-Montaño, F., Alías-Villegas, C., Bellogín, R. A., del Cerro, P., Espuny, M. R., Jiménez-Guerrero, I., et al. (2014). Plant growth promotion in cereal and leguminous agricultural important plants: From microorganism capacities to crop production. *Microbiol. Res.* 169, 325–336. doi: 10.1016/j.micres.2013.09.011

Palacios, O. A., Bashan, Y., and De-Bashan, E. L. (2014). Proven and potential involvement of vitamins in interactions of plants with plant growth-promoting bacteria—an overview. *Biol. Fert. Soils* 50, 415–432. doi: 10.1007/s00374-013-0894-3

Pan, J., Peng, F., Xue, X., You, Q., Zhang, W., Wang, T., et al. (2019). The growth promotion of two salt-tolerant plant groups with PGPR inoculation: a meta-analysis. *Sustain* 11, 378. doi: 10.3390/su11020378

Pearson, D. (1976). *General methods-determination of phosphate by the vanado-molybdate colorimetric method: in the chemical analysis of foods.* (Harlow: Longman Scientific and Technical). Available at: https://books.google.com.ar/books?id=fqsvAAAAYAAJ.

Pieterse, C. M. J., Pierik, R., and Van Wees, S. C. M. (2014). Different shades of JAZ during plant growth and defense. *New Phytol.* 204, 261–264. doi: 10.1111/nph.13029

Priya, V. K., Sarkar, S., and Sinha, S. (2014). Evolution of tryptophan biosynthetic pathway in microbial genomes: a comparative genetic study. *Syst. Synth. Biol.* 8, 59–72. doi: 10.1007/s11693-013-9127-1

Puehringer, S., Metlitzky, M., and Schwarzenbacher, R. (2008). The pyrroloquinoline quinone biosynthesis pathway revisited: a structural approach. *BMC Biochem.* 9, 8. doi: 10.1186/1471-2091-9-8

Reitz, M., Hoffmann-Hergarten, S., Hallmann, J., and Sikora, R. (2001). Induction of systemic resistance in potato by rhizobacterium Rhizobium etli strain G12 is not associated with accumulation of pathogenesis-related proteins and enhanced lignin biosynthesis. *J. Plant Dis. Prot.* 108 (1), 11–20.

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., and Peplies, J. (2015). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 929–931. doi: 10.1093/bioinformatics/btv681

Roser, M., and Ritchie, H. (2017). *Fertilizers and Pesticides* (Oxford: Our World in Data).

Ruiz Carrasco, K. B., Baroni Fornasiero, R., Tassoni, A., and Bagni, N. (2007). Identification of two phenotypes of Arabidopsis thaliana under *in vitro* salt stress conditions. *Biol. Plant* 51, 436–442. doi: 10.1007/s10535-007-0093-x

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. doi: 10.1038/nmeth.2019

Schulz-Bohm, K., Martín-Sánchez, L., and Garbeva, P. (2017). Microbial volatiles: small molecules with an important role in intra- and inter-kingdom interactions. *Front. Microbiol.* 8, 1–10. doi: 10.3389/fmicb.2017.02484

Singh, R. P., and Jha, P. N. (2017). The PGPR stenotrophomonas maltophilia SBP-9 augments resistance against biotic and abiotic stress in wheat plants. *Front. Microbiol.* 8, 1945. doi: 10.3389/fmicb.2017.01945

Sosa, E. J., Burguener, G., Lanzarotti, E., Defelipe, L., Radusky, L., Pardo, A. M., et al. (2018). Target-pathogen: a structural bioinformatics approach to prioritize drug targets in pathogens. *Nucleic Acids Res.* 46, D413–D418. doi: 10.1093/nar/gkx1015

Toyama, H., Chistoserdova, L., and Lidstrom, M. E. (1997). Sequence analysis of pqq genes required for biosynthesis of pyrroloquinoline quinone in Methylobacterium extorquens AM1 and the purification of a biosynthetic intermediate. *Microbiology* 143 (Pt 2), 595–602. doi: 10.1099/00221287-143-2-595

Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* 173, 697–703. doi: 10.1128/JB.173.2.697-703.1991

Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67 (5), 1613–1617. doi: 10.1099/ijsem.0.001755

Zhang, M., and Lidstrom, M. E. (2003). Promoters and transcripts for genes involved in methanol oxidation in Methylobacterium extorquens AM1. *Microbiology* 149, 1033–1040. doi: 10.1099/mic.0.26105-0

Transgene Biocontainment Strategies for Molecular Farming

*Michael Clark[1] and Maciej Maselko[1,2,3]**

[1] Applied Biosciences, Macquarie University, North Ryde, NSW, Australia, [2] CSIRO Health and Biosecurity, Canberra, ACT, Australia, [3] CSIRO Synthetic Biology Future Science Platform, Brisbane, QLD, Australia

Advances in plant synthetic biology promise to introduce novel agricultural products in the near future. 'Molecular farms' will include crops engineered to produce medications, vaccines, biofuels, industrial enzymes, and other high value compounds. These crops have the potential to reduce costs while dramatically increasing scales of synthesis and provide new economic opportunities to farmers. Current transgenic crops may be considered safe given their long-standing use, however, some applications of molecular farming may pose risks to human health and the environment. Unwanted gene flow from engineered crops could potentially contaminate the food supply, and affect wildlife. There is also potential for unwanted gene flow into engineered crops which may alter their ability to produce compounds of interest. Here, we briefly discuss the applications of molecular farming and explore the various genetic and physical methods that can be used for transgene biocontainment. As yet, no technology can be applied to all crop species, such that a combination of approaches may be necessary. Effective biocontainment is needed to enable large scale molecular farming.

Keywords: biocontainment, molecular farming, pharmaceuticals, plant synthetic biology, metabolic engineering, transgene, industrial enzymes, biofuel

MOLECULAR FARMING

The potential of engineered plants as low-input production platforms for large-scale production of pharmaceuticals is an area of active research. Examples of plant made pharmaceuticals (PMPs) with global markets include human insulin, human serum albumin (HSA) and HIV-neutralizing antibodies. There is a large need for human insulin due to the high incidence of diabetes world-wide, which includes a substantial undersupplied market in Asia. Plant production of insulin could meet this shortfall at a price diabetics in this region could afford (Stoger et al., 2014). Over 500 tons per year of HSA are necessary to treat fetal erythroblastosis, fluid loss due to burn injuries, hypoproteinemia, and ascites caused by cirrhosis of the liver (Chen et al., 2013). InVitria, a division of Ventria Bioscience has developed Optibumin, a rice-derived HSA that has already been commercialized (He et al., 2011)[1]. An estimated 5 tons of HIV-neutralizing antibody is needed to supply 10 million women with the minimal amount necessary to prevent HIV (Shattock and Moore, 2003). Two publicly funded projects, Pharma-Planta and Future-Pharma have produced HIV-neutralizing antibodies in tobacco and corn seeds for clinical trials. It is hoped their production platform could be used in affected areas to produce microbicides "in the region for the region" (Stoger et al., 2014). For all three examples, there is a greater demand for these products than there is a supply of them, and this is particularly the

---

[1] https://invitria.com/cell-culture-products/optibumin-lipid-reduced-recombinant-human-albumin-rhsa/

case in under-developed countries. The immense scalability of molecular farming could meet the demand at a price that matches the economic situation of the target areas.

Molecular farming also has the potential to enhance the production of pharmaceuticals naturally produced in plants such as the anti-cancer drug Taxol (paclitaxel) and artemisinin, a crucial anti-malarial compound. The plants that synthesize these compounds do so in low concentrations and grow slowly resulting in only minute quantities of the desired compound (Buyel, 2018). Taxol was originally extracted from the Pacific yew, *Taxus brevifolia,* where the bark from a single 100 year old tree yields about 300 mg of Taxol, enough for only one dose (Horwitz, 1994). Today Taxol is produced by Bristol-Myers Squibb using a semi-synthetic process starting with Taxol intermediates extracted from the Yew tree needles. Using the needles rather than the bark is non-destructive but the Yew tree is still slow growing and the intermediates require expensive purification (Howat et al., 2014). Similarly low amounts of artemisinin – 0.01–1.4% dry weight (DW) – accumulate in sweet wormwood, *Artemisia annua* (Ikram and Simonsen, 2017). Engineering the biosynthetic pathways for these compounds into heterologous plants optimized for molecular farming could boost supplies and reduce costs (Wurtzel et al., 2019).

Although the complete biosynthetic pathway for the production of Taxol hasn't been elucidated the biosynthetic pathway to produce the first committed product, taxadiene, has been engineered into *Nicotiana benthamiana*. The full biosynthetic pathway for artemisinin has also been engineered into *N. tabacum*. Both tobacco species are production platforms for molecular farming due to their fast growth and high biomass production. The most successful attempt to biosynthetically produce artemisinin took two large sections of the metabolic pathway for artemisinin and genetically engineered them separately into three different cellular compartments (chloroplast, nucleus, and mitochondria). The resulting heterologous expression of artemisinin at ∼0.8 mg/g DW was less than in the native plant, which can reach 31.4 mg/g DW (Zhang et al., 2009; Malhotra et al., 2016). This can in part be explained by the complexity of the gene expression and regulation of the biosynthetic pathway (Ikram and Simonsen, 2017). A mg/g comparison also doesn't reflect *N. tabacum's* faster growth and higher biomass production when compared to *A. annua*. The genetic engineering of *N. benthamiana* to express a taxadiene synthase gene, which produces taxadiene from geranylgeranyl diphosphate (GGPP), produced 11–27 μg/g DW taxadiene. Further suppression of the phytoene synthase gene and addition of methyl jasmonate increased taxadiene accumulation to 35 μg/g DW (Hasan et al., 2014). The successful *de novo* production of taxadiene could lead to the development of a heterologous plant system that biosynthesises Taxol. Future improvements in metabolic engineering could see a breakthrough in how these high value compounds are produced.

Using plants for the production of enzymes or other proteins impacts both the safety and the potential activity of the isolated products. Plant production is also free from human pathogens – a major concern in mammalian cell culture production systems –

and free from endotoxins, which are a risk in bacterial systems (Commandeur et al., 2002). Protein glycosylation patterns can be manipulated in plants, including to produce 'humanized' glycosylation patterns (Hanania et al., 2017; Mercx et al., 2017). This is important for complex glycoproteins such as monoclonal antibodies or membrane proteins as glycosylation can affect protein stability, subcellular targeting, biological activity, and immunogenicity (He et al., 2014). The glycosylation of asparagine or arginine side-chains is similar for plants and mammals until the glycan reaches the Golgi apparatus. In plants the side-chain can be modified by the attachment of an α(1,3)-linked fucose or β(1,2)-linked xylose, whereas in mammals there can be the attachment of an α(1,6)-linked fucose, β(1,4)-linked galactose or sialic acid (Gomord et al., 2010). In some cases, plant glycosylation produces proteins with higher pharmacological activity than proteins produced by bacterial or mammalian cells. For example, plant production systems produce taliglucerase alfa, a mannose-terminated glycoprotein for the treatment of Gaucher's disease, where terminal mannose residues are needed to bind to macrophage mannose receptors. In contrast, mammalian cell system production requires post-production glycosylation modifications to expose terminal α-mannose residues (Grabowski et al., 2014). There is, however, the possibility alternative glycosylation will increase the chance of immunogenicity. Several plant production systems have been engineered to give the recombinant protein human glycosylation patterns (Kallolimath et al., 2016).

Plants can also produce large volumes of industrial compounds. Examples of plant made industrial compounds (PMIs) include cellulases and amylases for bioethanol production, xylanases to enhance animal feed and oxidation/reduction enzymes such as laccases and peroxidases for paper manufacturing (Van Der Maarel et al., 2002; Bailey et al., 2004; Clough et al., 2006; Shen et al., 2012; Hood and Requesens, 2014). Currently, bioethanol is produced by using starch derived from corn. To enhance this process Syngenta's genetically modified (GM) corn, Enogen, expresses an α-amylase enzyme, which catalyzes the breakdown of starch into glucose (Que et al., 2014). Corn is also used as animal feed or human food, meaning that there is competition for agricultural space. Plant biotechnology could enable utilizing more of the cellulose and hemicellulose to be used to produce biofuels. The US company Agrivida increased ethanol production by 55% by engineering corn to express cell wall degrading enzymes *in planta* (Zhang et al., 2011).

Transgenic plants have been developed to be a source of fibrous animal proteins such as collagen, keratin, silk, and elastin (Börnke and Broer, 2010). The Israeli biotechnology company CollPlant developed a tobacco line to produce recombinant human collagen (Stein et al., 2009). Typically, medical collagen comes from animal or human cadavers which pose an infection risk from prions (Pammer et al., 1998). Additionally, the extraction process forms unwanted inter- and intra-molecular bonds, which reduce the solubility and the ability of the collagen to form into more desirable highly structured scaffolds (Zeugolis et al., 2008). Whereas, the plant-derived collagen is cross-link and pathogen free, so it can be modified for the desired application.

For maximal scalability and cheap production molecular farming is conducted with field grown crops. A good case study is the plant biotechnology company Infinite Enzyme, which uses field grown corn to heterologously produce 1.5 million kg of cellulase annually (the amount needed for a 190 million liter per year cellulosic biofuel facility). To produce and process field grown corn only $2 million in capital investment – for dry milling and defatting equipment – was required; with $11.7 million per year in operating costs ($7.8/kg enzyme). In contrast, a microbial fermentation system, which requires tanks and the associated infrastructure, would require $100 million in upfront capital investment. A further $15 million per year would also be needed in operating costs ($10/kg enzyme)[2]. However, the economic advantages of using a field grown crop, must be balanced out by the possibility of the transgene contaminating other crop production.

THE TROUBLED HISTORY OF TRANSGENE ESCAPE

While molecular farming has the potential to lower the cost of medications and industrially useful compounds, the growth of these technologies is contingent on the containment of the transgenes. Challenges of transgene biocontainment are not just hypothetical; there are two salient examples of the need for effective containment – the ProdiGene and StarLink affairs (Murphy, 2007). ProdiGene produced a transgenic corn that expressed a vaccine for preventing bacteria-induced diarrhea in pigs, and while the vaccine protein was non-toxic to humans, strict exclusion from the human food chain was required (Hileman, 2003). StarLink's corn crop was genetically engineered with a gene for resistance to the herbicide glufosinate, and it contained a variant of the pest control *Bacillus thuringiensis* (Bt) protein (Cry9C) – it also lacked approval for food use. In 2000, StarLink's transgenic corn contaminated millions of tons of non-transgenic corn throughout the United States. Government officials have said StarLink's developer, Aventis CropScience, failed to ensure farmers kept StarLink corn separate from other varieties[3]. The contaminated corn was recalled for disposal, costing Aventis an estimated $500 million (Murphy, 2007). In 2002, ProdiGene failed to eradicate plants that had seeded from their previous season's transgenic corn crop. This led to the contamination of non-transgenic soybeans. ProdiGene's failure to manage their transgenic corn crop resulted in 12,000 tons of soybean being destroyed. The combined cost to ProdiGene was about $3.5 million with an additional US government fine of $250,000 (Thayer, 2002).

The fallout from the ProdiGene and StarLink affairs was lasting. In response the molecular farming industry pushed for tighter regulations regarding the approval process for molecular farming crops (Murphy, 2007). In 2003, the Animal and Plant Health Inspection Service (APHIS) of the US Department of Agriculture (USDA) introduced the requirement that crops engineered to produce PMIs be grown under permit. Previously, a GM PMI producing crop could be cultivated under notification, which expedited the permitting procedure (Federal Register [FR], 2005). A full discussion of the interplay between regulation and molecular farming is beyond the scope of this review. Although, it is worth making the point that regulatory hurdles remain a barrier to molecular farming. For example, Syngenta's development of Enogen cost several 100 million dollars, a lot of which was due to it taking almost 6 years to pass USDA's regulatory review process (Wang and Ma, 2012). It is promising though that in 2011 Enogen met USDA's requirements to be fully deregulated. In doing so Enogen became the first plant genetically engineered for industry to be granted this status (Wang and Ma, 2012). The success of Enogen shows a pathway to the commercialization of a PMI production platform.

Inefficient transgene biocontainment has impacted international trade. Japan and South Korea halted imports of corn from the United States during the StarLink corn incident. Exports of wheat to Japan and South Korea were also briefly stopped in 2013, after a GM wheat event MON71800 – developed by Monsanto to be glyphosate-tolerant, was found growing in a field. Monsanto paid $2.1 million to farmers to compensate the loss of export income and reputational damage, and paid $250,000 to several wheat growers' associations[4]. In 2016, a sister event (the same DNA was inserted into a different genomic location) – MON71700 – was found to have contaminated a field in the state of Washington. The 22 plants descend from a field trial conducted by Monsanto from 1998 to 2001[5]. In both cases the reoccurrence of the GM wheat was unexplained. The precedent of a GM crop re-emerging more than a decade after a trial stokes public concern over food safety and biosecurity. Such concerns will continue to impact the adoption and development of plant biotechnologies (Murphy, 2007). In order to foster acceptance of transgenic plant production systems there must be proper containment and security at all levels of production.

IMPORTANCE OF BIOCONTAINMENT

There are concerns from the public and from within the scientific community that molecular farming could threaten non-GM agriculture, the environment, and human health. Without adequate biocontainment, neighboring non-GM crops or weeds could receive transgenes and transgenic seeds could contaminate seed storage (Mallory-Smith and Sanchez Olguin, 2011; Gressel, 2015). Contamination worries many in the food industry, who are not involved with molecular farming, but could suffer financially and in terms of public confidence if theirs or any other major edible crop became contaminated (Murphy, 2007). Contamination can impact international trade between countries that have legal restrictions on importing transgenic products (Lu,

---

[2]https://infiniteenzymes.com/technology-2/

[3]https://web.archive.org/web/20070711190925/http://archives.cnn.com/2000/FOOD/news/10/18/conagra.grain.ap/

[4]https://time.com/3582953/monsanto-wheat-farming-genetically-modified-settlement/

[5]https://monsanto.com/company/media/statements/statement-gmo-wheat-plants/

2003). There are also environmental concerns stemming from the possibility of crop-to-wild transgene flow. In most cases, the few resulting offspring from crop × wild crosses will be outcompeted due to being less locally adapted than the wild type (Gressel, 2015) although the transfer of herbicide resistance genes to weeds, including invasive species, could increase the difficulty of eradicating them. It is improbable, but a transgene could also spread from an engineered crop to a weed and then from that weed to another crop. In this way, weeds that contain the transgene could act as a reservoir for that transgene allowing spread to non-GM crops.

In some cases, molecular farming could potentially pose a risk of humans or animals being harmed through inadvertent exposure to an unsafe level of recombinant protein (Breyer et al., 2012). The majority of PMPs currently in production, such as antibodies, growth hormone, insulin and most other proteins, are expected to have no pharmacological effect when ingested (Goldstein and Thomas, 2004). Instead the gastro-intestinal tract will degrade most PMPs to harmless peptides or amino acids. However, many exceptions may exist in the future, and some plant pharmaceuticals, such as oral vaccines, are designed to be active when ingested. There is also potential for skin or eye contact and inhalation of the recombinant protein as well as the potential allergenicity of the plant itself (Breyer et al., 2012). The human health threats are heightened by the fact that a plant product could enter the human food or animal feed chain. An event that is more likely if the transgenic crop is also a food crop, as was seen for ProdiGene.

As well as potentially exposing humans or animals to a harmful compound, contamination can affect the quality of related crops. The North American Miller's Association were concerned that the transgene for amylase expression in Enogen could spread into other corn varieties and result in lower quality tortillas, corn puffs, and bread (Waltz, 2011). The advance of agriculture will likely see new crop varieties generating novel products such as cotton engineered to be red in color. In order to maintain the phenotypic integrity of transgenic and non-transgenic cultivars effective biocontainment will be required.

The potential economic, environmental, and health threats from molecular farming can be greatly reduced through controlling the flow of the transgene. It's also important to point out that the level of threat from transgene escape depends on the nature of the contamination. Trace mixing of seed that contains a toxic protein is unlikely to be harmful due to dilution. However, the introgression of a transgene, which expresses a toxic protein, into a neighboring crop or weed could seriously contaminate human food or animal feed chains. Although any contamination, regardless of risk, will likely impact public support for GM agriculture.

TRANSGENE CONTAINMENT TECHNOLOGIES

Gene flow is a process where the frequency of a gene changes in a population and can occur through gametes, an organism or groups of organisms moving from one population to another. The potential for there to be gene flow into or from a crop depends on the crop's pollination strategy, on the size of the crop, seed size and viability, and whether there are compatible species within pollination distance (Mallory-Smith and Sanchez Olguin, 2011). **Figure 1** details the three main ways that transgenes can spread into the environment. Volunteer plants – plants that have self-seeded from a previous season's crop – can contaminate the next season's crop if they are accidentally harvested alongside the intended crop (Michael et al., 2010). Transgenes may also spread in seeds that can be spilled during the harvest and transfer of seed. Lastly, cross-pollination can lead to either transgenes escaping into neighboring plants or introgression from neighboring plants into the transgenic crop (Gressel, 2015). As we are primarily concerned with the movement of genes into another population, pollen transfer is the form of gene flow that is of most concern.

There are essentially two approaches for minimizing gene flow: containment and mitigation. Containment aims to stop the flow of the gene from the crop and mitigation focuses on preventing the gene from establishing in a significant proportion of the population (Gressel, 2015). Containment can be physical or biological. Physical containment provides a barrier, such as a greenhouse, filters in the lab or isolation distances in the field. There are also efforts to conduct molecular farming underground, e.g., in unused mines, which provide an even higher degree of physical containment[6]. So far there are no documented cases of physical containment failing in the laboratory or greenhouse (Gressel, 2015). Whereas, the shortcomings of geographic isolation were shown when transgenes from GM glyphosate-resistant creeping bentgrass, *Agrostis stolonifera,* were found in non-agronomic bentgrass up to 3.8 km beyond the control area perimeter (Reichman et al., 2006). With the unreliability of geographic isolation in many situations it is preferable to avoid the use of crop plants grown for human or animal consumption.

Alternative plant production platforms have been developed to reduce the risk of contamination. Some examples of non-food and non-feed crops include tobacco (*N. benthamiana*), duckweed (*Lemna minor*), microalgae (*Chlamydomonas reinhardtii*), and moss (*Physcomitrella patens*) (Yao et al., 2015). As can be seen from **Table 1**, tobacco and moss are popular production platforms. The use of these plants prevent introgression of a transgene into a plant used for food or feed. If a crop plant is to be used, crops that can be crossed with weedy relatives, such as the sunflower, *Helianthus annuus,* should be avoided.

Sound biocontainment and rapid production of recombinant protein can be achieved using a transient expression system which does not result in a transgene integrated in the germline. One method to establish a transient expression system is agroinfiltration where the bacteria *Agrobacterium tumefaciens* – acting as a vector for the gene of interest – is injected or vacuum infiltrated into leaf cells (Whaley et al., 2011). Another approach is to use plant RNA viruses (Yusibov et al., 2006). Both of these approaches can be combined, where agroinfiltration

---

[6]https://www.wired.com/2004/05/drug-farms-forced-underground/
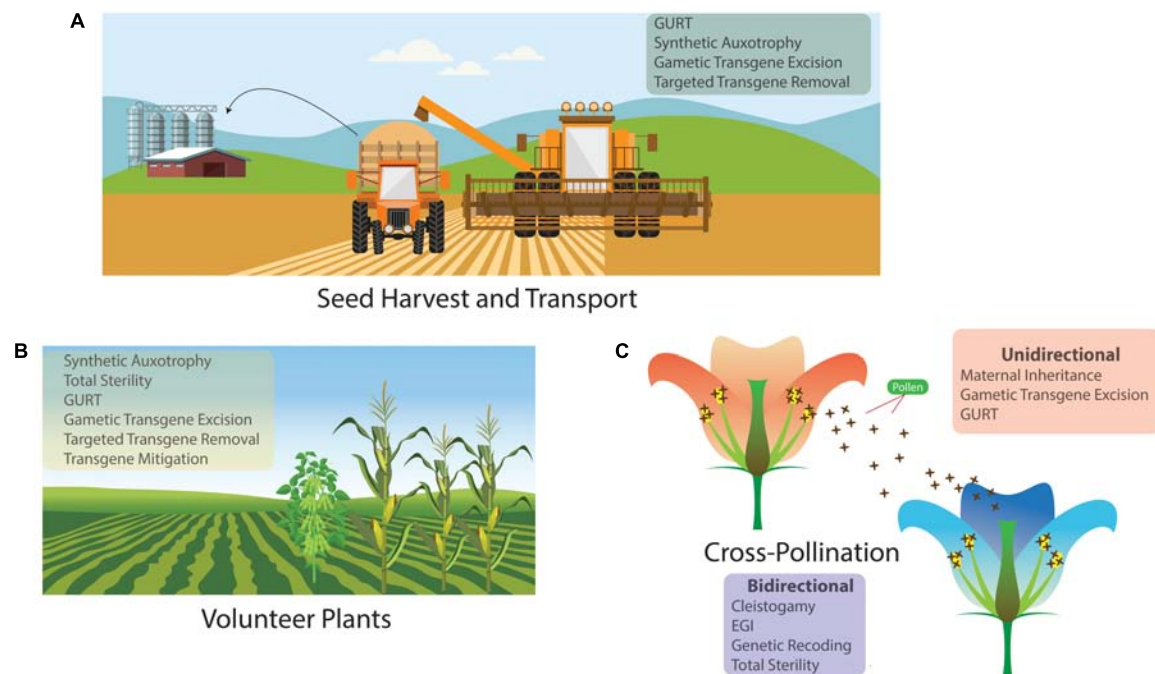
FIGURE 1 | The three main pathways for unwanted contamination or gene flow in an agricultural setting; with a list of the genetic biocontainment technologies that could be used to reduce the possibility of the gene flow occurring. **(A)** Seed dispersal during harvest and transport. **(B)** Contamination from volunteer plants. **(C)** Genetic biocontainment can limit pollen-mediated gene flow unidirectionally, where transgenes are prevented from spreading from the transgenic crop into neighboring plants, and it can operate bidirectionally, where gene flow into the transgenic crop is also limited.

is used to deliver RNA viral vectors into the leaves of a plant. This process, called 'magnifection' combines the transfection efficiency of *A. tumefaciens,* the post-translational modifications of a plant and the high expression yield obtained with viral vectors (Marillonnet et al., 2005). In all of these approaches the transferred DNA is expressed but not integrated into the germline. The tobacco *N. benthamiana* is most often used as the production platform due to the ease with which it can be transformed. Compared to the time it takes to establish a stable transgenic plant line – 6 months to a year – transient expression systems can produce recombinant protein within 3– 5 days (Yao et al., 2015). This is ideal for combating sudden viral epidemics, such as severe acute respiratory syndrome (SARS) or Ebola. Transient expression systems, as a consequence of not introducing transgenes into germline tissue, don't risk contaminating food through transgene outflow into non-GM crops or their wild relatives (Huafang Lai and Jake Stahnke, 2013). However, *Agrobacterium* infiltration is labor intensive, which was a barrier to transient expression supplying sufficient supplies of an Ebola vaccine (Yao et al., 2015).

Whole-plant production platforms remain attractive due to their scalability but for some applications *in vitro* systems are preferable. Current *in vitro* technologies include plant–cell suspension and hairy root cultures. Plant–cell suspensions are typically derived from new tissue formed over a plant callus, which has been cultivated on solidified media. The clumps that easily break apart can be transferred to liquid media. If a homogenous culture forms, the fermentation of the plant

cells can be conducted using similar techniques to fermenting lower eukaryotes (Fischer et al., 1999). Cell suspension cultures have sound containment and have a quick development cycle but are a much less scalable production platform, when compared to transgenic plants (Santos et al., 2016). Hairy root cultures are differentiated cultures of transformed roots generated by infection with *Agrobacterium rhizogenes* (Häkkinen et al., 2014). Hairy root culture can be grown with simple defined media like undifferentiated cells, but it has greater genetic stability and it is highly scalable. These features make it suitable for producing pharmaceutical proteins at an industrial-scale (Guillon et al., 2006). However, *in vitro* techniques require sophisticated and sterile laboratory settings. If the scalability and low-cost potential of plant production of PMPs or PMIs is to be realized, plants need to be grown in fields.

The higher contamination risk from growing plants in fields can be reduced by genetic containment which may exploit existing reproductive limitations or introduce them via genetic engineering. Many genetic approaches for containing plant transgenes have been investigated including cleistogamy, maternal inheritance, gametic transgene excision, synthetic auxotrophy, total sterility, and genetic use restriction technologies (GURT or Terminator). Other genetic containment technologies in development could be extended to plants, such as engineered genetic incompatibility (EGI), genetic recoding and targeted transgene removal (see **Table 2**). Many of these technologies work well for specific types of plants and can be enhanced by pairing them with other technologies.

TABLE 1 | Examples of plant made pharmaceuticals.

Product	Disease	Plant production host	Clinical trial status	Company	References
HIV/HSV microbicide MB66	HIV/HSV	Tobacco	Phase I	Mapp Biopharmaceutical, United States	https://mappbio.com/product-development/
Zmapp^{TM}	Ebola Zaire virus	Tobacco	Phase II/III	Mapp Biopharmaceutical, United States	https://mappbio.com/product-development/
VEN BETA	Gastroenteritis	Rice	Preclinical phase	Ventria Bioscience	https://ventria.com/
VEN120	Inflammatory bowel disease	Rice	Phase II	Ventria Bioscience	https://ventria.com/
Moss-aGal (agalsidase)	Fabry disease	Moss	Phase I	Greenovation Biopharmaceuticals	https://www.greenovation.com/home.html
Moss-FH	C3 Glomerulopathy	Moss	Preclinical phase	Greenovation Biopharmaceuticals	https://www.greenovation.com/home.html
H1N1 vaccine	Seasonal influenza	Tobacco	Phase III	Medicago, United States	https://www.medicago.com/en/pipeline/
H5N1 vaccine	Pandemic influenza	Tobacco	Phase II	Medicago, United States	https://www.medicago.com/en/pipeline/
Rotavirus vaccine	Rotavirus vaccine	Tobacco	Phase I	Medicago, United States	https://www.medicago.com/en/pipeline/
Optibumin	Loss of albumin	Rice	On Market	InVitria, United States	https://invitria.com/
Non-Hodgkin's lymphoma vaccine	Non-Hodgkin's lymphoma	Tobacco	Phase I	Icon Genetics, Germany	https://www.icongenetics.com/
Vibrio cholerae	Cholera	Potato	Phase I	Arizona State University	Tacket, 2005
Heat-labile toxin B subunit of *Escherichia coli*	Diarrhea	Potato	Phase I	Arizona State University	Tacket, 2005
Capsid protein Norwalk virus	Diarrhea	Potato, Tomato	Phase I	Arizona State University	Zhang et al., 2006
Antibody against hepatitis B	Vaccine purification	Tobacco	On market	CIGB, Cuba	Kaiser, 2008
ISOkine^{TM}, BIOEFFECT^{TM} EGF Serum (human growth factors and cytokines)		Barley	On market	ORF genetics	https://orfgenetics.com/

TABLE 2 | The important features of genetic biocontainment technologies.

	Weakened by introgression	Mitigates volunteer plants	Difficult to engineer	Transgene will persist	Unidirectional cross-pollination control	Bidirectional cross-pollination control	Demonstrated in plants
Cleistogamy				X		X	X
Gametic transgene excision		X			X		X
Synthetic auxotrophy	X	X	X				X
Total sterility		X				X	X
GURT		X			X		X
Maternal inheritance			X	X	X		X
EGI				X		X	
Genetic recoding	X		X	X		X	
Targeted transgene removal		X			X		

Cleistogamy, where there is self-pollination within a closed flower, is a promising tool to limit gene flow. Currently it suffers from some flowers opening, which allows for cross-pollination. Cleistogamy requires that the plant's flower contain male and female parts and that there can be self-fertilization. Crops like rice have such flowers, but plants that have separate male and female flowers, like asparagus and spinach, or with unusual flower anatomies, such as corn, aren't suited for cleistogamy. It was

found that for imidazolinone herbicide resistant rice a few flowers opened enabling hybridization with weedy rice (Gealy, 2005). To combat this, rice was genetically engineered to enhance the percentage of cleistogamous flowers through incorporating the cleistogamous gene, 'superwoman1.' The engineered cultivar, in a variety of plots, had an outcrossing rate of 0.000% compared to the non-engineered cultivar, which ranged from 0.005 to 0.200% (Ohmori et al., 2012). The potential of cleistogamy is limited for GM food production as current practices tend to use higher-yielding hybrid rice varieties, which require parental lines that aren't cleistogamous (Gressel, 2015). However, this would not be an issue for the molecular farming of high value compounds where cleistogamy could be used to restrict pollen mediated gene flow.

Synthetic auxotrophy works by genetically engineering a strain to depend on an externally supplied compound. The dependence can come from deleting essential genes that are needed, for example, to synthesize amino acids or co-factors that are necessary for crucial biological functions (Moe-Behrens et al., 2013). So far, this approach has found little traction for use in plants. There are isolated cases such as the duckweed *Lemna*, which has been engineered to be dependent on the addition of isoleucine through inactivating threonine deaminase expression (Nguyen et al., 2012). However, the genetic redundancy that is a common feature of plant genomes increases the difficulty in engineering recessive auxotrophic mutations. For most plants there are likely multiple proteins that catalyze the same reaction, which requires a large number of genetic changes to confer metabolic dependence (Last et al., 1991). The addition of potentially expensive chemicals, in itself a drawback, also requires changes to normal cultivation techniques. Synthetic auxotrophy can also fail due to introgression of genes from non-transgenic plants, which could restore the knocked out metabolic pathway.

Another approach exploits the maternal inheritance of plastids (e.g., chloroplasts). For the vast majority of higher plants, which display maternal inheritance, transgenes located in the plastid genome are unlikely to be transmitted to other plants by pollination (Maliga, 2004). Plastid engineering has therefore been employed to locate the transgene in the plastid genome, however, the advance of plastid engineering has been stymied by poor transformation protocols for plants other than tobacco. Transformation relies on many essential factors unique to the species and sometimes unique to the cultivar (Lu et al., 2013). There must be detailed knowledge of the plastid genome sequence including the regions in between genes suitable for transgene integration, there also needs to be an optimized DNA delivery system, as well as effective antibiotic selection and selectable marker genes. For several years the chloroplast genome sequences have been available for monocots, such as wheat and corn, but the chloroplast hasn't been transformed due to the engineering complexity (Wani et al., 2015). This approach may also be less efficient than envisioned considering that species that were thought to strictly engage in maternal plastid inheritance still had about 0.4% plastid transmission via pollen (Avni and Edelman, 1991; Svab and Maliga, 2007). Additional problems are: proteins expressed in the chloroplast undergo different post-translational modifications, meaning that enzyme function might

be altered (Grabsztunowicz et al., 2017); plastid transformation can also be laborious and time-consuming (Ruf et al., 2001).

Total sterility offers a sound basis for genetic biocontainment. Several crops are already sterile or have sterile varieties, such as cassava (*Manihot esculenta*), potatoes (*Solanum tuberosum*), and banana (*Musa acuminata*) (Celis et al., 2004; Heslop-Harrison and Schwarzacher, 2007; Sayre et al., 2011). As long as the sterility is not leaky, these crops would be safe candidates for molecular farming. A totally sterile plant can also be engineered by deleting genes that encode for gamete production (Kwit et al., 2011). The downsides are that total sterility requires plants to be vegetatively propagated by either tubers, tissue culture, cuttings, or artificial seed (Gressel, 2015). Total sterility could be used with tuber or bulb propagated crops, leafy vegetable crops and forestry. Whereas crops that are harvested for compounds accumulated in seeds would not be candidates for total sterility.

Gametic transgene excision uses a site-specific recombination system to excise a transgene. Currently, the efficiency of the recombinase is quite low, where 99% excision is considered to be high performing (Moon et al., 2011). This level of efficiency is too low to restrict transgene escape, however, it could be used to excise selectable marker genes used in the engineering of transgenic plants (Hu et al., 2013). Farmers are also not able to collect seed containing the transgene for future seasons unless the recombinase can be externally controlled, which alters normal cultivation practices (Ryffel, 2014; Gressel, 2015).

Genetic use restriction technologies were originally developed to prevent farmers from infringing on patents by saving seed. They have been some of the most controversial GM biotechnologies due to the widespread perception that they were designed to entrench a multinational corporation seed monopoly (Lombardo, 2014). GURTs use a tightly controlled genetic system to regulate the expression of a target gene. There are typically four components to this genetic system: the target gene, the target gene's promoter, the trait switch and the genetic switch. The target gene needs to be activated by the promoter. In order to prevent leaky expression from unwanted promoter activity a blocker sequence separates the promoter from the target gene. The blocking sequence can in turn be removed through a cascade beginning with an external input, which will be amplified by the genetic switch. The amplified input becomes a biological signal that activates the trait switch. The trait switch usually encodes an enzyme, such as a site-specific recombinase that removes the blocker sequence (Lombardo, 2014). Without the blocker sequence there can then be transcription and expression of the target gene.

In part due to public opposition GURTs have never been commercialized. However, there is scope for GURTs to be used for biocontainment. For this, the GURT system would be linked to the transgene, so that when the GURT is activated there is expression of a disrupter gene that drives cell death. Disrupter genes typically encode for cytotoxins such as barnase and ribonuclease A (Mariani et al., 1990; Burgess et al., 2002; Gils et al., 2008; Zhang et al., 2012). There is no evidence that disrupter genes generate products that are toxic to humans or animals. However, it is possible that the potential health risk will add to the already controversial nature of using GURTs (Conner et al., 2003;

Gressel, 2010). The other disadvantages to GURT are that it is a more expensive system, requiring exogenous inputs and there is greater difficulty in propagating a GURT crop.

FUTURE BIOLOGICAL CONTAINMENT TECHNOLOGIES

The biocontainment technologies that have been developed in microbes could in some cases be extended to plants. Some of these technologies include, genetic recoding, targeted transgene removal and EGI.

Genetic recoding removes every instance of at least one codon for an amino acid in an organism's genome and replaces it with another. The codon that has been removed can be replaced with a synonymous codon or it can then encode for a non-standard amino acid (NSAA) (Mukai et al., 2017). If an essential gene was recoded to require an amino acid not found in nature this would increase the stringency of an auxotrophy. Further, the genetic recoding could create reproductive isolation and block gene flow with non-recoded organisms due to incompatible genetic codes. *Escherichia coli* has been recoded so that the UAG stop codon instead incorporated a NSAA in the cores of essential enzymes. This conferred a dependence on synthetic metabolites for proper protein function, such that the bacteria were less capable of mutational escape and metabolic supplementation (Mandell et al., 2015). Following on from this the genetic recoding of plant genomes could confer better biocontainment. Despite the advance of the technology, we are unlikely to see recoding of higher organisms with ease in the near future due to the scale of changes needed in large genomes.

Another strategy could be to precisely remove the engineered genes instead of killing the whole organism. The spread of transgenes from volunteer plants or inadvertent seed dispersal could be mitigated by using a CRISPR-based system to selectively remove the transgene after the desired protein has been produced. In one such method, a genetically encoded device, termed DNAi, responds to a transcriptional input by degrading DNA adjacent to a synthetic CRISPR array. The DNAi system was shown to be non-toxic when carried in *E. coli,* and when activated it was able to reduce the number of viable cells by 1.9×10^{-8} making it one of the most effective switches for programmed cell death (Caliando and Voigt, 2015). This same mechanism could be engineered so that with the addition of a transcriptional input the transgene is degraded. An advantage of this system is that the removal of the transgene applies little selective pressure toward deactivating the genetic machinery; whereas directing whole organism death selects for mutations that lead to an organism's survival.

The aforementioned biological containment technologies, with the exceptions of cleistogamy, genetic recoding, and total sterility, don't prevent the flow of genes into the transgenic plant. This is an important consideration as unwanted gene flow can alter important traits in a genetically engineered organism. In order to restrict gene flow in both directions, plants could be engineered to be genetically incompatible with related plants such that the hybrid is less fit – this is known as

underdominance. The model organism *Drosophila melanogaster* has been engineered such that engineered-WT hybrids display underdominance. This was achieved using a genetic construct to encode for two genes: the first encodes for a RNAi knockdown of the WT version of the gene *Rpl14*; the second gene is a refactored version of *Rpl4* such that it isn't susceptible to RNAi knockdown. When the engineered organism was mated with WT flies there was a marked fitness reduction in the heterozygotes (Reeves et al., 2014). However, in order to be effective for biocontainment the underdominance must result in total sterility or death of the hybrids.

An artificial reproductive barrier, where the hybrids are non-viable, has been engineered in *Saccharomyces cerevisiae* using EGI. This system utilizes programmable transcriptional activators (PTAs) to overexpress a gene leading to lethality. Lethality in the engineered organism is avoided by editing the target sequence of the PTA, such that the PTA is unable to bind and overexpress the gene (Maselko et al., 2017). When there is a cross between the WT and the engineered organism, the PTA targets the WT PTA binding sequence and drives lethal levels of gene expression. Attempts have also been made at constructing a synthetic species of *D. melanogaster*, where an artificial reproductive barrier is engineered, however the goal of complete genetic isolation wasn't achieved. The main difficulty proved to be getting strong activation of a lethal gene without the fitness costs associated with broad expression of the transactivating CRISPR machinery (Waters et al., 2018).

There is an inherent versatility to the use of PTAs, so that lethal overexpression of a target gene could theoretically be engineered in any sexually reproducing organism (Maselko et al., 2017). Proof of concept has so far only been established in *S. cerevisiae*. Although, it is conceivable that EGI could be extended to plants, where it could be used to generate many orthogonal strains of the same parent species which could each be used as production platforms for different compounds. If interbreeding can be prevented then the phenotypic integrity of transgenic cultivars could be protected. EGI could also be used to make synthetic auxotrophy more robust by preventing introgression from neighboring plants, which would otherwise compromise the auxotrophy.

TRANSGENE MITIGATION TECHNOLOGIES

Even the most stringent containment system can fail. Technologies are therefore needed to reduce the chances of a transgene becoming established after escape. Transgenic mitigation involves linking the transgene to genes that confer a selective disadvantage. Weedy traits such as a propensity toward shattering, bolting, and greater height can be targeted (Gressel, 2015). Transgenic mitigation reduced the reproductive fitness of transgenic-weed oilseed rape hybrids. A dwarfing mitigator gene was linked to a herbicide resistance transgene, which reduced the reproductive fitness of the transgenic-weed hybrid to 0.9% of the competing weed's reproductive fitness (Al-Ahmad and Gressel, 2006). However, there is the potential for the linkage

of the mitigator gene to the transgene to be broken through meiotic crossing over. Additionally, there can be mutation of the mitigator gene so that it ceases to confer the deleterious phenotype. Both of these issues can in some part be addressed through linking another mitigator gene to the transgene, such that there are mitigating genes either side of the transgene (Gressel, 2015).

CONCLUSION

Molecular farming has the potential to lower the cost of medication and industrial enzymes. However, in cases where the recombinant protein is potentially toxic, there are environmental and human health risks. The introgression of the transgene into a neighboring crop or weed may contaminate food or feed supplies. Any contamination event, such as in the high-profile cases of StarLink and ProdiGene, could jeopardize confidence in molecular farming. For these reasons there must be effective containment of transgenes.

There has been considerable progress in the development of biological containment technologies. For some species such as rice, cleistogamy could contain gene flow. For tubers and bulb propagated crops total sterility is practical. But for many species these technologies aren't applicable. There is some promise that technologies like EGI combined with synthetic auxotrophy could contain gene flow. Further work in this area is needed to ensure the safety and widespread adoption of field grown molecular farming crops.

AUTHOR CONTRIBUTIONS

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

FUNDING

REFERENCES

Al-Ahmad, H., and Gressel, J. (2006). Mitigation using a tandem construct containing a selectively unfit gene precludes establishment of *Brassica napus* transgenes in hybrids and backcrosses with weedy *Brassica rapa*. *Plant Biotechnol. J.* 4, 23–33. doi: 10.1111/j.1467-7652.2005.00153.x

Avni, A., and Edelman, M. (1991). Direct selection for paternal inheritance of chloroplasts in sexual progeny of Nicotiana. *Mol. Gen. Genet.* 225, 273–277. doi: 10.1007/bf00269859

Bailey, M. R., Woodard, S. L., Callaway, E., Beifuss, K., Magallanes-Lundback, M., Lane, J. R., et al. (2004). Improved recovery of active recombinant laccase from maize seed. *Appl. Microbiol. Biotechnol.* 63, 390–397. doi: 10.1007/s00253-003-1362-z

Börnke, F., and Broer, I. (2010). Tailoring plant metabolism for the production of novel polymers and platform chemicals. *Curr. Opin. Plant Biol.* 13, 353–361. doi: 10.1016/j.pbi.2010.01.005

Breyer, D., De Schrijver, A., Goossens, M., Pauwels, K., and Herman, P. (2012). "Biosafety of molecular farming in genetically modified plants," in *Molecular Farming in Plants: Recent Advances and Future Prospects*, eds A. Wang, and S. Ma, (Dordrecht: Springer), 259–274. doi: 10.1007/978-94-007-2217-0_12

Burgess, D. G., Ralston, E. J., Hanson, W. G., Heckert, M., Ho, M., Jenq, T., et al. (2002). A novel, two-component system for cell lethality and its use in engineering nuclear male-sterility in plants. *Plant J.* 31, 113–125. doi: 10.1046/j.1365-313X.2002.01330.x

Buyel, J. F. (2018). Plants as sources of natural and recombinant anti-cancer agents. *Biotechnol. Adv.* 36, 506–520. doi: 10.1016/j.biotechadv.2018.02.002

Caliando, B. J., and Voigt, C. A. (2015). Targeted DNA degradation using a CRISPR device stably carried in the host genome. *Nat. Commun.* 6:6989. doi: 10.1038/ncomms7989

Celis, C., Scurrah, M., Cowgill, S., Chumbiauca, S., Green, J., and Franco, J. (2004). Environmental biosafety and transgenic potato in a centre of diversity for this crop. *Nature* 432, 222–225. doi: 10.1038/nature03048

Chen, Z., He, Y., Shi, B., and Yang, D. (2013). Human serum albumin from recombinant DNA technology: challenges and strategies. *Biochim. Biophys. Acta* 2013, 5515–5525. doi: 10.1016/j.bbagen.2013.04.037

Clough, R. C., Pappu, K., Thompson, K., Beifuss, K., Lane, J., Delaney, D. E., et al. (2006). Manganese peroxidase from the white-rot fungus phanerochaete chrysosporium is enzymatically active and accumulates to high levels in transgenic maize seed. *Plant Biotechnol. J.* 4, 53–62. doi: 10.1111/j.1467-7652.2005.00157.x

Commandeur, U., Twyman, R. M., and Fischer, R. (2002). The biosafety of molecular farming in plants. *AgBiotechNet* 5:ABN10.

Conner, A. J., Glare, T. R., and Nap, J.-P. (2003). The release of genetically modified crops into the environment, Part II. Overview of Ecological Risk Assessment. *Plant J.* 33, 19–46. doi: 10.1046/j.0960-7412.2002.001607.x

Federal Register [FR] (2005). *Introductions of Plants Genetically Engineered To Produce Industrial Compounds*. Available online at: https://www.govinfo.gov/app/details/FR-2005-05-04/05-8860

Fischer, R., Emans, N., Schuster, F., Hellwig, S., and Drossard, J. (1999). Towards molecular farming in the future: using plant-cell-suspension cultures as bioreactors. *Biotechnol. Appl. Biochem.* 30, 109–112. doi: 10.1111/j.1470-8744.1999.tb00899.x

Gealy, D. (2005). "Gene movement between rice (*Oryza sativa*) and weedy rice (*Oryza sativa*) — a U.S. Temperate Rice Perspective," in *Crop Ferality and Volunteerism*, ed. J. Gressel, (Boca Raton, FL: CRC Press), 323–354. doi: 10.1201/9781420037999.ch20

Gils, M., Marillonnet, S., Werner, S., Grützner, R., Giritch, A., Engler, C., et al. (2008). A novel hybrid seed system for plants. *Plant Biotechnol. J.* 6, 226–235. doi: 10.1111/j.1467-7652.2007.00318.x

Goldstein, D. A., and Thomas, J. A. (2004). Biopharmaceuticals derived from genetically modified plants. *QJM* 97, 705–716. doi: 10.1093/qjmed/hch121

Gomord, V., Fitchette, A. C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D., et al. (2010). Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol. J* 8, 564–587. doi: 10.1111/j.1467-7652.2009.00497.x

Grabowski, G. A., Golembo, M., and Shaaltiel, Y. (2014). Taliglucerase alfa: an enzyme replacement therapy using plant cell expression technology. *Mol. Genet. Metab.* 112, 1–8. doi: 10.1016/j.ymgme.2014.02.011

Grabsztunowicz, M., Koskela, M. M., and Mulo, P. (2017). Post-translational modifications in regulation of chloroplast function: recent advances. *Front. Plant Sci.* 8:240. doi: 10.3389/fpls.2017.00240

Gressel, J. (2010). Needs for and environmental risks from transgenic crops in the developing world. *N. Biotechnol* 27, 522–527. doi: 10.1016/j.nbt.2010.05.015

Gressel, J. (2015). Dealing with transgene flow of crop protection traits from crops to their relatives. *Pest Manag. Sci.* 71, 658–667. doi: 10.1002/ps.3850

Guillon, S., Trémouillaux-Guiller, J., Pati, P. K., Rideau, M., and Gantet, P. (2006). Hairy root research: recent scenario and exciting prospects. *Curr. Opin. Plant Biol.* 9, 341–346. doi: 10.1016/j.pbi.2006.03.008

Häkkinen, S. T., Raven, N., Henquet, M., Laukkanen, M.-L., Anderlei, T., Pitkänen, J.-P., et al. (2014). Molecular farming in tobacco hairy roots by triggering the secretion of a pharmaceutical antibody. *Biotechnol. Bioeng.* 111, 336–346. doi: 10.1002/bit.25113

Hanania, U., Ariel, T., Tekoah, Y., Fux, L., Sheva, M., Gubbay, Y., et al. (2017). Establishment of a tobacco BY2 cell line devoid of plant-specific xylose and fucose as a platform for the production of biotherapeutic proteins. *Plant Biotechnol. J.* 15, 1120–1129. doi: 10.1111/pbi.12702

Hasan, M. M., Kim, H. S., Jeon, J. H., Kim, S. H., Moon, B. K., and Song, J. Y. (2014). Metabolic engineering of nicotiana benthamiana for the increased production of taxadiene. *Plant Cell Rep.* 33, 895–904. doi: 10.1007/s00299-014-1568-9

He, Y., Ning, T., Xie, T., Qiu, Q., Zhang, L., Sun, Y., et al. (2011). Large-scale production of functional human serum albumin from transgenic rice seeds. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19078–19083. doi: 10.1073/pnas.1109736108

He, Y., Wang, K., and Yan, N. (2014). The recombinant expression systems for structure determination of eukaryotic membrane proteins. *Protein Cell* 5, 658–672. doi: 10.1007/s13238-014-0086-4

Heslop-Harrison, J. S., and Schwarzacher, T. (2007). Domestication, genomics and the future for banana. *Ann. Bot.* 100, 1073–1084. doi: 10.1093/aob/mcm191

Hileman, B. (2003). ProdiGene and StarLink incidents provide ammunition to critics. *Chem. Eng. News* 81, 25–33.

Hood, E. E., and Requesens, D. V. (eds) (2014). "Commercial plant-produced recombinant cellulases for biomass conversion," in *Biotechnology in Agriculture and Forestry*, Vol. 68, (Berlin: Springer), 231–246. doi: 10.1007/978-3-662-43836-7_12

Horwitz, S. B. (1994). How to make taxol from scratch. *Nature* 367, 593–594. doi: 10.1038/367593a0

Howat, S., Park, B., Oh, I. S., Jin, Y. W., Lee, E. K., and Loake, G. J. (2014). Paclitaxel: biosynthesis, production and future prospects. *N. Biotechnol* 31, 242–245. doi: 10.1016/j.nbt.2014.02.010

Hu, Z., Ding, X., Hu, S., Sun, Y., and Xia, L. (2013). Tissue-specifically regulated site-specific excision of selectable marker genes in bivalent insecticidal, genetically-modified rice. *Biotechnol. Lett.* 35, 2177–2183. doi: 10.1007/s10529-013-1310-7

Huafang Lai, Q. C., and Jake Stahnke, J. H. (2013). Agroinfiltration as an effective and scalable strategy of gene delivery for production of pharmaceutical proteins. *Adv. Tech. Biol. Med.* 1:103. doi: 10.4172/atbm.1000103

Ikram, N. K. B. K., and Simonsen, H. T. (2017). A review of biotechnological artemisinin production in plants. *Front. Plant Sci* 8:1966. doi: 10.3389/fpls.2017.01966

Kaiser, J. (2008). Is the drought over for pharming. *Science* 320, 473–475. doi: 10.1126/science.320.5875.473

Kallolimath, S., Castilho, A., Strasser, R., Grünwald-Gruber, C., Altmann, F., Strubl, S., et al. (2016). Engineering of complex protein sialylation in plants. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9498–9503. doi: 10.1073/pnas.1604371113

Kwit, C., Moon, H. S., Warwick, S. I., and Stewart, C. N. (2011). Transgene introgression in crop relatives: molecular evidence and mitigation strategies. *Trends Biotechnol.* 29, 284–293. doi: 10.1016/j.tibtech.2011.02.003

Last, R. L., Bissinger, P. H., Mahoney, D. J., Radwanski, E. R., and Fink, G. R. (1991). Tryptophan mutants in *Arabidopsis*: the consequences of duplicated tryptophan synthase B genes. *Plant Cell* 3, 345. doi: 10.2307/3869210

Lombardo, L. (2014). Genetic use restriction technologies: a review. *Plant Biotechnol. J.* 12, 995–1005. doi: 10.1111/pbi.12242

Lu, B. R. (2003). Transgene containment by molecular means–is it possible and cost effective? *Environ. Biosaf. Res.* 2, 3–8. doi: 10.1051/ebr/2003000

Lu, Y., Rijzaani, H., Karcher, D., Ruf, S., and Bock, R. (2013). Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. *Proc. Natl. Acad. Sci. U.S.A.* 110, E623–E632. doi: 10.1073/pnas.1216898110

Malhotra, K., Subramaniyan, M., Rawat, K., Kalamuddin, M., Qureshi, M. I., and Malhotra, P. (2016). Compartmentalized etabolic engineering for Artemisinin biosynthesis and effective malaria treatment by oral delivery of plant cells. *Mol. Plant* 9, 1464–1477. doi: 10.1016/j.molp.2016.09.013

Maliga, P. (2004). Plastid transformation in higher plants. *Annu. Rev. Plant Biol.* 55, 289–313. doi: 10.1146/annurev.arplant.55.031903.141633

Mallory-Smith, C. A., and Sanchez Olguin, E. (2011). Gene flow from herbicide-resistant crops: it's not just for transgenes. *J. Agric. Food Chem.* 59, 5813–5818. doi: 10.1021/jf103389v

Mandell, D. J., Lajoie, M. J., Mee, M. T., Takeuchi, R., Kuznetsov, G., Norville, J. E., et al. (2015). Biocontainment of genetically modified organisms by synthetic protein design. *Nature* 518, 55–60. doi: 10.1038/nature14121

Mariani, C., De Beuckeleer, M., Truettner, J., Leemans, J., and Goldberg, R. B. (1990). Induction of male sterility in plants by a chimaeric ribonuclease gene. *Nature* 347, 737–741. doi: 10.1038/347737a0

Marillonnet, S., Thoeringer, C., Kandzia, R., Klimyuk, V., and Gleba, Y. (2005). Systemic *Agrobacterium Tumefaciens*-mediated transfection of viral replicons for efficient transient expression in plants. *Nat. Biotechnol.* 23, 718–723. doi: 10.1038/nbt1094

Maselko, M., Heinsch, S. C., Chacón, J. M., Harcombe, W. R., and Smanski, M. J. (2017). Engineering species-like barriers to sexual reproduction. *Nat. Commun.* 8:883. doi: 10.1038/s41467-017-01007-3

Mercx, S., Smargiasso, N., Chaumont, F., De Pauw, E., Boutry, M., and Navarre, C. (2017). Inactivation of the β(1,2)-Xylosyltransferase and the α(1,3)-Fucosyltransferase Genes in Nicotiana Tabacum BY-2 Cells by a Multiplex CRISPR/Cas9 Strategy Results in Glycoproteins without Plant-Specific Glycans. *Front. Plant Sci.* 8:403. doi: 10.3389/fpls.2017.00403

Michael, P. J., Owen, M. J., and Powles, S. B. (2010). Herbicide-resistant weed seeds contaminate grain sown in the western Australian Grainbelt. *Weed Sci.* 58, 466–472. doi: 10.1614/ws-d-09-00082.1

Moe-Behrens, G. H. G., Davis, R., and Haynes, K. A. (2013). Preparing synthetic biology for the world. *Front. Microbiol.* 4:5. doi: 10.3389/fmicb.2013.00005

Moon, H. S., Abercrombie, L. L., Eda, S., Blanvillain, R., Thomson, J. G., Ow, D. W., et al. (2011). Transgene excision in pollen using a Codon optimized serine resolvase CinH-RS2 site-specific recombination system. *Plant Mol. Biol.* 75, 621–631. doi: 10.1007/s11103-011-9756-2

Mukai, T., Lajoie, M. J., Englert, M., and Söll, D. (2017). Rewriting the genetic code. *Annu. Rev. Microbiol* 71, 557–577. doi: 10.1146/annurev-micro-090816-093247

Murphy, D. J. (2007). Improving containment strategies in biopharming. *Plant Biotechnol. J.* 5, 555–569. doi: 10.1111/j.1467-7652.2007.00278.x

Nguyen, L. V., Cox, K. M., Ke, J. S., Peele, C. G., and Dickey, L. F. (2012). Genetic engineering of a lemna isoleucine auxotroph. *Transgenic Res.* 21, 1071–1083. doi: 10.1007/s11248-012-9594-2

Ohmori, S., Tabuchi, H., Yatou, O., and Yoshida, H. (2012). Agronomic traits and gene containment capability of cleistogamous rice lines with the superwoman1-cleistogamy mutation. *Breed. Sci.* 62, 124–132. doi: 10.1270/jsbbs.62.124

Pammer, J., Weninger, W., and Tschachler, E. (1998). Human keratinocytes express cellular Prion-related protein *in vitro* and during inflammatory skin diseases. *Am. J. Pathol.* 153, 1353–1358. doi: 10.1016/S0002-9440(10)65720-3

Que, Q., Elumalai, S., Li, X., Zhong, H., Nalapalli, S., Schweiner, M., et al. (2014). Maize transformation technology development for commercial event generation. *Front. Plant Sci.* 5:379. doi: 10.3389/fpls.2014.00379

Reeves, R. G., Bryk, J., Altrock, P. M., Denton, J. A., and Reed, F. A. (2014). First steps towards underdominant genetic transformation of insect populations. *PLoS One* 9:e97557. doi: 10.1371/journal.pone.0097557

Reichman, J. R., Watrud, L. S., Lee, E. H., Burdick, C. A., Bollman, M. A., and Storm, M. (2006). Establishment of transgenic herbicide-resistant creeping bentgrass (*Agrostis stolonifera* L.) in nonagronomic habitats. *Mol. Ecol.* 15, 4243–4255. doi: 10.1111/j.1365-294X.2006.03072.x

Ruf, S., Hermann, M., Berger, I. J., Carrer, H., and Bock, R. (2001). Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nat. Biotechnol.* 19, 870–875. doi: 10.1038/nbt0901-870

Ryffel, G. U. (2014). Transgene flow: facts, speculations and possible countermeasures. *GM Crops Food* 5, 249–258. doi: 10.4161/21645698.2014.945883

Santos, R. B., Abranches, R., Fischer, R., Sack, M., and Holland, T. (2016). Putting the spotlight back on plant suspension cultures. *Front. Plant Sci.* 7:297. doi: 10.3389/fpls.2016.00297

Sayre, R., Beeching, J. R., Cahoon, E. B., Egesi, C., Fauquet, C., Fellman, J., et al. (2011). The BioCassava plus program: biofortification of cassava for sub-saharan Africa. *Annu. Rev. Plant Biol.* 62, 251–272. doi: 10.1146/annurev-arplant-042110-103751

Shattock, R. J., and Moore, J. P. (2003). Inhibiting sexual transmission of HIV-1 infection. *Nat. Rev. Microbiol.* 1, 25–34. doi: 10.1038/nrmicro729

Shen, B., Sun, X., Zuo, X., Shilling, T., Apgar, J., Ross, M., et al. (2012). Engineering a thermoregulated intein-modified xylanase into maize for consolidated lignocellulosic biomass processing. *Nat. Biotechnol.* 30, 1131–1136. doi: 10.1038/nbt.2402

Stein, H., Wilensky, M., Tsafrir, Y., Rosenthal, M., Amir, R., and Avraham, T. (2009). Production of bioactive, post-translationally modified, heterotrimeric, human recombinant type-I collagen in transgenic tobacco. *Biomacromolecules* 10, 2640–2645. doi: 10.1021/bm900571b

Stoger, E., Fischer, R., Moloney, M., and Ma, J. K.-C. (2014). Plant molecular pharming for the treatment of chronic and infectious diseases. *Annu. Rev. Plant Biol.* 65, 743–768. doi: 10.1146/annurev-arplant-050213-035850

Svab, Z., and Maliga, P. (2007). Exceptional transmission of plastids and mitochondria from the transplastomic pollen parent and its impact on transgene containment. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7003–7008. doi: 10.1073/pnas.0700063104

Tacket, C. O. (2005). Plant-Derived Vaccines against Diarrheal Diseases. *Vaccine* 23, 1866–1869. doi: 10.1016/j.vaccine.2004.11.019

Thayer, A. (2002). USDA fines prodigene; industry reacts to farm interests. *Chem. Eng. News* 80:12. doi: 10.1021/cen-v080n050.p012a

Van Der Maarel, M. J., Van Der Veen, B., Uitdehaag, J. C. M., Leemhuis, H., and Dijkhuizen, L. (2002). Properties and applications of starch-converting enzymes of the α-amylase family. *J. Biotechnol.* 94, 137–155. doi: 10.1016/S0168-1656(01)00407-2

Waltz, E. (2011). Amylase corn sparks worries. *Nat. Biotechnol.* 29, 294–294. doi: 10.1038/nbt0411-294

Wang, A., and Ma, S. (2012). *Molecular Farming in Plants: Recent Advances and Future Prospects.* Dordrecht: Springer, doi: 10.1007/978-94-007-2217-0

Wani, S. H., Sah, S. K., Sági, L., and Solymosi, K. (2015). Transplastomic plants for innovations in agriculture. A review. *Agron. Sustain. Dev.* 35, 1391–1430. doi: 10.1007/s13593-015-0310-5

Waters, A. J., Capriotti, P., Gaboriau, D. C. A., Papathanos, P. A., and Windbichler, N. (2018). Rationally-engineered reproductive barriers using CRISPR CRISPRa: an evaluation of the synthetic species concept in *Drosophila melanogaster. Sci. Rep.* 8:13125. doi: 10.1038/s41598-018-31433-2

Whaley, K. J., Hiatt, A., and Zeitlin, L. (2011). Emerging antibody products and Nicotiana manufacturing. *Hum. Vaccin* 7, 349–356. doi: 10.4161/HV.7.3.14266

Wurtzel, E. T., Vickers, C. E., Hanson, A. D., Millar, A. H., Cooper, M., and Voss-Fels, K. (2019). Revolutionizing agriculture with synthetic biology. *Nat. Plants* 5, 1207–1210. doi: 10.1038/s41477-019-0539-0

Yao, J., Weng, Y., Dickey, A., and Wang, K. Y. (2015). Plants as factories for human pharmaceuticals: applications and challenges. *Int. J. Mol. Sci* 16, 28549–28565. doi: 10.3390/ijms161226122

Yusibov, V., Rabindran, S., Commandeur, U., Twyman, R. M., and Fischer, R. (2006). The potential of plant virus vectors for vaccine production. *Drugs R D* 7, 203–217. doi: 10.2165/00126839-200607040-00001

Zeugolis, D. I., Li, B., Lareu, R. R., Chan, C. K., and Raghunath, M. (2008). Collagen solubility testing, a quality assurance step for reproducible electro-spun nano-fibre fabrication. A technical note. *J. Biomater. Sci. Polym. Ed.* 19, 1307–1317. doi: 10.1163/156856208786052344

Zhang, C., Norris-Caneda, K. H., Rottmann, W. H., Gulledge, J. E., Chang, S., Kwan, B. Y. H., et al. (2012). Control of pollen-mediated gene flow in transgenic trees. *Plant Physiol.* 159, 1319–1334. doi: 10.1104/pp.112.197228

Zhang, D., VanFossen, A. L., Pagano, R. M., Johnson, J. S., Parker, M. H., Pan, S., et al. (2011). Consolidated pretreatment and hydrolysis of plant biomass expressing cell wall degrading enzymes. *Bioenergy Res.* 4, 276–286. doi: 10.1007/s12155-011-9138-2

Zhang, L., Jing, F., Li, F., Li, M., Wang, Y., and Wang, G. (2009). Development of transgenic *Artemisia annua* (Chinese Wormwood) plants with an enhanced content of artemisinin, an effective anti-malarial drug, by hairpin-RNA-mediated gene silencing. *Biotechnol. Appl. Biochem.* 52:199. doi: 10.1042/ba20080068

Zhang, X., Buehner, N. A., Hutson, A. M., Estes, M. K., and Mason, H. S. (2006). Tomato is a highly effective vehicle for expression and oral immunization with Norwalk virus capsid protein. *Plant Biotechnol. J.* 4, 419–432. doi: 10.1111/j.1467-7652.2006.00191.x

An Interdisciplinary Approach to Study the Performance of Second-generation Genetically Modified Crops in Field Trials: A Case Study With Soybean and Wheat Carrying the Sunflower HaHB4 Transcription Factor

Fernanda Gabriela González[1], Nicolás Rigalli[2], Patricia Vivian Miranda[3,4], Martín Romagnoli[2], Karina Fabiana Ribichich[5], Federico Trucco[3], Margarita Portapila[2], María Elena Otegui[6]* and Raquel Lía Chan[5]*

[1] Estación Experimental Pergamino, INTA, CITNOBA, CONICET-UNNOBA, Pergamino, Argentina, [2] CIFASIS, Universidad Nacional de Rosario—CONICET, Rosario, Argentina, [3] Instituto de Agrobiotecnología Rosario (INDEAR)/BIOCERES, Rosario, Argentina, [4] CONICET Buenos Aires, Argentina, [5] Instituto de Agrobiotecnología del Litoral, Universidad Nacional del Litoral—CONICET, Facultad de Bioquímica y Ciencias Biológicas, Santa Fe, Argentina, [6] CONICET-INTA-FAUBA, Estación Experimental Pergamino, Facultad de Agronomía Universidad de Buenos Aires, Pergamino, Argentina

Research, production, and use of genetically modified (GM) crops have split the world between supporters and opponents. Up to now, this technology has been limited to the control of weeds and pests, whereas the second generation of GM crops is expected to assist farmers in abiotic stress tolerance or improved nutritional features. Aiming to analyze this subject holistically, in this presentation we address an advanced technology for drought-tolerant GM crops, upscaling from molecular details obtained in the laboratory to an extensive network of field trials as well as the impact of the introduction of this innovation into the market. Sunflower has divergent transcription factors, which could be key actors in the drought response orchestrating several signal transduction pathways, generating an improved performance to deal with water deficit. One of such factors, HaHB4, belongs to the homeodomain-leucine zipper family and was first introduced in Arabidopsis. Transformed plants had improved tolerance to water deficits, through the inhibition of ethylene sensitivity and not by stomata closure. Wheat and soybean plants expressing the *HaHB4* gene were obtained and cropped across a wide range of growing conditions exhibiting enhanced adaptation to drought-prone environments, the most important constraint affecting crop yield worldwide. The performance of wheat and soybean, however, differed slightly across mentioned environments; whereas the improved behavior of GM wheat respect to controls was less dependent on the temperature regime (cool or warm), differences between GM and wild-type soybeans were remarkably larger in warmer compared to cooler conditions. In

both species, these GM crops are good candidates to become market products in the near future. In anticipation of consumers' and other stakeholders' interest, spectral analyses of field crops have been conducted to differentiate these GM crops from wild type and commercial cultivars. In this paper, the potential impact of the release of such market products is discussed, considering the perspectives of different stakeholders.

INTRODUCTION

The challenge imposed by the expected increase in food demand by 2050 will not be accompanied by the necessary increase in the relative rate of yield progress of major grain crops (wheat, maize, rice, and soybean), even considering breakthroughs as photosynthesis improvement *via* bioengineering (Hall and Richards, 2013). Additionally, grain yield (GY) losses of 8–43% respect to present-day yields are estimated for these crops, as climate change is predicted to increase temperatures and the frequency of extreme events such as drought (IPCC, 2014). This scenario will be accompanied by a largely augmented demand of water for direct human use, all trends that will require technologies aimed to improve and secure crop production with maximum resource-use efficiency and low environmental impact.

The difference between potential and actual GY varies extensively depending on the production area and the unpredictable climate. Water deficits of variable duration and intensity are among the main determinants of mentioned GY losses (Aramburu et al., 2015; Rattalino Edreira et al., 2018), and breeding efforts to increase crop tolerance to abiotic stress represent an environment-friendly avenue to reduce this gap.

Researchers worldwide have been working during decades applying different breeding strategies to increase crops GY under unfavorable environments (Campos et al., 2004), in a process that traditionally takes a long way from the initial stage in nurseries or *gene discovery* in labs to final adoption by farmers (Hall and Richards, 2013).

The present study contributes to a better understanding of the possibilities, difficulties and significant time requirements that occur when a transgenic technology developed in a model plant such as Arabidopsis is upgraded to evolutionary distant and economic important crops like wheat or soybean. A successful case is presented here: wheat and soybean transformed with the sunflower transcription factor gene *HaHB4* (*Helianthus annuus HomeoBox 4*). Transformed cultivars of both species are expected to be released to the market soon (probably during 2020), and this success was achieved through common and cooperative efforts of molecular biologists and agronomists from public institutions and private companies, who were able to overcome the additional obstacle usually imposed by epistemological barriers (Sadras and Richards, 2014).

ALTHOUGH GENETICALLY MODIFIED CROPS HAVE BEEN ADOPTED WORLDWIDE, SECOND GENERATION IS ABSENT IN THE MARKET

Since 1996, genetically modified (GM) crops have been adopted by farmers worldwide because they increase food and feed production efficiently by generating plants with higher GY in reasonably short times. The main advantage of transgenic technologies is the possibility to overcome sexual incompatibilities between plants and species barriers allowing the introduction of genes from unrelated organisms such as bacteria, fungi or other plants and also from viruses. Genetically modified organisms (GMOs), however, triggered controversies both in adopting and non-adopting countries, although stringent regulatory processes for food/feed and environmental safety were implemented and applied. Furthermore, several countries have established mandatory GMO labeling whereas voluntary labeling is preferred by others (Kamle et al., 2017). Most of the recently developed commercial GM crops exhibit herbicide tolerance, insect resistance or both traits stacked.

The second generation of GM crops was projected to mitigate abiotic stress effects. However, such crops are not commercially available so far and this is due to several reasons. Firstly, most evaluated events failed to translate benefits observed in controlled environments to field conditions (Passioura, 2012). Additionally, a long process is needed to release a GM product to the market, and this is due to regulation requirements mostly as a consequence of the bad public perception about GMOs (Blancke et al., 2015; Fernbach et al., 2019). Huge investments are needed to accomplish such requirements and have limited often attempts to advance GMOs at different stages of development.

An exception to the rule is the maize hybrid expressing a bacterial RNA chaperone that was released for use in a limited, drought-prone region of the United States (Castiglioni et al., 2008). Other drought-tolerant GMOs were developed but not released, such as sugarcane expressing a betaine gene and exhibiting augmented sugar content under water deficits (GM Approval Database, 2019).

UPSCALING DROUGHT TOLERANCE FROM THE POT TO THE CROP

The advent of molecular genetics brought a pronounced increase in the number of studies involving plant transformation aimed to

improve crop performance under water deficit conditions. The initial excitement was soon followed by the striking evidence of serious difficulties to scale up from individual plants grown in pots to communal plants in the field (Passioura, 2012). Lack of success has been usually linked to two weaknesses. One was the lack of a clear understanding of the benefits/drawbacks of a gene/trait at the crop rather than at the single plant level. For instance, compensations that usually take place when moving from plants to crops are of high importance (Pedró et al., 2012). Also, breeders and agronomists do not deal usually with growing conditions for which survival traits may represent an advantage (i.e. very low yielding environments). Traits of value are expected to represent an actual benefit in GY under stressful conditions with no penalty in the high yielding ones. The second weakness was poor knowledge of the variability in drought scenarios (i.e. opportunity, extent and intensity of drought) and their frequency (Chapman et al., 2000) in what breeders describe as the target population of environments (Cooper et al., 2005). Such variability usually receives little if any attention by molecular biologists (Passioura, 2006).

The water budget of an environment depends upon rainfall distribution and soil characteristics, which combined with evaporative demand, regulate the capacity of plants to hold maximum transpiration or reduce it (Sadras and Milroy, 1996). Reduced transpiration leads to the occurrence of water deficits of variable intensity and duration (Chapman et al., 2000; Connor et al., 2011). Rainfall distribution, together with temperature patterns set the limits to crop choice by farmers in rainfed systems, basically between the monsoon climate type (prevalence of summer crops) and the Mediterranean climate type (prevalence of winter crops). There are, however, humid and sub-humid areas where total rainfall allows year-round cropping systems but many times with large intra-seasonal variability if there is an occurrence of drought (Harrington and Tow, 2011). Simultaneously to climate, the soil type (i.e. texture) and its condition (e.g. compaction) affect total plant-available soil water storage as well as the amount that is readily available to plants (Passioura, 1991; Dardanelli et al., 2004).

Assuming cycle duration and its partitioning between vegetative and reproductive phases have been optimized to the water budget of each environment (Passioura, 2006), variable conditions experienced by the soil–canopy–atmosphere continuum along the cycle may pose an additional challenge when breeding crops for drought-prone regions. Passioura (2006) proposed to consider three main issues when evaluating a trait for these regions. The trait should increase (i) water use by transpiration (WUt, in mm) of the limited water supply, (ii) transpired water use efficiency for biomass production (WUEt: biomass produced per unit of water transpired, in kg ha^{-1} mm^{-1}), and/or (iii) biomass allocation into harvestable products, namely harvest index in grain crops (HI: grain biomass/total biomass). These recommendations are based on our understanding of the physiological determination of GY at a crop level (Eq. 1)

$$GY\ \left(kg\ ha^{-1}\right) = WUt \times WUEt \times HI \qquad (1)$$

Therefore, the ability of a gene/trait to cope with water constraints should be analyzed within the conceptual framework of Eq. 1 and the characteristics of the target environment (**Table 1**). The latter will settle whether a trait is favorable or unfavorable for a breeding program.

THE STORY OF HaHB4, FROM THE BENCH TO THE FIELD

It is well known and documented that there are plant species exhibiting a high tolerance to certain abiotic stress factors and others more susceptible to them (Boscaiu et al., 2012). Moreover, it is possible to find different varieties of the same species with differential tolerance to a stress factor for which traits conferring tolerance have been pyramided along centuries, first by farmers and subsequently by professional crop breeders. For example, there are rice lines tolerant to drought, extreme temperatures or salinity (https://www.irri.org/climate-change-ready-rice). Among crops, the sunflower is a species exhibiting broad adaptation (Debaeke et al., 2017), a characteristic for which responsible genes have been identified. Unfortunately, genetic tools as characterized mutants are not available and the complex sunflower genome was revealed only very recently (Badouin et al., 2017). Due to the complexity of molecular networks displayed when plants sense stressing conditions, transcription factors (TFs), as master switches, were good candidates to start the research.

TFs are proteins able to recognize and bind specific short DNA sequences present in their target regulatory regions; i.e.

TABLE 1 | Crop level analysis of the aptitude to cope with drought-prone environments of some traits.

Trait	Physiological GY Determinant	Target Environment
High $^{12/13}$C discrimination (C3 species)	WUEt	Terminal drought
Low $^{12/13}$C discrimination (C3 species)	WUEt	Mild, transient stress
Spike photosynthesis (wheat)	WUEt	All
Leaf rolling	WUEt	All
Increased boundary layer by wax or pubescence	WUEt	Terminal drought
Radiation use efficiency (light and N distribution within the canopy)	WUEt	All
High stomata sensitivity (i.e. fast stomata closure)	WUt	Terminal drought
Low stomata sensitivity	WUt	No terminal drought
Increased root growth	WUt	Deep water source
Osmotic adjustment	WUt	Transient stress, no terminal drought
Stay-green	WUt	No terminal drought
High floret fertility, reduced grain abortion	HI	Transient stress at flowering, no terminal drought
Pre-anthesis partitioning to stem carbohydrates	HI	Terminal drought

Adapted from Passioura (2006); Reynolds et al. (2007), Tardieu (2012) and Sadras and Richards (2014).

promoters, enhancers, introns, etc. These proteins are particularly abundant in the plant kingdom, representing about 6% of total proteins (Riechmann, 2002; Ribichich et al., 2014). Plant TFs have been classified in families, subfamilies, and subgroups especially according to their DNA binding domains. However, other features are also important for such classification including gene structure, the presence of other motifs and domains as well as their role in plant development. Many TF families are shared between animals and plants and among them, 19 are more expanded in the plant kingdom suggesting a frequent adaptive response, besides a higher genome duplication rate (Shiu et al., 2005).

Among TFs families, there is the superfamily of homeodomain (HD) containing proteins. The HD was defined as a conserved 60 aminoacid sequence that folds in three alpha helixes bound by a loop and a turn (reviewed in Viola and González, 2016). This domain was discovered (and named as HD) in mutants of *Drosophila melanogaster* exhibiting the ectopic expression or mutation of a HD encoding gene which caused a homeotic effect, i.e. the change of a body segment by another. Examples of these TFs are *Antennapedia* and *Bithorax* (Gehring et al., 1994). In plants, HD TFs have not been assigned homeotic functions, but they play many roles in development, hormone signaling and the response to environmental factors (Viola and González, 2016).

The HD superfamily is divided into several subfamilies and among the latter, there is the HD-Zip (homeodomain-leucine zipper) family which is subdivided into four groups, named I to IV, according to their different structures and roles. Even though HD and leucine zipper form part of TFs in other kingdoms, their association in a sole protein is exclusive of plants and this characteristic is shared by the four groups (Ariel et al., 2007). Among these four subfamilies, members of the so-called HD-Zip I subfamily have been associated initially with abiotic stress response (Ariel et al., 2007). Subsequently, several works [revised by (Perotti et al., 2017)] described the role of particular members in developmental events not necessarily associated with stress and also in biotic responses. There are 17 HD-Zip I members in the model plant Arabidopsis, a number that varies among species (Perotti et al., 2017). HD-Zip I TFs have been identified in all the species in which genomes have been sequenced; however, a small portion has been functionally characterized (Perotti et al., 2017). Phylogenetic trees resolved these proteins from different species in six clades (Arce et al., 2011). Coming back to the sunflower, it is noteworthy that this species has several divergent HD-Zip I TFs that cannot be clustered in the trees constructed with proteins from model species or crops (Arce et al., 2011). Among these divergent TFs, HaHB4 presents an abnormally short carboxy-terminal and a short size. Taking only its HD-Zip domain, the closest Arabidopsis members to HaHB4 are AtHB7 and AtHB12, which have been shown to participate as positive regulators in ABA-dependent drought and salinity responses (Olsson et al., 2004; Valdés et al., 2012; Ré et al., 2014). The overexpression of *AtHB7* conferred drought tolerance in Arabidopsis and tomato plants (Olsson et al., 2004; Mishra et al., 2012; Ré et al., 2014).

At the beginning of the research reviewed here, and ignoring at that time the Arabidopsis genome as well as the functions of AtHB7 and AtHB12 and other members of the HD-Zip I family, the strategy to reveal the HaHB4 function was to study its binding specificity *in vitro* and its expression pattern in sunflower (Palena et al., 1999; Gago et al., 2002). The next step was to transform Arabidopsis plants overexpressing this TF (Dezar et al., 2005). Expression studies indicated that *HaHB4* is induced by water deficit and ABA (Gago et al., 2002). Arabidopsis transgenic plants, transformed with the sunflower TF, showed a drought-tolerant phenotype (Dezar et al., 2005). A deeper investigation about the mechanism triggered by this gene to confer drought tolerance indicated that it did not implicate stomata closure (which leads to drought tolerance but is usually accompanied by yield penalty in some environments) but a senescence delay *via* the inhibition of ethylene receptors (Manavella et al., 2006). Other signal transduction pathways are also regulated by this TF such as jasmonic acid enhancement, which leads to herbivory defense and inhibition of photosynthesis-related genes during darkness (Manavella et al., 2008a; Manavella et al., 2008b). However, the most important discovery was that Arabidopsis plants became, in certain form, 'myopic' to water deficit; they continue to grow when the stress was moderate and thus, the impact on productivity was reduced respect to control plants that exhibited stomata closure. In other words, when the plants were subjected to severe stress (not watered during 10-20 days), survival rates were much higher for GM plants expressing *HaHB4* than for wild-type controls (Dezar et al., 2005). Such a drought-tolerant phenotype was observed for many transgenic Arabidopsis plants expressing a variety of plant genes. However, the described trend did not hold when the same plants were grown under moderate water deficit (20–40% reduction of rosette area; Skirycz et al., 2011), a condition for which no clear trend was detected in yield penalty between GM and non-GM genotypes. By contrast, *HaHB4*-transgenics usually outyielded the wild-type controls across a wide range of field-tested growing conditions (**Figure 1**).

Described observations lead us to transform other species, particularly crops such as soybean and wheat, with constructs able to express *HaHB4* (González et al., 2019; Ribichich et al., 2020). Such GM crops expressing *HaHB4* outyielded their wild-type counterparts in a network of field trials that included a broad range of growing conditions, particularly in water balance and air temperature during critical reproductive stages (González et al., 2019; Ribichich et al., 2020). Yield data were supported by positive trends in its main physiological determinants (total biomass and biomass partitioning) as well as in floret fertility and grain numbers.

WHEAT AND SOYBEAN HaHB4: CROP PERFORMANCE AND YIELD IMPROVEMENT

Wheat and soybean expressing the cDNA corresponding to the sunflower *HaHB4* gene were tested in 37 and 27 field
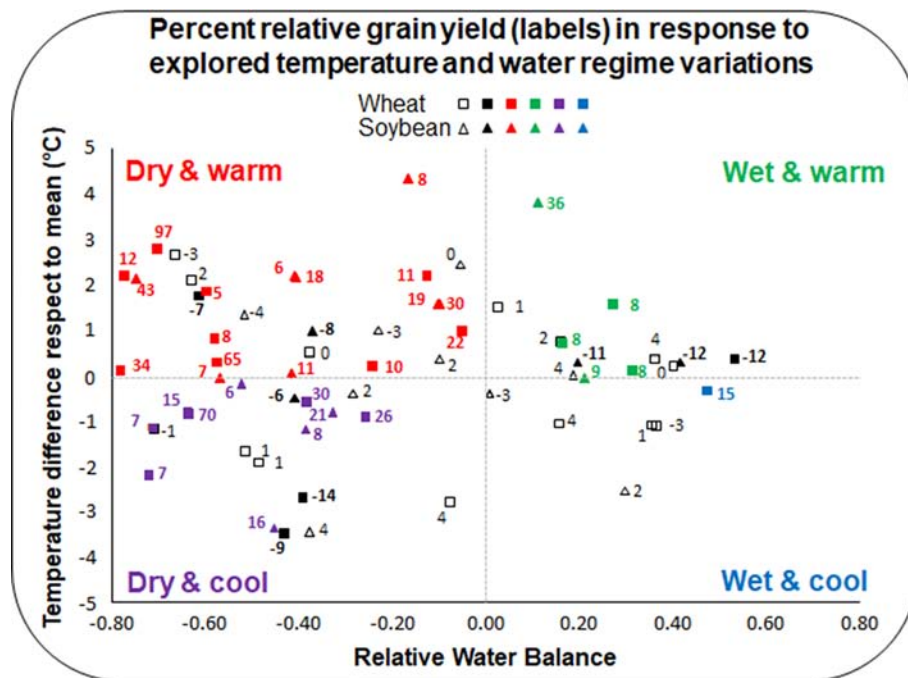
FIGURE 1 | Relative grain yield response of transgenic wheat and soybean lines across environments. For each species (triangles for soybean and squares for wheat), symbols represent the combination of (i) the difference in mean temperature of each site respect to the mean across environments (y axis), and (ii) the relative water balance (RWB) of each site (x axis), being RWB=(Rainfall+Irrigation-PET)/PET (PET: potential evapotranspiration). Variation in relative grain yield (RGY) was computed as RGY = (GYtg-GYwt)/GYwt (GYtg: grain yield transgenic; GYwt: grain yield wild-type) and expressed in percent (values next to symbols). Different colors represent cases with (i) RGY ≥ 5% (GYtg > GYwt), in bolded non-black colors that identified the corresponding environmental group, (ii) RGY ≤ −5% (GYwt > GYtg), in bolded black, and (iii) −5% < RGY < 5% (GYtg = GYwt), in plain black.

experiments, respectively (**Supplementary Table 1**; González et al., 2019; Ribichich et al., 2020)[1]. Genetic constructs used to transform crops shared *HaHB4* cDNA but not the promoters which were the UBI for wheat and the own *HaHB4* promoter for soybean. The parental wild-type cultivars (Cadenza for wheat and Williams 82 for soybean) together with the GM lines (IND-00412-7 for wheat and b10H for soybean) were sown in 13 or 14 sites during several years covering wide latitudinal (ca. 27°25′S to 39°50′S) and longitudinal (ca. 57°40′W to 65°28′W) ranges. Both crops experienced large differences in water balance as well as in mean and maximum temperatures (**Supplementary Table 1**). Considering all the range of tested environments (1,000–9,300 and 1,500–4,500 kg ha^{-1} average yield for wheat and soybean, respectively), the presence of *HaHB4* in the GM lines increased yield by 6% in wheat and by 4% in soybean, with no significant effect in crop phenology. Such a response is outstanding for commercial purposes because it allowed the incorporation of *HaHB4* to modern cultivars without altering the crop cycle, which has been already optimized to the target breeding area. When only the dry environments (i.e. negative water balance) were considered, the mean yield benefit from *HaHB4* increased to 16% in wheat and to 8.6% in soybean (**Figure 1**). Moreover,

yield improvement was even larger (20% for wheat and 11% for soybean) when the dry environment was associated with warm mean temperatures (>20 and >22°C, respectively), while it remained important (12% and 5% for wheat and soybean, respectively) for the dry-cool condition (**Figure 1**). Differences in yield were always associated with grain number (GN) produced per unit area for both crops. This trend was partially compensated by a decrease in individual grain weight (GW) in soybean, whereas no clear trade-off effect was registered in this grain yield component for wheat. This contrasting response of GW to the increase in GN agrees with differences in the source-sink balance experienced by each species during grain filling (Borrás et al., 2004). Such balance recognizes specific (e.g. plasticity in the establishment of maximum seed volume; soybean > wheat) as well as environmental (e.g. irradiance offer during grain filling; wheat > soybean) constraints.

To improve our understanding of *HaHB4* effects on the ecophysiological determination of GY, detailed measurements were performed in controlled field experiments. The simplest method to study yield determination is to evaluate the total biomass produced along the crop cycle and the proportion of that biomass allocated to grains as proposed in Eq. 1 and summarized in **Figure 2**. In both crops, the expression of *HaHB4* caused an increase in total biomass with no change in harvest index (HI). As HaHB4 had no impact on crop
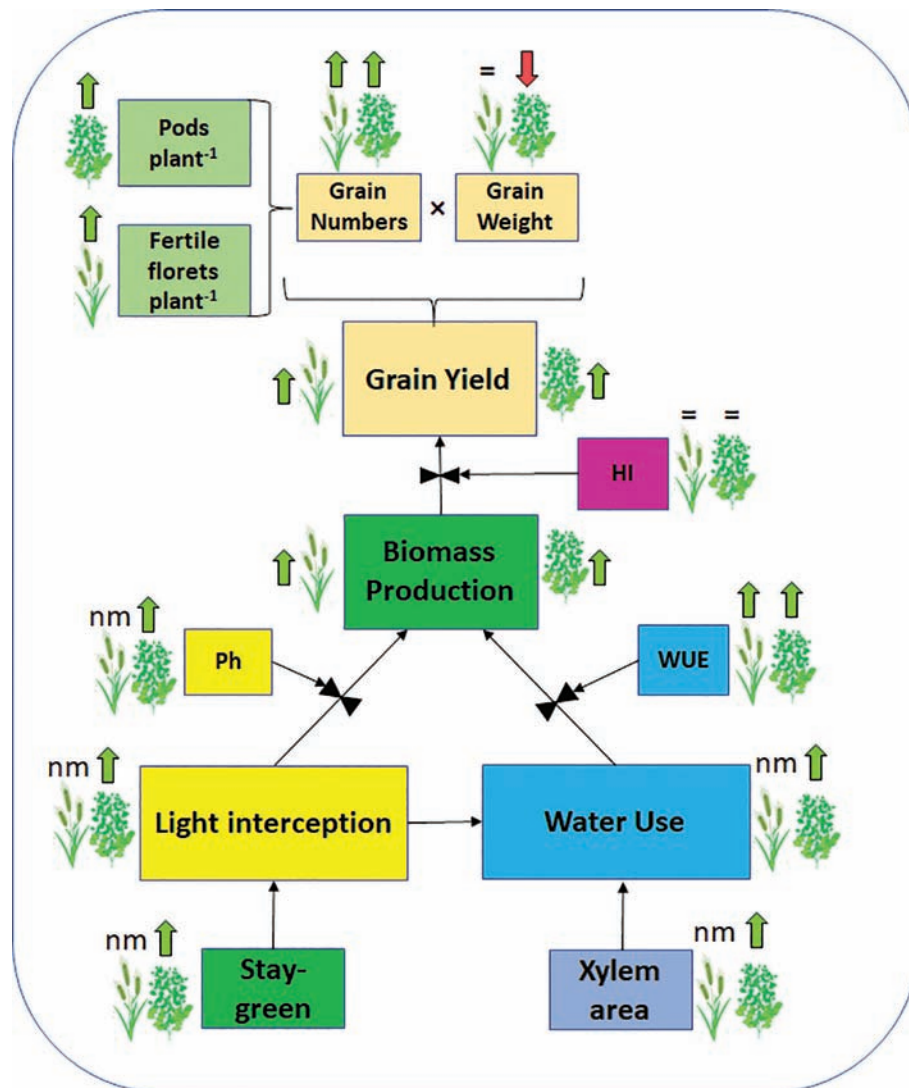
FIGURE 2 | Schematic representation of HaHB4 effects. Effects of HaHB4 on some of the physiological determinants of wheat and soybean grain yield as well as on grain yield components (grain numbers and grain weight). Arrows indicate positive (green upward) or negative (red downward) effects respect to the wild-type when grown under water deficit. The equal sign indicates no change and nm indicates not measured. HI, harvest index; Ph, photosynthesis; WUE, water use efficiency.

phenology, the observed increase in total biomass could be attributed to increased crop growth rate, particularly during those periods that are critical for the determination of the main driver of grain yield (i.e. GN). This period spans between (i) the start of stem elongation on ca. 20 days before anthesis and grain set at the beginning of grain filling in wheat (Fischer, 1975; Fischer, 1985; Kirby, 1988), and (ii) pod formation and the beginning of grain filling in soybean (Board and Qiang, 1995; Jiang and Egli, 1995). The crop growth rate of wheat GM line exceeded that of the wild-type parental line by 68% during the critical period. This trend was in line with the increase registered in fertile florets per plant observed in GM lines, suggesting an improved floret survival (González et al., 2011). The fertile florets per spike and the number of spikes per plant constitute the fertile

florets per plant, being the former similar (tiller spike) or higher (main stem spike), and the latter consistently higher, in the GM line compared to the wild type. For soybean, solar radiation interception and leaf photosynthesis were measured, both conducive to crop growth rate determination (Muchow et al., 1990). These two traits were higher for the GM line, the former during the entire critical period and the latter during grain filling. Improved light interception during the critical period resulted in more pods and branches per plant, whereas enhanced photosynthesis during grain filling was consistent with the clear visual observation of delayed senescence (**Figure 3**). Described stay-green improvement, which was also observed in model plant Arabidopsis (Manavella et al., 2006), probably prevented a complete trade-off between increased grain
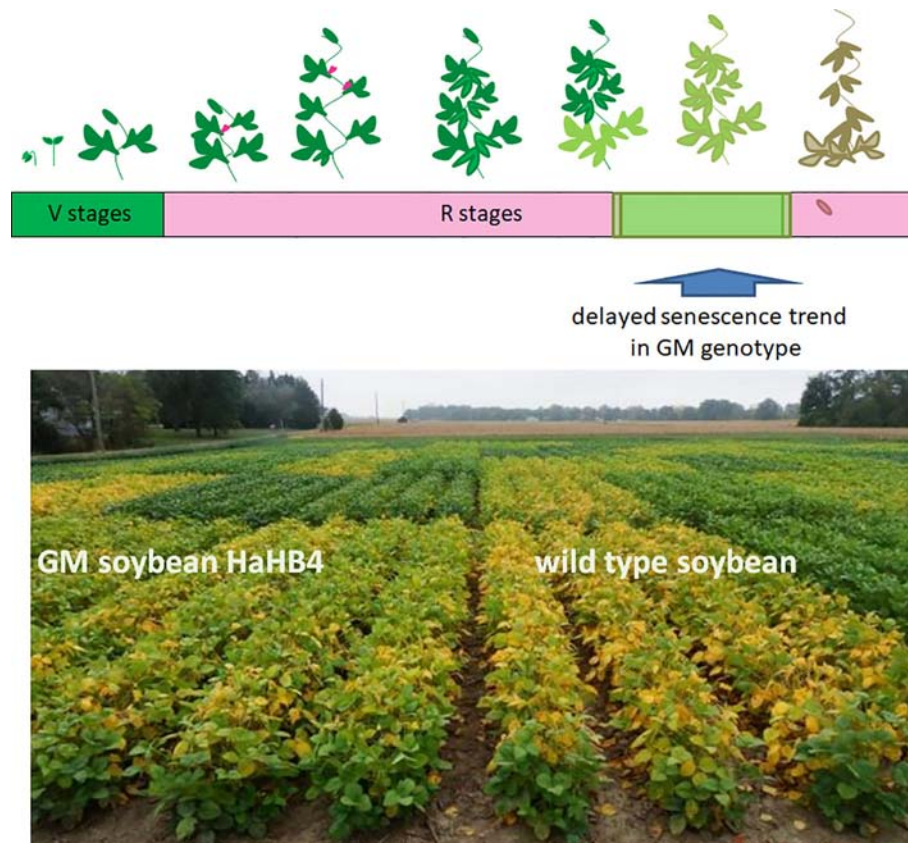
FIGURE 3 | GM soybean exhibits delayed senescence compared with its wild-type control. Upper panel: schematic representation of the soybean life cycle. Lower panel: illustrative picture of one of the field trials performed comparing the wild type genotype (right) with the transgenic *HaHB4* one (left).

numbers and final individual grain weight, which can be expected when seed growth takes place under the sharp decrease in irradiance that usually occurs in autumn (Borrás et al., 2004).

When water availability is limiting growth and yield, water use (WU) and water use efficiency (WUE) are the main physiological determinants of crop performance (Eq. 1). In the case of wheat, the average WUE (estimated as the yield produced per unit of rainfall) of the 37 field experiments was 9.4% greater in the GM line than in the wild type, and WUE increased by 14.2% for environments with less than 300 mm rainfall. For soybean, detailed measurements of crop evapotranspiration along the cycle in plots exposed to contrasting water regimes (WW: well-watered; WD: water deficit) showed that the GM line used more water under both conditions, being the difference even higher under irrigation (17.3 and 27.2% increase in water used for WD and WW, respectively). The enhanced water use of the GM line could not be confirmed in a greenhouse experiment, where both cultivars had almost identical water use in pots held at contrasting water regimes (field capacity and 60% field capacity). Such a response is not surprising under the severe root confinement usually experienced in pot experiments, which do not allow for correct comparisons of this type of traits.

Concurrently with mentioned differences in water use in the field, the hypocotyl diameter and xylem area were always larger in the GM line than in the wild-type soybean line, traits that have been associated with increased water conductivity and water use (Richards and Passioura, 1981). The fact that water use was reduced under water deficit, even in the GM line, suggests that some degree of stomata closure may have occurred, reducing water loss but with low impact in CO_2 exchange (Liu et al., 2005). This response is in line with the increased WUE ($\geq$22%) to produce biomass, and yield of the GM line when exposed to water deficit, and with the augmented photosynthetic rate observed in this germplasm during grain filling (commented in the previous paragraph). The results obtained in wheat and soybean crops are promising, showing that *HaHB4* may help to mitigate yield reductions in drought-prone environments.

THE LONG REGULATORY PROCESS

Developing new technology is a long journey, from the hypothesis to its verification in the plant model, to the posterior projection into agronomic relevant crops and, when successful, the selection of the best candidate fulfilling the

expected features (**Figure 4**). However, when all these stages are completed, another hard challenge starts. Since a precautionary approach is often favored for new technologies, a detailed safety assessment is required before their consumption or introduction into the environment.

While the approvals are obtained, these products are "regulated", which means that they must be under strict control to guarantee no accidental release into the environment until its safety is proven. The regulatory stage is a mandatory process during which a lot of information covering different aspects of the new product must be delivered and presented to the respective authorities.

Focusing on a GM crop, the final objective of the regulatory phase is intended to establish that the new plant is like the conventional version except on the trait that was intentionally introduced by the genetic modification. The assumption under this *comparative approach* is that, if the conventional plant is considered secure, the genetically modified version would be equally safe. Behind this comparative approach that includes evaluation of different crop aspects (agronomic, reproductive, environmental, compositional, nutritional, etc.), there is a long and thorough process that involves measuring more than 80 parameters in which equivalence to the non-GM counterpart needs to be verified (Ayala et al., 2019; González et al., 2019).

Complementing the comparative approach, there is a set of distinctive (new) features that must be characterized. These include:

- Studying the new expression product. Any new molecule introduced into the crop by the genetic modification (usually a protein) must be deeply studied. In the case of a protein, its function, biochemical profile, putative
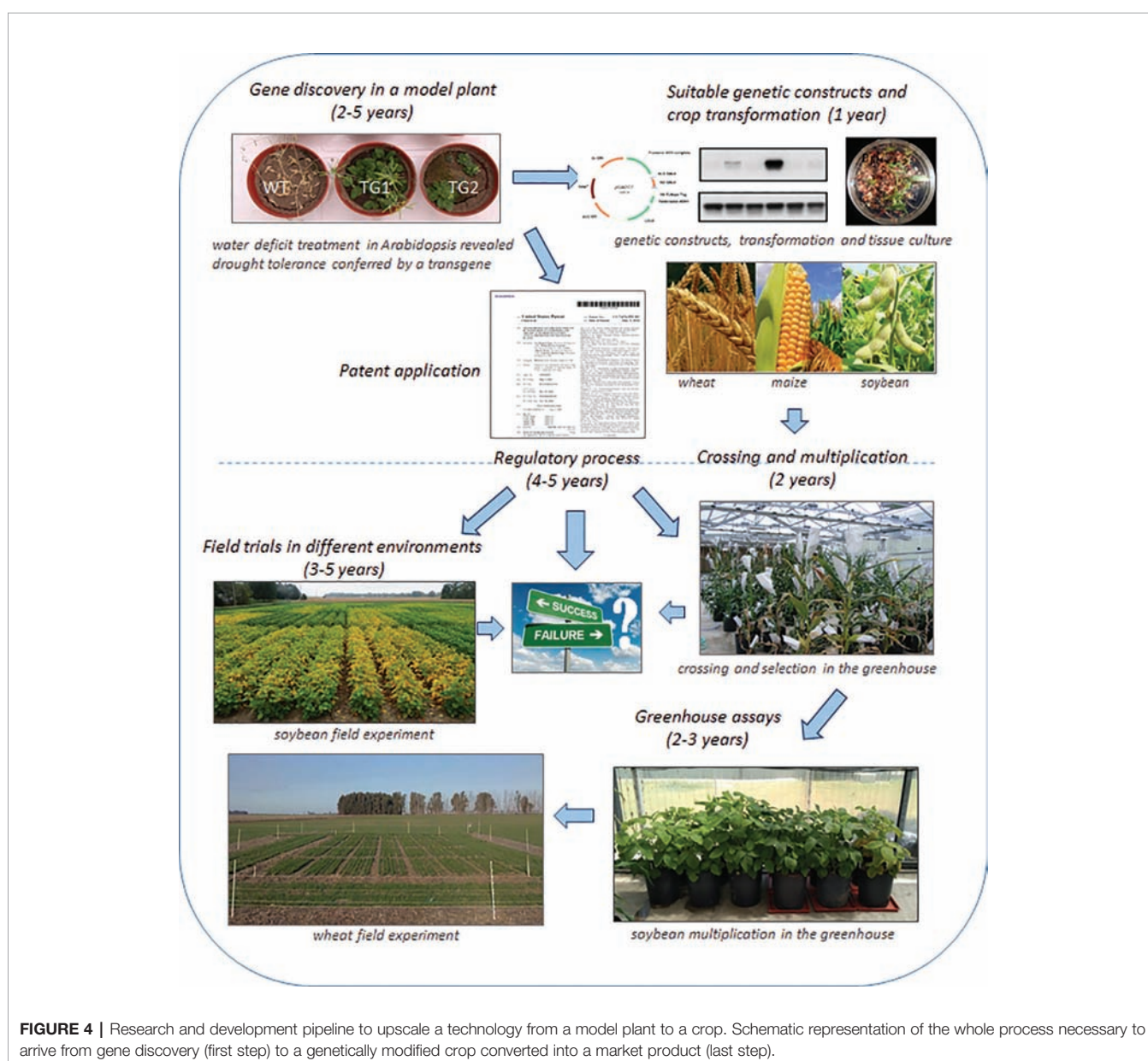
FIGURE 4 | Research and development pipeline to upscale a technology from a model plant to a crop. Schematic representation of the whole process necessary to arrive from gene discovery (first step) to a genetically modified crop converted into a market product (last step).

allergenicity, and toxicity and digestibility, among other attributes, must be known.

- Examining if the new molecule and metabolites derived from it can interact with endogenous plant pathways.
- Describing the genetic material inserted, its origin, location within the genome, full characterization of the insertion, analyzing whether endogenous genes may have been interrupted by the genetic modification, analyzing putative production of unexpected products and its potential risk for consumption, and stability of the genetic modification.
- Many of the studies are planned in such a way that if any unintended effect is produced, it would be necessarily revealed by these studies, and that is why a wide agronomic and compositional characterization of the crop is a major part of any regulatory process (Ayala et al., 2019).

Field trials including the GMO cultivars are developed under different agronomic conditions. Simultaneously, the non-GM parental line and several commercial varieties are grown to control any effect of the genetic modification and provide a reference range of natural variability, respectively. A wide set of agronomic parameters that define the crop characteristics are measured during the life cycle and samples of different tissues are taken for further analysis.

Particular attention is taken to any parameter that may reveal a tendency of the GMO to become a weed or be more invasive than its wild-type counterpart. That is why reproductive physiology, persistence in the environment, sensitivity to stresses, diseases, and plagues are analyzed.

All described studies are completed by the developer since they involve the use of material protected by intellectual property rights and require a great investment of money, other resources and time. The results are presented to the regulatory authority of different countries and evaluated by experts in different areas. A post-submission period of communication between regulators and the developer usually follows, during which requests for additional information or even new studies may be required. The conclusions from these scientific analyses are then considered by governmental authorities, which take into consideration these evidence-based conclusions together with many other local interests.

Regulatory field trials are usually conducted in the main areas where the crop is cultivated. However, countries like China, Brazil, Japan, Colombia, Bolivia, South Africa, and the USA require local trials even when the transportability of the data has been proven (Garcia-Alonso et al., 2014).

Destinations, where a new GMO is presented for approval, include those where the crop will be cultivated and/or where its products will be shipped. Information presented for approval is usually that related to the use of the new GMO in the country of destination (information required for cultivation is different from that associated with consumption or use of imported products), although some countries/regions are extremely precautious and do not follow this rule.

The regulatory road is usually tortuous. Although several countries share the essence of a science-based safety evaluation process, there are still dissimilar requirements among regulatory authorities. Besides, there is a lack of mutual recognition on safety assessments that leads to redundant evaluations and asynchronous approvals that significantly impairs commodities commercialization (Stein and Rodríguez-Cerezo, 2010). In particular, some countries or regions are politically decisive players in this field. Specifically, the European Union embodies the stringency of a regulatory process requiring data not necessarily related to science or intention of use. Alternatively, China regulatory authorities run their own set of experiments to verify the data already generated by the developer.

One distinctive feature of the HB4[®] technology (commercial name for HaHB4 introduced in different crops) is that the new expression product is a plant TF. Based on this singularity, this technology has faced additional scrutiny related to putative effects on non-target genes. Precedents on the extreme specificity of TF (an absolute requirement for the development of any individual), and evidenced on the different HaHB4-GM crops, do not support such speculation.

As a TF, the levels of *HaHB4* on the natural plant (sunflower) as well as in the GM crops expressing it are extremely low (at the nanogram per gram of dry weight; Alloatti et al., 2017). So low, that it can hardly be detected in the plant and, consequently, in its byproducts. If any concern could be raised by the expression of a foreign protein in a crop, in the case of HaHB4 it would be even lower than for other proteins.

Independently of the uniqueness of HaHB4, we have a familiarity with this protein, since sunflower has been in the animal and human diets for centuries. Besides, HaHB4 is similar to proteins already present in animals and plants, even working as TF.

Among the two HaHB4 crops close to reaching the market, HB4[®] wheat has faced some baseless negative reactions. This is not unexpected considering the anti-GMO movements. However, it is sadly remarkable that those in charge of evaluating the (environmental/commercial) benefits or the safety of new technology, let this putative future public perception determine their decisions (Fernbach et al., 2019). This reaction is supported by the statement that this would be the first GM wheat. However, it is not true. There is a precedent for GM wheat, which completed the approval process in four different countries though it was never commercialized (GM Approval Database, 2019). Concerning this, another false statement is done: this event was rejected by the regulatory authorities. The truth is that the developers shelved the glyphosate-tolerant wheat in 2004 amid market concern about rejection from foreign buyers (Ingwersen, 2019). Besides these two countries where a withdrawal took place (the United States and Australia), this former GM wheat is approved for food and feed use in two other countries (New Zealand and Colombia, GM Approval Database, 2019).

This unique and interrupted intent to introduce GM wheat into the market sustain the feeling that this crop has been kept aside from the improvements that genetic modification may provide (Wulff and Dhugga, 2018; Asseng et al., 2019). Despite the public's distaste for GM foods, especially in Europe, many

genetically engineered products have already reached the market (GM Approval Database, 2019). In addition to public resistance to GM crops in general, wheat faces its particular challenges since twenty percent of humans' calories and about the same percentage of protein intake come from wheat. In Argentina, 80% of the cultivated wheat is exported to Brazil. If buyers will not accept GM wheat, farmers will not grow it. According to a news release survey we conducted, based on articles selected through specific keywords, both in national and international media, stakeholders related to wheat exports speculate about potential market reactions if GM wheat is approved in Argentina, fearing a loss of customers based on not being able to deliver wheat that matches non-GM specifications. To get accurate, and credible information about identifying GM wheat, in the following section we perform a classification to distinguish between GM wheat and non-GM wheat, and between different "events" of GM wheat.

Another negative perception targeted GM wheat is that we are talking about human food and the direct consumption of a GMO. However, GM-derived products are already present in the food chain. For example, glyphosate-tolerant soybean MON-ØØ4Ø32-6 has been in the market for more than 20 years and soybean flour is regularly used as a supplement in several food products. Similarly, insect-resistant corn MON-ØØ81Ø-6 is widely cultivated and consumed since it was approved more than a couple of decades ago (GM Approval Database, 2019). Moreover, some of them are already consumed without prior processing (plum, cassava, apple). So, is there a real difference in approving this GMO or is it just a matter of perception?

Checking the history of GMOs, we can find that even in cases where there are no commercial interests behind a GM technology and its modification is attending a health issue as in the case of vitamin A-fortified golden rice (Dubock, 2014; Moghissi et al., 2018), lack of funding and anti-GMO movements can delay (or even prevent) a benefit to reach those who could be assisted by it.

TRANSGENIC CROPS CAN BE DIFFERENTIATED ON-FARM FROM THE NON-TRANSGENIC ONES BY SPECTRAL ANALYSES

Remote sensing techniques, such as spectrometry, are increasingly used for plant phenotyping. The spectrum of energy reflected by the plant is closely associated with absorption at certain wavelengths that are linked to specific characteristics or plant conditions. Spectrometers can acquire detailed information regarding the electromagnetic spectrum in a short time, making this technology ideal for assessing genotypes within a few hours. This would enable the estimation of multiple morpho-physiological and physicochemical traits, which would be otherwise impossible to evaluate due to the time and cost involved (Garriga et al., 2017).

For the estimation of plant traits, most previous studies have mainly resorted to the use of Vegetation Indices, while less attention was paid to the analysis of the full spectrum. The use of reflectance data and machine learning algorithms for plant phenotyping purposes using the full spectrum has been only recently addressed by the scientific community (Chlingaryan et al., 2018; El-Hendawy et al., 2019; Meacham-Hensold et al., 2019).

The capability of a spectrometer for the characterization of soybean, maize, and wheat in field experiments has been already explored by Rigalli et al. (2018). The objective of this work was to select different wavelengths intervals of the spectral reflectance curve (within the 632–1,125 nm range) as features for on-farm classification using machine learning methods. Two different classifications were presented, species selection and growth stage identification. An accuracy of 92% was reached for species classification, while 99% was obtained for stage classification. Besides, a new index was proposed that outperformed established vegetation indices under analysis, which showed the potential advantage of using this type of device. This fact indicated that a collection of field spectral data could be more representative of plant phenotyping than the information given by single vegetation indices.

In this section, we present spectral analyses of field-grown wheat and soybean crops through field-collected full-spectrum data, in an attempt to differentiate GM genotypes from their wild type and commercial cultivars (**Figures 5** and **6**). Thirty spectral reflectance curves (ten per plot, with three replications) were collected for each genotype. Each dataset contained spectra from two genotypes, giving 60 spectra per dataset to feed the machine learning algorithm. Detail of methods (ANN: Artificial Neural Networks; SVM: Support Vector Machine; RF: Random Forest) can be found in the supplementary material (**Supplementary File 1**) whereas details of experiments are summarized in **Supplementary Table 1**.

Transgenic Versus Wild Type Identification in Soybean and Wheat Genotypes

The best discrimination between GM and wild-type soybean cultivars was achieved with the ANN algorithm (**Table 2**), which reached values ≥70% in six out of 13 cases. Regarding the analyzed environments, the best performance was obtained for the water deficit condition (Environment II). Overall, the maximum standard deviation as a measure of uncertainty in the classification was 20%. In the case of wheat (**Table 3**), collected data belonged to a single environment and included two GM lines and the wild type at different growth stages. In this case, the precision of the classification increased when phenological differences between genotypes were reduced (bolded data in **Table 3**). The highest result for the GM/wild-type comparison (IND-00412-7(HB4)/Cadenza) was obtained on 130 days after sowing when both genotypes were at the boot stage. The outstanding precision value reached at this stage for wheat (96 ± 4% for ANN and 99 ± 8% for SVM) outperformed those registered for soybean (**Table 2**). This high precision allowed us to conclusively distinguish the presence of a GM wheat cultivar from its parental wild type through spectral analysis.
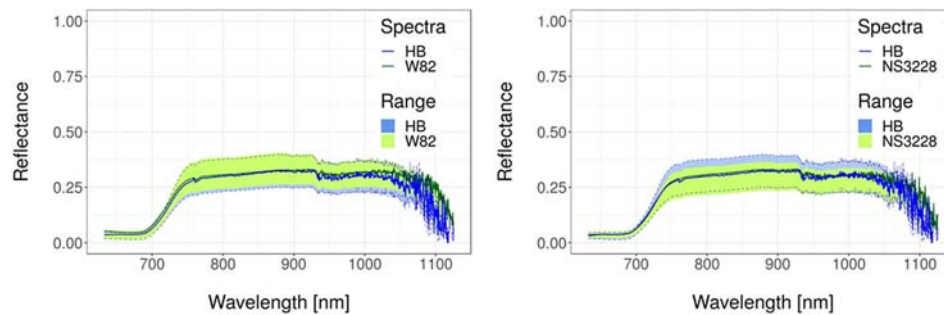
FIGURE 5 | Spectral reflectance curve datasets for soybean. Transgenic versus wild-type and transgenic versus commercial. Solid lines represent typical spectral reflectance curves. Shaded regions represent the data range for each genotype. Green regions represent the data range of the non-transgenic genotypes. Blue regions with dashed border lines represent the dataset range of the transgenic genotypes. Spectra were obtained as described in **Supplementary File 1** whereas details of experiments are in **Supplementary Table 1** and in González et al. (2020).
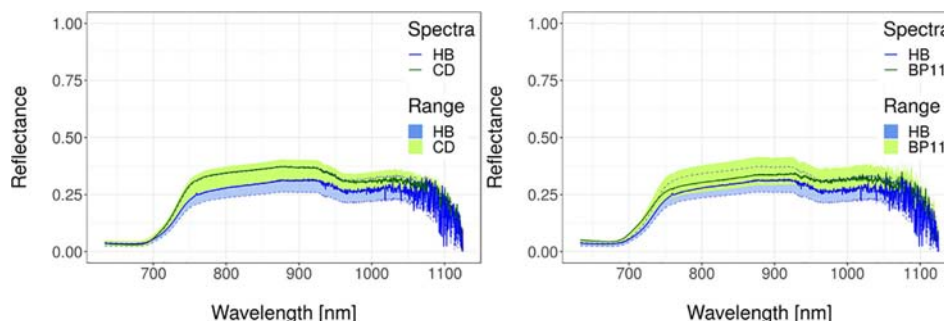
FIGURE 6 | Spectral reflectance data for wheat. Transgenic versus wild-type (left) and transgenic versus commercial (right). Solid lines represent typical spectral reflectance curves. Shaded regions represent the data range for each genotype. Green regions represent the data range of the non-transgenic genotypes. Blue regions with dashed border lines represent the dataset range of the transgenic genotypes. Spectra were obtained as described in **Supplementary File 1** whereas details of experiments are in **Supplementary Table 1** and in González et al. (2019). Transgenic is cultivar IND-00412-7 (HB), wild-type is cultivar Cadenza (CD), and commercial cultivar is Baguette Premium 11 (BP11).

TABLE 2 | Classification accuracy of three machine learning algorithms for discrimination between soybean genotypes.

	Environment I (Pergamino irrigated)					Environment II (Pergamino water deficit)					Environment III (IAL-Santa Fe)		
DAS[a]	**37**	45	**66**	87	100	37	**45**	**66**	**87**	100	**49**	**71**	**96**
Growth Stage	**V8**	R1	**R5**	R6	R7	V8	**R1**	**R5**	**R6**	R7	**R3**	**R5**	**R6**
ANN	0.69	0.65	**0.72**	0.57	0.55	0.68	**0.74**	**0.72**	**0.72**	0.58	**0.7**	0.52	**0.7**
SVM	0.66	0.64	**0.70**	0.57	0.54	0.67	**0.74**	**0.71**	0.65	0.56	0.68	0.52	**0.71**
RF	**0.71**	0.63	0.62	0.5	0.61	0.56	0.68	0.64	0.69	0.57	0.64	0.53	0.66

DAS, days after sowing; ANN, Artificial Neural Networks; SVM, Support Vector Machine; RF, Random Forest. Growth stages according to Fehr and Caviness (1977).
Tested soybean genotypes were b10H (GM) and W82 (wild-type parental). Results expressed as b10H/W82. Bolded data indicate results equal or higher than 70%.

Identification of Transgenic Versus Commercial Genotypes in Soybean and Wheat

Using the same approach, transgenic (b10H, OECD nomenclature IND-00412-5) vs commercial (NS3228) soybean genotypes were evaluated across three environments (**Table 4**). According to these results, ANN obtained again the best

performance, reaching values ≥70% in nine out of 13 cases. SVM also reached good results, showing values ≥70% in eight out of 13 cases. Classification accuracy across all environments was >70% in at least three developmental stages. Also, spectral reflectance data assessed at R5 (start of seed growth) showed classification accuracy values ≥79% for all analyzed environments. The performance improved under water deficit

TABLE 3 | Classification accuracy of three machine learning algorithms for discrimination between wheat genotypes.

DAS Growth Stage (GM/Wild-type)	IND-00412-7*/CD			IND-1015*/CD		
	81 1N-12/1N-15	130 BS+2/BS+2	136 Ant+2/Ant+1	81 1N-15/1N-15	130 Ant/BS+2	136 Ant+6/Ant+1
ANN[a]	0.72	**0.96**	**0.91**	**0.82**	0.79	0.68
SVM	0.77	**0.99**	**0.94**	**0.78**	0.74	0.68
RF	0.71	**0.72**	**0.82**	**0.65**	0.67	0.54

[a]1N, first node visible; Ant, anthesis; BS, boot stage. Other abbreviations as in **Table 2**.

Wheat genotypes used in this analysis were genotypes (i) IND-00412-7 (GM) and wild-type Cadenza (CD), and (ii) IND-1015 (GM) and CD. The growth stage is indicated as the nearer stage plus or minus calendar days from this stage. Bolded data indicate small variation between growth stages of the compared genotypes.

* IND-1015 come from the isogenic transgenic HaHB4 event obtained in Cadenza background (IND-00412-7), then introgressed in a pre-commercial advanced line.

TABLE 4 | Classification accuracy of three machine learning algorithms for discrimination between transgenic and commercial soybean genotypes.

	Environment I (Pergamino irrigated)					Environment II (Pergamino water deficit)					Environment III (IAL-Santa Fe)		
DAS	**37**	45	**66**	87	100	37	**45**	**66**	**87**	100	**49**	**71**	**96**
Growth Stage	**V8**	R1	**R5**	R6	R7	V8	R1	R5	R6	R7	R3	R5	R6
ANN[a]	**0.72**	0.67	**0.79**	**0.76**	0.56	**0.72**	0.68	**0.82**	**0.81**	0.6	**0.81**	**0.84**	**0.71**
SVM	**0.71**	0.64	**0.79**	**0.75**	0.56	0.68	0.69	**0.86**	**0.78**	0.59	**0.80**	**0.84**	**0.70**
RF	**0.72**	**0.70**	0.62	**0.70**	0.54	0.57	0.62	0.68	0.69	0.61	0.65	**0.73**	**0.77**

[a]Abbreviations as in **Table 2**.

Soybean genotypes tested in this analysis were HB (GM) and NS3228 (commercial variety). Results expressed as HB/NS3228 (bolded data indicate ≥0.70).

(Environment II) and high temperature (Environment III), reaching 82–86% and 84% (across methods), respectively. Standard deviations did not go beyond 16%.

As for wheat varieties, we confirmed that a reduced phenological difference increases the discrimination capability between transgenic and non-transgenic cultivars. Accuracy of ANN was always highest (>80%), with standard deviations lower than 18%. Results for SVM were very similar to those obtained by ANN, and with standard deviations of similar magnitude (**Tables 5**).

FUTURE PERSPECTIVES AND CONCLUDING REMARKS

An estimated one-quarter of greenhouse gas emissions are associated with anthropogenic activities linked, directly or indirectly, to agriculture. At the same time, increases in frequency and intensity of extreme weather have adversely affected food security and ecosystems, contributing to desertification and land degradation in many regions (IPCC, 2014). While climate change will likely impact a crop's yield and nutritional value, decreased agricultural outputs will fail to meet demands as population increases. Consequently, agriculture faces a major challenge: to enhance the resilience of global food systems and at the same time move towards carbon neutrality.

The unprecedented challenge of preserving our global environment today means we can no longer afford to increase agriculture production at the expense of environmental stability. This scenario leaves humanity with basically three avenues to reconcile agricultural productivity with environmental sustainability: reduce food waste, shift towards less meat intensive diets in the developed world, and use of the existing resources more sustainably.

Although second-generation GM crops have yet to reach global agriculture, they may contribute significantly to help us use existing resources more sustainably. An HaHB4 derived event for soybean has already been approved for cultivation in major agricultural territories like the United States, Brazil, and Argentina, with regulatory processes well advanced in other important geographies, such as China (developers' public

TABLE 5 | Classification accuracy of three machine learning algorithms for discrimination between transgenic and commercial wheat genotypes.

DAS	HB412/BP11			HB412/KP		
	81	130	136	81	130	136
Growth Stage (GM/Commercial)	1N[a]-12/1N	BS+2/Ant+11	Ant+1/Ant+17	1N-12/1N+2	BS+2/Ant+5	Ant+1/Ant+11
ANN	0.89	0.84	0.81	0.86	0.93	0.99
SVM	0.82	0.73	0.9	0.87	0.91	0.99
RF	0.74	0.83	0.62	0.54	0.72	0.84

[a]Abbreviations as in **Tables 2** and **3**.

Tested wheat genotypes were HB412 (GM) and commercial varieties Baguette Premium 11 (BP11) and Klein Pegaso (KP). The growth stage is indicated as the nearer stage plus or minus calendar days from this stage. Bolded data indicate small variation between growth stages of the compared genotypes.

information). A similar wheat event is also being considered. This technology is expected to be in thousands of hectares during the 2019-2020 crop cycle, with the potential to be over one million hectares in two cycles after that, subject to farmers' acceptance among other factors.

Drought tolerant wheat and soy crops, such as those described in this study, may yield more per unit of water used by plants. This resiliency may favor water-demanding double-cropping schemes that would otherwise be uneconomic. Sustainable intensification allowed by second-generation GM crops will result in improved carbon fixation, while less land is required to sustain current production outputs. These benefits may be of importance to a broader consumer audience, increasingly upset with our collective inability to preserve our terrestrial environment. Demonstrating these benefits at scale may generate an opportunity to re-signify GM perception derived from farmer-centric first-generation GMOs. In doing so, pressure could be mounted to streamline and synchronize global regulatory systems, making the process more affordable to a broader group of scientists and technology developers.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding authors.

REFERENCES

Alloatti, J., Blackford, C., Burachik, M., Chiozza, M., Dezar, C., Dicely, I., et al. (2017). *Petition for determination of non-regulated status for the new plant variety HB4 soybean (IND-00410-5) intended for environmental release and food and feed use.* (Riverdale, Madison, USA)

Aramburu, F., Pablo, J., Mercau, J. L., Taboada, M., Andrade, F. H., Hall, A. J., et al. (2015). Field crops research potential for crop production increase in Argentina through closure of existing yield gaps. *F. Crop Res.* 184, 145–154. doi: 10.1016/j.fcr.2015.10.001

Arce, A. L., Raineri, J., Capella, M., Cabello, J. V., and Chan, R. L. (2011). Uncharacterized conserved motifs outside the HD-Zip domain in HD-Zip subfamily I transcription factors; a potential source of functional diversity. *BMC Plant Biol.* 11, 42. doi: 10.1186/1471-2229-11-42

Ariel, F. D., Manavella, P. A., Dezar, C. A., and Chan, R. L. (2007). The true story of the HD-Zip family. *Trends Plant Sci.* 12, 419–426. doi: 10.1016/j.tplants.2007.08.003

Asseng, S., Martre, P., Maiorano, A., Rötter, R. P., O'Leary, G. J., Fitzgerald, G. J., et al. (2019). Climate change impact and adaptation for wheat protein. *Glob. Change Biol.* 25, 155–173. doi: 10.1111/gcb.14481

Ayala, F., Fedrigo, G. V., Burachik, M., and Miranda, P. V. (2019). Compositional equivalence of event IND-ØØ412-7 to non-transgenic wheat. *Transgenic Res.* 28, 165–176. doi: 10.1007/s11248-019-00111-y

Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., et al. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546, 148–152. doi: 10.1038/nature22380

Blancke, S., Van Breusegem, F., De Jaeger, G., Braeckman, J., and Van Montagu, M. (2015). Fatal attraction: the intuitive appeal of GMO opposition. *Trends Plant Sci.* 20, 414–418. doi: 10.1016/j.tplants.2015.03.011

Board, J. E., and Qiang, T. (1995). Assimilatory capacity effects on soybean yield components and pod number. *Crop Sci.* 35, 846–851. doi: 10.2135/cropsci1995.0011183X003500030035x

AUTHOR CONTRIBUTIONS

NR, MR and MP performed spectral analyses. PM and FT wrote about the regulatory process. FG, KR and MO wrote about ecophysiological aspects whereas RC described HaHB4 story. RC and MO conceived and designed the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00178/full#supplementary-material

Borrás, L., Slafer, G. A., and Otegui, M. E. (2004). Seed dry weight response to source-sink manipulations in wheat, maize and soybean: a quantitative reappraisal. *F. Crop Res.* 86, 131–146. doi: 10.1016/j.fcr.2003.08.002

Boscaiu, M., Donat, P., Llinares, J., and Vicente, O. (2012). Stress-tolerant wild plants: a source of knowledge and biotechnological tools for the genetic improvement of stress tolerance in crop plants. *Not. Bot. Horti. Agrobo.* 40, 323–327. doi: 10.15835/nbha4028199

Campos, H., Cooper, M., Habben, J. E., Edmeades, G. O., and Schussler, J. R. (2004). Improving drought tolerance in maize: a view from industry. *F. Crop Res.* 90, 19–34. doi: 10.1016/j.fcr.2004.07.003

Castiglioni, P., Warner, D., Bensen, R. J., Anstrom, D. C., Harrison, J., Stoecker, M., et al. (2008). Bacterial RNA chaperones confer abiotic stress tolerance in plants and improved grain yield in maize under water-limited conditions. *Plant Physiol.* 147, 446–455. doi: 10.1104/pp.108.118828

Chapman, S. C., Cooper, M., Hammer, G. L., and Butler, D. G. (2000). Genotype by environment interactions affecting grain sorghum. II. Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields. *Aust. J. Agric. Res.* 51, 209. doi: 10.1071/AR99021

Chlingaryan, A., Sukkarieh, S., and Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. *Comput. Electron. Agr.* 151, 61–69. doi: 10.1016/j.compag.2018.05.012

Connor, D., Loomis, R., and Cassman, K. (2011). "Photosynthesis," in *Crop Ecology: Productivity and Management in Agricultural Systems* (Cambridge: Cambridge University Press), 262–291. doi: 10.1017/CBO9780511974199.014

Cooper, M., Podlich, D. W., and Smith, O. S. (2005). Gene-to-phenotype models and complex trait genetics. *Aust. J. Agric. Res.* 56, 895–918. doi: 10.1071/AR05154

Dardanelli, J. L., Ritchie, J. T., Calmon, M., Andriani, J. M., and Collino, D. J. (2004). An empirical model for root water uptake. *F. Crop Res.* 87, 59–71. doi: 10.1016/j.fcr.2003.09.008

Debaeke, P., Casadebaigm, P., Flenetm, F., and Langlade, N. (2017). Sunflower crop and climate change: vulnerability, adaptation, and mitigation potential from case-studies in Europe. *OCL* 24 (1), D102. doi: 10.1051/ocl/2016052

Dezar, C. A., Gago, G. M., Gonzalez, D. H., and Chan, R. L. (2005). Hahb-4, a sunflower homeobox-leucine zipper gene, is a developmental regulator and confers drought tolerance to Arabidopsis thaliana plants. *Transgenic Res.* 14, 429–440. doi: 10.1007/s11248-005-5076-0

Dubock, A. (2014). The politics of Golden Rice. *GM Crops Food* 5, 210–222. doi: 10.4161/21645698.2014.967570

El-Hendawy, S. E., Al-Suhaibani, N. A., Hassan, W. M., Dewir, Y. H., Elsayed, S., Al-Ashkar, I., et al. (2019). Evaluation of wavelengths and spectral reflectance indices for high-throughput assessment of growth, water relations and ion contents of wheat irrigated with saline water. *Agric. Water Manage.* 212, 358–377. doi: 10.1016/j.agwat.2018.09.009

Fehr, W. R., and Caviness, C. E. (1977). *Stages of soybean development. Special Report* (Iowa, USA: Iowa State University), 87.

Fernbach, P. M., Light, N., Scott, S. E., Inbar, Y., and Rozin, P. (2019). Extreme opponents of genetically modified foods know the least but think they know the most. *Nat. Hum. Behav.* 3, 251–256. doi: 10.1038/s41562-018-0520-3

Fischer, R. A. (1975). Yield potential in a dwarf spring wheat and the effect of shading. *Crop Sci.* 5, 607–614. doi: 10.2135/cropsci1975.0011183X001500050002x

Fischer, R. A. (1985). Number of kernels in wheat crops and the influence of solar radiation and temperature. *J. Agric. Sci. Cambridge* 105, 447–461. doi: 10.1017/S0021859600056495

Gago, G. M., Almoguera, C., Jordano, J., Gonzalez, D. H., and Chan, R. L. (2002). Hahb-4, a homeobox-leucine zipper gene potentially involved in abscisic acid-dependent responses to water stress in sunflower. *Plant Cell Environ.* 25, 633–640. doi: 10.1046/j.1365-3040.2002.00853.x

Garcia-Alonso, M., Hendley, P., Bigler, F., Mayeregger, E., Parker, R., Rubinstein, C., et al. (2014). Transportability of confined field trial data for environmental risk assessment of genetically engineered plants: a conceptual framework. *Transgenic Res.* 23, 1025–1041. doi: 10.1007/s11248-014-9785-0

Garriga, M., Romero-Bravo, S., Estrada, F., Escobar, A., Matus, I. A., del Pozo, A., et al. (2017). Assessing wheat traits by spectral reflectance: do we really need to focus on predicted trait-values or directly identify the elite genotypes group? *Front. Plant Sci.* 8, 280. doi: 10.3389/fpls.2017.00280

Gehring, W. J., Affolter, M., and Bürglin, T. R. (1994). Homeodomain proteins. *Annu. Rev. Biochem.* 63, 487–526. doi: 10.1146/annurev.bi.63.070194.002415

GM Approval Database. (2019). http://www.isaaa.org/gmapprovaldatabase/.

González, F., Capella, M., Ribichich, K., Curín, F., Giacomelli, J., Ayala, F., et al. (2019). Wheat transgenic plants expressing the sunflower gene HaHB4 significantly outyielded their controls in field trials. *J. Exp. Bot.* 70, 1669–1681. doi: 10.1093/jxb/erz037

González, F. G., Miralles, D. J., and Slafer, G. A. (2011). Wheat floret survival as related to pre-anthesis spike growth. *J. Exp. Bot.* 62, 4889–4901. doi: 10.1093/jxb/err182

Hall, A. J., and Richards, R. A. (2013). Prognosis for genetic improvement of yield potential and water-limited yield of major grain crops. *F. Crop Res.* 143, 18–33. doi: 10.1016/j.fcr.2012.05.014

Harrington, L., and Tow, P. (2011). "Principles of a systems approach to agriculture: some definitions and concepts," in *Rainfed Farming Systems.* Eds. P. Tow, I. Cooper, I. Partridge and C. Birch, 45–74. doi: 10.1007/978-1-4020-9132-2

Ingwersen, J. (2019). *USDA investigates unapproved GMO wheat found in Washington state.*

IPCC (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Eds. Core Writing Team, R. K. Pachauri, and L. A. Meyer. (Geneva, Switzerland: IPCC), 151 pp.

Jiang, H., and Egli, D. B. (1995). Soybean seed number and crop growth rate during flowering. *Agron. J.* 87, 264–267. doi: 10.2134/agronj1995.00021962008700020020x

Kamle, M., Kumar, P., Patra, J. K., and Bajpai, V. K. (2017). Current perspectives on genetically modified crops and detection methods. *Biotech* 7, 219. doi: 10.1007/s13205-017-0809-3

Kirby, E. J. M. (1988). Analysis of leaf, stem and ear growth in wheat from terminal spikelet stage to anthesis. *Field Crops Res.* 18, 127–140. doi: 10.1016/0378-4290(88)90004-4

Liu, F., Andersen, M. N., Jacobsen, S. E., and Jensen, C. R. (2005). Stomatal control and water use efficiency of soybean (Glycine max L. Merr.) during progressive soil drying. *Environ. Exp. Bot.* 54, 33–40. doi: 10.1016/j.envexpbot.2004.05.002

Manavella, P. A., Arce, A. L., Dezar, C. A., Bitton, F., Renou, J. P., Crespi, M., et al. (2006). Cross-talk between ethylene and drought signalling pathways is mediated by the sunflower Hahb-4 transcription factor. *Plant J.* 48, 125–137. doi: 10.1111/j.1365-313X.2006.02865.x

Manavella, P. A., Dezar, C. A., Bonaventure, G., Baldwin, I. T., and Chan, R. L. (2008a). HAHB4, a sunflower HD-Zip protein, integrates signals from the jasmonic acid and ethylene pathways during wounding and biotic stress responses. *Plant J.* 56, 376–388. doi: 10.1111/j.1365-313X.2008.03604.x

Manavella, P. A., Dezar, C. A., Ariel, F. D., Drincovich, M. F., and Chan, R. L. (2008b). The sunflower HD-Zip transcription factor HAHB4 is up-regulated in darkness, reducing the transcription of photosynthesis-related genes. *J. Exp. Bot.* 59, 3143–3155. doi: 10.1093/jxb/ern170

Meacham-Hensold, K., Montes, C. M., Wu, J., Guan, K., Fu, P., Ainsworth, E. A., et al. (2019). High-throughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity. *Remote Sens. Environ.* 231, 111176. doi: 10.1016/j.rse.2019.04.029

Mishra, K. B., Iannacone, R., Petrozza, A., Mishra, A., Armentano, N., La Vecchia, G., et al. (2012). Engineered drought tolerance in tomato plants is reflected in chlorophyll fluorescence emission. *Plant Sci.* 182, 79–86. doi: 10.1016/j.plantsci.2011.03.022

Moghissi, A. A., Jaeger, L. M., Shafei, D., and Bloom, L. L. (2018). Regulatory science requirements of labeling of genetically modified food. *Crit. Rev. Biotechnol.* 38, 386–393. doi: 10.1080/07388551.2017.1356804

Muchow, R. C., Sinclair, T. R., and Bennett, J. M. (1990). Temperature and solar radiation effects on potential maize yield across locations. *Agron. J.* 82, 338. doi: 10.2134/agronj1990.00021962008200020033x

Olsson, A. S., Engström, P., and Soderman, E. (2004). The homeobox genes Athb12 and Athb7 encode potential regulators of growth in response to water deficit in Arabidopsis. *Plant Mol. Biol.* 55, 663–677. doi: 10.1007/s11103-004-1581-4

Palena, C. M., González, D. H., and Chan, R. L. (1999). A monomer-dimer equilibrium modulates the interaction of the sunflower homeodomain leucine-zipper protein Hahb-4 with DNA. *Biochem. J.* 341, 81–87.

Passioura, J. B. (1991). Soil structure and plant growth. *Aust. J. Soil Res.* 29, 717–728. doi: 10.1071/SR9910717

Passioura, J. (2006). Increasing crop productivity when water is scarce—from breeding to field management. *Agric. Water Manage.* 80, 176–196. doi: 10.1016/j.agwat.2005.07.012

Passioura, J. (2012). Phenotyping for drought tolerance in grain crops. *Funct. Integr. Genomics* 39, 851–859. doi: 10.1071/FP12079

Pedró, A., Savin, R., and Slafer, G. A. (2012). Crop productivity as related to single-plant traits at key phenological stages in durum wheat. *F. Crop Res.* 138, 42–51. doi: 10.1016/j.fcr.2012.09.016

Perotti, M. F., Ribone, P. A., and Chan, R. L. (2017). Plant transcription factors from the Homeodomain-Leucine Zipper family I. Role in development and stress responses. *IUBMB Life* 69, 280–289. doi: 10.1002/iub.1619.R

Ré, D. A., Capella, M., Bonaventure, G., and Chan, R. L. (2014). Arabidopsis AtHB7 and AtHB12 evolved divergently to fine tune processes associated with growth and responses to water stress. *BMC Plant Biol.* 14, 150. doi: 10.1186/1471-2229-14-150

Rattalino Edreira, J. I., Guilpart, N., Sadras, V., Cassman, K. G., van Ittersum, M. K., Schils, R. L. M., et al. (2018). Water productivity of rainfed maize and wheat: a local to global perspective. *Agric. For. Meteorol.* 259, 364–373. doi: 10.1016/j.agrformet.2018.05.019

Reynolds, M. P., Saint Pierre, C., Saad, A. S. I., Vargas, M., and Condon, A. G. (2007). Evaluating potential genetic gains in wheat associated with stress-adaptive trait expression in elite genetic resources under drought and heat stress. *Crop Sci.* 47 (S3), S172–S189. doi: 10.2135/cropsci2007.10.0022IPBS

Ribichich, K. F., Arce, A. L., and Chan, R. L. (2014). "Coping with drought and salinity stresses: role of transcription factors in crop improvement," in *Climate Change and Plant Abiotic Stress Tolerance.* Eds. N. Tuteja and S. S. Gill (Weinheim, Germany: Wiley-VCH Verlag GmbH & CO KGaA), 641–684.

Ribichich, K. F., Chiozza, M., Ávalos-Britez, S., Cabello, J. V., Arce, A. L., Watson, G., et al. (2020). Successful field performance in dry-warm environments of soybean

expressing the sunflower transcription factor HaHB4. *J. Exp. Bot.* doi: 10.1093/jxb/eraa064

Richards, R. A., and Passioura, J. B. (1981). Seminal root morphology and water use of wheat II. *Genet. Var. Crop Sci.* 21, 253–255. doi: 10.2135/cropsci1981.0011183X002100020012x

Riechmann, J. L. (2002). *Transcriptional Regulation: A Genomic Overview. Arabidopsis Book* Vol. 1 (USA: American Society of Plant Biologists), e0085. doi: 10.1199/tab.0085

Rigalli, N. F., Montero Bulacio, E., Romagnoli, M., Terissi, L., and Portapila, M. (2018). *Identification and characterization of crops through the analysis of spectral data with machine learning algorithms*, Argentinian Conference of Information Technology, M CAI-50 ISSN: 2525-0949. http://47jaiio.sadio.org.ar/sites/default/files/CAI-50.pdf.

Sadras, V. O., and Milroy, S. P. (1996). Soil-water thresholds for the responses of leaf expansion and gas exchange: a review. *F. Crop Res.* 47, 253–266. doi: 10.1016/0378-4290(96)00014-7

Sadras, V. O., and Richards, R. A. (2014). Improvement of crop yield in dry environments: Benchmarks, levels of organisation and the role of nitrogen. *J. Exp. Bot.* 65, 1981–1995. doi: 10.1093/jxb/eru061

Shiu, S.-H., Shih, M.-C., and Li, W.-H. (2005). Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol.* 139, 18–26. doi: 10.1104/pp.105.065110

Skirycz, A., Vandenbroucke, K., Clauw, P., Maleux, K., De Meyer, B., Dhondt, S., et al. (2011). Survival and growth of Arabidopsis plants given limited water are not equal. *Nat. Biotechnol.* 29, 212–214. doi: 10.1038/nbt.1800

Stein, A. J., and Rodríguez-Cerezo, E. (2010). International trade and the global pipeline of new GM crops. *Nat. Biotechnol.* 28, 23–25. doi: 10.1093/jxb/err269

Tardieu, F. (2012). Any trait or trait-related allele can confer drought tolerance: just design the right drought scenario. *J. Exp. Bot.* 63 (1), 25–31. doi: 10.1093/jxb/err269

Valdés, A., Övernäs, E., Johansson, H., Rada-Iglesias, A., and Engström, P. (2012). The homeodomain-leucine zipper (HD-Zip) class I transcription factors ATHB7 and ATHB12 modulate abscisic acid signalling by regulating protein phosphatase 2C and abscisic acid receptor gene activities. *Plant Mol. Biol.* 80, 405–418. doi: 10.1007/s11103-012-9956-4

Viola, I. L., and González, D. H. (2016). "Structure and Evolution of Plant Homeobox Genes," in *Plant Transcription Factors*. Ed. D. H. González (Amsterdam, Netherlands: Elsevier). Chapter 6.

Wulff, B. B. H., and Dhugga, K. S. (2018). Wheat-the cereal abandoned by GM. *Science* 361, 451–452. doi: 10.1126/science.aat5119

Advancing Agricultural Production With Machine Learning Analytics: Yield Determinants for California's Almond Orchards

Yufang Jin[1]*, Bin Chen[1], Bruce D. Lampinen[2] and Patrick H. Brown[2]

[1] Department of Land, Air and Water Resources, University of California, Davis, Davis, CA, United States, [2] Department of Plant Sciences, University of California, Davis, Davis, CA, United States

Agricultural productivity is subject to various stressors, including abiotic and biotic threats, many of which are exacerbated by a changing climate, thereby affecting long-term sustainability. The productivity of tree crops such as almond orchards, is particularly complex. To understand and mitigate these threats requires a collection of multi-layer large data sets, and advanced analytics is also critical to integrate these highly heterogeneous datasets to generate insights about the key constraints on the yields at tree and field scales. Here we used a machine learning approach to investigate the determinants of almond yield variation in California's almond orchards, based on a unique 10-year dataset of field measurements of light interception and almond yield along with meteorological data. We found that overall the maximum almond yield was highly dependent on light interception, e.g., with each one percent increase in light interception resulting in an increase of 57.9 lbs/acre in the potential yield. Light interception was highest for mature sites with higher long term mean spring incoming solar radiation (SRAD), and lowest for younger orchards when March maximum temperature was lower than 19°C. However, at any given level of light interception, actual yield often falls significantly below full yield potential, driven mostly by tree age, temperature profiles in June and winter, summer mean daily maximum vapor pressure deficit (VPD_{max}), and SRAD. Utilizing a full random forest model, 82% (±1%) of yield variation could be explained when using a sixfold cross validation, with a RMSE of 480 ± 9 lbs/acre. When excluding light interception from the predictors, overall orchard characteristics (such as age, location, and tree density) and inclusive meteorological variables could still explain 78% of yield variation. The model analysis also showed that warmer winter conditions often limited mature orchards from reaching maximum yield potential and summer VPD_{max} beyond 40 hPa significantly limited the yield. Our findings through the machine learning approach improved our understanding of the complex interaction between climate, canopy light interception, and almond nut production, and demonstrated a relatively robust predictability of almond yield. This will ultimately benefit data-driven climate adaptation and orchard nutrient management approaches.

Keywords: *Prunus dulcis*, yield gap, artificial intelligence, big data, light interceptioon, nutrient management

INTRODUCTION

Global food and fiber demand has been projected to double by the mid-century, driven mostly by increasing population and nutrition needs (Tilman et al., 2011; Davis et al., 2013). However, agricultural production has been shown vulnerable to multiple stresses including warming, droughts and floods, extreme weather variability (Rosenzweig et al., 2001; Reynolds and Tuberosa, 2008; Funk and Brown, 2009; Lesk et al., 2016), and degrading soils and water (Elliott et al., 2014). Growers face the grand challenges of increasing food production while minimizing environmental disruption, and improving the resilience of agriculture systems under changing climates (National Academies of Sciences Engineering Medicine, 2019). Optimizing food system requires a new approach that integrates existing datasets for new insights about yield determinants, and resolves the complex and interconnected physical and biological processes affecting yield across different scales. Recent technological advances in artificial intelligence provide promising tools to understand the constraints on potential yield and interpret and predict the variation of yield across space and time by harnessing many unique yet under-utilized datasets.

California's almond acreage has expanded rapidly in recent decades, from 283,280 hectares in 2005 to 538,232 hectares in 2017 (USDA-NASS, 2018), due to the increasing demand for almonds in domestic and international markets. Almond has become the second leading agricultural commodity in California, with a total farm gate value of 5.6 billion US dollars in 2017 (California Department of Food and Agriculture, 2017). California produces about 80% of the world's almonds and 100% of the U.S. commercial almond production. More than 95% of almond acreage is irrigated and growers rely heavily on surface irrigation deliveries and on groundwater when surface water is limited, as occured during the recent prolonged 2013–2017 drought in California (Faunt et al., 2016). Climate change, including warming and extreme weather, is another threat to almond production. The projected climatic conditions by the middle to end of the 21st century are predicted to threaten the long-term viability of the state's almond production (Luedeling et al., 2009). To optimize yield and ensure the almond industry remains economically viable and environmentally sustainable (Carletto et al., 2015; Tombesi et al., 2017), it is essential to understand key yield determinants and develop appropriate agricultural adaptation and management strategies.

Groundwater quality in California has also been degraded due to nitrogen leaching from agricultural fields (Burow et al., 2013; Baram et al., 2016). Facing with this serious challenge, the state of California has implemented legislatively mandated nitrogen (N) management strategies for all almond growers statewide to meet the goal of minimizing nitrogen losses to the environment. To optimize N management and ensure regulatory compliance, almond growers must now apply N in accordance with the estimated yield determined in each orchard in early spring, taking into account N available from all sources (e.g., fertilizer, composts and manures, and irrigation water nitrogen). Accurate yield prediction is thus critically important to help individual growers with the information required to manage inputs and

resources, to schedule on-farm activities and manage harvest and marketing agreements.

Almond yield varies by year and by location; however, the environmental and biophysical factors that underlie these differences are not well understood and have never been systematically characterized. Almond production is known to be highly dependent on a number of factors (Tombesi et al., 2010, 2017; Zarate-Valdez et al., 2015) including (a) biophysical attributes such as tree age, leaf area, tree vigor, and bloom intensity, (b) environmental conditions such as chilling and heat requirements, soil nutrition, and bee foraging activity, and (c) cropping history. To date, a detailed comprehensive assessment of each of these factors and a yield prediction algorithm has not been successfully achieved, especially at a finer spatial scale.

Among the variables that have been shown to impact yield in almond, canopy interception of photosynthetically active radiation (PAR), is directly related to maximum potential yield of almonds (Zarate-Valdez et al., 2015). Lampinen et al. (2012), reported that the maximum sustainable yield in the most productive commercial almond orchards is 56 kernel kg/ha per unit PAR intercepted by the canopy. Percent light interception at the orchard level is determined by canopy structure, e.g., total leaf area and health at the individual tree level, as well as row and tree spacing; while the location of the orchard (latitude) and cloud fraction affect the total amount of PAR incident on the canopy. Management activities such as cultivar selection, tree spacing, pruning practices, nutrition, and irrigation also have direct impacts on canopy interception and thus yield. As almond is a perennial crop, the multi-year photosynthetic accumulation and allocation to reproductive and vegetative organs from previous years also affect its yield (carry over effect), as well as spurs frequency.

Climate, such as temperature and water availability, is known to have an important role in crop growth and flowering, and thus influencing yield variation (Kerr et al., 2018; Pathak et al., 2018). A few prior studies have used relatively simple statistical analysis to understand how temperature and precipitation affected almond yield in California, but were largely limited by the spatial scale, e.g., from county to state levels (Lobell et al., 2007; Lobell and Field, 2011), and temporal coverage, resulting relatively small sample size for analysis (e.g., from tens to hundreds). At the scale of an individual plant, growth models developed by DeJong (2019) as well as knowledge of the role of flower number on yield potential (Tombesi et al., 2017) and modeled carbon budgets all contribute knowledge that can be integrated into a yield prediction model. However, these mechanistic approaches have not been systematically applied at any significant scale.

Moreover, nut production of almond trees is also highly dependent on bee pollination. Most almond cultivars are self-sterile, and two or more cultivars are usually inter-planted (Connell, 2000). Bee foraging activity is thus a crucial determinant of the final yield. In addition to being dependent on environmental variables such as temperature, solar radiation, and wind, bee activity is highly reliant on the timing and intensity of flowering, which in turn is also highly affected by weather conditions. Understanding these complicated impacts of environmental factors on almond nut production is therefore

rather challenging, especially at the individual field level, requiring a large spatial and temporal data set and more advanced analytical algorithms.

To address these issues and develop a yield prediction model and descriptor of key yield determinants on almond, we have obtained a 10-year collection of plant and field level biological measurements, management practices, and yield records from 33 locations across the main growing regions of California. Using an advanced machine learning algorithm, we integrated these data with two meteorological datasets to investigate the environmental, biological, and management factors that determine yield variability of almond. Specifically, we aim to answer the following scientific questions: (i) what are the limiting factors that affect yield at a given level of light interception? (ii) Is it possible to predict light interception with orchard age and environmental variables? and (iii) What are the overall impacts of environmental variables on actual yield when controlling for both light interception and the yield gap at a given light interception? An improved understanding of these questions is expected to guide and optimize the life-cycle management of almond production. There is considerable commercial interest in the ability to predict yield and identify production constraints effectively and, as a consequence, the models and information developed in this paper will also be useful to optimize management and hence sustainability.

MATERIALS AND METHODS

Study Area

Our study area focused on California's Central Valley, one of the most productive agricultural areas in the world. We have a 10-year collection of field measurements and yield records over a total of 33 individual almond orchards containing 7865 individual experimental plots (**Figure 1**). This region experiences a Mediterranean climate characterized by hot and dry summers and mild and wet winters. Typically, the rainless summer provides ample sunshine for almond growth and limits disease pressures. The cool and wet winter replenishes the soils and reservoirs in bordering mountainous areas, this and groundwater resources provide water for irrigation during the dry season.

Field Measurements

We collected canopy light interception and yield data over 33 almond orchards, that included a total of 7864 experimental plots, spanning the almond producing areas of the Sacramento and San Joaquin valleys of California, from 2009 to 2018 (Lampinen et al., 2012). The consistent practice of sample collection, supported by Almond Board of California, was designed to evaluate and understand almond production characteristics and drivers from a single tree to orchard scale, for the purpose of improving almond orchard management. For each plot, trees were randomly sampled over a full row length ranging from 50 to 150 individuals for canopy light interception measurement during May to August growing season. A mobile platform (MLB hereafter), consisting of a series of 18 ceptometer segments mounted on a Kawasaki mule utility vehicle, was used to measure PAR below the canopy

of both sides of almond trees (PAR_{below}). Simultaneously, a fixed light sensor recorded the full sun incoming PAR above the canopy (PAR_{above}). All PAR measurements were conducted at solar noon (± 1 h), and the light interception was calculated as the fractional PAR intercepted by the canopy:

$$LR = fPAR = 1 - \frac{PAR_{below}}{PAR_{above}} \quad (1)$$

For each individual experimental plot, average fPAR values of individual trees were calculated to represent the plot-level light interception.

Almond trees were harvested by shaking with a mechanical shaker and the nuts were collected after letting them dry on the ground for about 1 week. Fresh fruit weight was recorded for each individual experimental tree, and a 2 kg sample was used for dry fruit weight (hull plus shell plus kernel) and dry kernel yield (i.e., the yield value used in this study). For each experimental plot, we also recorded its specific orchard site, geographic location (latitude and longitude), planting year, cultivar composition, row and tree spacing (**Table 1**).

Climate and Weather Data

We used monthly climate record from the Parameter-Elevation Regressions on Independent Slope Model (PRISM) dataset (Daly et al., 2008), including monthly mean values of daily precipitation, daily maximum/minimum/mean temperature, and daily maximum VPD (VPD_{max}) (**Table 1**). PRISM uses weather station observations, a digital elevation model (DEM), and other spatial datasets to extrapolate the observations from weather stations to $\sim$ 4-km gridded estimates of monthly climatic variables over the United States (Daly et al., 2008, 2015).

We used the daily weather data at 1km scale from the Daymet Version 3 product, to quantify incoming shortwave radiation flux density (SRAD) at the surface and the duration of the daylight period (Dayl) (Thornton et al., 2017). We further derived the total number of extreme hot days for each month (HotDays). For each month, the threshold of daily T_{max} was set as the upper 10-percentile daily maximum temperature from 2009 to 2018, respectively, based on the daily DayMet T_{max} product. If the daily T_{max} for a certain day exceeded the extreme threshold value of the corresponding month, it was identified as a relatively hot day. All the monthly variables (except for Hotdays) from 2009 to 2018 were further aggregated to derive 10-year mean climatology at both seasonal (i.e., spring, summer, fall, and winter) and annual scales. Climate from both current year and preceding years were also explored for our analysis.

Yield Potential

Higher light interceptions usually lead to higher yields, but the yield also varies significantly with other environmental stressors (Lobell et al., 2007; Tombesi et al., 2010; Zhang et al., 2019). To understand the maximum yield potential that almond could reach at a given light interception, we grouped all plot-year samples by the associated light interception with an interval of 5%, and selected the upper 10-percentile samples within each light interception bin, as a proxy for the yield potential. The

FIGURE 1 | Study sites of almond orchards in California's Central Valley. **(A)** Orchard locations overlaid on the mean summer daily maximum Vapor Pressure Deficit (VPD$_{max}$) and mean winter Temperature (T$_{mean}$) averaged during 2009–2018. The inset shows the extent of Central Valley in the state of California overlaid on the 30-year mean monthly temperature during 1980–2010. **(B)** Annual time series of mean summer daily VPD$_{max}$ and winter T$_{mean}$ averaged over northern, middle, and southern study sites from 2009 to 2018. Also shown are the distributions of **(C)** cultivars and **(D)** orchard age for all plots at the sampling years.

light interception and its corresponding yield were then averaged over the subsamples for each group to model the upper bound of the yield at a given light interception percentage. A linear regression model was built with the interception set to zero. We conducted this analysis for all plots ($n = 7864$), and for a subset of plots ($n = 5581$) containing the most dominant cultivar, Non-pareil, respectively.

Environmental Stressors for Yield Gap

To further understand the factors that constrained the almond trees from reaching the maximum yield under a given level of light interception, we normalized the original yield by the modeled yield potentials, as follows:

$$y_n = \frac{y_o}{y_p} \qquad (2)$$

where y_o is the original yield, y_p is the modeled yield potential, and y_n is the final normalized yield, typically ranging from 0 to 1 (with very few samples beyond 1). Samples with y_n less than 1 indicated productivity under the yield potential. The deviation of y_n from 1 can therefore be used as a proxy for the yield gap.

We used the random forest machine learning approach to model and analyze the complex relationship between the normalized yield and a suite of meteorological variables, in order to understand what and how environmental stressors limit the yield at given light interception. Random forest is an ensemble learning technique to improve classification and regression trees method by combining a large set of decision trees (Liaw and Wiener, 2002; Belgiu and Drăguţ, 2016; Jeong et al., 2016). In random forest regression, each tree is built using a deterministic algorithm by selecting a random set of variables and a random sample from the training dataset. Specifically, the

TABLE 1 | Summary of input variables in this study.

Data source	Input variables	Variable name	Pearson's r (***p < 0.001)	
Biological variables of almond orchards				
Field measurement	Latitude	Lat	−0.17***	
	Longitude	Lon	0.16***	
	Cultivar	Cul	−0.15***	
	Tree age	Age	0.46***	
	Row spacing	Row	0.10***	
	Tree spacing	Tree	0.10***	
	Light interception	LI	0.60***	
Meteorological variables (averages over 12 individual months and 4 seasons from daily values)				
PRISM (4 km)	**(http://www.prism.oregonstate.edu/)**		**Current year**	**Previous year**
	Precipitation	PPT	−0.26 ∼0.20	−0.19 ∼0.05
	Maximum temperature	T_{max}[1]	−0.15 ∼0.09	−0.05 ∼0.19
	Minimum temperature	T_{min}[1]	−0.24 ∼0.19	0.07 ∼ 0.20
	Mean temperature	T_{mean}[1]	−0.19 ∼ 0.16	0.01 ∼ 0.20
	Maximum vapor pressure deficit	VPD_{max}[1]	−0.30 ∼0.06	−0.17 ∼0.08
Daymet (1 km)	**(https://daymet.ornl.gov/)**			
	Daylight duration	Dayl	0.08 ∼ 0.20	−0.21 ∼0.17
	Shortwave radiation flux density	SRAD	0.05 ∼ 0.24	−0.04 ∼0.30
	Extremely hot days	HotDays	−0.06 ∼−0.25	−0.09 ∼−0.19

The Pearson's correlation and its significance between each individual variable and the production were shown here. For time varying meteorological variables, the range of the statistics for each variable among monthly and seasonal parameters were included. [1] Monthly and seasonal variables were mean values of daily maximum, minimum, and mean averaged over each month or season.

"RandomForest" package within R environment software was used in this study[1].

Conceptually monthly and seasonal meteorological variables, during both the current year and the preceding year, may pose stresses at the different stages of plant growth, including flowering, leaf out, and fruit setting (Tombesi et al., 2010, 2017). Although a large set of explanatory variables is not an obstacle for the functioning of random forest model, the highly correlated meteorological variables may hinder the interpretation of the modeling results (Liaw and Wiener, 2002). We first used Pearson' correlation coefficient (r) to investigate how each individual independent variable was correlated with the yield gap, and how each individual weather variable correlated with each other among different time periods, thereby providing the basis for selecting a subset of more significant meteorological variables for building the model. In this study, we selected representative variables that are highly correlated with yield gap (i.e., r > 0.15) and less cross-correlated with other variables within the same category (i.e., r < 0.50) (**Supplementary Figures S1, S2** and **Supplementary Table S1**).

With random forest modeling, we ranked the variable importance based on how much the modeling accuracy decreased, or the increase in mean-square-error (i.e., IncMSE) of predictions, when a particular variable was excluded from the whole suite of input variables for model building (Grömping,

2009). The IncMSE of predictions, estimated with an out-of-bag cross validation, in percentage relative to the full model, is a robust and informative metric, e.g., higher values indicating that the corresponding variable is more important for yield prediction.

We further used partial dependence plots to understand how each of these variables affected the yield (Welling et al., 2016). Intuitively, partial dependence plots show the dependence between the target response and a set of explanatory features, marginalizing over the values of all other features. We can interpret the partial dependence as the expected target response as a function of the explanatory features.

To further examine what conditions or combinations of conditions are associated with relatively higher or much lower normalized yield, we used the regression tree model (i.e., "rpart" package within R environment[2]) to identify decision rules between explanatory variables and the target response that can best differentiate yield gaps, i.e., representative splitting nodes. We chose the decision tree with a highest predictive accuracy as the most representative tree in this study.

Determinants of Light Interception

As a dominant influential variable, light interception (or percentage of absorbed PAR) reflected the combined effects of canopy density, structure, and health status, which were again

---

[1]https://cran.r-project.org/web/packages/randomForest/

[2]https://cran.r-project.org/web/packages/rpart

FIGURE 2 | Almond yield vs. light interception percentages (LI) for **(A)** all samples (n = 7864) and **(C)** Non-pareil samples (n = 5581) in the experimental plots. Color represents the density of the samples. The mean values (red circles) of the upper 10-percentile samples (black crosses) within each 5% of LI interval were used for modeling the potential yield, with regression lines shown in red dashed lines in **(A,C)**. Also shown are the corresponding normalized yields (actual yield divided by the corresponding modeled yield potential) for **(B)** all cultivars and for **(D)** Non-pareil.

associated with tree age, row and tree spacing at a plot level, and meteorological conditions that affected tree physiology and development (Zarate-Valdez et al., 2015). To understand the dominant factors that affected light interception, we also analyzed the relationship between the light interception percentage and a suite of layers (including orchard characteristics, and current and preceding meteorological variables), using random forest model. Non-pareil was used as an example for this analysis (n = 5581), to exclude the potential confounding factors from different cultivars.

Drivers for Overall Almond Yield

Besides affecting the light interception via tree growth and health, environmental variables may also affect flower phenology, bee activities, pollination, fruit set, and production. To further examine the complex relationships between yield and biological and environmental controls, we built overall random forest models to predict almond yield at the plot level, driven by four sets of independent variables, respectively. Specifically, these included (A) biological variables including measured light interception percentage and cultivar composition
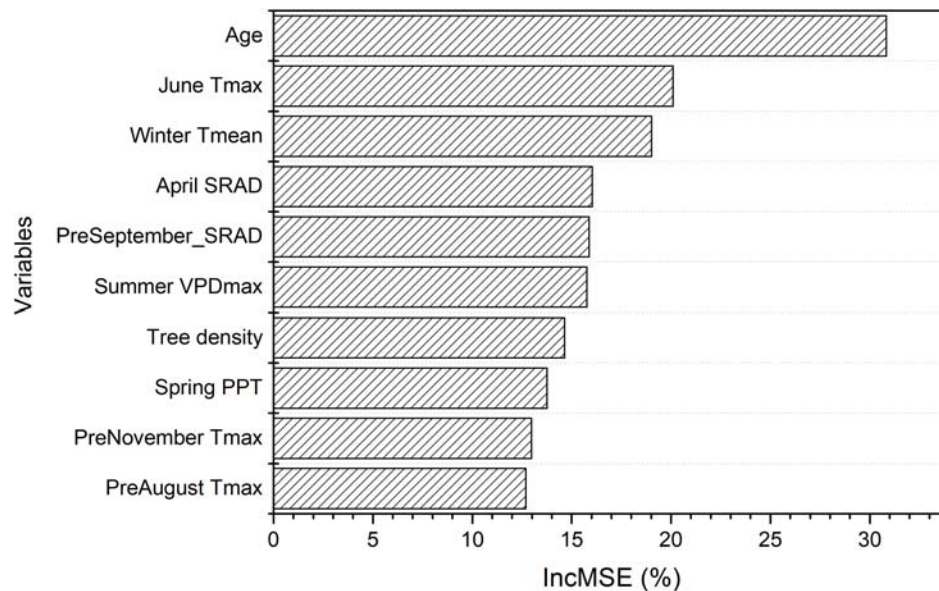
FIGURE 3 | Variable importance from the random forest model of yield gap, as measured by the increase in mean-square-error (IncMSE) of predictions when excluding each variable.

(**Table 1**), (B) biological variables and full meteorological variables (**Supplementary Figure S4**), (C) biological variables and selected meteorological variables, and (D) biological variables but excluding light interception and full meteorological variables. Model performance was evaluated and compared with a sixfold cross validation. The root mean square error (RMSE) and R-square (R^2) were used to quantify the models' accuracy. We also calculated a ratio of performance to interquartile distance (RPIQ), which accounts for both the prediction error and variation of observed values, and therefore it is more objective than the RMSE and easier to compare among models (Bellon-Maurel et al., 2010). A greater RPIQ represents a stronger predictive capacity of the model (Bellon-Maurel et al., 2010).

RESULTS

Controls on Almond Yield Potential

Overall almond yield highly depended on light interception (**Figure 2A**), as shown by the Pearson's correlation coefficient of 0.60 ($p < 0.001$) between the recorded yield and measured light interception percentage across all sample plots. Yield increased from 467.4 ± 432.6 lbs/acre to above 2907.6 ± 1084.2 lbs/acre, when LI increased from below 30% to above 70%. Across each 5% interval of light interception, we found a very strong linear relationship between the maximum yield, as represented by the upper 10-percentile samples, and the light interception (**Figures 2A,C**). The yield potential predicted by the linear regression model agreed well with the observation, with a R^2 of 0.95, when all cultivars were considered. In particular, we found that one percent of increase in light interception led to an increase of 57.9 lbs/acre in the potential yield, as shown by

the slope of the regression model (**Figure 2A**). Similar results were found when the analysis was restricted to the cultivar Non-pareil ($R^2 = 0.94$, slope = 57.7 lbs/acre per LI unit), further supporting that the yield potential was dominated by the light interception (**Figure 2C**).

Determinants on Almond Yield Gap

Actual almond nut production was found to vary significantly at a given level of light interception (**Figure 2A**), even for the same cultivar (**Figure 2C**). For example, Non-pareil trees had yields ranging from 2278 lbs/acre (lower quantile) to 3267 lbs/acre (upper quantile), and averaged 2790 ± 781 lbs/acre, when LI was between 70% to 75% (**Figure 2C**). Across all plots the majority of almond samples didn't reach yield potential (i.e., red dashed line) for any given light interception percentage (**Figures 2B,D**).

The random forest analysis, as described in section "Yield Potential," showed that the variation of yield gap, 1- actual yield normalized by the potential yield at the corresponding light interception, was mostly driven by tree age, mean June daily T_{max}, winter T_{mean}, SRAD, and mean summer daily VPD_{max}, among orchard characteristics and climate variables (**Figure 3**). Mature orchards (>5 years old) tended to have lower yield gap than younger orchards for the same amount of light interception and climate (**Figure 4A**). The partial dependence plots also showed that almond yield dropped significantly below the yield potential when the average winter temperature was higher than 10°C and April SRAD was lower than 450 W m^{-2} (**Figures 4C,D**). Daily T_{max} averaged in June, daily SRAD averaged over previous September, and daily VPD_{max} averaged in summer had a more gradual impact rather than a significant thresholding effect (**Figures 4B,E,F**).
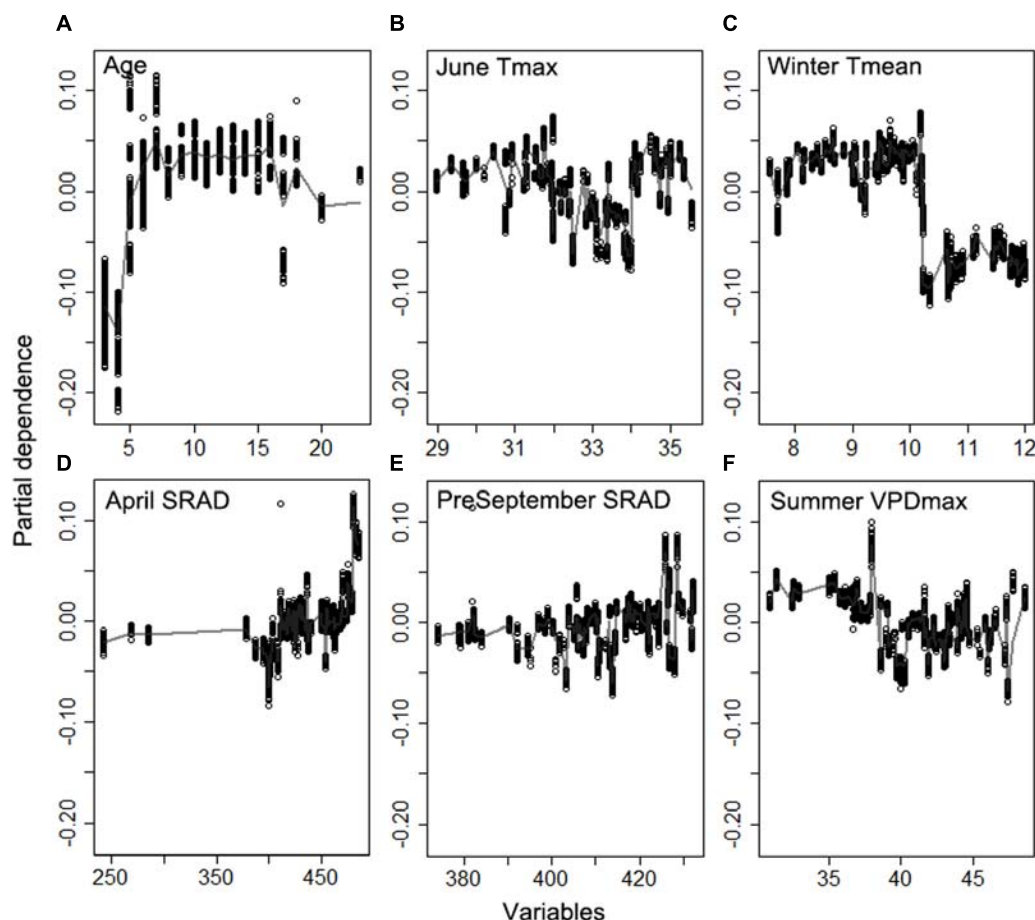
FIGURE 4 | Partial dependence of the normalized yield (actual yield divided by modeled yield potential) on the top six important variables **(A)** Age, **(B)** daily maximum temperature averaged over June, **(C)** daily mean temperature averaged over winter, **(D)** daily shortwave radiation flux density (SRAD) averaged over April, **(E)** daily SRAD averaged over previous September, **(F)** daily maximum vapor pressure deficit (VPD) averaged over summer for Non-pareil, ordered by variable importance.

A representative decision tree further supported that samples close to potential yield (i.e., yield gap > 0.90) were associated with mature orchards (i.e., age > 5) and when winter T_{mean} < 10.22°C, and April SRAD > 478.8 W m^{-2} (n = 223) (**Figure 5**). The largest yield gap nodes, e.g., with a normalized yield of 0.27 (n = 596), were found among mature orchards, and when winter T_{mean} was greater than 10.22°C and mean June daily T_{max} was lower than 34°C; another grouping of plots with large yield gaps (0.38, n = 520), were associated with young orchards, winter T_{mean} lower than 10.22°C, and June T_{max} lower than 32.19°C.

Determinants for Light Interception

The random forest model explained 82% of variation in light interception (**Supplementary Figure S3**), for Non-pareil (n = 5581), when using field based orchard characteristics and full set of meteorological variables as input. Age was the most important variable in determining light interception as expected, according to variable importance (**Figure 6**), and as shown by the high correlation (r = 0.63, p < 0.001) across all samples. The partial dependence plots further showed that light interception

increased significantly with tree age until 7 years old and then plateaued (**Figure 7**). Mean Fall daily T_{min} in previous year, long-term mean annual SRAD, February T_{min}, summer T_{max}, and long-term summer Dayl also affected current year light interception. Fall T_{min} lower than 10.5°C (**Figure 7B**) and long-term annual mean SRAD lower than 380 W m^{-2} (**Figure 7B** and **Supplementary Figure S5A**) reduced light interception. We also found that other long-term mean climatic variables such as summer Dayl, mean winter daily T_{max}, mean summer daily T_{min} had an important role, probably because they affected the general tree growth.

A representative decision tree further revealed that light interception in trees < 7 years old was influenced by a different set of determinant variables than trees older than 7 years. In trees younger than 7 years the lowest light interception nodes were associated with mean March daily T_{max} < 19.1°C (**Figure 8**). In orchards > 7 years old, long-term annual mean SRAD > 378.8 W m^{-2} (**Figure 8**), and the majority of them were distributed in middle to southern Central Valley (**Supplementary Figure S5B**). For young orchards, the highest LI (59%) were those samples distributed from norther to middle,

FIGURE 5 | A representative decision tree on the drivers of actual yield gaps (actual yield divided by modeled yield potential) for Non-pareil.

FIGURE 6 | Variable importance for driving light interception derived from the random forest model using field variables and full meteorological variables.
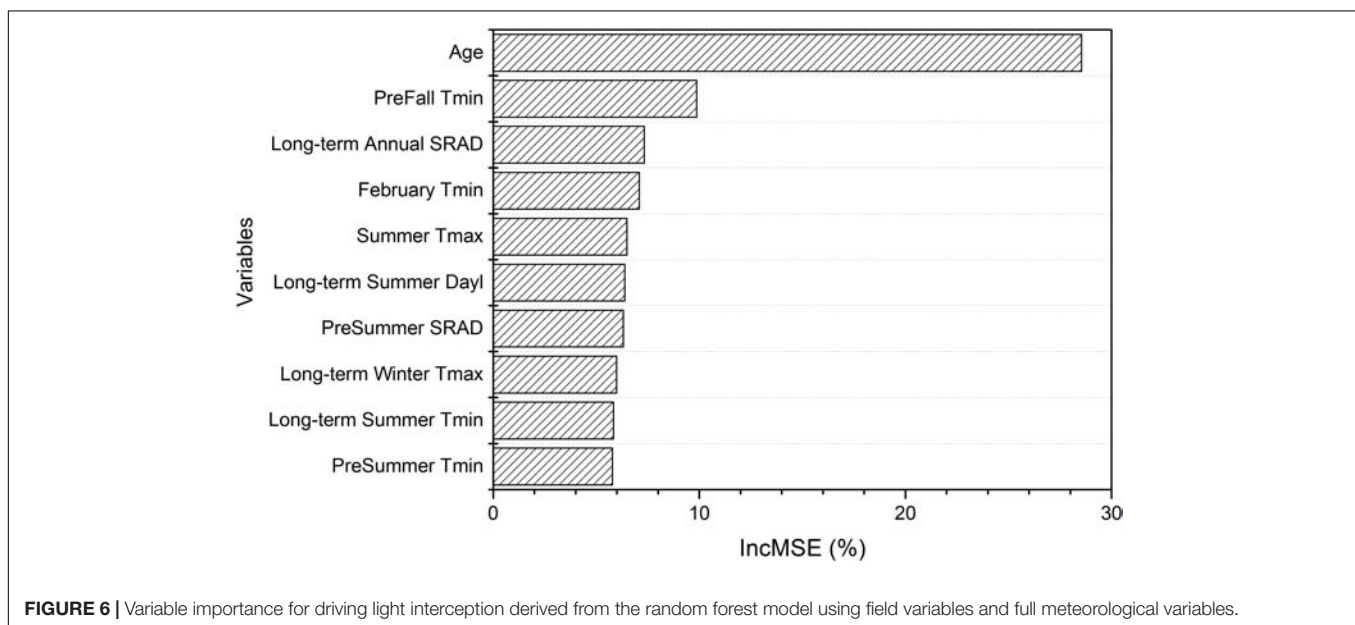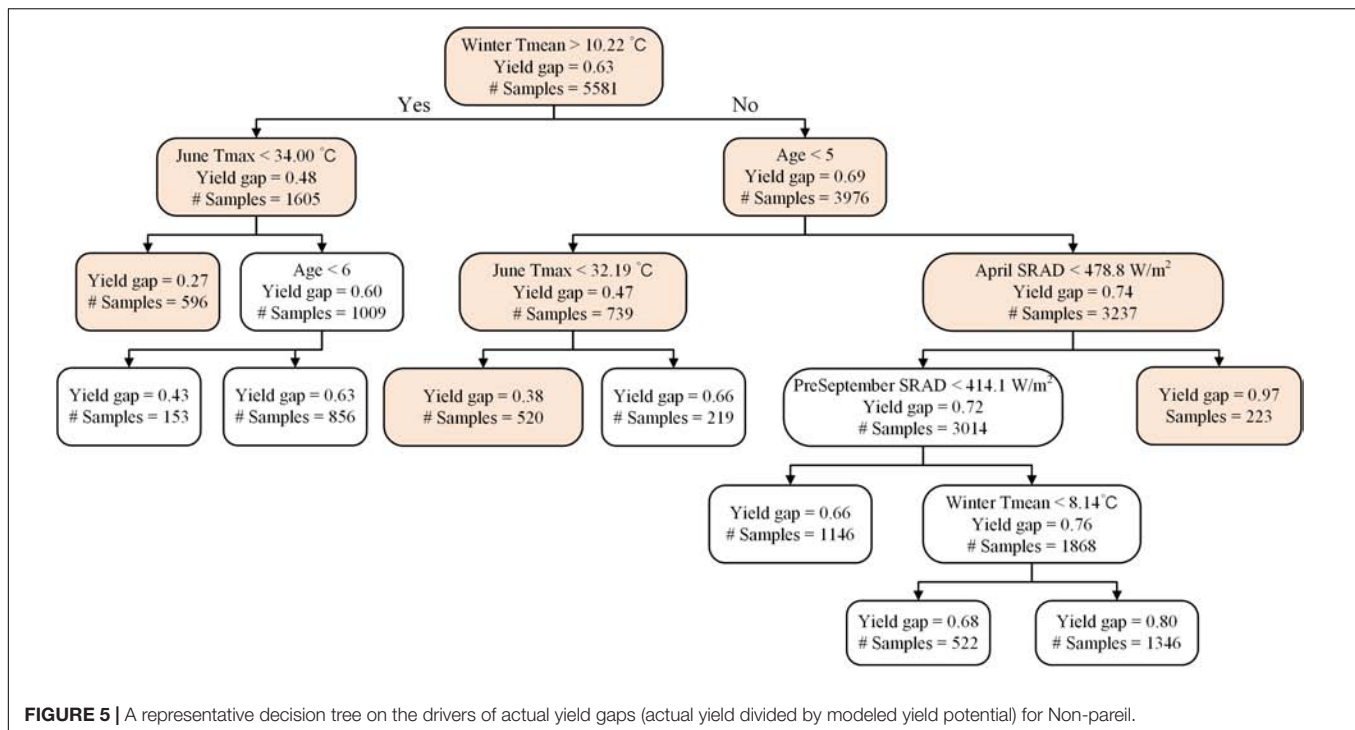
and southern Central Valley over various years (**Supplementary Figure S6A**). For mature orchards, the node with the lowest LI (53%) were 2013 samples clustered in the middle Central Valley (**Supplementary Figure S6B**).

Overall Yield Prediction and Determinants

The prediction results showed that all models were able to explain more than 78% of yield variation (**Figure 9**),

much higher than the linear yield prediction based only on field measured light interception (R^2 = 0.36), and the RF-based prediction using field measured light interception and orchard age (R^2 = 0.60). For example, when adding other orchard characteristics such as age and location (i.e., latitude and longitude), model (A) had a R^2 of 0.79 ± 0.01, RMSE of 530.64 ± 11.77 lbs/acre, and RPIQ of 3.12 ± 0.09, based on the random forest modeling with a sixfold cross validation. By further adding the whole suite of meteorological variables, the full model achieved
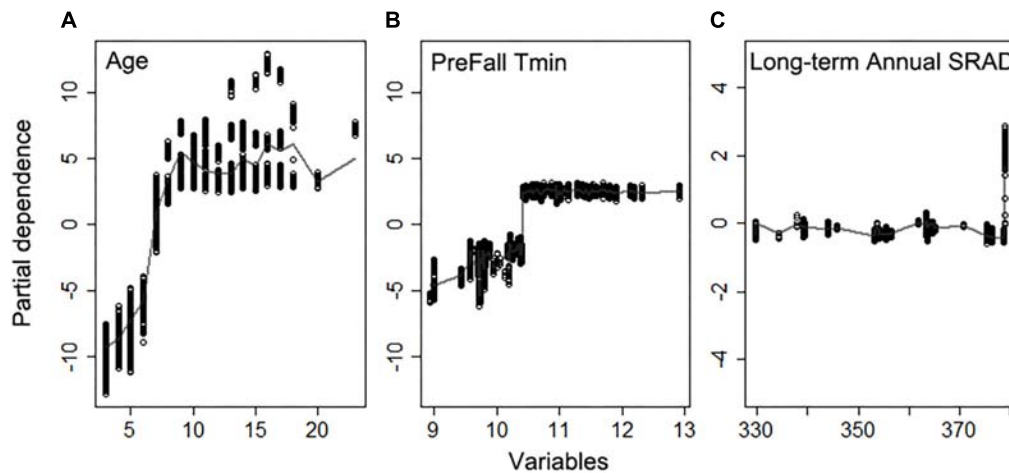
FIGURE 7 | Partial dependence of the light interception on top three important variables **(A)** Age, **(B)** daily minimum temperature averaged previous Fall, **(C)** long-term annual mean shortwave radiation flux density (SRAD) for Non-pareil, ordered by variable importance.
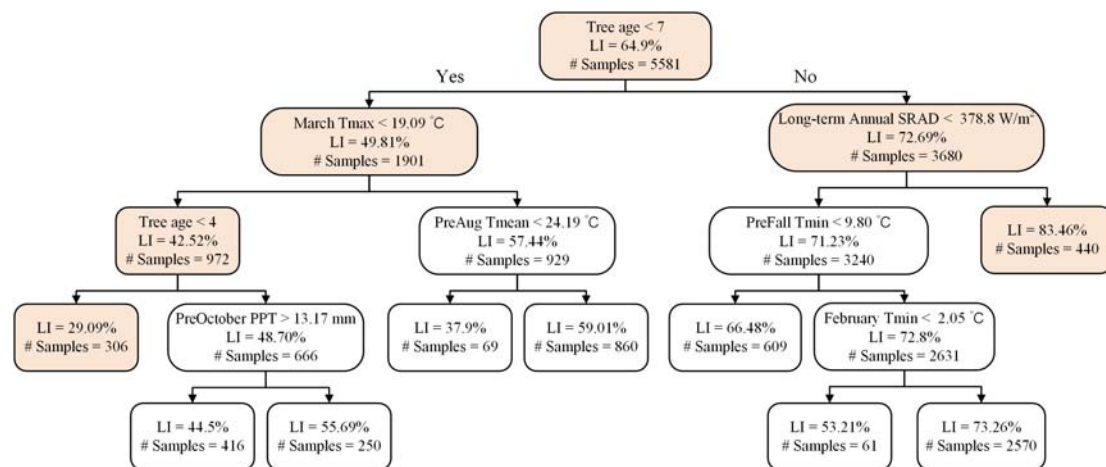
FIGURE 8 | A representative decision tree on the determinants of light interception for Non-pareil.

the more robust and higher accuracy, as shown by higher R^2 (0.82 ± 0.01), lower RMSE (480 ± 9 lbs/acre) and RPIQ (3.45 ± 0.17). After removing highly correlated meteorological variables, the reduced model with selected meteorological variables (**Supplementary Figure S7** and **Supplementary Table S2**) had a similar accuracy with that of full model (**Table 2**).

When excluding light interception, the overall orchard characteristics (like location and age, tree density) and environmental variables (Model D) could explain 78% of yield variation across samples, similar to the model (A) which uses all orchard characteristics plus tree level light interception.

Based on the model with field biological and selected meteorological variables, we found that cultivar, light interception, and age were most important in determining overall almond yield (**Figure 10**). The key meteorological variables that ranked relatively important were mean summer

daily VPD_{max}, mean winter daily T_{min}, April SRAD, and summer T_{mean} (**Figure 10**).

The partial dependence analysis further showed yield difference across different almond cultivars (**Figure 11A**). Among the most popular almond cultivars, Aldrich (Cultivar ID: 2), Monterey (Cultivar ID: 17), and Non-pareil (Cultivar ID: 18) had higher yields than Butte (Cultivar ID: 5) and Carmel (Cultivar ID: 7), with everything else being equal. Yield increased linearly with light interception, but dropped rapidly when the light interception was higher than approximately 82% (**Figure 11B**). Tree age was identified to play an important role mostly during the young stage (i.e., 1–6 years) of almond growth (**Figure 11C**); the impact from tree ages was quite stable after reaching the maturity, but yield decreased for plots over 20 years of age. The contribution from April SRAD to the yield kept stable from 250 to 470 Wm^{-2}, but rapidly increased after that threshold. In contrast, Mean summer daily VPD_{max} limited the yield

FIGURE 9 | Comparison of the predicted vs. reported yield, based on the one realization of cross validation, from four separate random forest models: **(A)** using only biological variables; **(B)** using biological variables and full meteorological variables; and **(C)** using biological variables and selected meteorological variables; and **(D)** using full model but excluding light interception. Red dashed line denotes the 1-to-1 line, and the gray solid line denotes the linear regression trend.

when it was higher than 40 hPa (**Figure 11D**). The additional meteorological variables such as mean winter daily T_{min} did have a slightly negative impact on the yield variation (**Figures 11E,F**).

Among mature orchards only, (i.e., tree ages from 7 to 18) ($n = 4337$), variable importance and partial dependence plots showed that light interception was the dominant control on the yield for mature almond trees (**Supplementary Figures S8, S9**),

the yield varied considerably across different cultivars. Almond cultivars with Monterey (2601 ± 458 lbs/acre) ($n = 191$) and Non-pareil (2401 ± 1152 lbs/acre) ($n = 3293$) were more productive than others (2052 ± 1110 lbs/acre, **Supplementary Figure S10**). The identified impacts from other meteorological variables were similar to those derived from the scenario using all almond samples (**Supplementary Figure S9**).

TABLE 2 | Yield prediction performance of random forest models driven by three sets of independent variables, based on the sixfold cross validation.

Models	RMSE (lbs/acres)	R^2	RPIQ
A. Biological variables (including field measurements of light interception)	530.6 ± 11.8	0.79 ± 0.01	3.12 ± 0.09
B. Full set of biological and meteorological variables	479.5 ± 9.3	0.82 ± 0.01	3.45 ± 0.17
C. Full model but with selected meteorological variables	476.9 ± 6.6	0.82 ± 0.01	3.50 ± 0.15
D. Full model but excluding light interception	536.7 ± 9.6	0.78 ± 0.01	3.08 ± 0.09
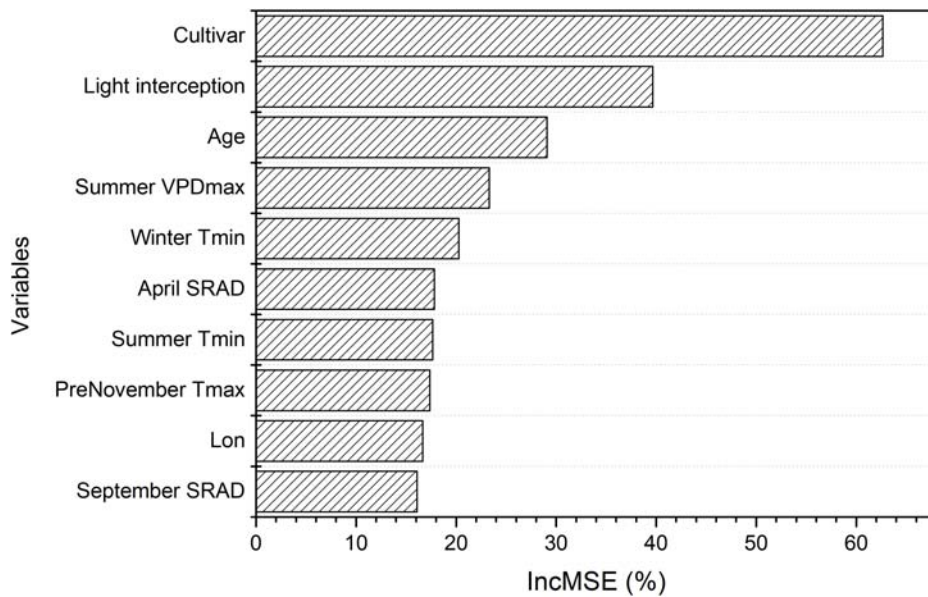
FIGURE 10 | Variable importance derived from the random forest model of overall yield using field variables and selected meteorological variables, which is measured by the increase in mean-square-error (IncMSE) of predictions estimated when excluding each variable.
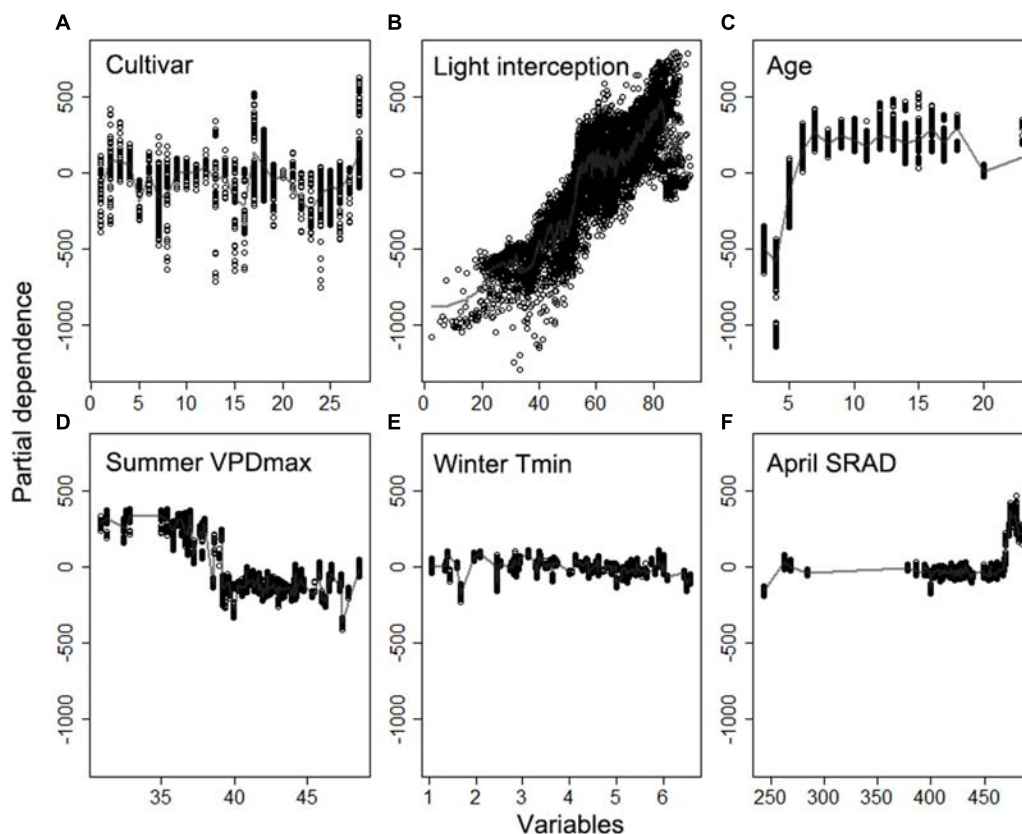
FIGURE 11 | Partial dependence of the yield on top six important variables **(A)** cultivar, **(B)** light interception, **(C)** Age, **(D)** daily maximum vapor pressure deficit (VPD) averaged over summer, **(E)** daily minimum temperature averaged winter, **(F)** daily shortwave radiation flux density (SRAD) averaged over April, derived from the model using biological variables and selected meteorological variables.

CONCLUSION AND DISCUSSION

Tree crops have rather complex processes in terms of nut production, involving physiology of tree growth, flowering phenology, bee activity, and etc (Connell, 2000; Zarate-Valdez et al., 2015; Tombesi et al., 2017; Chen et al., 2019a,b). A large dataset across the environmental gradients coupled with a more advanced data analytics, such as artificial intelligence (AI) including machine learning (ML) algorithms, are needed to understand the constraints on the yield gap at the plot and field scales (National Academies of Sciences Engineering Medicine, 2019). This study made use of a unique dataset of field measurements of light interception and almond yield records in California's almond orchards. We used random forest, a widely used ML approach, for interpreting and predicting the variations almond nut production. Our modeling experiments showed that the full random forest model explained about 82% ($\pm$1%) of yield variation using a sixfold cross validation, with a RMSE of 480 $\pm$ 9 lbs/acre). The RF-based prediction using only field measured light interception and orchard age ($R^2 = 0.60$); when excluding light interception, the overall orchard characteristics (like location and age, tree density) and environmental variables could still explain 78% of yield variation across samples. Cultivar, light interception, and age were most important in determining overall almond yield. Various climate variables were also found to play important roles in yield variation.

Both seasonal weather conditions during the current year and the previous year were found to affect the plant physiology and thus nut production from year to year at the field scale. Long term climate, on the other hand, determines the spatial variation in the almond yield at the regional scale. Our results showed that, at a given level of light interception, the departure of the actual almond nut production from the potential yield varied significantly, driven mostly by temperature in June and winter, mean summer daily VPD_{max}, and incoming solar radiation (SRAD) in addition to tree age. Warmer winter, e.g., limited the yield for the mature orchards from reaching the maximum yield. On the other hand, light interception fraction was found higher for mature sites with higher long term mean SRAD and lowest light interception for younger orchards and when March maximum temperature was lower than 19°C. For the overall almond yield, we also found that summer VPD_{max} limited the yield when it was beyond 40 hPa and warmer daily T_{min} also reduced the yield.

Further studies are needed to examine the stressors of extreme weather such as heatwaves on plant growth. We did find that the number of extreme hot days on the nut production had a negative impact, for example, extreme hot days in June either in preceding year or concurrent year had a considerable negative impact on yield ($r = -0.31$ and $r = -0.21$, **Supplementary Figures S1, S7**). However, when putting all other variables together, they didn't show as top six environmental controls, probably because these heat threats could be partially reflected by other monthly climatic variables such as VPD and temperature.

Our results showed that the light interception was found as the predominant control for the almond yield. Overall the almond yield was highly dependent on light interception, e.g., one percent of increase in light interception led to an increase of 57.9 lbs/acre in the potential yield. The mobile platform (MLB) has been used to measure the light interception at the tree and plot level. Recent advances in UAV technology makes it possible to measure the energy reflected by plants at the meter or sub-meter scale (Johansen et al., 2018; Tewes and Schellberg, 2018), and estimate the plant biomass (Bendig et al., 2015; Liu et al., 2019), therefore providing another cost-effective way to map the light interception across the field scale. Moreover, satellite observations with higher spatial and temporal resolutions have been increasing in recent years, such as at 3m by PlanetScope. The optical observations at the RGB and NIR have long been used to monitor plant growth and photosynthesis (Zhang et al., 2003; Chen et al., 2019a). An important next step is to calibrate the relationship between the field measured light interception with the optical remote sensing observations from UAVs and drones, and then map the light interception at a larger scale.

In addition to its impact on plant growth, weather condition also affects the timing and intensity of bloom and bee activity in February and March, and therefore the nut production later in the season. The bloom information derived from high resolution remote sensing observations (Chen et al., 2019b) can be integrated into the yield modeling. Yield is also largely impacted by growers' management practices including irrigation, nutrient, and canopy management such as pruning, weed management, and pests and disease control. The development of a large consistent database for location specific historic yield, orchard characteristics including the row orientation, and management history, is critical for future studies.

The machine learning approaches are expected to enhance both the explanatory power and the predictive capability, by bringing various big datasets together. A data-driven yield model based on advanced machine learning analytics, will allow researchers to query the causes and effects of location and year on productivity and to test current theories of the determinant of yields, a critical step in the development of improved sustainability practices. The prediction capability of the yield response to weather and climate, as shown by this study, is also expected to inform growers to adapt their management practices for plant protection under changing climate.

DATA AVAILABILITY STATEMENT

Publicly available datasets were generated in this study. This data can be found here: http://www.prism.oregonstate.edu/ and https://daymet.ornl.gov/.

AUTHOR CONTRIBUTIONS

YJ and PB conceived the project idea. BL collected the field data. BC compiled all the rest of spatial data, and built the models. BC and YJ performed the data analysis and wrote the manuscript.

All authors reviewed and edited the manuscript, and agreed with the submission.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00290/full#supplementary-material

REFERENCES

Baram, S., Couvreur, V., Harter, T., Read, M., Brown, P., Kandelous, M., et al. (2016). Estimating nitrate leaching to groundwater from orchards: comparing crop nitrogen excess, deep vadose zone data-driven estimates, and HYDRUS modeling. *Vadose Zone J.* 15:13.

Belgiu, M., and Drăguţ, L. (2016). Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31. doi: 10.1016/j.isprsjprs.2016.01.011

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., and McBratney, A. (2010). Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends Anal. Chem.* 29, 1073–1081. doi: 10.1016/j.trac.2010.05.006

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., et al. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Observ. Geoinform.* 39, 79–87. doi: 10.1016/j.jag.2015.02.012

Burow, K. R., Jurgens, B. C., Belitz, K., and Dubrovsky, N. M. (2013). Assessment of regional change in nitrate concentrations in groundwater in the Central Valley, California, USA, 1950s–2000s. *Environ. Earth Sci.* 69, 2609–2621. doi: 10.1007/s12665-012-2082-4

California Department of Food and Agriculture (2017). *California Agricultural Statistics Review.*

Carletto, G., Jolliffe, D., and Banerjee, R. (2015). *From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. Policy Research Working Paper no. WPS 7150.* Washington, DC: World Bank Group.

Chen, B., Jin, Y., and Brown, P. (2019a). Automatic mapping of planting year for tree crops with Landsat satellite time series stacks. *ISPRS J. Photogramm. Remote Sens.* 151, 176–188. doi: 10.1016/j.isprsjprs.2019.03.012

Chen, B., Jin, Y., and Brown, P. (2019b). An enhanced bloom index for quantifying floral phenology using multi-scale remote sensing observations. *ISPRS J. Photogramm. Remote Sens.* 156, 108–120. doi: 10.1016/j.isprsjprs.2019.08.006

Connell, J. H. (2000). Pollination of almonds: practices and problems. *Horttechnology* 10, 116–119. doi: 10.21273/horttech.10.1.116

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., et al. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol. J. R. Meteorol. Soc.* 28, 2031–2064. doi: 10.1002/joc.1688

Daly, C., Smith, J. I., and Olson, K. V. (2015). Mapping atmospheric moisture climatologies across the conterminous United States. *PLoS One* 10:e0141140. doi: 10.1371/journal.pone.0141140

Davis, K., Rulli, M., and D'Odorico, P. (2013). *Food Security in the 21st Century: Global Yield Projections and Agricultural Expansion.* Washington, DC: American Geophysical Union.

DeJong, T. (2019). Opportunities and challenges in fruit tree and orchard modelling. *Eur. J. Horticult. Sci.* 84, 117–123. doi: 10.17660/ejhs.2019/84.3.1

Elliott, J., Deryng, D., Müller, C., Frieler, K., Konzmann, M., Gerten, D., et al. (2014). Constraints and potentials of future irrigation water availability on agricultural production under climate change. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3239–3244. doi: 10.1073/pnas.1222474110

Faunt, C. C., Sneed, M., Traum, J., and Brandt, J. T. (2016). Water availability and land subsidence in the Central Valley, California, USA. *Hydrogeol. J.* 24, 675–684. doi: 10.1007/s10040-015-1339-x

Funk, C. C., and Brown, M. E. (2009). Declining global per capita agricultural production and warming oceans threaten food security. *Food Secur.* 1, 271–289. doi: 10.1007/s12571-009-0026-y

Grömping, U. (2009). Variable importance assessment in regression: linear regression versus random forest. *Am. Statis.* 63, 308–319. doi: 10.1198/tast.2009.08199

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., et al. (2016). Random forests for global and regional crop yield predictions. *PLoS One* 11:e0156571. doi: 10.1371/journal.pone.0156571

Johansen, K., Raharjo, T., and McCabe, M. (2018). Using multi-spectral UAV imagery to extract tree crop structural properties and assess pruning effects. *Remote Sens.* 10:854. doi: 10.3390/rs10060854

Kerr, A., Dialesandro, J., Steenwerth, K., Lopez-Brody, N., and Elias, E. (2018). Vulnerability of California specialty crops to projected mid-century temperature changes. *Clim. Change* 143, 419–436. doi: 10.1007/s10584-017-2011-3

Lampinen, B. D., Udompetaikul, V., Browne, G. T., Metcalf, S. G., Stewart, W. L., Contador, L., et al. (2012). A mobile platform for measuring canopy photosynthetically active radiation interception in orchard systems. *Horttechnology* 22, 237–244. doi: 10.21273/horttech.22.2.237

Lesk, C., Rowhani, P., and Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature* 529, 84–87. doi: 10.1038/nature16467

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News* 2, 18–22.

Liu, H., Dahlgren, R. A., Larsen, R. E., Devine, S. M., Roche, L. M., O'Geen, A. T., et al. (2019). Estimating rangeland forage production using remote sensing data from a small unmanned aerial system (sUAS) and planetscope satellite. *Remote Sens.* 11:595. doi: 10.3390/rs11050595

Lobell, D. B., Cahill, K. N., and Field, C. B. (2007). Historical effects of temperature and precipitation on California crop yields. *Clim. Change* 81, 187–203. doi: 10.1007/s10584-006-9141-3

Lobell, D. B., and Field, C. B. (2011). California perennial crops in a changing climate. *Clim. Change* 109, 317–333. doi: 10.1007/s10584-011-0303-6

Luedeling, E., Zhang, M., and Girvetz, E. H. (2009). Climatic changes lead to declining winter chill for fruit and nut trees in California during 1950–2099. *PLoS One* 4:e6166. doi: 10.1371/journal.pone.0006166

National Academies of Sciences Engineering Medicine (2019). *Science Breakthroughs to Advance Food and Agricultural Research By 2030.* Washington, DC: National Academies Press.

Pathak, T. B., Maskey, M. L., Dahlberg, J. A., Kearns, F., Bali, K. M., and Zaccaria, D. (2018). Climate change trends and impacts on california agriculture: a detailed review. *Agronomy.* 8:25. doi: 10.3390/agronomy8030025

Reynolds, M., and Tuberosa, R. (2008). Translational research impacting on crop productivity in drought-prone environments. *Curr. Opin. Plant Biol.* 11, 171–179. doi: 10.1016/j.pbi.2008.02.005

Rosenzweig, C., Iglesius, A., Yang, X.-B., Epstein, P. R., and Chivian, E. (2001). Climate change and extreme weather events-Implications for food production, plant diseases, and pests. *Glob. Change Hum. Health* 2, 90–104.

Tewes, A., and Schellberg, J. (2018). Towards remote estimation of radiation use efficiency in maize using uav-based low-cost camera imagery. *Agronomy* 8:16. doi: 10.3390/agronomy8020016

Thornton, P., Thornton, M., Mayer, B., Wei, Y., Devarakonda, R., Vose, R., et al. (2017). *Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3*. Oak Ridge, TN: ORNL DAAC.

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U.S.A.* 108:20260.

Tombesi, S., Lampinen, B. D., Metcalf, S., and DeJong, T. M. (2017). Yield in almond is related more to the abundance of flowers than the relative number of flowers that set fruit. *Calif. Agric.* 71, 68–74. doi: 10.3733/ca.2016a0024

Tombesi, S., Scalia, R., Connell, J., Lampinen, B., and DeJong, T. (2010). Fruit development in almond is influenced by early spring temperatures in California. *J. Horticult. Sci. Biotechnol.* 85, 317–322. doi: 10.1080/14620316.2010.11512674

USDA-NASS (2018). *Noncitrus Fruits and Nuts 2017 Summary*. Washington, DC: USDA-NASS.

Welling, S. H., Refsgaard, H. H., Brockhoff, P. B., and Clemmensen, L. H. (2016). Forest floor visualizations of random forests. *arXiv* [Preprint]. arXiv:1605.09196.

Zarate-Valdez, J. L., Muhammad, S., Saa, S., Lampinen, B. D., and Brown, P. H. (2015). Light interception, leaf nitrogen and yield prediction in almonds: a case study. *Eur. J. Agron.* 66, 1–7. doi: 10.1016/j.eja.2015.02.004

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., et al. (2003). Monitoring vegetation phenology using MODIS. *Remote Sens. Environ.* 84, 471–475.

Zhang, Z., Jin, Y., Chen, B., and Brown, P. (2019). California almond yield prediction at the orchard level with a machine learning approach. *Front. Plant Sci.* 10:809. doi: 10.3389/fpls.2019.00809

frontiers
in Plant Science

Check for updates

Emerging Technologies in Algal Biotechnology: Toward the Establishment of a Sustainable, Algae-Based Bioeconomy

Michele Fabris[1,2]*, Raffaela M. Abbriano[1], Mathieu Pernice[1], Donna L. Sutherland[1], Audrey S. Commault[1], Christopher C. Hall[1], Leen Labeeuw[1], Janice I. McCauley[1], Unnikrishnan Kuzhiuparambil[1], Parijat Ray[1], Tim Kahlke[1] and Peter J. Ralph[1]

[1] Climate Change Cluster (C3), University of Technology Sydney, Ultimo, NSW, Australia, [2] CSIRO Synthetic Biology Future Science Platform, Brisbane, QLD, Australia

Mankind has recognized the value of land plants as renewable sources of food, medicine, and materials for millennia. Throughout human history, agricultural methods were continuously modified and improved to meet the changing needs of civilization. Today, our rapidly growing population requires further innovation to address the practical limitations and serious environmental concerns associated with current industrial and agricultural practices. Microalgae are a diverse group of unicellular photosynthetic organisms that are emerging as next-generation resources with the potential to address urgent industrial and agricultural demands. The extensive biological diversity of algae can be leveraged to produce a wealth of valuable bioproducts, either naturally or via genetic manipulation. Microalgae additionally possess a set of intrinsic advantages, such as low production costs, no requirement for arable land, and the capacity to grow rapidly in both large-scale outdoor systems and scalable, fully contained photobioreactors. Here, we review technical advancements, novel fields of application, and products in the field of algal biotechnology to illustrate how algae could present high-tech, low-cost, and environmentally friendly solutions to many current and future needs of our society. We discuss how emerging technologies such as synthetic biology, high-throughput phenomics, and the application of internet of things (IoT) automation to algal manufacturing technology can advance the understanding of algal biology and, ultimately, drive the establishment of an algal-based bioeconomy.

Keywords: microalgae, synthetic biology, phenomics, industry 4.0, bioproducts, food, bioremediation, feedstock

INTRODUCTION

By 2050, it is estimated that the world population will exceed 10 billion people (United Nations, 2019). Agriculture is already nearly maximally exploited, most arable land is already in use, and issues such as climate change and urban expansion pose important challenges to the future of agriculture (Foley et al., 2011). Simply increasing the intensity of agriculture, farming, fishing, and fossil oil extraction will not be sufficient to meet future demands. Rising global temperatures, extreme weather, changing climatic patterns, and loss of cultivable land will require drastic

changes in current agrotechnology (Wurtzel et al., 2019) to minimize environmental impact through sustainable sourcing of commodities such as food, bioproducts, and bulk chemicals. Implementation of high-tech engineering and molecular genetics approaches, in the forms of phenomics and genetic engineering, has effectively improved the productivity, cost-effectiveness, and environmental impact of agricultural crops such as soy, corn, wheat, and rice (Mir et al., 2019). At the same time, plant-derived alternatives for animal-based foods such as meat and dairy, and commodities derived from petroleum such as plastics, are being developed (Zhu et al., 2016). Despite the clear advantages the of these solutions, the use of food crops to replace less sustainable manufacturing practices will eventually contribute to increased agricultural demand and face the same challenges that have been characterizing the "fuel vs. food debate." Therefore, new solutions and additional resources are required to meet the increasing demands.

Photosynthetic microalgae are microbes that have colonized every habitat on Earth, and exhibit extraordinary biological diversity, estimated to be greater than 200,000 species (Guiry, 2012), which reflects an enormous range of ecological adaptations. Unlike other microbes often exploited for bio-based manufacturing, such as yeast and bacteria, phototrophic algae have the advantage to use sunlight to fix atmospheric carbon, reducing their reliance on sugars for fermentation. Naturally thriving in environments with intermittent and scarce nutrient availability, many species of microalgae have evolved efficient metabolic adaptations to grow rapidly under favorable conditions (Smetacek, 1999; Litchman, 2007). As a result, algae often have a higher photosynthetic efficiency than plants (Bhola et al., 2014), which translates into a higher capacity to generate biomass (Benedetti et al., 2018).

When grown at large scale – in either a pond or photobioreactor – microalgae are more water-efficient than crop plants (Demirbas, 2009) and can be cultivated on non-arable land with minimal use of freshwater (Demirbas, 2009), or even grow in seawater or wastewater. Thus, many geographical areas that are not suitable or sufficiently fertile for crop cultivation could be effectively used for large-scale algal cultivation. Many algal species are naturally efficient producers of carbohydrates, lipids, proteins, pigments, as well as a range of commercial secondary metabolites that are currently sourced from conventional agriculture (Koyande et al., 2019). Also, microalgae are emerging as a next-generation, cell-sized biofactories for the sustainable manufacturing of a myriad of products (Rasala and Mayfield, 2015; Vavitsas et al., 2018), following the example of established microbial platforms such as yeasts and bacteria. In this respect, microalgal biofactories have the potential to be less expensive and more sustainable platforms that may be naturally predisposed to produce certain plant-derived products (Vavitsas et al., 2018).

Currently algae are used for a relatively small number of industrial applications. Recent works have described in details the transition of the focus from algal-based bioenergy to high-value bioproducts, and the model of algae-based biorefineries (Laurens et al., 2017). In this review, we describe how recent landmark achievements have demonstrated the untapped commercial potential of algae-based applications. Specifically,

we outline how cutting-edge technology developments such as automation, synthetic biology and phenomics can leverage the already naturally promising capabilities of microalgae in the coming years. By highlighting recent key achievements and unsolved knowledge gaps in the field – both in terms of technology advancements and applications – we describe the future development of microalgae as next-generation, low-cost, sustainable, scalable, and high productivity crop system. We anticipate that this will contribute to generate an algal-based bioeconomy, which will contribute to solutions to the imminent challenges caused by our growing society.

TECHNOLOGY DEVELOPMENT

While crop plants have been bred and selected for millennia to isolate specific traits and to obtain highly productive strains, all present microalgal species are effectively environmental isolates. To maximize productivity and increase the industrial potential of microalgae, it is key to optimize both the organism and environment that supports its growth. In the following sections, we describe how this can be achieved through the latest technology developments in algal cultivation and harvesting, automation, phenotyping, and synthetic biology (**Figure 1**).

Algal Cultivation

One of the most attractive intrinsic features of many algal species is that they are capable of rapidly and inexpensively generating large amounts of biomass compared to plants (Brennan and Owende, 2010). In nature, microalgae are capable of reaching high biomass concentrations under eutrophic conditions but, from a mass culture point of view, even these concentrations are not sufficient. In the past decade, there has been a large body of research focused on optimizing conditions that maximally promote algal growth rates, or elicit enhanced production of a specific product, under artificial growth conditions. However, one of the biggest limitations in algal mass cultivation is creating a cost-effective production system. In this regard, a diverse range of algal cultivation techniques can offer differing levels of control over the growth and product yield, with different associated capital and operating costs.

Several factors can limit microalgal growth in mass culture, including light availability, temperature and pH as well as both the concentration and ratio of the major nutrients, carbon, nitrogen and phosphorus (Sutherland et al., 2015). Some algae are capable of growing autotrophically as well as mixo- or heterotrophically which allows them to avoid light limitation constraints in dense culture, but does require the addition of organic carbon sources. As with any supplementation, adding organic carbon to the growth medium increases material input costs, but may achieve higher cell densities (Venkata Mohan et al., 2015). In principle, algae have the same basic requirements as plants, in that they need biologically available nitrogen and phosphorus, as well as trace nutrients (i.e. sulfur, calcium, iron, silicon.), and management of the pH levels to maximize nutrient availability (White and Ryan, 2015). The water source can affect what nutrients need to be additionally supplied, while water
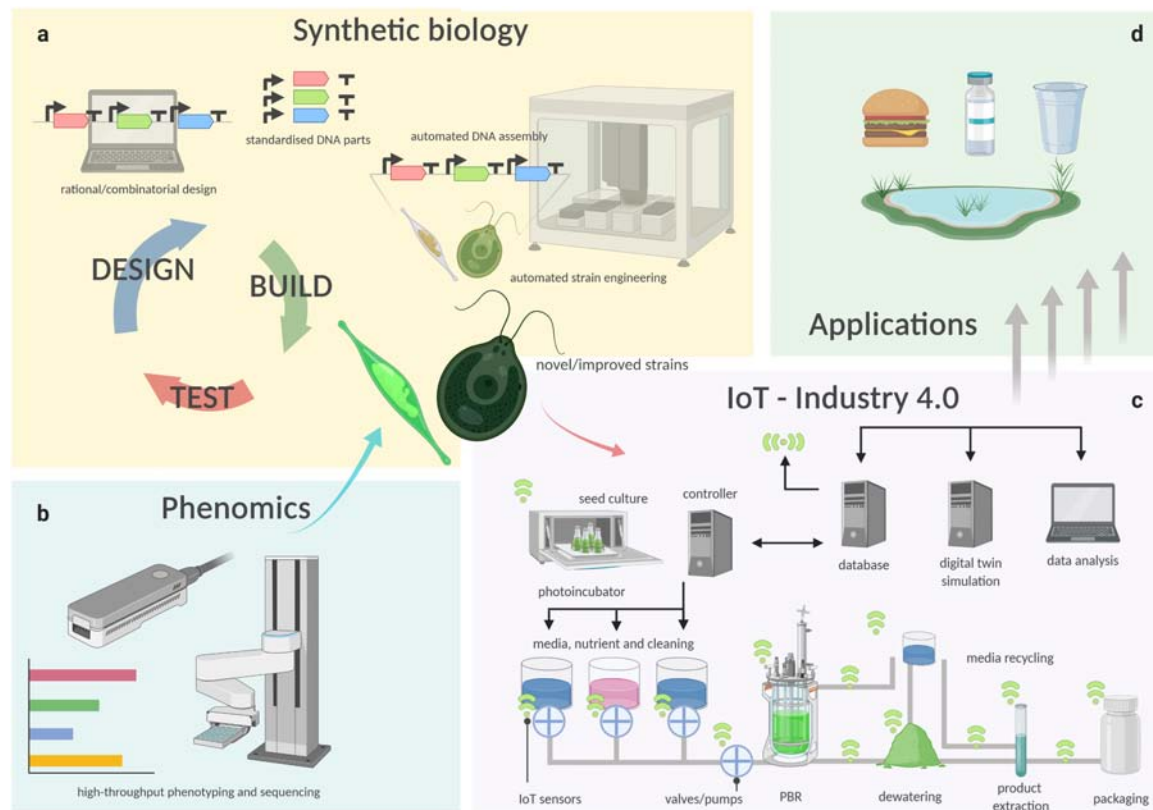
FIGURE 1 | Schematic representation of how technologies such as synthetic biology **(a)**, phenomics **(b)**, cultivation technology and IoT **(c)**, are connected in a semi-automated pipeline for the manufacture of bioproducts from microalgae **(d)**. Biological functions are encoded into instructions through rational or combinatorial design of genetic constructs, which are then used to generate thousands of new microalgal genotypes with iteration of the *design-build-test* cycle **(a)**. Either natural isolates or engineered strains (*test* phase), are phenotyped in different, controlled conditions by high-throughput analyses. **(b)** Novel or improved strains with superior traits are then isolated and utilized for industrial production. **(c)** Using Industry 4.0 principles, in which a controller, such as an industrial programmable logic computer (PLC), receives information and logs its operation to a database computer. The database collects data from a network of plug-and-play sensors, which inform a digital twin simulation of the facility. The digital twin predicts the future demand and yield of the algae culture and updates the controller to optimize the process to match the predicted demand.

availability and recovery is key in determining what algal species can be selected. Algae can grow on different water sources, such as marine, fresh, or waste water. Wastewater is naturally rich in nutrients, but has additional contaminants that could cause culture crashes. There is a large diversity of marine algae and ready availability of seawater. However, seawater requires the addition of fertilizers and, in open systems, it is subjected to evaporation. This causes the salinity to be altered and to require monitoring. Fresh water may also require additional nutrients, but may also increase the strain on water supplies in water scarce regions. Recycling the water can aid in reducing these issues and improve the economic viability. The overall water requirement is still lower than traditional plant based crops (Rawat et al., 2013) leading algae to be desirable alternatives for cultivation.

Traditionally, microalgae have been grown in simple open ponds (Becker, 1994), but research and technological advances over the past several decades have led to a diversity of high-productivity bioreactor designs. Large-scale autotrophic algal production designs accommodate suspended or attached growth in either open or closed systems, or a hybrid of these, reviewed extensively elsewhere (Ugwu et al., 2008; Brennan and Owende, 2010; Harun et al., 2010; Christenson and Sims, 2011; Olivieri et al., 2014).

Besides stagnant ponds, the cheapest option for large-scale microalgal production is a shallow open pond raceway design that includes basic mixing. Compared to other photo-bioreactor (PBR) designs, they also have lower energy requirements, lower capital and operating costs, and can be built at a large scale (Brennan and Owende, 2010; Borowitzka and Vonshak, 2017). However, they generally have the lowest areal productivity (<10 g m^{-2} d^{-1} compared to >20 g m^{-2} d^{-1} in some PBRs) (De Vree et al., 2015). Some advances have been made improving productivity by modifying the design of open systems, such as high rate open ponds (HRAP), where increased baffles or more complex geometries improve the overall mixing pattern and ensure algae remain in the illuminated part of the water column (Christenson and Sims, 2011; Craggs et al., 2012). Companies worldwide are investing in this system, For example in Hawaii, this has generated over

$US 10 million gross profits from biomass grown in open ponds (Maeda et al., 2018).

Another improvement in outdoor open cultivation are bioreactors for attached growth, such as algae turf scrubbers (ATS), or motorized wheels with biofilm growth (Wang and Lan, 2018). Novel biofilm-based algal cultivation in particular has seen increased research in recent years, in part due to higher harvested solid content (10–20% compared to <0.02% for suspended systems), which leads to lower harvesting costs. While biofilm growth is not suitable for all algal species, and can lead to complex mixed algae-bacteria communities (and therefore less suited for high-value, single products), it has been investigated for wastewater remediation (Gross et al., 2015) (section "Algal Biodegradation of Emerging Contaminants"). Closed suspended growth PBRs, such as flat-plate, tubular, or bag reactors have increased operating control, better mixing, and less chance of contamination compared to open systems and are suitable for genetically modified organisms, but also have significantly higher capital and operational costs (Gupta et al., 2015). Closed systems can be operated using artificial light (at increased costs), which can be tailored to the specific algae to increase productivity (Schulze et al., 2014; Glemser et al., 2016). In addition, genetically modified organisms (GMOs) may have regulatory limitations that prevent them from growing in open systems where they can be released into the wild. As such, closed systems may become increasingly common in the future. Despite these advantages, most of the current production is done in open pond systems (on the order of thousands of tonnes per annum) for products such as biofuels, animal feed, and nutraceuticals, while closed systems (hundreds of tonnes per annum) are used primarily for high-value products (Posten, 2009; Borowitzka and Vonshak, 2017).

Various techno-economic assessments have reviewed the feasibility of large-scale algal production (Laurens et al., 2017). The capital cost of an open pond can range from ∼$US 6/m^2 (Craggs et al., 2012) to US$ 50/m^2 (Huntley et al., 2015), while enclosed PBR systems can to cost up to 3 to 30 times more (Panis and Carreon, 2016). Operating cost can instead vary from US$ 0.8/kg dry weight (DW) to up to $8/kg DW for various systems and applications. Biofuels in particular have received most attention and are currently not price competitive yet, with production costs at ∼ US$ 3/L compared to <US$ 1/L of producing fuel from fossil oil (Sun et al., 2011; Laurens et al., 2017; Roles et al., 2020). The areal/volumetric productivity is generally one of the largest uncertainties as well as drivers for success (Jonker and Faaij, 2013; Chauton et al., 2015; Panis and Carreon, 2016; Hoffman et al., 2017). Strain selection is key to optimize productivities for the final product. This could require designing novel strains through genetic engineering or synthetic biology (section "Synthetic Biology") that have been thoroughly profiled and selected using a phenomics approach (section "Phenomics"). On the cost side of the equation, dewatering is one of the major expenses in algal processing currently, consisting in up to 20–30% of the final cost, as reviewed extensively by Christenson and Sims (2011), Milledge and Heaven (2013), Gerardo et al. (2015) and Fasaei et al. (2018).

Algal suspensions are generally very dilute; therefore increasing the biomass content in the cultivation stage can

substantially reduce costs, which is a major advantage for closed and attached growth PBRs compared to open ponds (Fasaei et al., 2018). Advances have been made in harvesting technology, by employing for example cross-flow filtration (Gerardo et al., 2014), cheaper flocculants ('T Lam et al., 2018; Nguyen et al., 2019), bio-flocculants (Ummalyma et al., 2017), microfluidics at lab scale (Kim et al., 2018), and novel techniques such as pulsed electric field, ultrasound, and electroflocculation, that have yet to be demonstrated at industrial scale (Milledge and Heaven, 2013; 'T Lam et al., 2018). However, while some harvesting processes can reduce the energy costs – for example filtration has a lower energy requirement compared to centrifugation – they can lead to a higher operating cost (e.g. filtration is subject to membrane fouling) (Bilad et al., 2014). Flocculants, chemicals added to cause algal cells to aggregate, can be inexpensive and have a long history in wastewater treatment, but can be hard to recover and affect downstream processing and media recycling (Milledge and Heaven, 2013). To address these short-falls, some systems combine multiple harvesting steps (e.g. flocculation combined with dissolved air flotation to remove the aggregates) (Pragya et al., 2013), while others are looking to bypass the harvesting step completely by modifying the algae to secrete the compound of interest (Christenson and Sims, 2011). The final harvesting step depends on the required product, the type of algae, and the specific cultivation strategies. As such, there is no one-size-fits-all solution to harvesting of the algal cultures.

In many cases, the final products can fit into existing industrial processes, for example transesterification for biofuels and extraction of high-value products (Greenwell et al., 2010; Khanra et al., 2018). There is also ongoing research for improvements in extraction of the algal products, such as by using supercritical extraction, pressure or microwave assisted extraction, ionic liquids, novel (less toxic) solvents, enzyme assisted extraction, or aqueous biphasic systems (Kadam et al., 2013; Kumar et al., 2015; Chew et al., 2017; 'T Lam et al., 2018; Khanra et al., 2018). Many of these novel and "green" extraction processes are dependent on the desired product, and are still only in use at the lab or pilot phase; getting them to an industrial scale would require significant investment and further research (Michalak and Chojnacka, 2015).

There is currently no single "best practice" method to cultivate algae, especially at scale. Final design of the system is dependent on the final product, the geographical location, as well as local resources available (e.g. accessibility to water, to CO_2, and to waste streams). Modeling and lab scale experiments have suggested novel innovations in designing and operating the process, but the final consideration is the cost: some processes may have a larger up-front capital cost, but reduce the overall operating cost (e.g. HRAP ponds), while others may have very low capital costs but affect further downstream processing (e.g. chemical flocculation). As such, due consideration for the overall cost will guide the final design and operation of the system.

Industry 4.0 Approach to Algal Biorefineries

Regardless how the biomass is produced, if the downstream processing can be performed in an integrated biorefinery that allows the greatest number of products and co-products
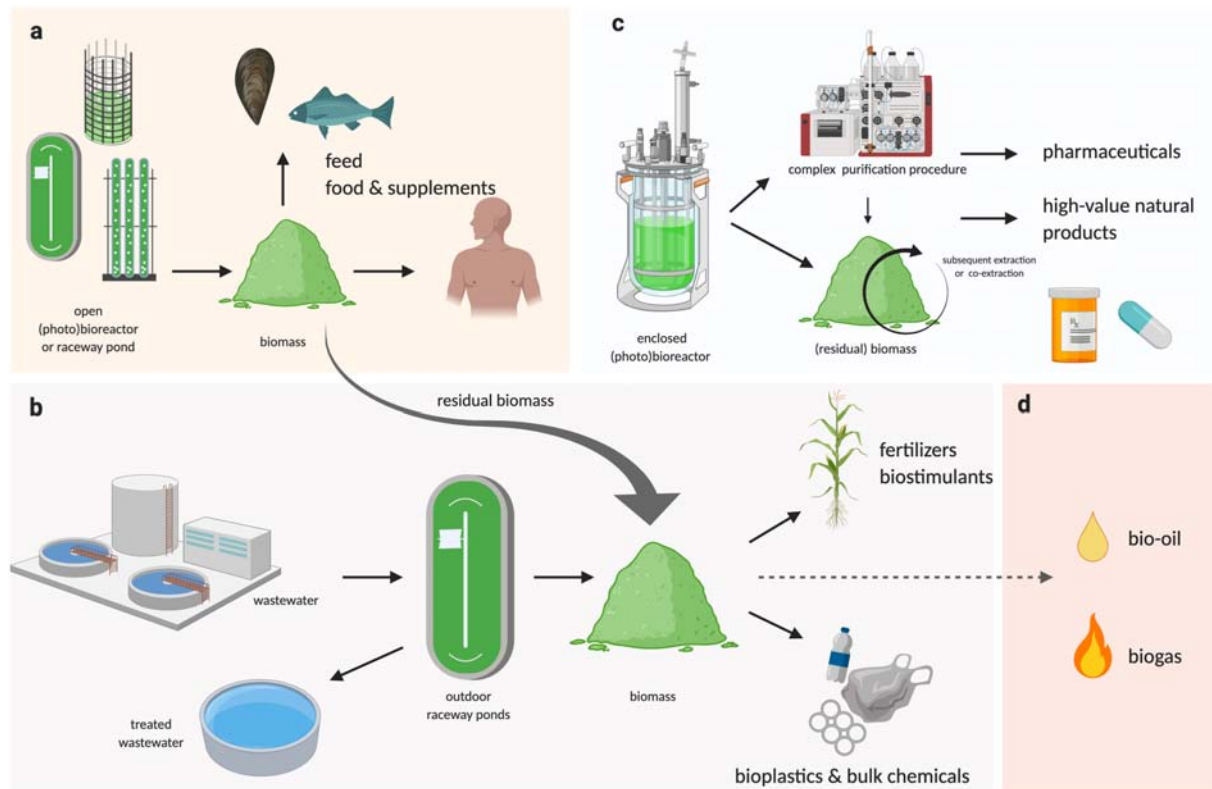
FIGURE 2 | Schematic representation of a multi-product algal bio-refinery model. **(a)** Algal biomass cultivated at large scale in outdoor raceway ponds or large PBRs can be used as feed or food supplements, where the residual biomass and/or biomass generated from bioremediation processes can be used for industrial applications **(b)**, as well as for bioenergy **(d**, not reviewed here), while pharmaceuticals or other high-value products requiring complex and controlled extraction procedures could be co-extracted or subsequently extracted from algal biomass grown in enclosed bioreactors (BRs) or photobioreactors (PBRs) **(c)**.

to be extracted (**Figure 2**), and with the least amounts of residual/waste, it will ensure the maximum return on investment for downstream processing. Industry 4.0 is an advanced manufacturing approach based on machine-to-machine communication technologies, also known as "the Internet of Things," or IoT (Atzori et al., 2010), whereby automation, sensors, and machine learning create a self-adapting manufacturing processes able to adjust in real time to changes in the process itself (Kagermann et al., 2011; Kagermann et al., 2013). In a microalgal biorefinery, this means that not only can the algal cultivation and harvesting system be automated to reduce operating costs, but a network of plug-and-play IoT sensors could allow the operators to monitor the algae growth and productivity in real time (**Figure 1c**; Whitmore et al., 2015). The concept of Industry 4.0 goes a step further by building a simulation, or digital twin, of the facility and the algal culture from the sensor data. The simulation can make real-time predictions of future cellular yield and adjust operations to meet expected product demand and to reduce waste (**Figure 1c**; Tuegel et al., 2011; Uhlemann et al., 2017; Tao et al., 2018). For example, a fully realized Industry 4.0 microalgal biorefinery (**Figures 1c**, **2**) would link the controlled cellular yield of specific components with automated serial downstream extraction of several co-products that are driven by current

demand, rather than traditional linear production stockpiling that awaits demand (Kagermann et al., 2013). Biorefineries can be located at regional hubs to service surrounding producers.

Phenomics

Phenomics is defined as "the acquisition of high-dimensional phenotypic data in an organism-wide scale" (Houle et al., 2010). Algal phenomics is currently in very early stages of developments, however, it holds great potential in microalgal agriculture for food security (section "Food and Nutraceuticals"), bioproducts sourcing (sections "Food and nutraceuticals," "Feedstocks," "High-Value Products," and "Biopolymers, Bioplastics, and Bulk Chemicals"), bioremediation (section "Algal Biodegradation of Emerging Contaminants"), and carbon sequestration. By creating a database of $GxE = P$ [where G, genome, E, environment(s) and P, phenotype(s)] interactions for a given algal species, researchers can screen natural and artificial diversity for the combination of gene alleles that will combine essential phenotypes (Furbank and Tester, 2011), such as fast growth and high product yield.

Recent advancements from the field of plant phenomics highlight the potential impact of phenomics techniques and technologies in microalgae. For example, a recent phenomics-based study on *Arabidopsis thaliana* yielded a mutant exhibiting both increased pathogen defense and photosynthetic growth,

breaking the supposed trade-off between growth and defense that was a near-dogma among plant researchers (Campos et al., 2016; Cruz et al., 2016). In microalgae, such a phenomics approach could create equally dramatic combinations of useful phenotypes, such as a strains that use quorum sensing (Das et al., 2019) to trigger autoflocculation (González-Fernández and Ballesteros, 2013) and induction of product synthesis (e.g. carotenoids) (Gong and Bassi, 2016) only when the culture reaches harvest density.

A major limitation to present-day microalgal phenomics is the lack of searchable phenomics databases. Plant and yeast researchers, for example, can design or even perform *in silico* experiments using The Arabidopsis Information Resource (TAIR) (Lamesch et al., 2011) and PROPHECY (Fernandez-Ricaud et al., 2005; Fernandez-Ricaud et al., 2016) databases, respectively. Such tools accelerate research by showing how different genes can be related by a shared phenotype (Ohyama et al., 2008), or even differentiate the functions of seemingly redundant gene copies (Yadav et al., 2007). To bring this power to the field of algal research, there will need to be investments in building a comprehensive phenotypic database. The field of plant phenomics has already created the standards for data sharing, knowledge retrieval, and ontology annotation (Oellrich et al., 2015; Munir and Sheraz Anjum, 2018; Neveu et al., 2019), which can be adapted to algae. The data analysis tools presently used for model microbes (e.g. yeast) can also be applied to microalgae to measure phenotype data from high-throughput algae culture formats (agar plates, microplates, etc.) using standard microbiology sensors such as fluorescence and absorption spectrophotometers (Fernandez-Ricaud et al., 2005; Fernandez-Ricaud et al., 2016), hyperspectral cameras (Roitsch et al., 2019), and flow cytometers (Cagnon et al., 2013). Even morphological phenotypes can be automatically digitized via machine learning (ML) approaches such as image processing with Support Vector Machines (SVNs) and Convoluted Neural Networks (CNNs), as demonstrated in Mohanty et al. (2016) and Sladojevic et al. (2016).

One challenge for developing microalgal phenomics databases will be choosing the growth environments that cause the microalgae to display a range of phenotypes based on their genetic predispositions. For example, much research has been done using high and low concentrations of carbon dioxide to learn about the roles of genes in photosynthesis (Suzuki et al., 1999; Vance and Spalding, 2005; Duanmu et al., 2009), but elucidating phenotypes related to stress responses and repair cycles can be more challenging (Cruz et al., 2016; Tietz et al., 2017). The risk of poorly chosen algal phenomics reference environments could result in insufficient segregation of phenotypes (Thomas, 1993), or worse, the environments might be so different from large scale cultivation as to render the measured phenotypes misleading and irrelevant to large scale enterprises (Rawat et al., 2013).

Another obstacle to algal phenomics is the limited capacity to manipulate the genetics of many non-model microalgal species. Many microalgae have complex life cycles (Graham et al., 2009) and their genomes are often large and highly repetitive, defying typical shot-gun sequencing techniques for genome sequencing

and assembly (Paajanen et al., 2017). However, 3rd generation methods that allow long-read sequencing – such as Nanopore and PacBio sequencing – are breaking the log-jam. For example, PacBio sequencing has proven capable of mapping *trans*-gene integration sites in plants (Liu et al., 2019). These tools will be key in mapping genotypes of mutant libraries. Novel gene editing technologies such CRISPR-Cas9 are increasingly used to create large, genome-wide knock-out libraries in important crops such as tomatoes and rice (Jacobs et al., 2017; Meng et al., 2017). In the future, a combination of these gene-editing methods with 3rd generation sequencing technologies will enable cost and time-effective creation and mapping of knock-out libraries of important microalgal species. With a reference genome in hand, linking a phenotype to the relevant gene alleles traditionally relies on statistical analysis of the progeny after cross-breeding two individuals or populations with different phenotypes for example through Quantitative Trait Loci (QTL) mapping. With each successive generation, the genomic regions responsible for the phenotypes can be narrowed down until one has a testable list of candidate gene loci (van Bezouw et al., 2019). Given that many microalgae either do not breed at all or only under often unknown environmental conditions, novel approaches must be developed to fully utilize the power of phenomic mutant screens.

Once these challenges have been solved, algal phenomics will have a big impact on algal biotechnology by enabling the development of microalgae as new bioproducts and pharmaceutical workhorses (section "High-Value Products"). In this case, the yield of the desired product (or a suitable proxy) is treated as one phenotype in a phenomics search for both productivity and reliability among engineered strains. In addition, the maturation of algal phenomics tools will likely change the very mindset of algal biotechnology researchers. Presently, algal biotech researchers pick a single strain that synthetises the product of interest, and then they try to optimize the culture environment to improve the productivity, often resulting in expensive and complicated PBR design (Vasumathi et al., 2012; Wang et al., 2012; Melnicki et al., 2013; Lucker et al., 2014). In the era of algal phenomics, researchers will instead define their production culture environment, and optimize the algae to that environment, as much as a plant breeder would do for field crops (Donald, 1968; Jordan et al., 2011).

Synthetic Biology

Synthetic biology applies engineering principles to the rational design of living organisms. Within this discipline, a biological system is viewed as a collection of characterized genetic parts that can be modified and reassembled to alter existing functions or to build them *de novo* in alternative host organisms. Genetic designs are revised through iterations of a design-build-test-learn cycle to achieve optimized metabolic configurations for biotechnological applications (Khalil and Collins, 2010; Nielsen and Keasling, 2016). Synthetic biology applied to microalgae will combine this powerful new approach with the benefits of a photosynthetic microbial host to generate novel production strains tailored to suit future environmental challenges (**Figure 1a**).

Tools for the genetic engineering of microalgae are evolving rapidly, enabled by the increased availability of sequenced

genomes across multiple algal lineages. The sequencing of microalgal genomes has facilitated genetic tool development in the green alga *Chlamydomonas reinhardtii* (Mussgnug, 2015), stramenopiles *Phaeodactylum tricornutum* (Huang and Daboussi, 2017) and *Nannochloropsis gaditana* (Poliner et al., 2018), and cyanobacteria *Synechocystis* sp. PCC 6803 (Hagemann and Hess, 2018), among others. In addition, methods for the genetic transformation in microalgae have been optimized for many species and include natural transformation, electroporation, bead beating, biolistic transformation, and conjugative plasmid transfer (Qin et al., 2012).

Genomic data from these species have facilitated the identification of native genetic elements necessary for genetic engineering and successful transformation. Several constitutive and inducible endogenous promoter/terminator pairs have been demonstrated to effectively express transgenes in model species (Wang et al., 2012; Ramarajan et al., 2019), including bidirectional promoters for gene stacking or co-expression with a selectable marker (Poliner et al., 2018). In addition, heterologous or synthetic promoter sequences have been characterized (Berla et al., 2013; Zhou et al., 2014; Scranton et al., 2016). Other regulatory effectors, such as ligand-binding riboswitches, have been identified and developed as tools to regulate gene expression in cyanobacteria and *C. reinhardtii* (Moulin et al., 2013; Nakahira et al., 2013). Additional characterization of sequences that regulate transcription will be crucial to transition from stepwise genetic engineering to targeting multiple sites, introducing multi-gene pathways, or building independent synthetic circuits. In addition to sequences that modulate transcription, the molecular toolkit in microalgae also includes a useful suite of selectable markers, reporter genes, protein tags, and peptide sequences for ribosomal skipping or protein localization (Vavitsas et al., 2019). To standardize these commonly used genetic parts and to facilitate collaboration, the scientific community has adopted Type IIS restriction endonuclease cloning systems. This approach allows for efficient modular assembly of complex plasmids from a library of domesticated parts, and is being widely implemented in several models, including in plants (Patron et al., 2015). Suites of parts specific to microalgae have been developed to be compatible with a common syntax to benefit from existing part registries. Type IIS cloning systems specific to microalgae include the MoClo toolkit for *C. reinhardtii* (Crozet et al., 2018), CyanoGate for cyanobacteria (Vasudevan et al., 2019), and uLoop for diatoms (Pollak et al., 2019).

Multiple molecular techniques are available to modify native gene expression or target specific areas of the genome in microalgae. Gene knockdown by introduction of antisense, artificial small RNAs, and CRISPRi has been implemented in multiple systems (De Riso et al., 2009; Zhao et al., 2009; Yao et al., 2016; Wei et al., 2017; Sun et al., 2018). Site-specific genetic manipulation by homologous recombination (HR) is routine in cyanobacteria (Zang et al., 2007), the chloroplast genome of *C. reinhardtii* (Esland et al., 2018), and the nuclear genome of *Nannochloropsis* (Kilian et al., 2011). In contrast, HR occurs at a low frequency in the nuclear genome of *C. reinhardtii* and *P. tricornutum*, but can be induced in the presence of double-strand DNA breaks by targeted endonucleases, enabling targeted

gene knockout and/or knock-in (Shin et al., 2016; Greiner et al., 2017; Kroth et al., 2018). Although technology for precision genome editing, including zinc-finger nucleases, transcription activator-like effector nucleases (TALENs), or CRISPR/Cas9, has been reported in many microalgae (Sizova et al., 2013; Weyman et al., 2015; Li et al., 2016; Nymark et al., 2016; Ajjawi et al., 2017), several challenges related to targeting, efficiency, and toxicity remain to be fully overcome. Strategies to circumvent these issues include transient Cas9 expression (Guzmán-Zapata et al., 2019), direct ribonucleoprotein (RNP) delivery (Baek et al., 2016; Shin et al., 2016) and use of Cas variants (Ungerer and Pakrasi, 2016). Marker-free and multiplex gene knock-out remains a challenge in some microalgae, although the use of multiple sgRNAs to multiplex genome editing targets has been shown to be feasible in diatoms (Serif et al., 2018).

Despite the rapid advances in the genetic tools available in microalgae, the field trails behind other established chassis microorganisms such as *Escherichia coli* and *Saccharomyces cerevisiae*. These model systems benefit from decades of intense study, resulting in diverse suites of characterized genetic parts and tools, known metabolic features, and well-annotated genomes. Approaches such as protein engineering and directed evolution that have been effectively implemented in these traditional hosts (Abatemarco et al., 2013) could also be applied in microalgae to hasten their development as chassis organisms.

Advances in synthetic biology are also enabling the design of entire microbial genomes (Hutchison et al., 2016; Richardson et al., 2017). While still on the horizon for eukaryotic algae, the development of self-replicating episomes in diatoms (Karas et al., 2015) has demonstrated that a synthetic sequence can be faithfully maintained in the diatom nucleus without integration into the native genome. This innovation is a step toward the design and assembly of independent, artificial chromosomes in microalgae. Reconstruction of a native *P. tricornutum* chromosome has already been demonstrated in yeast (Karas et al., 2013), and it is possible that a similar approach could be used to construct completely refactored chromosomal sequences.

Currently, genetic engineers are limited by the number of designs that they can feasibly assemble and test. However, it is anticipated that increased integration of computational design and automation with biology will rapidly shift this paradigm. Computational modeling can be used to predict non-intuitive approaches to optimize metabolic flux through heterologous pathways, as was demonstrated by the optimization of terpenoid production in cyanobacteria (Lin et al., 2017). Novel biological designs or complex combinatorial libraries can be rapidly assembled and evaluated in automated, high-throughput biofoundries, which are attracting investment from research institutions across the globe (Hillson et al., 2019). To evaluate clone libraries at scale, strain development must also be accompanied by improved technology for small molecule detection, such as the development of novel biosensors, as well as advancements in multi-dimensional phenotyping (section "Phenomics").

Synthetic biology is not limited to the production of existing natural compounds, since the deconstruction of biology into its basic genetic components permits systems to be redesigned free

from pre-existing constraints. An exciting avenue of synthetic biology will be the creation of novel, new-to-nature compounds with potential new functions and applications (Moses et al., 2014; Arendt et al., 2016; Luo et al., 2019). Synthetic biology can also be leveraged to improve agricultural outcomes for the cultivation of microalgae, including optimization of photosynthetic efficiency and improving carbon utilization (Gimpel et al., 2013; Erb and Zarzycki, 2016). For example, scenarios for the synthetic redesign of more efficient photosynthetic carbon fixation have been computationally predicted (Bar-Even et al., 2010). Given these advancements, the application of synthetic biology to microalgae has enormous potential to reinvent conventional animal and plant-based industries (e.g. food, high-value products, and chemicals) through innovations to minimize cost and environmental impact.

APPLICATIONS

Food and Nutraceuticals

Increasing the current capacity of microalgae to provide a source of nutrients, minerals, trace elements and other bioactive compounds is an active area of research that establishes a precedent for the development for new health products (Plaza et al., 2008; Lordan et al., 2011; Wells et al., 2017; Barkia et al., 2019). The microalgal industry has yet to reach its full potential, with an estimated global net worth of $US1-1.5 billion (Pulz and Gross, 2004). Due to a history of safe production and consumption, the cyanobacteria *Spirulina* sp., along with the green algae *Chlorella* sp. and *C. reinhardtii* are internationally recognized as "generally regarded as safe" or GRAS, a certification legislated under the United States Food and Drug Administration (FDA, 2019). Other certified GRAS species include the green algae *Haematococcus* sp. and *Dunaliella* sp. (FDA, 2019).

There are many commercial food markets that can be occupied by ingredients and products derived from microalgal biomass. For example, microalgal biomass can be a source of bulk protein, carbohydrates, and lipids (Koyande et al., 2019). Microalgal protein is a particularly promising avenue to contribute to the future of sustainably based agriculture. Currently, the majority of global protein intake is attributed to higher plants (Billen et al., 2014; Henchion et al., 2017; Caporgno and Mathys, 2018), but plants require large amounts of arable land, water, and use of herbicides and fungicides (Dahman et al., 2019). Algal-sourced protein can be a sustainable alternative soy-based protein, due to its higher protein content and favorable amino acid profile, making it a high-quality protein for human nutrition (Spolaore et al., 2006; Kent et al., 2015). Recent studies show promising results with regard to improved physico-chemical and nutritional properties of *Spirulina* protein blends (Grahl et al., 2018; Palanisamy et al., 2019).

Some microalgae species are also a source of bioactive secondary metabolites that may ameliorate disease symptoms or causes, such as inflammation (Montero-Lobato et al., 2018) or provide protection to neuro-degenerative diseases (Olasehinde et al., 2017). Dried algal biomass from GRAS-certified species is most commonly consumed as a powder and already marketed

as dietary supplement to improve health, is often added to other foods, such as blended beverages (Vigani et al., 2015). Powdered dietary supplements have been assessed in a number of clinical trials with some promising outcomes (Merchant, 2001; Karkos et al., 2011). For example, both *Spirulina* sp. and *Chlorella* sp. have clinically shown the ability to positively affect lipid profiles, various immune variables, and have antioxidant capacities (Mao et al., 2005; Park et al., 2008; Ryu et al., 2014; Kim et al., 2016; Garcia et al., 2017).

With an increased public preference for naturally sourced food additives, microalgae pigments offer an appealing alternative to synthetic pigments. Naturally derived pigments are a group of compounds that are inherently bioactive. They act as radical scavengers and can reduce oxidative damage (Singh et al., 2005), and therefore have appeal as dietary supplements or fortifying ingredients to promote human health (Stahl and Sies, 2005). This is in contrast to synthetic pigments, some of which are raising increasing concerns regarding their toxicities and subsequent adverse health effects, whilst also not providing any nutritional value (Oplatowska-Stachowiak and Elliott, 2017).

Microalgae are major producers of pigments such as fat-soluble chlorophylls, carotenoids (carotenes and xanthophylls) and water soluble phycobilins e.g. phycocyanin (Begum et al., 2016). *Haematococcus* sp. and *Dunaliella* sp. are two species that can accumulate significant levels bioactive pigment molecules such as astaxanthin (Guerin et al., 2003) and β-carotene (Tafreshi and Shariati, 2009), respectively. Astaxanthin is notable for its brilliant red color that brightens the flesh of seafood (Kidd, 2011). Humans do not synthesize astaxanthin, and dietary intake is almost exclusively via seafood (Kidd, 2011). Astaxanthin is presently mostly produced synthetically. Current production costs of microalgal-derived astaxanthin are still higher than those of synthetic (EUR 1540/kg and EUR 880/kg, respectively) (Panis and Carreon, 2016), although studies have estimated that these costs could be theoretically reduced to US$ 500 – US$ 800/kg (Li et al., 2011). Another common commercial pigment from microalgae is phycocyanin, derived from *Spirulina* sp. (Vigani et al., 2015). This deep natural blue pigment is utilized as a natural food colorant for food items such as chewing gum, ice sherbets, popsicles, candies, soft drinks, dairy products, and jellies (Begum et al., 2016).

Phytosterols are compounds often used as food supplement and cholesterol-lowering agents and are currently extracted from plants with suboptimal yields due to a complex extraction process (Ras et al., 2014). The sterol composition of algae is extremely diverse and comprises molecules typically synthetized by plants (e.g. brassicasterol and stigmasterol), animals (e.g. cholesterol) and fungi (e.g. ergosterol) (Rampen et al., 2010; Miller et al., 2012; Fabris et al., 2014; Lu et al., 2014), as well as novel and uncharacterized triterpenoids (Commault et al., 2019) and sterols, such as gymnodinosterol and brevesterol (Giner et al., 2003). Stigmasterol in particular is commonly used as cholesterol-lowering agents in food supplements (Batta et al., 2006). Several algal species naturally produce equal or greater amounts of phytosterols than plants, which are usually in the range of 0.025–0.4% of plant dry biomass (Piironen et al., 2003). Therefore diatoms and other algal groups have the potential to be

alternative, low-cost, and more sustainable source of phytosterols (Ahmed et al., 2015; Jaramillo-Madrid et al., 2019). For example, the model diatom *P. tricornutum* produces up to 0.32% d.w. of phytosterols and the haptophyte *Pavlova lutheri* can accumulate phytosterols up to 5.1% d.w (Ahmed et al., 2015).

Essential polyunsaturated fatty acids (PUFA) such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), play crucial roles in human health. DHA is necessary for neural development (Swanson et al., 2012) and is routinely utilized in infant formulas, fortified food and beverages and dietary supplements (Ratledge, 2004). Presently, most DHA and EPA is supplied by wild-catch and captive based fisheries (Lenihan-Geels et al., 2013). As primary producers of essential PUFAs in nature, with DHA concentrations up to 50% of total biomass (Ratledge, 2004), microalgae represent a promising and more sustainable alternative (Ryckebosch et al., 2014). Presently, the production costs associated with microalgal derived EPA/DHA reach US$ 40/kg EPA + DHA, but technological advancements could possibly lower this to ~US$ 10/kg EPA + DHA, which is competitive if compared to fish oil (~US$ 8/kg EPA + DHA) (Chauton et al., 2015).

Currently, three different commercial fermentation processes are used to produce DHA, with each utilizing different microorganisms (Ratledge, 2004). Martek Biosciences Corp (Netherlands) led the infant formula DHA market until 2011, when it was acquired by Dutch State Mines (DSM) in 2011. DSM utilizes the dinoflagellate microalgae *Crypthecodinium cohnii*, which accumulates DHA up to 60% of the total fatty acids fraction (Jacobsen et al., 2013) for use in infant formula, and *Schizochytrium* sp., a heterotrophic protist which can yield about 40% (w/w) of DHA (Ratledge et al., 2010). Oil from *Schizochytrium* sp. has been traditionally used for improving animal feeds, but there is also a market push toward human nutritional supplements (Ratledge and Cohen, 2008). The thraustochytrid *Ulkenia* sp. utilized by Nutrinova GmbH (Germany) produces up to 46% (w/w) DHA. In contrast to phototrophic algae, thraustochytrids are grown heterotrophically in stainless steel fermenters using complex organic substances, including by-products from other processes (e.g. sugars, organic acids) as a sole carbon and energy source (Chang et al., 2014). The oils obtained from *Schizochytrium* and *Ulkenia* are defined as "novel foods" under the European Union (EFSA Panel on Dietetic Products, Nutrition and Allergies [NDA], 2014) and in 2017 the Food Standards Australia New Zealand (FSANZ, 2019) approved the use of *Schizochytrium*-derived DHA-rich oil for use in infant formula products (Schedule 25, Permitted novel foods). Since 2017, several GRAS notices have been approved for *Schizochytrium*-derived oils by the Food Safety Authority (FDA), indicating a growing momentum in the utilization of microbe derived oils.

In contrast to DHA, much less progress has been made in the utilization of phototrophic microalgae for developing an alternative to fish oil for EPA and other important fatty acids. High-quality EPA is found in marine microalgae across a number of classes including Bacillariophyceae (diatoms), Chlorophyceae, Chrysophyceae, Cryptophyceae, Eustigamatophyceae and Prasinophyceae (Wen and Chen, 2003).

While EPA production to date has focused on photoautotrophic growth, it is not yet economical, but emerging production and processing technologies may lead to sufficient enhancement of EPA production in microalgae to achieve market viability (Vazhappilly and Chen, 1998).

Whilst the microalgal industry does currently contribute biomass for food and nutrition, its scope is limited to a handful of algae and applications. Progress toward a greater utilization of algal products faces numerous technological challenges. Extensive screening and biological evaluation is needed to optimize production of specific metabolites and gain an understanding of how algal dietary value is affected by geographical region and growth season (Wells et al., 2017). Emerging technologies such as phenomics (section "Phenomics") are useful to survey multiple quantitative traits and to provide feedback on culture optimization. Furthermore, progress in cultivation technology (section "Algal Cultivation") and bioprocessing is needed to ensure such processes are economical viable and can compete with traditional and synthetic sources.

Further challenges include compliance with legislation, cost of production, and consumer perception. The latter will need to necessarily address issues regarding the association between microalgae and toxic cyanobacterial blooms, and their portrayal by the media. New algal strains without a documented history of safe consumption must be assessed and approved as "novel food" under EU and AUS legislative regulations (Sidari and Tofalo, 2019) or obtain a Food and Drug Administration (FDA) GRAS certification (Caporgno and Mathys, 2018), to be considered as future agricultural sources for food and nutrition. Factors such as health and nutritional benefits, taste, safety, freshness, and sustainability may persuade adoption of such products. Barriers such as lack of knowledge and familiarity must also be recognized before achieving consumer confidence. Future research will need to validate health benefits scientifically via robust clinical *in vivo* studies (i.e. random controlled trials), as well as directing efforts toward positive re-enforcement between new microalgal derived products and existing bio-products to help overcome negative consumer perceptions.

Thus, with efforts being made in emerging technologies such as phenomics and bioprocessing, microalgae is anticipated as promising future agricultural crop to cater for the increasing demands of future human and animal nutrition or other high value ingredients.

Feedstocks

While the potential of algae as next-generation of biological resources is still emerging, some well-established industrial sectors are already routinely using them as feedstock. Among these, the aquaculture industry has used algae for the production of "aquafeed" for decades (Hemaiswarya et al., 2011). The rapid growth rates and balanced nutritional value of microalgae are ideal for aquafeed, and aquaculture production facilities commonly utilize microalgae either directly as live feed or indirectly as algal meal, consisting in the residual biomass left after extraction of lipids (Borowitzka, 1999).

The recent demand for algal meal is mainly driven by the increasing consumer demand in more sustainable food products.

Currently, a major proportion of conventional agriculture and aquaculture utilizes fishmeal, the crude flour obtained after cooking, drying, and grinding fish parts, which has a high protein and PUFA content, and relatively low cost production (approx. $US1,500 per ton, source[1]). Fishmeal has been used historically as a feed for farmed seafood, poultry, and pigs, and even as a fertilizer (Hardy and Tacon, 2002). However, fishmeal is now widely recognized as unsustainable, as its production is largely based on by-catch, leading to depletion of ecosystems and the collapse of local fisheries. Therefore, more sustainable ingredients are increasingly considered as alternatives to fishmeal, including soybean meal (Alvarez et al., 2007), cottonseed meal, insects meal, legumes, and algae (Hardy and Tacon, 2002). While algal feed represents one of the most promising alternative to fishmeal, because of the low land, freshwater and carbon footprints (Kim et al., 2019), matching its production with the low cost and large scale of conventional fishmeal production (6 to 7 million metric tons per annum) has proven challenging. In this respect, future developments in large-scale culture systems, as detailed in section "Algal Cultivation," but also in new business models incorporating multiple products (**Figure 2**), could help to solve these challenges and to achieve the potential of algal meal as an emerging feedstock. For example, large feeding trials have showed that, even after extraction of a vast majority of its PUFA content (which can be sold separately as high-value nutraceutical, section "Food and Nutraceuticals"), the residual biomass of *Nannochloropsis* has great potential as aquafeed for Atlantic salmon, common carp and whiteleg shrimp (Kiron et al., 2012). This clearly demonstrates the potential of algal meal, especially when integrated in a multi-product biorefinery business model, as an emerging and viable feedstock (**Figure 2**).

The most historical and natural application of algae for production of feedstock is as direct live feed. Indeed, algae have been widely used as direct live feed during juvenile stages of abalone, crustaceans, fish species and bivalves for decades (Benemann, 1992). Among these, bivalve hatcheries require the most microalgal production in comparison to any other form of food in aquaculture, due to bivalves being obligate filter-feeders throughout their entire life (Guedes and Malcata, 2012). Consequently, mass production of microalgae can account for >30% of a bivalve hatchery operating costs, indicating it is a major financial consideration for this sector of aquaculture industry (Guedes and Malcata, 2012). Approximately 20 algal species have been identified in the 1980s as most suitable live feed for aquaculture industry (Laing, 1987). From these selected species, the genera *Chaetoceros, Tisochrysis, Pavlova* and *Tetraselmis*, are considered some of the most suitable for the rearing of bivalves, with their specific size being one of the most important attribute (Guedes and Malcata, 2012). For example, *Tisochrysis lutea* (3–7.5 µm) is utilized throughout the production of many bivalves, from larvae and juveniles through to adults, as they are primarily an appropriate size but also nutritionally valuable and robust in culture (Bendif et al., 2013). While *T. lutea* is suitable for all growth phases of bivalves, some other species such as the diatom *Chaetoceros*

muelleri are slightly larger (5–8 µm) and therefore unsuitable for the early juvenile phase (Pacheco-Vega and Sánchez-Saavedra, 2009). As a result, the efficiency of feed profiles for bivalves has largely been optimized for the aquaculture industry by mixing a range of microalgae species that are nutritionally diverse and covering a range of sizes (Heasman et al., 2000). The biological differences between microalgae species means that their photosynthetic demands (light and CO_2) and nutritional requirements are likely to differ significantly (Ihnken et al., 2011). However, a similar set of standard growth conditions are generally set for all cultures (Heasman et al., 2000). This "standardization" operated by the aquaculture industry implies that a largely un-optimized "one size fits all" culture system is employed, regardless of efficiency (Guedes and Malcata, 2012). Consequently, future biotechnological research (sections "Algal Cultivation," "Phenomics," and "Synthetic Biology") integrating (i) differences in optimal growth conditions between microalgae species, (ii) strain selection, and (iii) new cultivation technology, especially the next generation of nearly fully automated photobioreactors, are likely to increase microalgal yields to an extent that was previously unimaginable.

High-Value Products

Most of the high-value products that are currently sourced from higher plants are also naturally produced by algae, or could be produced by algae through genetic engineering and synthetic biology. Given the vast diversity of microalgae, they naturally produce an extremely wide – and largely uncharacterized – range of natural products, potentially useful for human consumption and use. Some products are already synthetized efficiently, while the yields of others can be maximized to meet industrial requirements by the integration of advanced strain and bioprocess engineering (sections "Algal Cultivation," "Phenomics," and "Synthetic Biology"). Only a minute fraction of all algal species, consisting of mostly model species, are currently profiled for their biochemical capabilities (Sasso et al., 2012). Therefore, the full potential of algae in this context can only be estimated. Besides the above-mentioned food supplements, pigments and PUFAs (section "Food and Nutraceuticals"), a substantial number of different high-value products are already being sourced from algae.

Plant biostimulants (PBs) are a heterogeneous class of compounds that include phytohormones, small molecules, and polymers, which are used to improve crop performances and protect from abiotic stresses (Drobek et al., 2019). Extracts and bioactive compounds derived from wild type microalgae species are increasingly being explored as source of PBs (Chiaiese et al., 2018). Although this application is still at early stages, it could be developed and result particularly advantageous if included in a multi-purpose algal biorefinery (**Figure 2**). As the biochemical composition of microalgae greatly varies depending on species and culture conditions, the potential of algae-based applications in this context could be leveraged by more detailed knowledge on algal biochemistry and physiology, to allow better choice of species and growth conditions to promote the production of specific PBs (Drobek et al., 2019). On the other hand, the potential of engineered algal strains to enable the

production of higher amounts or specific molecules is completely unexplored. Advanced genetics and synthetic biology techniques could soon enable the development of novel algal strains designed to produce specific metabolites, phytohormones, or peptides with PBs activity.

Microalgae can be genetically engineered to synthesize a myriad of high-value products. Terpenoids are the largest class of natural products that include countless bioactive plant secondary metabolites with applications as cosmetics, biofuels, nutraceuticals, and as life-saving pharmaceuticals in high demand (Vickers et al., 2017). As plant secondary metabolites, these compounds are typically produced in trace amounts; therefore industrial extraction *ex planta* requires very large quantities of biomass, with high economic and environmental costs. This could be averted with microalgae engineered to produce these chemicals in higher concentrations than is possible in plants (Arendt et al., 2016; Moses et al., 2017; Vavitsas et al., 2018; Lauersen, 2019). Cyanobacteria, for example, have been widely used for heterologous plant-derived terpenoid engineering, extensively reviewed in Chaves and Melis (2018) and Lin and Pakrasi (2019), and proof-of-concept works have unveiled the potential of engineered eukaryotic microalgae such as *C. reinhardtii* in producing high value terpenoids such as the food flavoring and aromas sesquiterpenoid patchulol and (E)-α-bisabolene (Lauersen et al., 2016; Wichmann et al., 2018) and diterpenoids such as casbene, taxadiene, and 13R(+)manoylnyl oxide (Lauersen et al., 2018), as well as lambdane diterpenoids (Papaefthimiou et al., 2019), which are relevant precursors of plant-derived therapeutic and cosmetic products. Similarly, the diatom *P. tricornutum* is currently being explored for similar applications and demonstrated its potential in producing triterpenoid lupeol and traces of betulin, precursors of the topoisomerase inhibitor betulinic acid (D'Adamo et al., 2019), commonly used in anticancer and antiviral pharmaceutical preparations and naturally produced in trace amounts from the bark of plant species, such as the white birch tree (Pisha et al., 1995).

Plant monoterpenoids are particularly challenging compounds to produce in conventional microbial hosts such as *S. cerevisiae* and *E. coli* because these organisms do not naturally accumulate pools of the precursor geranyl diphosphate (GPP) (Vickers et al., 2017). It has been recently demonstrated that *P. tricornutum* naturally accumulates cytosolic pools of GPP and that these can be efficiently converted into the monoterpenoid geraniol (0.309 mg/L), through the episomal expression of a heterologous plant geraniol synthase enzyme (Fabris et al., 2020). Geraniol has several commercial applications as component of essential oils, flavouring agent and insect repellent, and is the key precursor of the monoterpenoid indole alkaloids (MIAs), a diverse group of bioactive plant metabolites that include the anticancer agents vinblastine and vincristine (Miettinen et al., 2014). This is a relevant demonstration that diatoms might harbour an important intrinsic advantage over conventional terpenoid production hosts for the synthesis of this challenging class of compounds, and that extrachromosomal episomes are suitable for metabolic engineering applications in diatoms, which is seminal for more complex, engineering approaches.

Exciting progress in algal genetics and synthetic biology, including key technologies for the assembly and expression of multi-gene constructs and tools for targeted gene editing (section "Synthetic Biology"), as well as advances in metabolic systems biology, will rapidly enable the expression of more complex metabolic pathways (Slattery et al., 2018) and increase understanding of the resulting interactions with endogenous algal metabolism, resulting in tailored engineering efforts that will go beyond simple proofs of concept and result in industrially relevant product yields.

Other products that could be sourced from engineered algal strains include industrial recombinant enzymes (Rasala et al., 2012; Laursen et al., 2013) and protein-based therapeutics (Rasala et al., 2010; Gimpel et al., 2015). These recombinant protein drugs are generally produced in microbes such as *E. coli*, yeasts or mammalian cell lines such as Chinese Hamster Ovarian (CHO) cells. The latter, in particular, are associated with extremely high cost of production, mostly due to complex growth media composition (US$10 – 500/L) (Xu et al., 2017). Therefore, in the last two decades, an increasing research effort has been put into developing robust, alternative production hosts. Plant and algae-based expression systems are envisioned as a valid, low-cost solution for producing therapeutics in countries and areas that lack resources for costly mammalian-based fermentation systems (Taunt et al., 2018), with the advantage of being immune to most pathogens and contaminations that affect animal hosts (Specht and Mayfield, 2014). In this search, it has been demonstrated the suitability of *C. reinhardtii* to produce – predominantly in the chloroplast – functional recombinant therapeutics, including a fully assembled human antibody, immunoglobulin G (IgG) (Tran et al., 2009), vaccine subunits (Gregory et al., 2013), vaccine antigens (Demurtas et al., 2013), immuno-conjugated cytotoxins for cancer targeted treatments (Tran et al., 2013), and single domain antibodies (VHH) (Barrera et al., 2015). Diatoms such as *P. tricornutum* have also been used to successfully and efficiently produce and secrete fully assembled antibodies (Hempel et al., 2011b; Hempel et al., 2017), while the silicified *Thalassiosira pseudonana* has been engineered for targeted drug delivery *in vivo*, by displaying a recombinant IgG binding domain on the silica frustules, turning the whole cell into a drug delivery vector effective on tumor models (Delalat et al., 2015). The production of edible vaccines is another developing field where algae-based expression systems are finding a relevant niche (Specht and Mayfield, 2014), with particular relevance to the poultry and aquaculture industry. However, challenges will need to be addressed to make algae the preferred production hosts for therapeutics. In addition to the cultivation challenges already mentioned, the production of recombinant therapeutics in algae is currently hindered by overall low expression levels, and it is expected that developments in algal genetics and synthetic biology (section "Synthetic Biology") will enable more competitive yields. Strategies involving innovative genetic design – for example the insertion of intronic sequences in the transgene of interest – could be used to significantly improve the expression of recombinant proteins (Baier et al., 2018). Also,

to be suitable for therapeutics production, algal production hosts need to exhibit the correct post-translational modifications, such as protein glycosylation, to avoid adverse immune reactions in inoculated animals. Little is known about the *N*-glycosylation properties of microalgae, but large-scale profiling of glycosylation properties of diverse non-model species and genetic engineering may possibly offer possibilities for algae to become a preferred production platform for glycoproteins (Mathieu-Rivet et al., 2014).

From these convincing proof-of-concepts examples, the (enhanced) production of endogenous and heterologous high-value products in microalgae will enormously benefit from the technological developments reviewed in section "Technology Development." More complex synthetic biology approaches, in combination with detailed knowledge on novel or engineered strains from high-throughput phenomics approaches, and advancements in cultivation technology (**Figure 1**), will address the main bottlenecks of low yields and upscaling, and open the doors to the cost-effective production of a much wider diversity of bioproducts. This will alleviate the environmental impact imposed by current practices involving inefficient bioproduct sourcing from plants or from other high-cost and less environmentally friendly production methods.

Biopolymers, Bioplastics, and Bulk Chemicals

The demand of plastic and plastic-based products have grown significantly in last few decades, which has placed a major strain on the remaining petrochemical resources of our planet. The increasing production of these petrochemical-based plastics has also generated concern regarding plastic pollution worldwide, mostly in marine ecosystems due to their persistence in environment as non-biodegradable materials (Tetu et al., 2019). Therefore, alternatives to petrochemical-based plastics sources are in high demand, as they would make plastic production sustainable while mitigating the issue of plastic pollution. Algae have the potential to be an economically viable feedstock for bioplastics production, as the biomass can be sold at US$ 970/tonne, which is within the current standard range for other sources of bioplastics (US$ 800 – 1200/tonne (Beckstrom et al., 2020).

Microalgal biomass components such as starch, carbohydrates, and lipids can be converted into plastics (Noreen et al., 2016). There are currently three main approaches to produce bioplastics from microalgae, including: (i) direct use of microalgae as bioplastics, (ii) blending of microalgae with existing petroleum-based plastics or bioplastics, and (iii) genetic engineering of microalgae to produce bioplastic polymer precursors. In the first approach, Zeller et al. (2013) have reported production of bioplastics and thermoplastic blends directly from *S. platensis* and *C. vulgaris*, while Wang et al. (2016) described the preparation of thermoplastics by blending a heterogeneous population of planktonic algae. However, the most common approach to making microalgae-based bioplastics is to blend the biomass with existing petrochemical-based plastics, such

as polyethylene, polypropylene, polyvinyl chloride. Shi et al. (2012) described the processing of microalgae-corn starch-based thermoplastics using *Nannochloropsis* and *Spirulina*, and further blending with polyethylene and polypropylene. *Chlorella* sp. biomass was blended with polyethylene and polypropylene and was found to possess good thermoplastics processability because of the presence of natural cellulosic type materials (Zhang et al., 2000a). The properties and processing of PVC-*Chlorella* composite has also been reported (Zhang et al., 2000b).

With the increasing demand of bioplastics in the market, considerable research effort has been directed in investigating the blending of algal biomass with other bio-derived plastics components. A recent study has reported of addition of green, brown, and red algal biomass to polylactic acid plastics (Bulota and Budtova, 2015) with no pre-treatment other than drying and sieving. Polyhydroxyalkanoates (PHAs), one of the widely studied biodegradable polyesters with high mechanical strength and melting point, is naturally produced in certain bacteria, including some cyanobacteria (Sudesh et al., 2000). Microbial production of PHAs generally occurs under stressful environmental conditions (Bassas et al., 2008; Balaji et al., 2013). PHA is generally extracted by three subsequent steps of disrupting the cells (by chemical, physical or biological treatment), recovery of PHAs, and purification (Fiorese et al., 2009). However, with the continuous increase in interest in PHAs production, metabolic engineering, and synthetic biology (section "Synthetic Biology") could enable the heterologous synthesis of PHA precursors in eukaryotic microalgae as demonstrated in diatoms (Hempel et al., 2011a).

Biomass-derived chemicals, such as 5-hydroxymethylfurfural (5-HMF), levulinic acid, furfurals, sugar alcohols, lactic acid, succinic acid, and phenols, are considered platform chemicals. These platform chemicals are used for the production of a variety of important chemicals on an industrial scale (Kohli et al., 2019). Bio-based bulk chemicals possess a clear substitution potential for fossil oil-based bulk chemicals. However, current biomass feedstocks for industrial use are typically derived from plant material, posing challenges such as destruction of rainforests, competitive food consumption, and other adverse environmental impacts. Microalgae, with its superior areal productivity to traditional agricultural crops and high concentration of lipid, carbohydrate, and proteins, have appeared as an alternative and attractive candidate for the production of bulk chemicals, including bio-based platform chemicals and bio-based solvents (Wijffels et al., 2010). Catalytic valorization is an emerging field that can be applied to the production of value added chemicals from microalgae. Even though the technology readiness for commercialization is still a challenge, the field is active with several research groups working on algae and catalytic systems for the conversion of algal biomass to value added platform chemicals. For example, *Chlorococcum* sp. was reported to be converted into 1,2-propanediol (1,2-PDO) and ethylene glycol (EG) in water over nickel-based catalysts (Miao et al., 2015), while the hydrolysis of *Scenedesmus* sp. over the Sn-Beta catalyst was used to produce lactate (Zan et al., 2018). This was achieved via formic acid induced controlled release hydrolysis, with an

achieved yield of 83%. Another recent study demonstrated the conversion of algal polysaccharides from *Phorphyridium cruentum* and *C. vulgaris* to monosaccharides, HMF, and furfural in the neat deep eutectic solvent (DES) or in the biphasic system ChCl/oxalic acid/methyl isobutyl ketone (Bodachivskyi et al., 2019).

Microalgal biomass that has the lipids already extracted is good source for carbohydrates. The reported yields are up to 80% of the cell mass and hence, could be useful upon hydrolysis to generate fermentable sugars. A recent study has reported hydrolyses of lipid extracted *C. vulgaris* biomass using solid acid catalysts to obtain monosaccharides such as glucose, galactose, xylose, rhamnose, mannose, and 2,3 butanediol (Seon et al., 2019). These monosaccharides can be used for microbial fermentation to produce many useful products, such as lactic acid, hydrogen gas, and ethanol. 2,3 butanediol is a value-added chemical with great potential for the industrial production of synthetic rubber, plastic, and biosolvent (Soo-Jung et al., 2017; Seon et al., 2019). In another study, microalgal hydrolysate from *C. vulgaris was* converted into ethanol via continuous immobilized yeast fermentation at a yield of 89% (Kim et al., 2014).

Several challenges will need to be addressed in terms of low product yield and relatively high costs of such biochemical conversion processes. However, the integration of these application in a multi-product biorefinery approach (**Figure 2**), could improve the overall economic feasibility of bioplastic and bulk chemical production from microalgal biomass.

Algal Biodegradation of Emerging Contaminants

Emerging contaminants (EC) are primarily synthetic organic chemicals, such as pharmaceuticals, herbicides, pesticides, and flame retardants, whose presence in the environments are of concern due to their potential risks to ecosystems and human health, at environmentally relevant concentrations (Petrie et al., 2015; Tran et al., 2018; Sutherland and Ralph, 2019). There is increasing concern over the presence of ECs in agricultural land- and water-scapes. With climate change and expanding populations, accumulating ECs due to agricultural intensification and increased water reuse could lead to unpredictable long-term consequences for humans and the environment (Martinez-Piernas et al., 2018). While direct application can be managed through improved on-farm best management practices, indirect application is reliant on improvements in wastewater treatment that would reduce, transform, or eliminate ECs.

Wastewater treatment using microalgae for nutrient removal is a well-established technology that has lower capital and operational costs, and is more efficient than traditional wastewater treatment systems (Benemann, 2008; Craggs et al., 2012). However, there have been few studies to date on the use of microalgae for bioremediation of ECs despite their potential for detoxifying organic and inorganic pollutants. Coupling of nutrient and EC removal by microalgae has the potential to provide more cost-effective and efficient wastewater treatment as

well as meeting both environmental and human health protection goals (Sutherland and Ralph, 2019).

While still in its infancy, microalgal biodegradation provides one of the most promising technologies to transform, neutralize, or eliminate ECs from agricultural runoff. Unlike other remediation techniques, such as activated carbon adsorption filters, which simply concentrates the EC and removes it from one environment to another environment, biodegradation involves the transformation of complex compounds into simpler breakdown molecules through catalytic metabolic degradation (Sutherland and Ralph, 2019). Microalgal degradation of ECs can occur via two main mechanisms. The first mechanism involves direct metabolic degradation of the EC by the microalga. In this case, the microalga employs mixotrophic growth strategies and the EC serves as the carbon source or electron donor/acceptor (Tiwari et al., 2017). The second mechanism involves indirect, or co-metabolism, where the EC is degraded by enzymes that are catalyzing other substrates present (Tiwari et al., 2017). Microalgae possess a large number of enzymes that play a role in cellular protection through the deactivation and/or degradation of a range of organic compounds that induce cellular stress in microalgae (Wang et al., 2019). Microalgal degradation of ECs relies on a complex enzymatic process involving a number of enzymes, including: superoxide dismutase, catalase, glutamyl-tRNA reductase, malate/pyruvate dehydrogenase, mono(di)oxygenase, pyrophosphatase, carboxylase/decarboxylase, dehydratase, alkaline and acid phosphatase, transferase, and hydrolases (Elbaz et al., 2010; Xiong et al., 2018; Wang et al., 2019). Several of these enzymes, including superoxide dismutase and catalase, have shown increased activity in several freshwater microalgal species, when the cells were exposed to the veterinary antibiotics Florfenicol and Ofloxacin (Wang et al., 2019).

In one bioremediation study, the green algae *Scenedesmus obliquus* and *Chlorella pyrenoidosa* were found to enzymatically degrade progesterone and norgestrel by reduction (hydrogenation), hydroxylation, oxidation (dehydrogenation) and side-chain breakdown (Peng et al., 2014). In another study, co-metabolic removal of the antibiotic ciprofloxacin by the green alga *Chlamydomonas mexicana* was observed, but the enzymatic mechanisms involved in its metabolism were not identified (Xiong et al., 2017). Due to the complexity of enzymatic biodegradation processes, simply screening microalgal strains for EC biodegradation activity remains the most viable strategy for developing new bioremediation strains (Sutherland and Ralph, 2019).

One of the challenges with screening microalgae for EC biodegradation is the large number of both ECs and microalgal species. Currently, there are approximately 200 known ECs in the environment, while there are thousands of recognized algal species (Pradhan and Rai, 2001; Guiry, 2012). Therefore, there is a need for the development of cost effective high through-put screening methods that allow for rapid screening of a wide range of microalgal species against a wide range of ECs. A microalgal phenomics facility (section "Phenomics," **Figure 1b**) would provide the necessary cost-effective and

efficient high through-put screening to help rapidly develop microalgal biodegradation technology.

Another challenge with screening microalgae for EC biodegradation is that the enzymes responsible for degrading the EC may not be active at the time of screening (Sutherland and Ralph, 2019). This is due to both the production and maintenance of these complex enzymes being metabolically expensive, which comes at the cost of growth and reproduction of the cell (Sutherland and Ralph, 2019). For example, both the cellular energy budget and growth rates were significantly reduced in the microalga *Raphidocelis subcapitata*, following the induction of superoxide dismutase production by the cells exposed to four different antibiotics (Aderemi et al., 2018). For some microalgae, pre-acclimation to sub-toxic concentrations of the EC may be required to initiate enzyme production in order to screen for biodegradation potential (Sutherland and Ralph, 2019). For example, microalgal biodegradation of several different antibiotics was enhanced following pre-exposure of the microalgal strain to low levels of the antibiotic due to increased production of antioxidants, including xanthophylls, by the cells (Chen et al., 2015; Xiong et al., 2017). Biodegradation may also lead to intermediary products that could be similarly, or more toxic, than the parent compound. Identification of the breakdown products with specific assays, coupled with toxicological screening is an important step that needs to be included in microalgal biodegradation assessments.

For microalgal species with demonstrated biodegradation capability, the induction of elevated Phase I and Phase II enzyme production can further enhance the EC degradation process, both improving its efficiency and effectiveness. This can be induced through genetic means, such as synthetic biology, targeted gene editing, or genetic engineering. For example, Zhang et al. (2018) used random mutagenesis and site-directed mutagenesis to increase the production of the degrading enzyme, laccase, by 31- to 37-fold in the white-rot fungus *Cerrena unicolor* BBP6. Similar approaches could be used to increase the biocatalytic activities of microalgal laccases.

Synthetic biology approaches (section "Synthetic Biology," **Figure 1a**), can be used to engineer microalgae and overexpress entire artificial degrading pathways that include enzymes, such as fungal laccases, peroxidases, cellulases, and ligninases, to further increase the potential of algal bioremediation. These pathways can be either expressed in the host in the same configuration as in the source organism, or even in new-to-nature combinations, picking enzymes from multiple organisms and assembling new degradation pathways, both by rational design, and by random/combinatorial assembly and screening (Tay et al., 2017). While there are currently limited studies on genetic engineering of microalgae for bioremediation purposes, Chiaiese et al. (2011) successfully demonstrated fungal laccase POX A1b expression in the green alga, *Chlorella emersonii*, which enhanced microalgal biodegradation of phenols by up to about 40%. However, while genetically engineering microalgae for enhanced biodegradation appears promising, the potential environmental risks intrinsic to the use of genetically modified organisms (GMO) that would limits their application in outdoors

settings need to be evaluated (Szyjka et al., 2017). In addition to this, for many countries, the legislation around the limited use, or the total ban of, GMOs means that transgenic microalgae for ECs biodegradation would not be a viable option, at present.

While microalgae have the demonstrated ability to biodegrade ECs associated with agricultural practises, further research is needed to exploit microalgal biodegradation, through enhanced enzyme expression and optimized growth conditions. When coupled with nutrient removal, such as HRAPs, microalgal treatment of EC can be a cost-effective viable option for the reduction of contaminant pollution in waterways (Sutherland and Ralph, 2019).

CONCLUSION

Agriculture is one of the most ancient human practices and it has always been essential to our civilization. Agriculture and human society have co-evolved, reciprocally influencing each other. Over millennia humans isolated, bred, and generated new species to satisfy needs that have been steadily increasing in size and diversity. In modern times, agriculture technology has seen impressive improvements in yield, efficiency, and product differentiation thanks to developments in cultivation technology, genetics, and phenomics. Although algae-derived applications have been present in human history, the push to develop these organisms as industrial resources is a very recent objective. Compared to conventional agriculture crops, algae-based practices are an extremely young application field, and all current industrial algal strains are relatively uncharacterized. However, decades of foundational research on algal biochemistry and physiology (not reviewed here), may be leveraged to expedite the use of algae in biotechnology (Hildebrand et al., 2013). Efforts to progress the understanding of diverse algal traits has recently been bolstered by the advent of genome sequencing projects and functional genetic tools, revealing novel aspects of algal metabolism relevant to industrial applications (Moellering and Benning, 2010; Allen et al., 2011; Fabris et al., 2012; Kirst et al., 2012; Radakovits et al., 2012; Fabris et al., 2014; Abbriano et al., 2018; Luo et al., 2018; Pollier et al., 2019; Smith et al., 2019). Moreover, it is expected that knowledge on algal traits will be increasingly generated by the implementation of advanced synthetic and molecular biology approaches combined with phenomics. Presently, however, the relatively few algal species employed in commercial applications largely consist of natural isolates with minimal selections, breeding or genetic engineering (if any) to better perform in industrial settings or for improved yields. Despite this, as illustrated by the achievements highlighted in this review, algae already find applications in many industrial fields and sectors, often with the clear potential of replacing more energy, cost, and environmentally intensive solutions. Evaluating the current progress and achievement of algal biotechnology and industry from this perspective is at the same time both impressive and encouraging, and this needs to be kept into account when drawing the trajectory of future developments of this field. The emerging technologies that we described will drastically accelerate the process of industrialization of algae, providing

knowledge and tools to deliver highly productive, algae-based solutions to a diversity of societal needs. This includes deeper understanding of algal biology, genetics, and biochemical capabilities, which will drive the optimization of both the organisms and the environment in which it is cultivated. This will allow in the near future the move toward *ad hoc*, highly productive strains, either as novel natural isolates or genetically engineered strains, and efficient cultivation systems with minimal environmental impact. We envision that high-tech algae-based solutions will find applications in almost every industrial sector, including ones essential to meeting the increasing needs of human society, such as food, pharmaceutical and bulk chemicals manufacture, while ensuring minimal environmental impact and lower production costs. The development of highly efficient algal biorefineries (’T Lam et al., 2018; **Figure 2**) will allow co-sourcing different products, minimizing waste and maximizing the productivity, improving the economics of processes otherwise low-efficient. As such, we anticipate that the progress of algae biotechnology will have a disruptive effect to the current industrial landscape, and will prompt the emergence of a scalable, sustainable, and efficient algae-based bio-economy, which will be key in overcoming challenges and limitations that conventional agriculture will face in the years ahead.

AUTHOR CONTRIBUTIONS

FUNDING

REFERENCES

Abatemarco, J., Hill, A., and Alper, H. S. (2013). Expanding the metabolic engineering toolbox with directed evolution. *Biotechnol. J.* 8, 1397–1410. doi: 10.1002/biot.201300021

Abbriano, R., Vardar, N., Yee, D., and Hildebrand, M. (2018). Manipulation of a glycolytic regulator alters growth and carbon partitioning in the marine diatom *Thalassiosira pseudonana. Algal Res.* 32, 250–258. doi: 10.1016/j.algal.2018.03.018

Aderemi, A. O., Novais, S. C., Lemos, M. F. L., Alves, L. M., Hunter, C., and Pahl, O. (2018). Oxidative stress responses and cellular energy allocation changes in microalgae following exposure to widely used human antibiotics. *Aquat. Toxicol.* 203, 130–139. doi: 10.1016/j.aquatox.2018.08.008

Ahmed, F., Zhou, W., and Schenk, P. M. (2015). Pavlova lutheri is a high-level producer of phytosterols. *Algal Res.* 10, 210–217. doi: 10.1016/j.algal.2015.05.013

Ajjawi, I., Verruto, J., Aqui, M., Soriaga, L. B., Coppersmith, J., Kwok, K., et al. (2017). Lipid production in *Nannochloropsis gaditana* is doubled by decreasing expression of a single transcriptional regulator. *Nat. Biotechnol.* 35, 647–652. doi: 10.1038/nbt.3865

Allen, A. E., Dupont, C. L., Obornik, M., Horak, A., Nunes-Nesi, A., Mccrow, J. P., et al. (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473, 203–207. doi: 10.1038/nature10074

Alvarez, J., Hernández-Llamas, A., Galindo, J., Fraga, I., García, T., and Colmenares, H. (2007). Substitution of fishmeal with soybean meal in practical diets for juvenile white shrimp *Litopenaeus schmitti* (Pérez-Farfante & Kensley 1997). *Aquac. Res.* 38, 689–695. doi: 10.1111/j.1365-2109.2007.01654.x

Arendt, P., Pollier, J., Callewaert, N., and Goossens, A. (2016). Synthetic biology for production of natural and new-to-nature terpenoids in photosynthetic organisms. *Plant J.* 87, 16–37. doi: 10.1111/tpj.13138

Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: a survey. *Comput. Netw.* 54, 2787–2805.

Baek, K., Kim, D. H., Jeong, J., Sim, S. J., Melis, A., Kim, J.-S., et al. (2016). DNA-free two-gene knockout in *Chlamydomonas reinhardtii* via CRISPR-Cas9 ribonucleoproteins. *Sci. Rep.* 6:30620. doi: 10.1038/srep30620

Baier, T., Wichmann, J., Kruse, O., and Lauersen, K. J. (2018). Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga *Chlamydomonas reinhardtii. Nucleic Acids Res.* 46, 6909–6919. doi: 10.1093/nar/gky532

Balaji, S., Gopi, K., and Muthuvelan, B. (2013). A review on production of poly β hydroxybutyrates from cyanobacteria for the production of bio plastics. *Algal Res.* 2, 278–285. doi: 10.1016/j.algal.2013.03.002

Bar-Even, A., Noor, E., Lewis, N. E., and Milo, R. (2010). Design and analysis of synthetic carbon fixation pathways. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8889–8894. doi: 10.1073/pnas.0907176107

Barkia, I., Saari, N., and Manning, S. R. (2019). Microalgae for high-value products towards human health and nutrition. *Mar. Drugs* 17:304. doi: 10.3390/md17050304

Barrera, D. J., Rosenberg, J. N., Chiu, J. G., Chang, Y.-N., Debatis, M., Ngoi, S.-M., et al. (2015). Algal chloroplast produced camelid VHH antitoxins are capable of neutralizing botulinum neurotoxin. *Plant Biotechnol. J.* 13, 117–124. doi: 10.1111/pbi.12244

Bassas, M., Diaz, J., Rodriguez, E., Espuny, M. J., Prieto, M. J., and Manresa, A. (2008). Microscopic examination in vivo and in vitro of natural and cross-linked polyunsaturated mclPHA. *Appl. Microbiol. Biotechnol.* 78, 587–596. doi: 10.1007/s00253-008-1350-4

Batta, A. K., Xu, G., Honda, A., Miyazaki, T., and Salen, G. (2006). Stigmasterol reduces plasma cholesterol levels and inhibits hepatic synthesis and intestinal absorption in the rat. *Metabolism* 55, 292–299. doi: 10.1016/j.metabol.2005.08.024

Becker, E. W. (1994). *Microalgae: Biotechnology and Microbiology*. New York, NY: Cambridge University Press.

Beckstrom, B. D., Wilson, M. H., Crocker, M., and Quinn, J. C. (2020). Bioplastic feedstock production from microalgae with fuel co-products: A techno-economic and life cycle impact assessment. *Algal Res.* 46:101769. doi: 10.1016/j.algal.2019.101769

Begum, H., Yusoff, F. M. D., Banerjee, S., Khatoon, H., and Shariff, M. (2016). Availability and utilization of pigments from microalgae. *Crit. Rev. Food Sci. Nutr.* 56, 2209–2222. doi: 10.1080/10408398.2013.764841

Bendif, E. M., Probert, I., Schroeder, D. C., and De Vargas, C. (2013). On the description of *Tisochrysis lutea* gen. nov. sp. nov. and *Isochrysis nuda* sp. nov. in the Isochrysidales, and the transfer of *Dicrateria* to the Prymnesiales (Haptophyta). *J. Appl. Phycol.* 25, 1763–1776. doi: 10.1007/s10811-013-0037-0

Benedetti, M., Vecchi, V., Barera, S., and Dall'osto, L. (2018). Biomass from microalgae: the potential of domestication towards sustainable biofactories. *Microb. Cell Fact.* 17:173. doi: 10.1186/s12934-018-1019-3

Benemann, J. R. (1992). Microalgae aquaculture feeds. *J. Appl. Phycol.* 4, 233–245. doi: 10.1007/bf02161209

Benemann, J. R. (2008). *Opportunities and Challenges in Algae Biofuels Production*. Walnut Creek, CA: Benemann Associates.

Berla, B., Saha, R., Immethun, C., Maranas, C., Moon, T. S., and Pakrasi, H. (2013). Synthetic biology of cyanobacteria: unique challenges and opportunities. *Front. Microbiol.* 4:246. doi: 10.3389/fmicb.2013.00246

Bhola, V., Swalaha, F., Ranjith Kumar, R., Singh, M., and Bux, F. (2014). Overview of the potential of microalgae for CO2 sequestration. *Int. J. Environ. Sci. Technol.* 11, 2103–2118.

Bilad, M. R., Arafat, H. A., and Vankelecom, I. F. J. (2014). Membrane technology in microalgae cultivation and harvesting: A review. *Biotechnol. Adv.* 32, 1283–1300. doi: 10.1016/j.biotechadv.2014.07.008

Billen, G., Lassaletta, L., and Garnier, J. (2014). A biogeochemical view of the global agro-food system: nitrogen flows associated with protein production, consumption and trade. *Glob. Food Secur.* 2, 209–219. doi: 10.1016/j.gfs.2014.08.003

Bodachivskyi, I., Kuzhiumparambil, U., and Williams, D. B. G. (2019). The role of the molecular formula of ZnCl2·nH2O on its catalyst activity: a systematic study of zinc chloride hydrates in the catalytic valorisation of cellulosic biomass. *Catal. Sci. Technol.* 9, 4693–4701. doi: 10.1039/c9cy00846b

Borowitzka, M. A. (1999). Commercial production of microalgae: ponds, tanks, tubes and fermenters. *J. Biotechnol.* 70, 313–321. doi: 10.1016/s0168-1656(99)00083-8

Borowitzka, M. A., and Vonshak, A. (2017). Scaling up microalgal cultures to commercial scale. *Eur. J Phycol.* 52, 407–418. doi: 10.1186/s13068-016-0541-y

Brennan, L., and Owende, P. (2010). Biofuels from microalgae-A review of technologies for production, processing, and extractions of biofuels and co-products. *Renew. Sustain. Energy Rev.* 14, 557–577. doi: 10.1016/j.rser.2009.10.009

Bulota, M., and Budtova, T. (2015). PLA/algae composites: morphology and mechanical properties. *Compos. Part A Appl. Sci. Manuf.* 73, 109–115. doi: 10.1016/j.msec.2015.02.020

Cagnon, C., Mirabella, B., Nguyen, H. M., Beyly-Adriano, A., Bouvet, S., Cuiné, S., et al. (2013). Development of a forward genetic screen to isolate oil mutants in the green microalga *Chlamydomonas reinhardtii*. *Biotechnol. Biofuels* 6, 178–178. doi: 10.1186/1754-6834-6-178

Campos, M. L., Yoshida, Y., Major, I. T., De Oliveira Ferreira, D., Weraduwage, S. M., Froehlich, J. E., et al. (2016). Rewiring of jasmonate and phytochrome B signalling uncouples plant growth-defense tradeoffs. *Nat. Commun.* 7, 12570–12570. doi: 10.1038/ncomms12570

Caporgno, M. P., and Mathys, A. (2018). Trends in microalgae incorporation into innovative food products with potential health benefits. *Front. Nutr.* 5:58. doi: 10.3389/fnut.2018.00058

Chang, K. J. L., Nichols, C. M., Blackburn, S. I., Dunstan, G. A., Koutoulis, A., and Nichols, P. D. (2014). Comparison of *Thraustochytrids aurantiochytrium* sp., *Schizochytrium* sp., *Thraustochytrium* sp., and Ulkenia sp. for production of biodiesel, long-chain omega-3 oils, and exopolysaccharide. *Mar. Biotechnol.* 16, 396–411. doi: 10.1007/s10126-014-9560-5

Chauton, M. S., Reitan, K. I., Norsker, N. H., Tveterås, R., and Kleivdal, H. T. (2015). A techno-economic analysis of industrial production of marine microalgae as a source of EPA and DHA-rich raw material for aquafeed: research challenges and possibilities. *Aquaculture* 436, 95–103. doi: 10.1016/j.aquaculture.2014.10.038

Chaves, J. E., and Melis, A. (2018). Engineering isoprene synthesis in cyanobacteria. *FEBS Lett.* 592, 2059–2069. doi: 10.1002/1873-3468.13052

Chen, J., Zheng, F., and Guo, R. (2015). Algal feedback and removal efficiency in a sequencing batch reactor algae process (SBAR) to treat the antibiotic cefradine. *PLoS One* 10:e0133273. doi: 10.1371/journal.pone.0133273

Chew, K. W., Yap, J. Y., Show, P. L., Suan, N. H., Juan, J. C., Ling, T. C., et al. (2017). Microalgae biorefinery: High value products perspectives. *Bioresour. Technol.* 229, 53–62. doi: 10.1016/j.biortech.2017.01.006

Chiaiese, P., Corrado, G., Colla, G., Kyriacou, M. C., and Rouphael, Y. (2018). Renewable sources of plant biostimulation: microalgae as a sustainable means to improve crop performance. *Front. Plant Sci.* 9:1782. doi: 10.3389/fpls.2018.01782

Chiaiese, P., Palomba, F., Tatino, F., Lanzillo, C., Pinto, G., Pollio, A., et al. (2011). Engineered tobacco and microalgae secreting the fungal laccase POXA1b reduce phenol content in olive oil mill wastewater. *Enzyme Microb. Technol.* 49, 540–546. doi: 10.1016/j.enzmictec.2011.06.002

Christenson, L., and Sims, R. (2011). Production and harvesting of microalgae for wastewater treatment, biofuels, and bioproducts. *Biotechnol. Adv.* 29, 686–702. doi: 10.1016/j.biotechadv.2011.05.015

Commault, A. S., Fabris, M., Kuzhiumparambil, U., Adriaans, J., Pernice, M., and Ralph, P. J. (2019). Methyl jasmonate treatment affects the regulation of the 2-C-methyl-D-erythritol 4-phosphate pathway and early steps of the triterpenoid biosynthesis in *Chlamydomonas reinhardtii*. *Algal Res.* 39:101462. doi: 10.1016/j.algal.2019.101462

Craggs, R., Sutherland, D., and Campbell, H. (2012). Hectare-scale demonstration of high rate algal ponds for enhanced wastewater treatment and biofuel production. *J. Appl. Phycol.* 24, 329–337. doi: 10.1007/s10811-012-9810-8

Crozet, P., Navarro, F. J., Willmund, F., Mehrshahi, P., Bakowski, K., Lauersen, K. J., et al. (2018). Birth of a photosynthetic chassis: a moclo toolkit enabling synthetic biology in the microalga *Chlamydomonas reinhardtii*. *ACS Synth. Biol.* 7, 2074–2086. doi: 10.1021/acssynbio.8b00251

Cruz, J. A., Savage, L. J., Zegarac, R., Hall, C. C., Satoh-Cruz, M., Davis, G. A., et al. (2016). Dynamic Environmental photosynthetic imaging reveals emergent phenotypes. *Cell Syst.* 2, 365–377. doi: 10.1016/j.cels.2016.06.001

D'Adamo, S., Schiano Di Visconte, G., Lowe, G., Szaub-Newton, J., Beacham, T., Landels, A., et al. (2019). Engineering the unicellular alga *Phaeodactylum tricornutum* for high-value plant triterpenoid production. *Plant Biotechnol. J.* 17, 75–87. doi: 10.1111/pbi.12948

Dahman, Y., Syed, K., Begum, S., Roy, P., and Mohtasebi, B. (2019). "Biofules: their characteristics and analysis," in *Biomass, Biopolymer-Based Materials, and Bioenery*, eds D. Verma, E. Fortunati, S. Jain, and X. Zhange (United Kingdom: Woodhead Pubishing).

Das, S., Das, S., and Ghangrekar, M. M. (2019). Quorum-sensing mediated signals: a promising multi-functional modulators for separately enhancing algal yield and power generation in microbial fuel cell. *Bioresour. Technol.* 294:122138. doi: 10.1016/j.biortech.2019.122138

De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C., and Falciatore, A. (2009). Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res.* 37:e96. doi: 10.1093/nar/gkp448

De Vree, J. H., Bosma, R., Janssen, M., Barbosa, M. J., and Wijffels, R. H. (2015). Comparison of four outdoor pilot-scale photobioreactors. *Biotechnol. Biofuels* 8, 1–12. doi: 10.1186/s13068-015-0400-2

Delalat, B., Sheppard, V. C., Rasi Ghaemi, S., Rao, S., Prestidge, C. A., Mcphee, G., et al. (2015). Targeted drug delivery using genetically engineered diatom biosilica. *Nat. Commun.* 6:8791. doi: 10.1038/ncomms9791

Demirbas, A. (2009). Progress and recent trends in biodiesel fuels. *Energy Convers. Manag.* 50, 14–34. doi: 10.1016/j.enconman.2008.09.001

Demurtas, O. C., Massa, S., Ferrante, P., Venuti, A., Franconi, R., and Giuliano, G. (2013). A *Chlamydomonas*-derived human papillomavirus 16 e7 vaccine induces specific tumor protection. *PLoS One* 8:e61473. doi: 10.1371/journal.pone.0061473

Donald, C. M. (1968). The breeding of crop ideotypes. *Euphytica* 17, 385–403. doi: 10.1007/bf00056241

Drobek, M., Frąc, M., and Cybulska, J. (2019). Plant biostimulants: importance of the quality and yield of horticultural crops and the improvement of plant tolerance to abiotic stress—a review. *Agronomy* 9:335. doi: 10.3390/agronomy9060335

Duanmu, D., Miller, A. R., Horken, K. M., Weeks, D. P., and Spalding, M. H. (2009). Knockdown of limiting-CO2-induced gene HLA3 decreases HCO3- transport and photosynthetic Ci affinity in *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5990–5995. doi: 10.1073/pnas.0812885106

EFSA Panel on Dietetic Products, Nutrition and Allergies (NDA) (2014). Scientific opinion on the extension of use for DHA and EPA-rich algal oil from *Schizochytrium* sp. as a novel food ingredient. *EFSA J.* 12:3843.

Elbaz, A., Wei, Y. Y., Meng, Q., Zheng, Q., and Yang, Z. M. (2010). Mercury-induced oxidative stress and impact on antioxidant enzymes in *Chlamydomonas reinhardtii*. *Ecotoxicology* 19, 1285–1293. doi: 10.1007/s10646-010-0514-z

Erb, T. J., and Zarzycki, J. (2016). Biochemical and synthetic biology approaches to improve photosynthetic CO2-fixation. *Curr. Opin. Chem. Biol.* 34, 72–79. doi: 10.1016/j.cbpa.2016.06.026

Esland, L., Larrea-Alvarez, M., and Purton, S. (2018). Selectable markers and reporter genes for engineering the chloroplast of *Chlamydomonas reinhardtii*. *Biology* 7:E46. doi: 10.3390/biology7040046

Fabris, M., George, J., Kuzhiumparambil, U., Lawson, C. A., Jaramillo Madrid, A. C., Abbriano, R. M., et al., (2020). Extrachromosomal genetic engineering of the marine diatom *Phaeodactylum tricornutum* enables the heterologous

production of monoterpenoids. *ACS Synth. Biol.* [Online ahead of print]. doi: 10.1021/acssynbio.9b00455

Fabris, M., Matthijs, M., Carbonelle, S., Moses, T., Pollier, J., Dasseville, R., et al. (2014). Tracking the sterol biosynthesis pathway of the diatom *Phaeodactylum tricornutum*. *New Phytol.* 204, 521–535. doi: 10.1111/nph.12917

Fabris, M., Matthijs, M., Rombauts, S., Vyverman, W., Goossens, A., and Baart, G. J. (2012). The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *Plant J.* 70, 1004–1014. doi: 10.1111/j.1365-313X.2012.04941.x

Fasaei, F., Bitter, J. H., Slegers, P. M., and Van Boxtel, A. J. B. (2018). Techno-economic evaluation of microalgae harvesting and dewatering systems. *Algal Res.* 31, 347–362. doi: 10.1016/j.algal.2017.11.038

FDA (2019). Available online at: https://www.fda.gov/food/food-ingredients-packaging/generally-recognized-safe-gras (accessed October, 2019).

Fernandez-Ricaud, L., Kourtchenko, O., Zackrisson, M., Warringer, J., and Blomberg, A. (2016). PRECOG: a tool for automated extraction and visualization of fitness components in microbial growth phenomics. *BMC Bioinformatics* 17:249. doi: 10.1186/s12859-016-1134-2

Fernandez-Ricaud, L., Warringer, J., Ericson, E., Pylvänäinen, I., Kemp, G. J. L., Nerman, O., et al. (2005). PROPHECY—a database for high-resolution phenomics. *Nucleic Acids Res.* 33, D369–D373.

Fiorese, M. L., Freitas, F., Pais, J., Ramos, A. M., De Aragão, G. M. F., and Reis, M. A. M. (2009). Recovery of polyhydroxybutyrate (PHB) from *Cupriavidus necator* biomass by solvent extraction with 1,2-propylene carbonate. *Eng. Life Sci.* 9, 454–461. doi: 10.1002/elsc.200900034

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., et al. (2011). Solutions for a cultivated planet. *Nature* 478, 337–342. doi: 10.1038/nature10452

FSANZ (2019). Available online at: https://www.foodstandards.gov.au (accessed October, 2019).

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

Garcia, J. L., De Vicente, M., and Galan, B. (2017). Microalgae, old sustainable food and fashion nutraceuticals. *Microb. Biotechnol.* 10, 1017–1024. doi: 10.1111/1751-7915.12800

Gerardo, M. L., Oatley-Radcliffe, D. L., and Lovitt, R. W. (2014). Integration of membrane technology in microalgae biorefineries. *J Memb. Sci.* 464, 86–99. doi: 10.1016/j.memsci.2014.04.010

Gerardo, M. L., Van Den Hende, S., Vervaeren, H., Coward, T., and Skill, S. C. (2015). Harvesting of microalgae within a biorefinery approach: a review of the developments and case studies from pilot-plants. *Algal Res.* 11, 248–262. doi: 10.1016/j.algal.2015.06.019

Gimpel, J. A., Henriquez, V., and Mayfield, S. P. (2015). In metabolic engineering of eukaryotic microalgae: potential and challenges come with great diversity. *Front. Microbiol.* 6:1376. doi: 10.3389/fmicb.2015.01376

Gimpel, J. A., Specht, E. A., Georgianna, D. R., and Mayfield, S. P. (2013). Advances in microalgae engineering and synthetic biology applications for biofuel production. *Curr. Opin. Chem. Biol.* 17, 489–495. doi: 10.1016/j.cbpa.2013.03.038

Giner, J.-L., Faraldos, J. A., and Boyer, G. L. (2003). Novel sterols of the toxic dinoflagellate *Karenia brevis* (Dinophyceae): a defensive function for unusual marine sterols?1. *J. Phycol.* 39, 315–319. doi: 10.1046/j.1529-8817.2003.01254.x

Glemser, M., Heining, M., Schmidt, J., Becker, A., Garbe, D., and Buchholz, R. (2016). Application of light-emitting diodes (LEDs) in cultivation of phototrophic microalgae : current state and perspectives. *Appl. Microbiol. Biotechnol.* 100, 1077–1088. doi: 10.1007/s00253-015-7144-6

Gong, M., and Bassi, A. (2016). Carotenoids from microalgae: a review of recent developments. *Biotechnol. Adv.* 34, 1396–1412. doi: 10.1016/j.biotechadv.2016.10.005

González-Fernández, C., and Ballesteros, M. (2013). Microalgae autoflocculation: an alternative to high-energy consuming harvesting methods. *J. Appl. Phycol.* 25, 991–999. doi: 10.1007/s10811-012-9957-3

Graham, L. E., Graham, J. M., and Wilcox, L. W. (2009). *Algae*. San Francisco: Pearson Benjamin Cummings.

Grahl, S., Palanisamy, M., Strack, M., Meier-Dinkel, L., Toepfl, S., and Morlein, D. (2018). Towards more sustainable meat alternatives: how technicalvparameters

affect the sensoryvproperties of extrusion products derivedvfrom soy and algae. *J. Clean. Prod.* 198, 962–971. doi: 10.1016/j.jclepro.2018.07.041

Greenwell, H. C., Laurens, L. M. L., Shields, R. J., Lovitt, R. W., and Flynn, K. J. (2010). Placing microalgae on the biofuels priority list: a review of the technological challenges. *J. R. Soc. Interface* 7, 703–726. doi: 10.1098/rsif.2009.0322

Gregory, J. A., Topol, A. B., Doerner, D. Z., and Mayfield, S. (2013). Alga-produced cholera toxin-Pfs25 fusion proteins as oral vaccines. *Appl. Environ. Microbiol.* 79, 3917–3925. doi: 10.1128/AEM.00714-13

Greiner, A., Kelterborn, S., Evers, H., Kreimer, G., Sizova, I., and Hegemann, P. (2017). Targeting of photoreceptor genes in *Chlamydomonas reinhardtii* via zinc-finger nucleases and CRISPR/Cas9. *Plant Cell* 29, 2498–2518. doi: 10.1105/tpc.17.00659

Gross, M., Jarboe, D., and Wen, Z. (2015). Biofilm-based algal cultivation systems. *Appl. Microbiol. Biotechnol.* 99, 5781–5789. doi: 10.1007/s00253-015-6736-5

Guedes, A., and Malcata, F. (2012). "Nutritional value and uses of microalgae in aquaculture," in *Aquaculture*, ed. Z. Muchlisin (Rijeka: InTech), 402.

Guerin, M., Huntley, M. E., and Olaizola, M. (2003). *Haematococcus* astaxanthin: applications for human health and nutrition. *Trends Biotechnol.* 21, 210–216. doi: 10.1016/s0167-7799(03)00078-7

Guiry, M. D. (2012). How many species of algae are there? *J. Phycol.* 48, 1057–1063. doi: 10.1111/j.1529-8817.2012.01222.x

Gupta, P. L., Lee, S. M., and Choi, H. J. (2015). A mini review: photobioreactors for large scale algal cultivation. *World J. Microbiol. Biotechnol.* 31, 1409–1417. doi: 10.1007/s11274-015-1892-4

Guzmán-Zapata, D., Sandoval-Vargas, J. M., Macedo-Osorio, K. S., Salgado-Manjarrez, E., Castrejón-Flores, J. L., Oliver-Salvador, M. D. C., et al. (2019). Efficient editing of the nuclear APT reporter gene in *Chlamydomonas reinhardtii* via expression of a CRISPR-Cas9 module. *Int. J. Mol. Sci.* 20:1247. doi: 10.3390/ijms20051247

Hagemann, M., and Hess, W. R. (2018). Systems and synthetic biology for the biotechnological application of cyanobacteria. *Curr. Opin. Biotechnol.* 49, 94–99. doi: 10.1016/j.copbio.2017.07.008

Hardy, R., and Tacon, A. (2002). "Fish meal: historical uses, production trends and future outlook for sustainable supplies," in *Responsible Marine Aquaculture*, eds R. Stickney and J. P. Mcve (New York, NY: CABI publishing), 391.

Harun, R., Singh, M., Forde, G. M., and Danquah, M. K. (2010). Bioprocess engineering of microalgae to produce a variety of consumer products. *Renew. Sustain. Energy Rev.* 14, 1037–1047. doi: 10.1016/j.rser.2009.11.004

Heasman, M., Diemar, J., O'connor, W., Sushames, T., and Foulkes, L. (2000). Development of extended shelf-life microalgae concentrate diets harvested by centrifugation for bivalve molluscs – a summary. *Aquac. Res.* 31, 637–659. doi: 10.1046/j.1365-2109.2000.00492.x

Hemaiswarya, S., Raja, R., Ravi Kumar, R., Ganesan, V., and Anbazhagan, C. (2011). Microalgae: a sustainable feed source for aquaculture. *World J. Microbiol. Biotechnol.* 27, 1737–1746. doi: 10.1007/s11274-010-0632-z

Hempel, F., Bozarth, A. S., Lindenkamp, N., Klingl, A., Zauner, S., Linne, U., et al. (2011a). Microalgae as bioreactors for bioplastic production. *Microb. Cell Fact.* 10:81. doi: 10.1186/1475-2859-10-81

Hempel, F., Lau, J., Klingl, A., and Maier, U. G. (2011b). Algae as protein factories: Expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*. *PLoS One* 6:e28424. doi: 10.1371/journal.pone.0028424

Hempel, F., Maurer, M., Brockmann, B., Mayer, C., Biedenkopf, N., Kelterbaum, A., et al. (2017). From hybridomas to a robust microalgal-based production platform: molecular design of a diatom secreting monoclonal antibodies directed against the Marburg virus nucleoprotein. *Microb. Cell Fact.* 16:131. doi: 10.1186/s12934-017-0745-2

Henchion, M., Hayes, M., Mullen, A. M., Fenelon, M., and Tiwari, B. (2017). Future protein supply and demand: strategies and factors influencing a sustainable equilibrium. *Foods* 6:E53. doi: 10.3390/foods6070053

Hildebrand, M., Abbriano, R. M., Polle, J. E., Traller, J. C., Trentacoste, E. M., Smith, S. R., et al. (2013). Metabolic and cellular organization in evolutionarily diverse microalgae as related to biofuels production. *Curr. Opin. Chem. Biol.* 17, 506–514. doi: 10.1016/j.cbpa.2013.02.027

Hillson, N., Caddick, M., Cai, Y., Carrasco, J. A., Chang, M. W., Curach, N. C., et al. (2019). Building a global alliance of biofoundries. *Nat. Commun.* 10:2040.

Hoffman, J., Pate, R. C., Drennen, T., and Quinn, J. C. (2017). Techno-economic assessment of open microalgae production systems. *Algal Res.* 23, 51–57. doi: 10.1021/es304135b

Houle, D., Govindaraju, D. R., and Omholt, S. (2010). Phenomics: the next challenge. *Nat. Rev. Genet.* 11, 855–866. doi: 10.1038/nrg2897

Huang, W., and Daboussi, F. (2017). Genetic and metabolic engineering in diatoms. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372:20160411. doi: 10.1098/rstb.2016.0411

Huntley, M. E., Johnson, Z. I., Brown, S. L., Sills, D. L., Gerber, L., Archibald, I., et al. (2015). Demonstrated large-scale production of marine microalgae for fuels and feed. *Algal Res.* 10, 249–265. doi: 10.1016/j.algal.2015.04.016

Hutchison, C. A., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science* 351:aad6253. doi: 10.1126/science.aad6253

Ihnken, S., Roberts, S., and Beardall, J. (2011). Differential responses of growth and photosynthesis in the marine diatom *Chaetoceros muelleri* to CO2 and light availability. *Phycologia* 50, 182–193. doi: 10.2216/10-11.1

Jacobs, T. B., Zhang, N., Patel, D., and Martin, G. B. (2017). Generation of a collection of mutant tomato lines using pooled CRISPR libraries. *Plant Physiol.* 174, 2023–2037. doi: 10.1104/pp.17.00489

Jacobsen, C., Horne, A. F., Sørensen, A.-D. M., and Nielsen, N. S. (2013). *Food Enrichment with Omega-3 Fatty Acids : Food Enrichment with Omega-3 Fatty Acids.* Cambridge: Elsevier Science & Technology.

Jaramillo-Madrid, A. C., Ashworth, J., Fabris, M., and Ralph, P. J. (2019). Phytosterol biosynthesis and production by diatoms (*Bacillariophyceae*). *Phytochemistry* 163, 46–57. doi: 10.1016/j.phytochem.2019.03.018

Jonker, J. G. G., and Faaij, A. P. C. (2013). Techno-economic assessment of micro-algae as feedstock for renewable bio-energy production. *Appl. Energy* 102, 461–475. doi: 10.1016/j.apenergy.2012.07.053

Jordan, D. R., Mace, E. S., Cruickshank, A. W., Hunt, C. H., and Henzell, R. G. (2011). Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. *Crop Sci.* 51, 1444–1457. doi: 10.2135/cropsci2010.06.0326

Kadam, S. U., Tiwari, B. K., and O'donnell, C. P. (2013). Application of novel extraction technologies for bioactives from marine algae. *J. Agric. Food Chem.* 61, 4667–4675. doi: 10.1021/jf400819p

Kagermann, H., Lukas, W.-D., and Wahlster, W. (2011). *Industrie 4.0: Mit dem Internet der Dinge auf dem Weg zur 4. industriellen Revolution. VDI Nachrichten.* Available online at: http://www.vdi-nachrichten.com/Technik-Gesellschaft/Industrie-40-Mit-Internet-Dinge-Weg-4-industriellen-Revolution (accessed May 26, 2016).

Kagermann, H., Wahlster, W., and Helbig, J. (2013). *Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Securing the Future of German Manufacturing Industry; Final Report of the Industrie 4.0 Working Group.* Frankfurt: acatech – National Academy of Science and Engineering.

Karas, B. J., Diner, R. E., Lefebvre, S. C., Mcquaid, J., Phillips, A. P., Noddings, C. M., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* 6:6925. doi: 10.1038/ncomms7925

Karas, B. J., Molparia, B., Jablanovic, J., Hermann, W. J., Lin, Y.-C., Dupont, C. L., et al. (2013). Assembly of eukaryotic algal chromosomes in yeast. *J. Biol. Eng.* 7:30. doi: 10.1186/1754-1611-7-30

Karkos, P. D., Leong, S. C., Karkos, C. D., Sivaji, N., and Assimakopoulos, D. A. (2011). Spirulina in clinical practice: evidence-based human applications. *Evid. Based Complement. Altern. Med.* 2011, 1–4.

Kent, M., Welladsen, H. M., Mangott, A., and Li, Y. (2015). Nutritional evauulation of Australian microalgae as potential human health supplements. *PLoS One* 10:e0118985. doi: 10.1371/journal.pone.0118985

Khalil, A. S., and Collins, J. J. (2010). Synthetic biology: applications come of age. *Nat. Rev. Genet.* 11, 367–379. doi: 10.1038/nrg2775

Khanra, S., Mondal, M., Halder, G., Tiwari, O. N., Gayen, K., and Bhowmick, T. K. (2018). Downstream processing of microalgae for pigments, protein and carbohydrate in industrial application: a review. *Food Bioprod. Process.* 110, 60–84. doi: 10.1016/j.fbp.2018.02.002

Kidd, P. (2011). Astaxanthin, cell membrane nutrient with diverse clinical benefits and anti-aging potential. *Altern. Med. Rev.* 16, 355–364.

Kilian, O., Benemann, C. S., Niyogi, K. K., and Vick, B. (2011). High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. *Proc. Natl. Acad. Sci. U.S.A.* 108, 21265–21269. doi: 10.1073/pnas.1105861108

Kim, H. S., Devarenne, T. P., and Han, A. (2018). Microfluidic systems for microalgal biotechnology: a review. *Algal Res.* 30, 149–161. doi: 10.1093/femsle/fnx238

Kim, K. H., Choi, I. S., Kim, H. M., Wi, S. G., and Bae, H.-J. (2014). Bioethanol production from the nutrient stress-induced microalga *Chlorella vulgaris* by enzymatic hydrolysis and immobilized yeast fermentation. *Bioresour. Technol.* 153, 47–54. doi: 10.1016/j.biortech.2013.11.059

Kim, S., Kim, J., Lim, Y., Kim, Y. J., Kim, J. Y., and Kwon, O. (2016). A dietary cholesterol challenge study to assess *Chlorella* supplementation in maintaining healthy lipid levels in adults: a double-blinded, randomized, placebo- controlled study. *Nutr. J.* 15, 1–7. doi: 10.1186/s12937-016-0174-9

Kim, S. W., Less, J. F., Wang, L., Yan, T., Kiron, V., Kaushik, S. J., et al. (2019). Meeting global feed protein demand: challenge, opportunity, and strategy. *Annu. Rev. Anim. Biosci.* 7, 221–243. doi: 10.1146/annurev-animal-030117-014838

Kiron, V., Wutiporn, P., Huntley, M., Archibald, I., and De Scheemaker, G. (2012). Marine microalgae from biorefinery as a potential feed protein source for Atlantic salmon, common carp and whiteleg shrimp. *Aquac. Nutr.* 18, 521–531. doi: 10.1111/j.1365-2095.2011.00923.x

Kirst, H., Garcia-Cerdan, J. G., Zurbriggen, A., Ruehle, T., and Melis, A. (2012). Truncated photosystem chlorophyll antenna size in the green microalga <em<*Chlamydomonas reinhardtii* upon Deletion of the TLA3-RP43 Gene. *Plant Physiol.* 160, 2251–2260. doi: 10.1104/pp.112.206672

Kohli, K., Prajapati, R., and Sharma, B. K. (2019). Bio-based chemicals from renewable biomass for integrated biorefineries. *Energies* 12:233. doi: 10.1002/biot.201100069

Koyande, A. K., Chew, K. W., Rambabu, K., Tao, Y., Chu, D.-T., and Showa, P.-L. (2019). Microalgae: a potential alteranive to health supplementation for humans. *Food Sci. Hum. Wellness* 8, 16–24. doi: 10.1017/S1751731116001543

Kroth, P. G., Bones, A. M., Daboussi, F., Ferrante, M. I., Jaubert, M., Kolot, M., et al. (2018). Genome editing in diatoms: achievements and goals. *Plant Cell Rep.* 37, 1401–1408. doi: 10.1007/s00299-018-2334-1

Kumar, R. R., Rao, P. H., and Arumugam, M. (2015). Lipid extraction methods from microalgae: a comprehensive review. *Front. Energy Res.* 3:61. doi: 10.3389/fenrg.2014.00061

Laing, I. (1987). The use of artificial diets in rearing bivalve spat. *Aquac.* 65, 243–249. doi: 10.1016/0044-8486(87)90237-7

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2011). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202—-D1210. doi: 10.1093/nar/gkr1090

Lauersen, K. J. (2019). Eukaryotic microalgae as hosts for light-driven heterologous isoprenoid production. *Planta* 249, 155–180. doi: 10.1007/s00425-018-3048-x

Lauersen, K. J., Baier, T., Wichmann, J., Wordenweber, R., Mussgnug, J. H., Hubner, W., et al. (2016). Efficient phototrophic production of a high-value sesquiterpenoid from the eukaryotic microalga *Chlamydomonas reinhardtii*. *Metab. Eng.* 38, 331–343. doi: 10.1016/j.ymben.2016.07.013

Lauersen, K. J., Vanderveer, T. L., Berger, H., Kaluza, I., Mussgnug, J. H., Walker, V. K., et al. (2013). Ice recrystallization inhibition mediated by a nuclear-expressed and -secreted recombinant ice-binding protein in the microalga *Chlamydomonas reinhardtii*. *Appl. Microbiol. Biotechnol.* 97, 9763–9772. doi: 10.1007/s00253-013-5226-x

Lauersen, K. J., Wichmann, J., Baier, T., Kampranis, S. C., Pateraki, I., Moller, B. L., et al. (2018). Phototrophic production of heterologous diterpenoids and a hydroxy-functionalized derivative from *Chlamydomonas reinhardtii*. *Metab. Eng.* 49, 116–127. doi: 10.1016/j.ymben.2018.07.005

Laurens, L. M. L., Chen-Glasser, M., and Mcmillan, J. D. (2017). A perspective on renewable bioenergy from photosynthetic algae as feedstock for biofuels and bioproducts. *Algal Res.* 24, 261–264. doi: 10.1016/j.algal.2017.04.002

Lenihan-Geels, G., Bishop, S. K., and Ferguson, R. L. (2013). Alternative sources of omega-3 fats: can we find a sustainable substitute for fish?. *Nutrients* 5, 1301–1315. doi: 10.3390/nu5041301

Li, H., Shen, C. R., Huang, C. H., Sung, L. Y., Wu, M. Y., and Hu, Y. C. (2016). CRISPR-Cas9 for the genome engineering of cyanobacteria and succinate production. *Metab. Eng.* 38, 293–302. doi: 10.1016/j.ymben.2016.09.006

Li, J., Zhu, D., Niu, J., Shen, S., and Wang, G. (2011). An economic assessment of astaxanthin production by large scale cultivation of *Haematococcus*

pluvialis. *Biotechnol. Adv.* 29, 568–574. doi: 10.1016/j.biotechadv.2011. 04.001

Lin, P. C., and Pakrasi, H. B. (2019). Engineering cyanobacteria for production of terpenoids. *Planta* 249, 145–154. doi: 10.1007/s00425-018-3047-y

Lin, P.-C., Saha, R., Zhang, F., and Pakrasi, H. B. (2017). Metabolic engineering of the pentose phosphate pathway for enhanced limonene production in the cyanobacterium *Synechocystis* sp. *PCC* 6803. *Sci. Rep.* 7:17503. doi: 10.1038/s41598-017-17831-y

Litchman, E. (2007). "CHAPTER 16 - resource competition and the ecological success of phytoplankton," in *Evolution of Primary Producers in the Sea*, eds P. G. Falkowski and A. H. Knoll (Burlington: Academic Press), 351–375. doi: 10.1016/b978-012370518-1/50017-5

Liu, J., Nannas, N. J., Fu, F. F., Shi, J., Aspinwall, B., Parrott, W. A., et al. (2019). Genome-scale sequence disruption following biolistic transformation in rice and maize. *Plant Cell* 31, 368–383. doi: 10.1105/tpc.18.00613

Lordan, S., Ross, R. P., and Stanton, C. (2011). Marine bioactives as functional food ingredients: potential to reduce the incidence of chronic diseases. *Mar. Drugs* 9, 1056–1100. doi: 10.3390/md9061056

Lu, Y., Zhou, W., Wei, L., Li, J., Jia, J., Li, F., et al. (2014). Regulation of the cholesterol biosynthetic pathway and its integration with fatty acid biosynthesis in the oleaginous microalga *Nannochloropsis oceanica*. *Biotechnol. Biofuels* 7:81. doi: 10.1186/1754-6834-7-81

Lucker, B. F., Hall, C. C., Zegarac, R., and Kramer, D. M. (2014). The environmental photobioreactor (ePBR): An algal culturing platform for simulating dynamic natural environments. *Algal Res.* 6(Pt B), 242–249. doi: 10.1016/j.algal.2013.12. 007

Luo, Q., Bian, C., Tao, M., Huang, Y., Zheng, Y., Lv, Y., et al. (2018). Genome and transcriptome sequencing of the astaxanthin-producing green microalga, *Haematococcus pluvialis*. *Genome Biol. Evol.* 11, 166–173. doi: 10.1093/gbe/evy263

Luo, X., Reiter, M. A., D'espaux, L., Wong, J., Denby, C. M., Lechner, A., et al. (2019). Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* 567, 123–126. doi: 10.1038/s41586-019-0978-9

Maeda, Y., Yoshino, T., Matsunaga, T., Matsumoto, M., and Tanaka, T. (2018). Marine microalgae for production of biofuels and chemicals. *Curr. Opin. Biotechnol.* 50, 111–120. doi: 10.1016/j.copbio.2017.11.018

Mao, T. K., Van De Water, J., and Gershwin, M. E. (2005). Effects of a spirulina-based dietary supplement on cytokine production from allergic rhinitis patients. *J. Med. Food* 8, 27–30. doi: 10.1089/jmf.2005.8.27

Martinez-Piernas, A. B., Plaza-Bolanos, P., Garcia-Gomez, E., Fernandez-Ibanez, P., and Aguera, A. (2018). Determination of organic microcontaminants in agricultural soils irrigated with reclaimed wastewater: Target and suspect approaches. *Anal. Chim. Acta* 1030, 115–124. doi: 10.1016/j.aca.2018.05.049

Mathieu-Rivet, E., Kiefer-Meyer, M. C., Vanier, G., Ovide, C., Burel, C., Lerouge, P., et al. (2014). Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Front. Plant Sci.* 5:359. doi: 10.3389/fpls.2014.00359

Melnicki, M. R., Pinchuk, G. E., Hill, E. A., Kucek, L. A., Stolyar, S. M., Fredrickson, J. K., et al. (2013). Feedback-controlled LED photobioreactor for photophysiological studies of cyanobacteria. *Bioresour. Technol.* 134, 127–133. doi: 10.1016/j.biortech.2013.01.079

Meng, X., Yu, H., Zhang, Y., Zhuang, F., Song, X., Gao, S., et al. (2017). Construction of a Genome-Wide Mutant Library in Rice Using CRISPR/Cas9. *Mol. Plant* 10, 1238–1241. doi: 10.1016/j.molp.2017.06.006

Merchant, R. E. (2001). A review of recent clinical trials of the nutritional supplement *Chorella pyrenoidosa* in the treatment of Fibromyalgia, hypertension and ulcerative colitis. *Altern. Ther. Health Med.* 7, 79–91.

Miao, G., Zhu, C. C., Wang, J. J., Tan, Z. C., Wang, L., Liu, J. L., et al. (2015). Efficient one-pot production of 1,2-propanediol and ethylene glycol from microalgae (*Chlorococcum* sp.) in water. *Green Chem.* 17, 2538–2544. doi: 10.1039/c4gc02467b

Michalak, I., and Chojnacka, K. (2015). Algae as production systems of bioactive compounds. *Eng. Life Sci.* 15, 160–176. doi: 10.1002/elsc.20140 0191

Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., et al., (2014). The seco-iridoid pathway from Catharanthus roseus. *Nat. Commun.* 5:3606. doi: 10.1038/ncomms4606

Milledge, J. J., and Heaven, S. (2013). A review of the harvesting of micro-algae for biofuel production. *Rev. Environ. Sci. Biotechnol.* 12, 165–178. doi: 10.1007/s11157-012-9301-z

Miller, M. B., Haubrich, B. A., Wang, Q., Snell, W. J., and Nes, W. D. (2012). Evolutionarily conserved Delta(25(27))-olefin ergosterol biosynthesis pathway in the alga *Chlamydomonas reinhardtii*. *J. Lipid Res.* 53, 1636–1645. doi: 10. 1194/jlr.M027482

Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., and Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Sci.* 282, 60–72. doi: 10.1016/j.plantsci.2019.01.007

Moellering, E. R., and Benning, C. (2010). RNA interference silencing of a major lipid droplet protein affects lipid droplet size in chlamydomonas reinhardtii. *Eukaryot. Cell* 9:97. doi: 10.1128/EC.002 03-09

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7:1419. doi: 10.3389/fpls. 2016.01419

Montero-Lobato, Z., Vázquez, M., Navarro, F., Fuentes, J. L., Bermejo, E., Garbayo, I., et al. (2018). Chemically-induced production of anti-inflammatory molecules in microalgae. *Mar. Drugs* 16, 1–21.

Moses, T., Mehrshahi, P., Smith, A. G., and Goossens, A. (2017). Synthetic biology approaches for the production of plant metabolites in unicellular organisms. *J. Exp. Bot.* 68, 4057–4074. doi: 10.1093/jxb/erx119

Moses, T., Pollier, J., Almagro, L., Buyst, D., Van Montagu, M., Pedreno, M. A., et al. (2014). Combinatorial biosynthesis of sapogenins and saponins in *Saccharomyces cerevisiae* using a C-16alpha hydroxylase from *Bupleurum falcatum*. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1634–1639. doi: 10.1073/pnas. 1323369111

Moulin, M., Nguyen, G. T. D. T., Scaife, M. A., Smith, A. G., and Fitzpatrick, T. B. (2013). Analysis of Chlamydomonas thiamin metabolism in vivo reveals riboswitch plasticity. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14622–14627. doi: 10.1073/pnas.1307741110

Munir, K., and Sheraz Anjum, M. (2018). The use of ontologies for effective knowledge modelling and information retrieval. *Appl. Comput. Inform.* 14, 116–126. doi: 10.1016/j.aci.2017.07.003

Mussgnug, J. H. (2015). Genetic tools and techniques for *Chlamydomonas reinhardtii*. *Appl. Microbiol. Biotechnol.* 99, 5407–5418. doi: 10.1007/s00253-015-6698-7

Nakahira, Y., Ogawa, A., Asano, H., Oyama, T., and Tozawa, Y. (2013). Theophylline-dependent riboswitch as a novel genetic tool for strict regulation of protein expression in Cyanobacterium *Synechococcus elongatus* PCC 7942. *Plant Cell Physiol.* 54, 1724–1735. doi: 10.1093/pcp/pct115

Neveu, P., Tireau, A., Hilgert, N., Nègre, V., Mineau-Cesari, J., Brichet, N., et al. (2019). Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven phenotyping hybrid information system. *New Phytol.* 221, 588–601. doi: 10.1111/nph.15385

Nguyen, L. N., Labeeuw, L., Commault, A. S., Emmerton, B., Ralph, P. J., Hasan, A., et al. (2019). Validation of a cationic polyacrylamide flocculant for the harvesting fresh and seawater microalgal biomass. *Environ. Technol. Innovat.* 16:100466. doi: 10.1016/j.eti.2019.100466

Nielsen, J., and Keasling, J. D. (2016). Engineering cellular metabolism. *Cell* 164, 1185–1197. doi: 10.1016/j.cell.2016.02.004

Noreen, A., Zia, K. M., Zuber, M., Ali, M., and Mujahid, M. (2016). A critical review of algal biomass: A versatile platform of bio-based polyesters from renewable resources. *Int. J. Biol. Macromol.* 86, 937–949. doi: 10.1016/j.ijbiomac.2016.01. 067

Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6:24951. doi: 10.1038/srep24951

Oellrich, A., Walls, R. L., Cannon, E. K. S., Cannon, S. B., Cooper, L., and Gardiner, J. (2015). An ontology approach to comparative phenomics in plants. *Plant Methods* 11:10. doi: 10.1186/s13007-015-0053-y

Ohyama, K., Ogawa, M., and Matsubayashi, Y. (2008). Identification of a biologically active, small, secreted peptide in Arabidopsis by in silico gene screening, followed by LC-MS-based structure analysis. *Plant J.* 55, 152–160. doi: 10.1111/j.1365-313X.2008.03464.x

Olasehinde, T. A., Olaniran, A. O., and Okoh, A. I. (2017). Therapeutic potential of microalgae in the treatment of Alzheimer's disease. *Molecules* 22, 1–18.

Olivieri, G., Salatino, P., and Marzocchella, A. (2014). Advances in photobioreactors for intensive microalgal production: Configurations, operating strategies and applications. *J. Chem. Technol. Biotechnol.* 89, 178–195. doi: 10.1002/jctb.4218

Oplatowska-Stachowiak, M., and Elliott, C. T. (2017). Food colors: existing and emerging food safety concerns. *Crit. Rev. Food Sci. Nutr.* 57, 524–548. doi: 10.1080/10408398.2014.889652

Paajanen, P., Strauss, J., Van Oosterhout, C., Mcmullan, M., Clark, M. D., and Mock, T. (2017). Building a locally diploid genome and transcriptome of the diatom *Fragilariopsis cylindrus*. *Sci. Data* 4, 170149–170149. doi: 10.1038/sdata.2017.149

Pacheco-Vega, J. M., and Sánchez-Saavedra, M. D. P. (2009). The Biochemical Composition of *Chaetoceros muelleri* (Lemmermann Grown) with an agricultural fertilizer. *J. World Aquac. Soc.* 40, 556–560. doi: 10.1111/j.1749-7345.2009.00276.x

Palanisamy, M., Topfl, S., Berger, R. G., and Hertel, C. (2019). Physico-chemical and nutritional properties of meat analogues based on Spirulina/lupin protein mixtures. *Eur. Food Res. Technol.* 245, 1889–1898. doi: 10.1007/s00217-019-03298-w

Panis, G., and Carreon, J. R. (2016). Commercial astaxanthin production derived by green alga *Haematococcus pluvialis*: a microalgae process model and a techno-economic assessment all through production line. *Algal Res.* 18, 175–190. doi: 10.1016/j.algal.2016.06.007

Papaefthimiou, D., Diretto, G., Demurtas, O. C., Mini, P., Ferrante, P., Giuliano, G., et al. (2019). Heterologous production of labdane-type diterpenes in the green alga *Chlamydomonas reinhardtii*. *Phytochemistry* 167:112082. doi: 10.1016/j.phytochem.2019.112082

Park, H. J., Lee, Y. J., Ryu, H. K., Kim, M. H., Chung, H. W., and Kim, W. Y. (2008). A randomized double-blind, placebo-controlled study to establishthe effects of Spirulina in elderly koreans. *Ann. Nutr. Metab.* 52, 322–328. doi: 10.1159/000151486

Patron, N. J., Orzaez, D., Marillonnet, S., Warzecha, H., Matthewman, C., and Youles, M. (2015). Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytol.* 208, 13–19. doi: 10.1111/nph.13532

Peng, F. Q., Ying, G. G., Yang, B., Liu, S., Lai, H. J., Liu, Y. S., et al. (2014). Biotransformation of progesterone and norgestrel by two freshwater microalgae (*Scenedesmus obliquus* and *Chlorella pyrenoidosa*): transformation kinetics and products identification. *Chemosphere* 95, 581–588. doi: 10.1016/j.chemosphere.2013.10.013

Petrie, B., Barden, R., and Kasprzyk-Hordern, B. (2015). A review on emerging contaminants in wastewaters and the environment: current knowledge, understudied areas and recommendations for future monitoring. *Water Res.* 72, 3–27. doi: 10.1016/j.watres.2014.08.053

Piironen, V., Toivo, J., Puupponen-Pimiä, R., and Lampi, A.-M. (2003). Plant sterols in vegetables, fruits and berries. *J. Sci. Food Agric.* 83, 330–337. doi: 10.1002/jsfa.1316

Pisha, E., Chai, H., Lee, I.-S., Chagwedera, T. E., Farnsworth, N. R., Cordell, G. A., et al. (1995). Discovery of betulinic acid as a selective inhibitor of human melanoma that functions by induction of apoptosis. *Nat. Med.* 1, 1046–1051. doi: 10.1038/nm1095-1046

Plaza, M., Cifuentes, A., and Ibáñez, E. (2008). In the search of new functional food ingredients from algae. *Trends Food Sci. Technol.* 19, 31–39. doi: 10.1016/j.tifs.2007.07.012

Poliner, E., Takeuchi, T., Du, Z. Y., Benning, C., and Farre, E. M. (2018). Nontransgenic marker-free gene disruption by an episomal CRISPR system in the oleaginous microalga, *Nannochloropsis oceanica* CCMP1779. *ACS Synth. Biol.* 7, 962–968. doi: 10.1021/acssynbio.7b00362

Pollak, B., Matute, T., Nuñez, I., Cerda, A., Lopez, C., Vargas, V., et al. (2019). Universal Loop assembly (uLoop): open, efficient, and species-agnostic DNA fabrication. *bioRxiv* [Preprint]. doi: 10.1101/744854

Pollier, J., Vancaester, E., Kuzhiumparambil, U., Vickers, C. E., Vandepoele, K., Goossens, A., et al. (2019). A widespread alternative squalene epoxidase participates in eukaryote steroid biosynthesis. *Nat. Microbiol.* 4, 226–233. doi: 10.1038/s41564-018-0305-5

Posten, C. (2009). Design principles of photo-bioreactors for cultivation of microalgae. *Eng. Life Sci.* 9, 165–177. doi: 10.1002/elsc.200900003

Pradhan, S., and Rai, L. C. (2001). Biotechnological potential of *Microcystis* sp. in Cu, Zn and Cd biosorption from single and multimetallic systems. *Biometals* 14, 67–74.

Pragya, N., Pandey, K. K., and Sahoo, P. K. (2013). A review on harvesting, oil extraction and biofuels production technologies from microalgae. *Renew. Sustain. Energy Rev.* 24, 159–171. doi: 10.1016/j.rser.2013.03.034

Pulz, O., and Gross, W. (2004). Valuable products from biotechnology of microalgae. *Appl. Microbiol. Biotechnol.* 65, 635–648. doi: 10.1007/s00253-004-1647-x

Qin, S., Lin, H., and Jiang, P. (2012). Advances in genetic engineering of marine algae. *Biotechnol. Adv.* 30, 1602–1613. doi: 10.1016/j.biotechadv.2012.05.004

Radakovits, R., Jinkerson, R. E., Fuerstenberg, S. I., Tae, H., Settlage, R. E., Boore, J. L., et al. (2012). Draft genome sequence and genetic transformation of the oleaginous alga Nannochloropis gaditana. *Nat. Commun.* 3:686. doi: 10.1038/ncomms1688

Ramarajan, M., Fabris, M., Abbriano, R. M., Pernice, M., and Ralph, P. J. (2019). Novel endogenous promoters for genetic engineering of the marine microalga *Nannochloropsis gaditana* CCMP526. *Algal Res.* 44:101708. doi: 10.1016/j.algal.2019.101708

Rampen, S. W., Abbas, B. A., Schouten, S., and Sinninghe Damste, J. S. (2010). A comprehensive study of sterols in marine diatoms (Bacillariophyta): Implications for their use as tracers for diatom productivity. *Limnol. Oceanogr.* 55, 91–105. doi: 10.4319/lo.2010.55.1.0091

Ras, R. T., Geleijnse, J. M., and Trautwein, E. A. (2014). LDL-cholesterol-lowering effect of plant sterols and stanols across different dose ranges: a meta-analysis of randomised controlled studies. *Br. J. Nutr.* 112, 214–219. doi: 10.1017/S0007114514000750

Rasala, B. A., Lee, P. A., Shen, Z., Briggs, S. P., Mendez, M., and Mayfield, S. P. (2012). Robust expression and secretion of Xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide. *PLoS One* 7:e43349. doi: 10.1371/journal.pone.0043349

Rasala, B. A., and Mayfield, S. P. (2015). Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. *Photosynth. Res.* 123, 227–239. doi: 10.1007/s11120-014-9994-7

Rasala, B. A., Muto, M., Lee, P. A., Jager, M., Cardoso, R. M., Behnke, C. A., et al. (2010). Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. *Plant Biotechnol. J.* 8, 719–733. doi: 10.1111/j.1467-7652.2010.00503.x

Ratledge, C. (2004). Fatty acid biosynthesis in microorganisms being used for singe cell oil production. *Biochimie* 86, 807–815. doi: 10.1016/j.biochi.2004.09.017

Ratledge, C., and Cohen, Z. (2008). Microbial and algal oils: do they have a future for biodiesel or as commodity oils? *Lipid Technol.* 20, 155–160. doi: 10.1002/lite.200800044

Ratledge, C., Streekstra, H., Cohen, Z., and Fichtali, J. (2010). "Downstream processing, extraction and purification of single cell oils," in *Single Cell Oils*, eds Z. Cohen and C. Ratledge (US: AOCS Press).

Rawat, I., Kumar, R. R., Mutanda, T., and Bux, F. (2013). Biodiesel from microalgae: A critical evaluation from laboratory to large scale production. *Appl. Energy* 103, 444–467. doi: 10.1016/j.apenergy.2012.10.004

Richardson, S. M., Mitchell, L. A., Stracquadanio, G., Yang, K., Dymond, J. S., Dicarlo, J. E., et al. (2017). Design of a synthetic yeast genome. *Science* 355, 1040–1044. doi: 10.1126/science.aaf4557

Roitsch, T., Cabrera-Bosquet, L., Fournier, A., Ghamkhar, K., Jiménez-Berni, J., Pinto, F., et al. (2019). New sensors and data-driven approaches—A path to next generation phenomics. *Plant Sci.* 282, 2–10. doi: 10.1016/j.plantsci.2019.01.011

Roles, J., Yarnold, J., Wolf, J., Stephens, E., Hussey, K., and Hankamer, B. (2020). Charting a development path to deliver cost competitive microalgae-based fuels. *Algal Res.* 45:101721. doi: 10.1016/j.algal.2019.101721

Ryckebosch, E., Bruneel, C., Termonte-Verhalle, R., Goiris, K., Muylaert, K., and Foubert, I. (2014). Nutritional evauation of microalgae oils rich in omega-3 long chain polyunsaturated fatty acids as an alternative for fish oil. *Food Chem.* 160, 393–400. doi: 10.1016/j.foodchem.2014.03.087

Ryu, N. H., Lim, Y., Park, J. E., Kim, J., Kim, J. Y., Kwon, S. W., et al. (2014). Impact of daily *Chlorella* consumption on serum lipid and carotenoid profiles in mildly hypercholesterolemic adults: a double-blinded, randomized, placebo-controlled study. *Nutr. J.* 13, 1–8. doi: 10.1186/1475-2891-13-57

Sasso, S., Pohnert, G., Lohr, M., Mittag, M., and Hertweck, C. (2012). Microalgae in the postgenomic era: a blooming reservoir for new natural products. *FEMS Microbiol. Rev.* 36, 761–785. doi: 10.1111/j.1574-6976.2011.00304.x

Schulze, P. S. C., Barreira, L. A., Pereira, H. G. C., Perales, J. A., and Varela, J. C. S. (2014). Light emitting diodes (LEDs) applied to microalgal production. *Trends Biotechnol.* 32, 422–430. doi: 10.1016/j.tibtech.2014.06.001

Scranton, M. A., Ostrand, J. T., Georgianna, D. R., Lofgren, S. M., Li, D., Ellis, R. C., et al. (2016). Synthetic promoters capable of driving robust nuclear gene expression in the green alga *Chlamydomonas reinhardtii*. *Algal Res.* 15, 135–142. doi: 10.1016/j.algal.2016.02.011

Seon, G., Joo, H. W., Kim, Y. J., Park, J., and Chang, Y. K. (2019). Hydrolysis of lipid-extracted *Chlorella vulgaris* by simultaneous use of solid and liquid acids. *Biotechnol. Prog.* 35:e2729. doi: 10.1002/btpr.2729

Serif, M., Dubois, G., Finoux, A. L., Teste, M. A., Jallet, D., and Daboussi, F. (2018). One-step generation of multiple gene knock-outs in the diatom *Phaeodactylum tricornutum* by DNA-free genome editing. *Nat. Commun.* 9:3924. doi: 10.1038/s41467-018-06378-9

Shi, B., Wideman, G., and Wang, J. H. (2012). A new approach of BioCO2 fixation by thermoplastic processing of Microalgae. *J. Polym. Environ.* 20, 124–131. doi: 10.1007/s10924-011-0329-x

Shin, S. E., Lim, J. M., Koh, H. G., Kim, E. K., Kang, N. K., Jeon, S., et al. (2016). CRISPR/Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*. *Sci. Rep.* 6:27810. doi: 10.1038/srep27810

Sidari, R., and Tofalo, R. (2019). A comprehensive overview on microalgal-fortified/based food and beverages. *Food Rev. Int.* 35, 778–805. doi: 10.1080/87559129.2019.1608557

Singh, S., Kate, B. N., and Banerjee, U. C. (2005). Bioactive compounds from cyanobacteria and microalgae: an overview. *Crit. Rev. Biotechnol.* 25, 73–95. doi: 10.1080/07388550500248498

Sizova, I., Greiner, A., Awasthi, M., Kateriya, S., and Hegemann, P. (2013). Nuclear gene targeting in *Chlamydomonas* using engineered zinc-finger nucleases. *Plant J.* 73, 873–882. doi: 10.1111/tpj.12066

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., and Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Comput. Intell. Neurosci.* 2016:11. doi: 10.1155/2016/3289801

Slattery, S. S., Diamond, A., Wang, H., Therrien, J. A., Lant, J. T., Jazey, T., et al. (2018). An expanded plasmid-based genetic toolbox enables Cas9 genome editing and stable maintenance of synthetic pathways in *Phaeodactylum tricornutum*. *ACS Synth. Biol.* 7, 328–338. doi: 10.1021/acssynbio.7b00191

Smetacek, V. (1999). Diatoms and the ocean carbon cycle. *Protist* 150, 25–32. doi: 10.1016/s1434-4610(99)70006-4

Smith, S. R., Dupont, C. L., Mccarthy, J. K., Broddrick, J. T., Obornik, M., Horak, A., et al. (2019). Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat. Commun.* 10:4552. doi: 10.1038/s41467-019-12407-y

Soo-Jung, K., Jin-Woo, K., Ye-Gi, L., Yong-Cheo, P., and Jin-Ho, S. (2017). Metabolic engineering of *Saccharomyces cerevisiae* for 2,3-butanediol production. *Appl. Microbiol. Biotechnol.* 101, 2241–2250.

Specht, E., and Mayfield, S. (2014). Algae-based oral recombinant vaccines. *Front. Microbiol.* 5:60. doi: 10.3389/fmicb.2014.00060

Spolaore, P., Joannis-Cassan, C., Duran, E., and Isambert, A. (2006). Commercial applications of microalgae. *J. Biosci. Bioeng.* 101, 87–96. doi: 10.1263/jbb.101.87

Stahl, W., and Sies, H. (2005). Bioactivity and protective effects of natural carotenoids. *Biochim. Biophys. Acta* 1740, 101–107. doi: 10.1016/j.bbadis.2004.12.006

Sudesh, K., Abe, H., and Doi, Y. (2000). Synthesis, structure and properties of polyhydroxyalkanoates: biological polyesters. *Prog. Polym. Sci.* 25, 1503–1555. doi: 10.1016/s0079-6700(00)00035-6

Sun, A., Davis, R., Starbuck, M., Ben-Amotz, A., Pate, R., and Pienkos, P. T. (2011). Comparative cost analysis of algal oil production for biofuels. *Energy* 36, 5169–5179. doi: 10.1016/j.jbiotec.2012.02.008

Sun, T., Li, S., Song, X., Pei, G., Diao, J., Cui, J., et al. (2018). Re-direction of carbon flux to key precursor malonyl-CoA via artificial small RNAs in photosynthetic *Synechocystis* sp. PCC 6803. *Biotechnol. Biofuels* 11:26. doi: 10.1186/s13068-018-1032-0

Sutherland, D. L., Howard-Williams, C., Turnbull, M. H., Broady, P. A., and Craggs, R. J. (2015). Enhancing microalgal photosynthesis and productivity in wastewater treatment high rate algal ponds for biofuel production. *Bioresour. Technol.* 184, 222–229. doi: 10.1016/j.biortech.2014.10.074

Sutherland, D. L., and Ralph, P. J. (2019). Microalgal bioremediation of emerging contaminants - Opportunities and challenges. *Water Res.* 164:114921. doi: 10.1016/j.watres.2019.114921

Suzuki, K., Mamedov, T. G., and Ikawa, T. (1999). A mutant of *Chlamydomonas reinhardtii* with reduced rate of photorespiration. *Plant Cell Physiol.* 40, 792–799. doi: 10.1093/oxfordjournals.pcp.a029607

Swanson, D., Block, R., and Mousa, S. A. (2012). Omega-3 fatty acids EPA and DHA: health benefits throughout life. *Adv. Nutr.* 3, 1–7. doi: 10.3945/an.111.000893

Szyjka, S. J., Mandal, S., Schoepp, N. G., Tyler, B. M., Yohn, C. B., Poon, Y. S., et al. (2017). Evaluation of phenotype stability and ecological risk of a genetically engineered alga in open pond production. *Algal Res.* 24, 378–386. doi: 10.1016/j.algal.2017.04.006

'T Lam, G. P., Vermuë, M. H., Eppink, M. H. M., Wijffels, R. H., and Van Den Berg, C. (2018). Multi-product microalgae biorefineries: from concept towards reality. *Trends Biotechnol.* 36, 216–227. doi: 10.1016/j.tibtech.2017.10.011

Tafreshi, A. H., and Shariati, M. (2009). Dunaliella biotechnology: methods and applications. *J. Appl. Microbiol.* 107, 14–35.

Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., and Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. *Int. J. Adv. Manuf. Technol.* 94, 3563–3576. doi: 10.1007/s00170-017-0233-1

Taunt, H. N., Stoffels, L., and Purton, S. (2018). Green biologics: The algal chloroplast as a platform for making biopharmaceuticals. *Bioengineered* 9, 48–54. doi: 10.1080/21655979.2017.1377867

Tay, P. K. R., Nguyen, P. Q., and Joshi, N. S. (2017). A synthetic circuit for mercury bioremediation using self-assembling functional amyloids. *ACS Synth. Biol.* 6, 1841–1850. doi: 10.1021/acssynbio.7b00137

Tetu, S. G., Sarker, I., Schrameyer, V., Pickford, R., Elbourne, L. D. H., Moore, L. R., et al. (2019). Plastic leachates impair growth and oxygen production in *Prochlorococcus*, the ocean's most abundant photosynthetic bacteria. *Commun. Biol.* 2:184.

Thomas, J. H. (1993). Thinking about genetic redundancy. *Trends Genet.* 9, 395–399. doi: 10.1016/0168-9525(93)90140-d

Tietz, S., Hall, C. C., Cruz, J. A., and Kramer, D. M. (2017). NPQ(T): a chlorophyll fluorescence parameter for rapid estimation and imaging of non-photochemical quenching of excitons in photosystem-II-associated antenna complexes. *Plant Cell Environ.* 40, 1243–1255. doi: 10.1111/pce.12924

Tiwari, B., Sellamuthu, B., Ouarda, Y., Drogui, P., Tyagi, R. D., and Buelna, G. (2017). Review on fate and mechanism of removal of pharmaceutical pollutants from wastewater using biological approach. *Bioresour. Technol.* 224, 1–12. doi: 10.1016/j.biortech.2016.11.042

Tran, M., Van, C., Barrera, D. J., Pettersson, P. L., Peinado, C. D., Bui, J., et al. (2013). Production of unique immunotoxin cancer therapeutics in algal chloroplasts. *Proc. Natl. Acad. Sci. U.S.A.* 110, E15–E22. doi: 10.1073/pnas.1214638110

Tran, M., Zhou, B., Pettersson, P. L., Gonzalez, M. J., and Mayfield, S. P. (2009). Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. *Biotechnol. Bioeng.* 104, 663–673. doi: 10.1002/bit.22446

Tran, N. H., Reinhard, M., and Gin, K. Y. (2018). Occurrence and fate of emerging contaminants in municipal wastewater treatment plants from different geographical regions-a review. *Water Res.* 133, 182–207. doi: 10.1016/j.watres.2017.12.029

Tuegel, E. J., Ingraffea, A. R., Eason, T. G., and Spottswood, S. M. (2011). Reengineering aircraft structural life prediction using a digital twin. *Int. J Aerosp. Eng.* 2011:154798.

Uhlemann, T. H. J., Lehmann, C., and Steinhilper, R. (2017). The digital twin: realizing the cyber-physical production system for industry 4.0. *Proc. CIRP* 61, 335–340. doi: 10.1016/j.procir.2016.11.152

Ummalyma, S. B., Gnansounou, E., Sukumaran, R. K., Sindhu, R., Pandey, A., and Sahoo, D. (2017). Bioflocculation: an alternative strategy for harvesting of microalgae – An overview. *Bioresour. Technol.* 242, 227–235. doi: 10.1016/j.biortech.2017.02.097

Ungerer, J., and Pakrasi, H. B. (2016). Cpf1 is a versatile tool for crispr genome editing across diverse species of Cyanobacteria. *Sci. Rep.* 6:39681. doi: 10.1038/srep39681

United Nations (2019). Available: https://population.un.org/wpp/Publications/ (accessed October, 2019).

van Bezouw, R. F. H. M., Keurentjes, J. J. B., Harbinson, J., and Aarts, M. G. M. (2019). Converging phenomics and genomics to study natural variation in plant photosynthetic efficiency. *Plant J.* 97, 112–133. doi: 10.1111/tpj.14190

Vance, P., and Spalding, M. H. (2005). Growth, photosynthesis, and gene expression in *Chlamydomonas* over a range of CO2 concentrations and CO2/O2 ratios: CO2 regulates multiple acclimation states. *Can. J. Bot.* 83, 796–809. doi: 10.1139/b05-064

Vasudevan, R., Gale, G. A. R., Schiavon, A. A., Puzorjov, A., Malin, J., Gillespie, M. J., et al. (2019). CyanoGate: a modular cloning suite for engineering cyanobacteria based on the Plant MoClo syntax. *Plant Physiol.* 180, 39–55. doi: 10.1104/pp.18.01401

Vasumathi, K. K., Premalatha, M., and Subramanian, P. (2012). Parameters influencing the design of photobioreactor for the growth of microalgae. *Renew. Sustain. Energy Rev.* 16, 5443–5450. doi: 10.1016/j.rser.2012.06.013

Vavitsas, K., Crozet, P., Vinde, M. H., Davies, F., Lemaire, S. D., and Vickers, C. E. (2019). The synthetic biology toolkit for photosynthetic microorganisms. *Plant Physiol.* 181, 14–27. doi: 10.1104/pp.19.00345

Vavitsas, K., Fabris, M., and Vickers, C. E. (2018). Terpenoid metabolic engineering in photosynthetic microorganisms. *Genes* 9:E520.

Vazhappilly, R., and Chen, F. (1998). Eicosapentaenoic acid and docosahexaenoic acid production potential of microalgae and their heterotrophic hrowth. *J. Am. Oil Chem. Soc.* 75, 393–397. doi: 10.1007/s11746-998-0057-0

Venkata Mohan, S., Rohit, M. V., Chiranjeevi, P., Chandra, R., and Navaneeth, B. (2015). Heterotrophic microalgae cultivation to synergize biodiesel production with waste remediation: Progress and perspectives. *Bioresour. Technol.* 184, 169–178. doi: 10.1016/j.biortech.2014.10.056

Vickers, C. E., Williams, T. C., Peng, B., and Cherry, J. (2017). Recent advances in synthetic biology for engineering isoprenoid production in yeast. *Curr. Opin. Chem. Biol.* 40, 47–56. doi: 10.1016/j.cbpa.2017.05.017

Vigani, M., Parisi, C., Rodríguez-Cerezo, E., Barbosa, M. J., Sijtsma, L., Ploeg, M., et al. (2015). Food and feed products from micro-algae: market opportunities and challenges for the EU. *Trends Food Sci. Technol.* 42, 81–92. doi: 10.1016/j.tifs.2014.12.004

Wang, B., Wang, J., Zhang, W., and Meldrum, D. R. (2012). Application of synthetic biology in cyanobacteria and algae. *Front. Microbiol.* 3:344. doi: 10.3389/fmicb.2012.00344

Wang, C., Dong, D., Zhang, L., Song, Z., Hua, X., and Guo, Z. (2019). Response of freshwater biofilms to antibiotic florfenicol and ofloxacin stress: role of extracellular polymeric substances. *Int. J. Environ. Res. Public Health* 16:715. doi: 10.3390/ijerph16050715

Wang, C., and Lan, C. Q. (2018). Effects of shear stress on microalgae – A review. *Biotechnol. Adv.* 36, 986–1002. doi: 10.1016/j.biotechadv.2018.03.001

Wang, K., Mandal, A., Ayton, E., Hunt, R., Zeller, M. A., and Sharma, S. (2016). "Chapter 6 - Modification of protein rich algal-biomass to form bioplastics and odor removal," in *Protein Byproducts*, ed. G. Singh Dhillon (Cambridge, MA: Academic Press), 107–117. doi: 10.1016/b978-0-12-802391-4.00006-9

Wei, L., Xin, Y., Wang, Q., Yang, J., Hu, H., and Xu, J. (2017). RNAi-based targeted gene knockdown in the model oleaginous microalgae *Nannochloropsis oceanica*. *Plant J.* 89, 1236–1250. doi: 10.1111/tpj.13411

Wells, M. L., Potin, P., Craigie, J. S., Raven, J. A., Merchant, S. S., Helliwell, K. E., et al. (2017). Algae as nutritional and functional food sources: revisiting our understanding. *J. Appl. Phycol.* 29, 949–982. doi: 10.1007/s10811-016-0974-5

Wen, Z.-Y., and Chen, F. (2003). Hetertrophic production of eicosapentaenoic acid by microalgae. *Biotechnol. Adv.* 21, 273–294. doi: 10.1016/s0734-9750(03)00051-x

Weyman, P. D., Beeri, K., Lefebvre, S. C., Rivera, J., Mccarthy, J. K., Heuberger, A. L., et al. (2015). Inactivation of *Phaeodactylum tricornutum* urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnol. J.* 13, 460–470. doi: 10.1111/pbi.12254

White, R. L., and Ryan, R. A. (2015). Long-term cultivation of algae in open-raceway ponds: lessons from the field. *Ind. Biotechnol.* 11, 213–220. doi: 10.1089/ind.2015.0006

Whitmore, A., Agarwal, A., and Da Xu, L. (2015). The internet of things—A survey of topics and trends. *Inform. Syst. Front.* 17, 261–274. doi: 10.1007/s10796-014-9489-2

Wichmann, J., Baier, T., Wentnagel, E., Lauersen, K. J., and Kruse, O. (2018). Tailored carbon partitioning for phototrophic production of (E)-alpha-bisabolene from the green microalga *Chlamydomonas reinhardtii*. *Metab. Eng.* 45, 211–222. doi: 10.1016/j.ymben.2017.12.010

Wijffels, R. H., Barbosa, M. J., and Eppink, M. H. M. (2010). Microalgae for the production of bulk chemicals and biofuels. *Biofuels Bioprod. Biorefin.* 4, 287–295. doi: 10.1002/bbb.215

Wurtzel, E. T., Vickers, C. E., Hanson, A. D., Millar, A. H., Cooper, M., Voss-Fels, K. P., et al. (2019). Revolutionizing agriculture with synthetic biology. *Nat. Plants* 5, 1207–1210. doi: 10.1038/s41477-019-0539-0

Xiong, J. Q., Kurade, M. B., and Jeon, B. H. (2017). Biodegradation of levofloxacin by an acclimated freshwater microalga, *Chlorella vulgaris*. *Chem. Eng. J.* 313, 1251–1257. doi: 10.1016/j.cej.2016.11.017

Xiong, J. Q., Kurade, M. B., and Jeon, B. H. (2018). Can microalgae remove pharmaceutical contaminants from water? *Trends Biotechnol.* 36, 30–44. doi: 10.1016/j.tibtech.2017.09.003

Xu, S., Gavin, J., Jiang, R., and Chen, H. (2017). Bioreactor productivity and media cost comparison for different intensified cell culture processes. *Biotechnol. Prog.* 33, 867–878. doi: 10.1002/btpr.2415

Yadav, J., Muend, S., Zhang, Y., and Rao, R. (2007). A phenomics approach in yeast links proton and calcium pump function in the Golgi. *Mol. Biol. Cell* 18, 1480–1489. doi: 10.1091/mbc.e06-11-1049

Yao, L., Cengic, I., Anfelt, J., and Hudson, E. P. (2016). Multiple gene repression in cyanobacteria using CRISPRi. *ACS Synth. Biol.* 5, 207–212. doi: 10.1021/acssynbio.5b00264

Zan, Y., Sun, Y., Kong, L., Miao, G., Bao, L., Wang, H., et al. (2018). Cover feature: formic acid-induced controlled-release hydrolysis of microalgae (*Scenedesmus*) to lactic acid over Sn-beta catalyst (ChemSusChem 15/2018). *ChemSusChem* 11, 2466–2466. doi: 10.1002/cssc.201801670

Zang, X., Liu, B., Liu, S., Arunakumara, K. K., and Zhang, X. (2007). Optimum conditions for transformation of *Synechocystis* sp. PCC 6803. *J. Microbiol.* 45, 241–245.

Zeller, M. A., Hunt, R., Jones, A., and Sharma, S. (2013). Bioplastics and their thermoplastic blends from Spirulina and *Chlorella microalgae*. *J. Appl. Polym. Sci.* 130, 3263–3275. doi: 10.1002/app.39559

Zhang, F., Kabeya, H., Kitagawa, R., Hirotsu, T., Yamashita, M., and Otsuki, T. (2000a). An exploratory research of PVC-*Chlorella* composite material (PCCM) as effective utilization of *Chlorella* biologically fixing CO2. *J. Mater. Sci.* 35, 2603–2609.

Zhang, F., Kabeya, H., Kitagawa, R., Hirotsu, T., Yamashita, M., and Otsuki, T. (2000b). A novel polyethylene–*Chlorella* composite. I. Characterization of *Chlorella* biologically fixing CO2. *J. Appl. Polym. Sci.* 77, 2278–2284. doi: 10.1002/1097-4628(20000906)77:10<2278::aid-app21>3.0.co;2-z

Zhang, J., Ma, F., Zhang, X., and Geng, A. (2018). Directed evolution of a homodimeric laccase from *Cerrena unicolor* BBP6 by random mutagenesis and *in vivo* assembly. *Int. J. Mol. Sci.* 19:2989. doi: 10.3390/ijms19102989

Zhao, T., Wang, W., Bai, X., and Qi, Y. (2009). Gene silencing by artificial microRNAs in *Chlamydomonas*. *Plant J.* 58, 157–164. doi: 10.1111/j.1365-313X.2008.03758.x

Zhou, J., Zhang, H., Meng, H., Zhu, Y., Bao, G., Zhang, Y., et al. (2014). Discovery of a super-strong promoter enables efficient production of heterologous proteins in cyanobacteria. *Sci. Rep.* 4:4500. doi: 10.1038/srep04500

Zhu, Y., Romain, C., and Williams, C. K. (2016). Sustainable polymers from renewable resources. *Nature* 540, 354–362.

frontiers
in Plant Science

Application of Transposon Insertion Sequencing to Agricultural Science

Belinda K. Fabian[1,2], Sasha G. Tetu[1,2]* and Ian T. Paulsen[1,2]*

[1] ARC Centre of Excellence in Synthetic Biology, Macquarie University, Sydney, NSW, Australia, [2] Department of Molecular Sciences, Macquarie University, Sydney, NSW, Australia

Many plant-associated bacteria have the ability to positively affect plant growth and there is growing interest in utilizing such bacteria in agricultural settings to reduce reliance on pesticides and fertilizers. However, our capacity to utilize microbes in this way is currently limited due to patchy understanding of bacterial–plant interactions at a molecular level. Traditional methods of studying molecular interactions have sought to characterize the function of one gene at a time, but the slow pace of this work means the functions of the vast majority of bacterial genes remain unknown or poorly understood. New approaches to improve and speed up investigations into the functions of bacterial genes in agricultural systems will facilitate efforts to optimize microbial communities and develop microbe-based products. Techniques enabling high-throughput gene functional analysis, such as transposon insertion sequencing analyses, have great potential to be widely applied to determine key aspects of plant-bacterial interactions. Transposon insertion sequencing combines saturation transposon mutagenesis and high-throughput sequencing to simultaneously investigate the function of all the non-essential genes in a bacterial genome. This technique can be used for both *in vitro* and *in vivo* studies to identify genes involved in microbe-plant interactions, stress tolerance and pathogen virulence. The information provided by such investigations will rapidly accelerate the rate of bacterial gene functional determination and provide insights into the genes and pathways that underlie biotic interactions, metabolism, and survival of agriculturally relevant bacteria. This knowledge could be used to select the most appropriate plant growth promoting bacteria for a specific set of conditions, formulating crop inoculants, or developing crop protection products. This review provides an overview of transposon insertion sequencing, outlines how this approach has been applied to study plant-associated bacteria, and proposes new applications of these techniques for the benefit of agriculture.

Keywords: biocontrol, plant growth promoting bacteria, fertilizer, microbiome, pesticide, transposon insertion sequencing, transposon mutagenesis

INTRODUCTION

A major component of agricultural ecosystems is the microbiota (bacteria, archaea, protists, fungi, and viruses) present on plants and in soils. Interactions between microbes and plants can be detrimental, commensal, or favorable (Buée et al., 2009). The benefits provided by microbes in agricultural systems include breaking down organic matter, fixing nutrients, making nutrients

available for plant use, nutrient recycling, defense against plant pathogens and improving abiotic stress tolerance (Naik et al., 2019). To survive and thrive in their niche, plant-associated microbes need to be able to evade host defenses, use the nutrients available from the host and successfully compete with other microbes (Gonzalez-Mula et al., 2019).

To maximize crop production and minimize losses due to biotic and abiotic stresses it is increasingly clear that we need to take advantage of the benefits that beneficial microbes can provide. However, this is currently difficult to achieve with only a limited understanding of microbial gene functions and their interactions with plants on a molecular level. An increasingly common approach when studying agricultural microbiota is to ascertain microbial community composition using metagenomics (Shelake et al., 2019). This allows for the collection of genetic information about more members of the community than just the microbes that are able to be cultivated (Levy et al., 2018). Decreasing sequencing costs have facilitated the rapid generation of vast amounts of genomic data. However, collecting this sequence information does not readily translate into understanding what functions are being performed within a microbial community and what genes are involved (van Opijnen and Camilli, 2013; Vorholt et al., 2017).

The traditional methods for experimental demonstration of gene function, such as gene knockouts and over-expression, happen at a much slower pace than the rate at which genomic information is accumulating (Chang et al., 2016). The laborious and often problematic nature of this gene function characterization (see section " Saturation Transposon Insertion Mutagenesis and High Throughput DNA Sequencing") means only a small proportion of genes in a small number of microbes from agricultural systems have been experimentally investigated (Price et al., 2018a). For microbes that have not been experimentally characterized, sequence homology, genome location or similarity of domains is often used to infer gene function, but predictions of gene or organism function generated by these computational approaches are often erroneous, especially when the closest characterized relative is separated by a large phylogenetic distance (Mutalik et al., 2019). There are many circumstances where even these methods geared at prediction of gene function do not yield any functional information and genes remain annotated as hypothetical proteins. The combination of these issues results in the functions of many microbial genes remaining unknown or poorly understood (Van Opijnen and Camilli, 2012; Jarboe, 2018). This is currently limiting our understanding of the role many microbes play in agricultural and other systems and hampers interpretation of microbiome survey data and efforts to formulate optimal plant associated microbial communities (Shelake et al., 2019).

Most of what is currently known about genes involved in beneficial plant interactions has been determined by comparing genomes and individually testing genes that correlate with a phenotype of interest (Levy et al., 2018). A high throughput method of gathering gene functional information and relating genotype to phenotype is required to rapidly advance our understanding of the complex molecular interactions that occur

between microbes and plants (van Opijnen and Camilli, 2013; Perry et al., 2016) and move beyond cataloging microbial community members toward a mechanistic understanding of agricultural microbiomes (Poole et al., 2018).

THE NEED FOR ALTERNATIVES TO PESTICIDES AND FERTILIZERS

To provide food and fiber for the rapidly growing global population, crop yields need to increase dramatically (Godfray et al., 2012; Pardey et al., 2014). In the past, agricultural intensification has been enabled by the use of pesticides, fertilizers, machinery, and increased plowing depth and sowing density. While this has previously been adequate to increase crop yields and meet global demand (Emmerson et al., 2016), these methods are no longer delivering sufficient increases in yields and are contributing to biodiversity declines which can have flow on effects for agricultural yields, for example, pollination services that are critical for food production (Gallai et al., 2009; Tscharntke et al., 2012). Declines in the rates at which crop yields are increasing, the changing climate and abiotic stresses (especially water availability and salinity) mean that new approaches are necessary to meet demand for food and fiber (Ray et al., 2012; Backer et al., 2018). The major ways to meet this demand while holding agricultural land area stable are to increase crop yields and/or dramatically reduce the quantity of crops lost to disease (Phalan et al., 2011; Godfray et al., 2012).

One path to increasing biomass accumulation and increasing crop yields is through the addition of fertilizers to crop lands. When fertilizers with elevated levels of nitrogen and phosphorus are added to fields they can create an imbalance in the inorganic nutrients present in the soil (Gosal et al., 2018). The addition of phosphorus in this way results in a large proportion binding to soil particles (immobilization) and not being biologically available for plant use (Khan et al., 2007). Excessive use of fertilizers can lead to surrounding lands being contaminated through run-off, changes to the physico-chemical properties of soil and detrimental impacts on the native soil microbiota which can have flow on effects for the resilience of agricultural soils and crop yields (Jangid et al., 2008; Santhanam et al., 2015; Naik et al., 2019). For these reasons, there is increasing recognition that alternatives to inorganic fertilizers are needed (Cordell et al., 2009).

Every year 20–40% of global food production is lost worldwide to plant pests and diseases despite efforts to control crop diseases using pesticides and other crop management techniques (FAO, 2016). Crop diseases are mainly controlled through the use of pesticides and farming practices, such as crop rotation, integrated pest management and the development of disease resistant crop varieties (Syed Ab Rahman et al., 2018). Continuously using pesticides as the major weapon against crop pathogens ultimately results in the pathogens developing resistance and pesticides losing their effectiveness. This has been observed in *Zymoseptoria tritici*, the fungal causative agent for Septoria tritici blotch, progressively evolving resistance to fungicides in Europe since the early 2000s, and since 2011 in Tasmania, Australia

(McDonald et al., 2019). The Fungicide Resistance Action Committee (FRAC), a pesticide industry expert group, lists over 429 instances of plant pathogens becoming resistant to fungicides (Fungicide Resistance Action Committee [FRAC], 2018).

In many instances, farmers have to apply higher doses and multiple pesticides to combat this rising resistance; eventually plant pathogens will no longer be controllable using existing pest control methods (Lucas et al., 2015). In addition to rising pathogen resistance, the application of pesticides to crops may cause detrimental off-target effects for many other members of these ecosystems. Commensal and beneficial microbes are often killed and insects that are protecting crops from herbivory or providing essential services such as pollination are damaged or removed from these habitats (Bünemann et al., 2006; Geiger et al., 2010). With the combination of the slowing rate of discovery and development of new pesticides (Borel, 2017), resistance developed by pathogens and the harm pesticides cause to the environment it is widely agreed that new methods of disease control are needed (Syed Ab Rahman et al., 2018).

BOOSTING CROP YIELDS VIA APPLICATION OF BENEFICIAL BACTERIA

Many plant-associated bacteria are now understood to have the capacity to positively affect plant growth and there is growing interest in utilizing such bacteria in agricultural settings to reduce reliance on pesticides and fertilizers (Naik et al., 2019). These plant growth promoting bacteria colonize the surface of plants and the thin layer of soil surrounding the plant roots (rhizosphere; Zhang et al., 2017). These bacteria use the nutrients exuded from plant roots to fuel their growth and in turn can stimulate plant growth (Buée et al., 2009).

Plant growth is promoted by such bacteria in a number of direct and indirect ways (**Figure 1**). Firstly, beneficial bacteria can assist with biofertilization, increasing the bioavailability of nutrients for plant use. Examples include atmospheric nitrogen fixation (Ke et al., 2019), solubilizing phosphate (Arkhipova et al., 2019; Wu et al., 2019) and the production of iron binding siderophores which can be taken up by plants (Flores-Félix et al., 2015). The second way rhizobacteria can stimulate plant growth is through the production of phytohormones (also referred to as plant growth regulators). These hormones can stimulate plant growth in the same way as endogenous plant hormones. For example, the endophytic bacteria *Sphingomonas* sp. LK11 releases gibberellins which increases tomato plant shoot length and root weight (Khan et al., 2014). Thirdly, beneficial bacteria can ameliorate the effect of abiotic stresses and lead to increased tolerance of abiotic stresses. The presence of *Pseudomonas putida* AKMP7 has been shown to promote wheat growth under heat stress (Ali et al., 2011) and under waterlogged conditions *Achromobacter xylosoxidans* Fd2 induces waterlogging tolerance which results in 46.5% higher yield from Holy basil plants (*Ocimum sanctum*; Barnawal et al., 2012). The final direct method of promoting plant growth is rhizoremediation. This is when soil bacteria degrade pollutants and attenuate the effects of

toxins in the soil. The removal of these contaminants from the environment reduces the stress on the plant and results in increased plant growth (Correa-Garcia et al., 2018). For example, rice growth in heavy metal contaminated soils is boosted by cadmium-resistant *Ochrobactrum* sp. and lead- and arsenic-resistant *Bacillus* sp. (Pandey et al., 2013).

The indirect method for promoting plant growth is when beneficial microbes are used to control plant pathogens, an eco-friendly and sustainable process termed biocontrol (Handelsman and Stabb, 1996). Biocontrol bacteria reduce the impacts that pathogens have on plants through three interacting mechanisms (**Figure 1**). Firstly, the beneficial microbes grow in the same spaces as the pathogens and compete with them for nutrients and niches. For example, the biocontrol bacteria *Chryseobacterium* sp. WR21 controls bacterial wilt disease by more effectively competing for tomato root exudates compared to the pathogenic bacteria *Ralstonia solanacearum* (Huang et al., 2017). The second method for reducing the impact of plant pathogens is through the induction of plant systemic resistance. *Bacillus amyloliquefaciens* SQR9 primes the immune system of Arabidopsis against the pathogens *Pseudomonas syringae* pv. tomato DC3000 and *Botrytis cinerea* (Wu et al., 2018). The final technique used by plant growth promoting microbes is antibiosis through the production of antibiotics or metabolites. For example, the biocontrol bacteria *Pseudomonas protegens* Pf-5 produces the antibiotic pyoluteorin to protect cotton plants against infection by the oomycete pathogen *Pythium ultimum* (Howell and Stipanovic, 1980).

Consistent and reliable performance in protecting crops against pathogens is an essential characteristic of any plant growth promoting bacteria before consideration for development into a commercial crop protection or crop enhancement product. As a commercial product a bacterial inoculum will be applied in a broad range of settings with diverse physico-chemical parameters. Beneficial bacteria need to not only cope with a range of environments and climatic conditions, but actively enhance plant growth under such conditions. In addition to colonizing the crop host, the bacteria need to compete with existing soil microbiota for resources and be safe for the environment (Rosconi et al., 2016). The bacterial inoculum may be applied in conjunction with chemical pesticides and fertilizers, so it is advantageous if it is able to survive and maintain plant growth promoting activity in the presence of these additions and must also be able to survive packaging, transport and storage before it is applied to a crop (Glick, 2015; Backer et al., 2018). Ideally, the biocontrol agent should also be able to tolerate a range of environmental stresses, for example, drought, salinity, temperature or heavy metal pollution (Compant et al., 2019).

Over 30 bacterial strains have been successfully commercialized into crop inoculants (not including the many *Rhizobia* spp. used for legume crop inoculation; Glick, 2015), but there are many more beneficial strains of bacteria that have potential for development. Even for the beneficial bacteria that have been developed into commercial biocontrol agents, there is often no strong functional understanding of their mechanisms of action (Bardin et al., 2015). There is a clear need for high throughput functional analyses to investigate beneficial bacteria
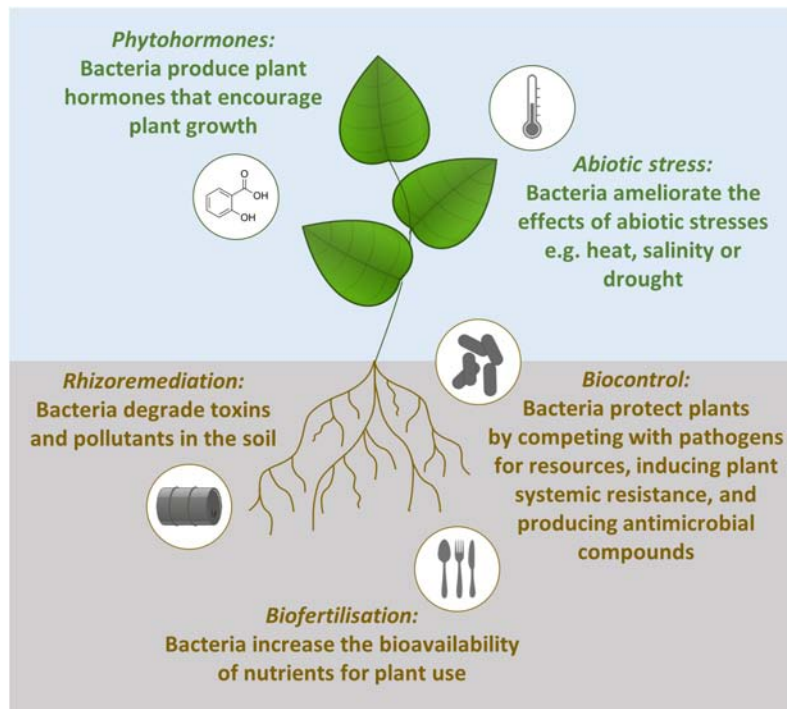
FIGURE 1 | Mechanisms of plant growth promotion by beneficial bacteria. Phytohormone production, abiotic stress tolerance, rhizoremediation and biofertilization are direct methods of plant growth promotion. In contrast, biocontrol is an indirect method of plant growth promotion operating through competition with pathogens, priming plant defenses and producing antimicrobial compounds.

with potential for development into commercial agents (Backer et al., 2018). This would permit much faster identification of the metabolic pathways they use to gather resources, their modes of action in a variety of environments and hosts, and the roadblocks that are limiting their current application.

SATURATION TRANSPOSON INSERTION MUTAGENESIS AND HIGH THROUGHPUT DNA SEQUENCING

To further our understanding of plant-microbe and microbe-microbe interactions a faster method of determining microbial gene function is needed. The traditional methods of ascribing functions to genes (for example, making single gene knockouts or heterologous expression) are laborious and very slow as mutants are screened individually (Royet et al., 2019). In addition, these methods can be problematic, for example, gene expression in a heterologous bacterial host may exhibit an altered phenotype due to the different genomic context of the original organism (Levy et al., 2018). These methods are also reductionist; mutants are studied in isolation meaning that interactions between genes cannot be identified and genes may not appear important in pure culture even though they may be crucial *in vivo* (Poole et al., 2018).

In the last 10 years, the combination of saturation transposon mutagenesis with high throughput sequencing has enabled identification of the suite of genes and pathways that are

important for plant growth promotion in a single experiment (Cole et al., 2017). The fundamental element that underlies this methodology is the transposon; a class of genetic elements that can move to different locations within a genome (Hoffman and Jendrisak, 2002). When a transposon inserts into a genome it may interrupt or modify the function of a gene or regulatory element resulting in a change in the organism's phenotype, allowing phenotypic changes to be linked to specific gene disruptions (Hoffman et al., 2000; Reznikoff, 2008).

The first step in transposon insertion sequencing is generating a saturated mutant library by using a transposable element to generate a large population of cells with mutations at many locations throughout the genome (Barquist et al., 2013). The choice of transposable element depends on the organism; most commonly used are the Tn5 or *mariner* transposons (reviewed in Chao et al., 2016). Tn5 transposons insert randomly into the genome, while *mariner* transposons insert at TA sites. TA site occurrence is relatively regular across the genome but can vary at local scales. Knowing how many possible mariner transposons insertion sites there are in a genome allows for statistical calculations of transposon insertion saturation which is not possible in Tn5 based mutant libraries. On the other hand, as Tn5 transposons do not require a specific site for insertion into the genome these libraries can potentially have higher transposon insertion density (Chao et al., 2016).

A transposon insertion into an essential gene is most often fatal and mutants of these genes will not be included in the resulting mutant library (essentiality is relative to growth

conditions; Barquist et al., 2013). Insertions within non-essential genes result in recoverable cells which may show phenotypic or fitness differences under some growth conditions (Hoffman et al., 2000). The goal of this approach is to generate a saturated mutant library of cells which collectively have transposon insertions throughout the genome, enabling assessment of the importance of each gene to fitness under a range of testable growth conditions (Barquist et al., 2016). For example, simultaneous fitness assessment of the hundreds of thousands of mutants in a library is possible by subjecting the mutant library to a defined condition such as exposure to a pesticide or colonization of a host (Chao et al., 2016; Calero et al., 2017; **Figure 2**).

Multiple methods for transposon insertion sequencing have arisen in the past decade, including TraDIS (Langridge et al., 2009), Tn-Seq (van Opijnen et al., 2009), HITS (Gawronski et al., 2009), and INSeq (Goodman et al., 2009). These methods have been further refined, for example Tn-Seq Circle (Gallagher et al., 2011), RB-TnSeq (Wetmore et al., 2015) and improvements to INSeq (Goodman et al., 2011) to reduce the time and resources required to carry out these studies or improve the effectiveness of the procedure. Other developments, such as TraDISort, enable physical enrichment of desired mutants via cell sorting (Hassan et al., 2016). All of these methods use similar approaches (**Figure 2**): (1) construct a saturated mutant library by transposon mutagenesis of a bacteria with a transposable element containing an antibiotic resistance cassette; (2) subject the library to challenge assays that affect the fitness of the cells containing mutated conditionally essential genes; (3) DNA extraction from recovered mutants and enrichment for transposon-chromosome junctions; (4) massively parallel sequencing of the mutant library input (pre-challenge) and output (post-challenge) pools; (5) bioinformatic analysis to map the location of each transposon insertion site in the genome; and (6) comparison of the number of reads at each transposon insertion site to determine the conditionally essential genes for the condition of interest.

The differences between the transposon insertion sequencing methods lie in the construction of the mutant library and the enrichment for transposon-chromosome junctions before sequencing. Tn-Seq and INSeq mutant libraries are constructed using mariner transposons. The inclusion of *Mme*I restriction sites at each end of the transposon allows for the chromosomal DNA either side of the transposon to be cut at a fixed length of 20 base pairs during preparation of Tn-Seq and INSeq libraries before sequencing. Adapter ligation follows and PCR and purification steps are carried out. The original INSeq protocol includes a PAGE gel purification step, while Tn-Seq uses agarose gel purification (van Opijnen and Camilli, 2013). Improvements to INSeq purification steps include the use of a biotinylated primer for PCR and magnetic bead purification to decrease the time and cost involved as well as increasing the technique's sensitivity (Goodman et al., 2011). RB-TnSeq uses random barcodes with the transposon inserted into each cell. This allows for the chromosomal location of the transposon insertion to be mapped from the initial sequencing run and each subsequent sequencing run takes less time and funding (Wetmore et al., 2015). In HITS and TraDIS there is no restriction site included in the transposon, instead

random shearing of the genomic DNA is employed, resulting in variable length DNA fragments. In TraDIS, adapters are ligated to all DNA fragments and enrichment of fragments of interest (those containing a transposon-chromosome junction) is achieved through PCR. The HITS technique includes these steps as well as an affinity purification step to purify the PCR amplicons of interest containing the transposon (Gawronski et al., 2009). The details of each method and the advantages and disadvantages have been comprehensively reviewed by Barquist et al. (2013) and van Opijnen and Camilli (2013).

A complementary approach to loss-of-function assays is dual-barcoded shotgun expression library sequencing (Dub-Seq), where gain-of-function overexpression libraries are used in competition assays to gather insights into gene function and the fitness of mutants that have increased gene dosage (Mutalik et al., 2019). Combining data from both gain- and loss-of-function approaches can lead to increased precision of biological insights and help to overcome the limitations when either approach is used in isolation (Mutalik et al., 2019).

The advantage of transposon insertion sequencing approaches is that gene functions can be identified without prior knowledge of their likely roles (Hassan et al., 2016). When using traditional methods of determining gene function, some information about a gene is necessary to establish a starting point for characterization; this can be problematic in situations when sequence databases and genome context do not give many (or any) clues about the function of hypothetical proteins. In contrast, transposon insertion sequencing approaches allow for the simultaneous collection of gene functional information for a large number of genes, even when no other information is available. By comparing the number of transposon-insertion reads in the mutant library before and after a selection pressure is applied the genes involved in coping with that pressure can be identified (diCenzo et al., 2018). Changes in the relative abundances of each mutant shows which genes have been positively or negatively affected by the challenge conditions and therefore their possible functions (Paulsen et al., 2017). Combining information from multiple challenge assays produces multi-dimensional data and leads to a greater functional understanding of both annotated and hypothetical proteins. Such information is useful in selecting candidate genes which are indicated to be required for improved fitness under the conditions of interest for further, targeted functional characterization.

Using a transposon insertion sequencing approach allows for the identification of the genetic factors likely to affect bacterial fitness in a complex condition or environment (Barquist et al., 2013). Transposon insertion sequencing techniques can be used to investigate the suite of genes required for *in vivo* growth in a single experiment, such as assays on plants or plant tissues (Cole et al., 2017). Employing the saturated transposon library in these types of complex, multi-faceted scenarios provides comprehensive information about the genes required to cope with *in vivo* selection pressures, how genes interact and create a complex phenotype, and can reveal the range of modes of action used by bacteria (van Opijnen and Camilli, 2013; Kohl et al., 2019).

FIGURE 2 | Transposon insertion sequencing methodology for agriculturally relevant bacteria. High throughput sequencing of the starting population of bacterial mutants in the saturated transposon insertion library determines the locations of the transposon insertions in the genome. An aliquot of the mutant library is grown with a selective pressure (for example, persistence on root surfaces in either sterilized or natural soil, pesticide tolerance, abiotic stress), the surviving cells are recovered, and the transposon-chromosome junctions are sequenced. Changes in the abundances of each mutant in the starting pool compared with the challenge pool indicate which genes increase or decrease bacterial fitness under the challenge conditions. This provides information on the possible functions of these genes that can then be investigated by targeted follow up experiments.

To ensure transposon insertion sequencing is used appropriately there are some considerations that need to be taken into account in the design of these high throughput studies. Transposon insertion sequencing allows a mutant library of bacteria to be studied in conjunction with the existing microbiome of an organism to identify genes related to interacting with the host and the host's microbiome (including beneficial, commensal and pathogenic microbes; Barquist et al., 2013). This technique has been successfully applied in animal models. For example, a *Salmonella enterica* serovar Typhimurium saturated mutant library was screened using chickens, pigs and cattle to identify the genes required for infection in each host (Chaudhuri et al., 2013). When using a dense mutant library in agricultural studies, experiments need to be designed to avoid bottlenecking the population due to insufficient plant surface area or resources to support bacterial growth (van Opijnen and Camilli, 2013). A related consideration when working with bacterial inoculants as part of an existing microbiome is the challenge of recovering enough mutant cells and extracting enough DNA to be able to conduct the high-throughput sequencing. Failure to recover sufficient library after a challenge could lead to biasing of the samples and insufficient sequence depth for reliable results. When analyzing data from transposon insertion sequencing experiments it is important to consider that fitness effects may not be solely due to the gene interrupted by the transposon insertion. If a transposon inserts into an operon or promoter region there could be downstream effects on gene expression.

There are also some limitations to where and how transposon insertion sequencing approaches can be applied. As with other molecular techniques, this approach is only feasible for genetically tractable microbes which can be cultured in the laboratory (Levy et al., 2018). Additional limitations include the inability to identify genes for which there is functional redundancy in the genome (Mutalik et al., 2019) or those encoding "public goods" (products that are secreted and shared with the community; Royet et al., 2019) as these functional deficiencies are unlikely to reduce the fitness of the individuals affected whilst they are in a mixed population (Cole et al., 2017).

The majority of initial studies using transposon insertion sequencing approaches focused on human pathogens, but these methods are starting to be used to examine other microbial lifestyles. Recent studies of human and animal pathogens include susceptibility and resistance of *Escherichia coli* to phage infection (Cowley et al., 2018), the mechanism of antibiotic binding in *Staphylococcus aureus* (Santiago et al., 2018), and resistance of *Borrelia burgdorferi* to reactive oxygen and nitrogen species (Ramsey et al., 2017). Non-pathogenic microbes are also being explored using transposon insertion sequencing methods, for example, regulation and biosynthesis of holdfast assembly has been investigated in *Caulobacter crescentus* (Hershey et al., 2019), circadian clock proteins have been identified in cyanobacteria (Welkie et al., 2018) and previously unknown bacterial amino acid biosynthesis genes have been identified in heterotrophic bacteria (Price et al., 2018b).

TRANSPOSON INSERTION SEQUENCING AND THE FUTURE OF AGRICULTURE

Current Uses of Transposon Insertion Sequencing for Plant-Associated Microbes

Transposon insertion sequencing techniques have been applied to study a relatively small number of plant-associated bacteria (**Table 1**). One application has been the discovery of genes critical

TABLE 1 | Summary of transposon insertion sequencing studies of plant growth promoting and plant pathogenic bacteria.

Microbe	Plant host	Method	Main finding	References
Plant and soil colonization				
Pseudomonas aeruginosa PGPR2	Corn	INSeq	Identified 108 genes that contribute to fitness during root colonization-amino acid catabolism, stress adaptation, detoxification, signal transduction, and transport functions	Sivakumar et al., 2019
Pseudomonas simiae WCS417r	*Arabidopsis thaliana*	RB-TnSeq	Identified 115 genes required for colonization of the root system and 243 genes where disruption positively affects colonization – large number of these encode proteins for amino acid transport, suggesting that auxotrophy confers a selective advantage in the plant-associated environment rich with exuded amino acids and sugars	Cole et al., 2017
Pseudomonas syringae pv. *syringae* B728a	Green bean, *Phaseolus vulgaris*	RB-TnSeq	Genes for amino acid and polysaccharide biosynthesis are important for fitness on the leaf surface and in the leaf interior (apoplast). Genes for type III secretion system and syringomycin synthesis are important in the apoplast	Helmann et al., 2019a
Bacillus thuringiensis	Clover, *Triflorium hybridum*	TraDIS	Identified genes required for survival in soil – ability to form spores was a key characteristic for multiplying and colonizing the soil	Bishop et al., 2014
Pseudomonas sp. WCS365	*Arabidopsis thaliana*	Tn-Seq	Identified novel factors required to evade plant defenses by conducting screens on wild-type and immunocompromised plants – identified 231 genes that increase rhizosphere fitness and verified seven genes involved in avoiding plant defenses	Liu et al., 2018
Rhizobial symbiosis				
Rhizobium leguminosarum bv. *viciae* 3841 *Sinorhizobium meliloti* RM1021 *Agrobacterium tumefaciens* UBAPF2	Legumes	INSeq	Used a mariner transposon to mutagenize three members of *Rhizobiaceae* at high frequency. Generated a saturated mutant library in RLV3841 on rich media and identified essential genes	Perry and Yost, 2014
Rhizobium leguminosarum bv. *viciae* 3841	Legumes	INSeq	Identified genes required for growth on minimal mannitol containing media. Compared to genes required for growth on rich media and found ~10% of chromosomal genes required under both growth conditions. Identified plasmid genes encoding functional activities important to central physiology	Perry et al., 2016
Rhizobium leguminosarum bv. *viciae* 3841	Legumes	INSeq	Identified genes important for growth on minimal media supplemented with glucose or succinate at both 1 or 21% O_2. Analyzed central carbon metabolism pathways to examine roles in glucose and succinate metabolism and quantify the ability of pathways to compensate for single mutations	Wheatley et al., 2017
Sinorhizobium meliloti RmP3499	Legumes	Tn-seq	Evaluated the interactions between the chromosomal genes and the extra-chromosomal replicons when grown in defined and rich media – ten percent of chromosomal genes found to functionally interact with the replicons	diCenzo et al., 2018
Sinorhizobium meliloti 1021	*Medicago truncatula*	Tn-seq	Identified genes and pathways that contribute to sensitivity to NCR247 peptide, a plant signaling peptide with antimicrobial activity which plays a key role during rhizobial differentiation in the nodule – found 78 genes and several pathways that affect sensitivity to NCR247	Arnold et al., 2017
Sinorhizobium meliloti Rm2011	Legumes	Signature tagged mutagenesis combined with next generation sequencing	Identified an additional 4,473 transposon insertion sites in a mutant library, bringing the total number of mutants with known insertion sites to 9,562 – 59% of predicted protein-coding genes have a transposon insertion	Serrania et al., 2017

(Continued)

134

TABLE 1 | Continued

Microbe	Plant host	Method	Main finding	References
Pathogenicity and virulence				
Dickeya dadantii	Chicory	Tn-Seq	Identified metabolic pathways and genes required for growth of a necrotrophic pathogen *in planta* – the uridine monophosphate, purine and leucine, cysteine, and lysine biosynthetic pathways are essential for bacterial survival in the plant; RsmC and GcpA are important for regulating the infection process; and glycosylation of flagellin confers fitness during plant infection	Royet et al., 2019
Salmonella enterica serovar Typhimurium ATCC 14028	Campari tomato, *Solanum lycopersicum* cv. Campari	Tn-Seq	When colonizing tomatoes, *Salmonella enterica* uses a distinct set of plant-associated genes which only partly overlap with the genes used for virulence in animals and by other phytopathogens during plant infection	de Moraes et al., 2017
Pantoea stewartii ssp. *stewartii*	Sweetcorn, *Zea mays* cv. Jubilee	Tn-Seq	*In planta* study showed 486 genes are important for survival (reduced fitness when mutated) and loss of six genes increased fitness. O*mpC, lon* and *ompA* were found to play a role in virulence	Duong et al., 2018
Agrobacterium tumefaciens	Tomato, *Solanum lycopersicum*	Tn-Seq	Combined plant metabolomics, transcriptomics and Tn-seq to identify genes and pathways involved in host exploitation. On sucrose the *pgi, pycA, cisY*, and *sdhCDA* genes are crucial and sucrose breakdown involves the Entner–Doudoroff and tricarboxylic acid (TCA) pathways. When growing on GHB (γ-hydroxybutyric acid) as a carbon source the *blc* genes are crucial and breakdown requires the TCA cycle and gluconeogenesis	Gonzalez-Mula et al., 2019
Liberibacter crescens BT-1	Mountain papaya	TraDIS	Identified 238 essential genes in common between *L. crescens* and *L. asiaticus* (unculturable pathogen of citrus) as potential antimicrobial targets	Lai et al., 2016
Pseudomonas syringae pv. *actinidiae*	Kiwifruit, *Actinidia* spp.	TraDIS	Identified auxotrophic, motility and lipopolysaccharide (LPS) biosynthesis mutants using a "phenotype of interest" library. Used a "mutant of interest" library to identify a putative LPS mutant	Mesarich et al., 2017
Bacterial physiology				
32 bacteria, including eight plant-associated bacteria: *Azospirillum brasilense* Sp245 *Burkholderia phytofirmans* PsJN *Desulfovibrio vulgaris* Miyazaki F *Dyella japonica* UNC79MFTsu3.2 *Herbaspirillum seropedicae* SmR1 *Klebsiella michiganensis* M5a1 *Pseudomonas simiae* WCS417 *Sinorhizobium meliloti* 1021	Wheat roots Onion roots Paddy field Arabidopsis root Cereal root Rice rhizosphere Wheat rhizosphere Alfalfa	RB-TnSeq	Across all 32 bacteria, identified phenotypes for 11,779 genes that are not annotated with a specific function, including 4,135 genes that encode proteins that do not belong to any characterized family in either Pfam or TIGRFAMs. Combined functional associations with comparative sequence analysis to identify putative DNA repair proteins and proposed specific functions for transporter proteins, catabolic enzymes, and domains of unknown function	Price et al., 2018a
Pseudomonas putida KT2440		Tn-Seq	Transposon insertion libraries generated in *P. putida* KT2440 and *P. putida* KT2440 Δ*fcs*. *P. putida* found to be significantly more tolerant than *E. coli* to p-coumaric acid, a phenolic acid present in soils and plants. Genes involved in maintenance of membrane structure and efflux of solvent compounds were important when exposed to p-coumaric acid	Calero et al., 2017
Pseudomonas syringae pv. *syringae* B728a	Bean, *Phaseolus vulgaris*	RB-TnSeq	Identified and used a hyper-susceptible mutant to identify the substrates of redundant transporters. Identified genes that contribute to tolerance of acridine orange, acriflavine and berberine	Helmann et al., 2019b
Herbaspirillum seropedicae SmR1	Rice, maize and sorghum	Tn-Seq	Constructed transposon insertion libraries – identified 395 genes essential for growth in rich media	Rosconi et al., 2016

135

for leaf, root and rhizosphere colonization (Cole et al., 2017; Liu et al., 2018; Helmann et al., 2019a; Sivakumar et al., 2019). An RB-TnSeq study of *P. syringae* found 31 genes important for colonization of common bean leaf surfaces, including amino acid and polysaccharide biosynthesis genes. Sixty-five genes were important for fitness in apoplastic environments, including those that make up the type III secretion system and syringomycin synthesis genes (Helmann et al., 2019a). Further *in planta* investigation of single gene mutants confirmed that Δ*eftA* and Δ*Psyr_0920* mutants have decreased fitness in the apoplast, suggesting that these glycosyltransferase domain-containing genes are important for the evasion of plant surveillance systems (Helmann et al., 2019a). Investigations of *Pseudomonas simiae* found a large number of amino acid transporter genes are important for survival in the rhizosphere, suggesting that utilizing the amino acids exuded from plant roots confers a selective advantage over microbes that devote resources to the synthesis of these compounds (Cole et al., 2017). Inoculation of wild type and immunocompromised Arabidopsis plants with *Pseudomonas* sp. WC365 identified 231 genes involved in rhizosphere competence (Liu et al., 2018). Follow-up experiments with targeted single gene knockout mutants confirmed the role of the mutants Δ*spuC* and Δ*morA* in inducing pattern-triggered immunity of the Arabidopsis host and inhibiting plant growth. Point mutations in the conserved GGDEF motif of MorA showed that this protein acts as a phosphodiesterase which inhibits biofilm formation (Liu et al., 2018).

Another application of transposon insertion sequencing examined interactions between plants and bacteria using legumes and their nitrogen-fixing symbionts, *Rhizobium leguminosarum* and *Sinorhizobium meliloti*. These studies identified genes important for surviving the environmental conditions inside root nodules (Wheatley et al., 2017) and genes that confer resistance to plant antimicrobial peptides (Arnold et al., 2017). Investigations of a previously unknown, but highly conserved, gene *smc03872* showed that it protects *S. meliloti* against the antimicrobial activities of the NCR247 peptide of the host plant *Medicago truncatula*. Site specific mutations in this gene revealed that SMc03872 is a lipoprotein with peptidase activity that can provide its protective effect as long it is anchored in a membrane (Arnold et al., 2017).

Transposon insertion sequencing in pathogenic bacteria can be used to rapidly identify virulence factors, essential genes and modes of action of bacterial pathogens (**Table 1**). An *in planta* Tn-Seq study of *S. enterica* showed that the colonization of tomatoes uses a distinct set of plant-associated genes which only partly overlap with the genes used for virulence in animals and by other phytopathogens during plant infection (de Moraes et al., 2017). Determining which genes are important for plant host colonization by this bacteria is important as the populations of plant-associated bacteria can act as reservoirs for infection of animal hosts, including humans. Another pathogenic lifestyle is the necrotrophic growth of rot-causing pathogens such as *Dickeya dadantii*. A Tn-Seq study of this pathogen showed that the biosynthetic pathways for uridine monophosphate, purines, and the amino acids leucine, cysteine and lysine, are essential for survival on chicory plants (Royet et al., 2019). The study also

determined that the RsmC and GcpA regulators are important in the infection process and glycosylation of flagellin confers fitness during plant infection. Transposon insertion sequencing studies can also identify potential antimicrobial targets, such as genes that are essential for growth of pathogenic bacteria but are not essential in other bacterial strains.

Transposon insertion sequencing studies have also provided insights into the fundamental physiology of plant pathogenic and plant growth promoting bacteria (**Table 1**). As part of a larger study, Price et al. (2018a) created saturated mutant libraries of eight plant-associated bacteria and conducted fitness testing using multiple carbon and nitrogen sources as well as up to 55 different stress conditions. This study recorded phenotypes for thousands of genes and proposed specific functions for transporter proteins, catabolic enzymes, and domains of unknown function. Transporter substrates (Helmann et al., 2019b) and abiotic and biotic stress tolerance genes (Calero et al., 2017) can also be investigated using these techniques. A Tn-Seq study in *P. putida* KT2440 revealed that genes related to membrane stability are important for p-coumaric acid tolerance (Calero et al., 2017). Follow up studies with single gene knockout mutations showed strong involvement of the cytochrome c maturation system (*ccm* operon) and the *ttg2* operon (encodes an ABC transporter) in this tolerance (Calero et al., 2017). Applying these techniques to a broad range of agriculturally relevant bacteria has the potential to inform efforts to cultivate beneficial plant-microbe interactions and reduce the burden associated with agricultural pathogens.

Future Avenues to Explore With Transposon Insertion Sequencing

There are a wide variety of ways that the use of transposon insertion sequencing techniques in bacteria could be expanded beyond these current applications to rapidly increase our understanding of bacterial functions relevant for plant growth promotion and agriculture. Bacterial growth in the rhizosphere is influenced by the soil type and structure, soil organic matter, macronutrient levels and moisture levels (Shelake et al., 2019). These physico-chemical properties of the soil can vary at the centimeter scale so a bacterial inoculum applied to a crop may encounter a wide range of these conditions (Fierer, 2017). Bacteria residing in the phyllosphere (above ground plant surfaces) have to be able to cope with widely fluctuating environmental conditions, including light levels, water availability, temperature and UV radiation (Shelake et al., 2019). Transposon insertion sequencing techniques could be applied to plant-associated bacteria to determine which genes are involved in fundamental processes required for living in these environments and what genetic factors make some bacteria highly successful at colonizing these environments.

Bacterial chemotaxis and motility-related genes are essential for detecting signals produced by plants and moving toward chemical attractants or away from deterrents (Scharf et al., 2016). Flagella play a critical role in attachment to plant surfaces and the initial formation of biofilms; both essential processes for biocontrol bacteria to protect plants against pathogens (Nian

et al., 2007; Rudrappa et al., 2008). In an agricultural context, bacterial cells with enhanced motility and the ability to form biofilms are better competitors for resources and have greater colonization efficiency, leading to greater enhancement of plant growth (Barahona et al., 2010; Backer et al., 2018; Pinski et al., 2019). Transposon insertion sequencing has successfully been used to identify 14 non-flagellar genes and intergenic regions involved in enhanced motility of *E. coli* EC958, showing that genes outside the flagellar and chemotaxis regulons are important for motility (Kakkanat et al., 2017). This approach could be applied to identify genes important for motility in agriculturally relevant bacteria.

Transposon insertion sequencing can even provide new insights into processes that are extremely well studied. For example, despite decades of sporulation research, a recent Tn-Seq study of *Bacillus subtilis* led to the identification of 24 additional genes involved in sporulation and mutants in which sporulation is delayed or accelerated (Meeske et al., 2016). The genes rhizobial bacteria use to adapt to rhizospheric and root endophytic conditions have been extensively studied, but recent transposon insertion sequencing studies have revealed new insights into the genes important for utilizing carbon compounds exuded by roots and surviving in low oxygen environments inside root nodules (Wheatley et al., 2017).

Most genetic studies of plant-associated bacteria are a snapshot at a particular moment in time and cannot shed light on changing interactions, such as across plant life stages. Executing a succession of transposon insertion sequencing experiments could give insights into the different stages of a plant growth promoting bacteria's interaction with its host (diCenzo et al., 2019). An *in vivo* transposon insertion sequencing experiment of host infection over a 2-week time series was conducted using *Edwardsiella piscicida*, a fish pathogen. This study examined the fitness of mutants over the course of host infection and identified genes that contribute to colonization of the fish host (Yang et al., 2017). When compared with traditional endpoint transposon insertion sequencing studies the time series experiment identified more genes affecting *in vivo* fitness and yielded new insights into possible targets for vaccine development (Yang et al., 2017). The selective pressures bacteria experience during plant colonization or infection are not constant over time, so using time-series transposon insertion sequencing studies of plant associated bacteria could provide insights into the relative importance of particular genes during these dynamic processes.

When living on plant surfaces plant growth promoting bacteria need to be able to cope with biotic and abiotic stresses, such as plant toxins, anthropogenic pollutants, salinity, acidic or alkaline conditions and temperature. Some of these stresses are commonly experienced in combination, for example, drought and heat stress are often experienced together due to the changing climate. Stress physiologically affects both plants and microbes and can have flow-on effects for plant-microbe interactions (Naik et al., 2019). Transposon insertion sequencing has been successfully used in non-plant-associated organisms to study bacterial responses to stress conditions. A Tn-seq study using high osmolarity, reactive oxygen species and temperature successfully identified individual *E. coli* genes that confer stress resistance as well as gene combinations that work synergistically to impart improved stress resistance (Lennen and Herrgard, 2014). Transposon insertion sequencing analysis of the pathogen *B. burgdorferi* uncovered 66 genes not previously known to be involved in resistance to reactive oxygen and nitrogen species (Ramsey et al., 2017). Conducting *in vitro* or *in planta* transposon insertion sequencing studies that incorporate rhizospheric or phyllospheric stresses could rapidly advance our understanding of the genes employed by plant growth promoting bacteria in stressful conditions.

To extend our knowledge of how genes interact to create a phenotype of interest, transposon mutant libraries could be created using plant-associated bacteria with a genetic background where a specific gene of interest is inactivated. This would allow interactions with the inactivated gene, redundant genes and the components of regulatory networks to be identified (Barquist et al., 2013; van Opijnen and Camilli, 2013). To determine the interaction profiles of five genes of interest in the human pathogen *Streptococcus pneumoniae*, independent transposon mutant libraries were constructed using cells with a background knockout of one of the five genes (van Opijnen et al., 2009). This study revealed interactions that exacerbated or improved fitness defects compared to the single mutants and showed that one of the genes of interest is a master regulator of complex carbohydrate metabolism. Using this technique in plant-associated bacteria could reveal previously unknown gene interactions and shed light on gene networks crucial for growth on plant surfaces.

POTENTIAL AGRICULTURAL APPLICATIONS

Using transposon insertion sequencing allows us to move beyond traditional methods of molecular genetic analyses where bacterial phenotypes and single mutations are linked and perform en masse identification of genes and pathways involved in plant growth promotion (diCenzo et al., 2019). These high-throughput techniques can rapidly accelerate the rate of gene functional determination and provide rapid insights into the genes and pathways that underlie biotic interactions, metabolism and survival of agriculturally relevant bacteria. This will dramatically increase our knowledge of mechanisms that beneficial bacteria use to promote plant growth and the genes and pathways that plant pathogens rely on when causing disease.

Understanding the growth requirements of plant growth promoting bacteria could assist with developing formulations for maximum efficiency when applied to crops (Kohl et al., 2019). For example, if a beneficial bacteria is known to be able to metabolize a particular compound the bacterial inoculant could be formulated to include that compound to create optimal growth conditions for the beneficial bacteria (Syed Ab Rahman et al., 2018; Rocha et al., 2019). This would increase the likelihood of it being able to effectively compete with the native microbiota, colonize the plant surface and persist over

time (Naik et al., 2019). A similar effect could be achieved if plant breeding selected crop genotypes that have the ability to support beneficial bacteria, for example by exuding a specific chemical compound from the roots (Syed Ab Rahman et al., 2018; Compant et al., 2019).

By conducting a series of high-throughput transposon insertion sequencing experiments a large amount of information could be leveraged in the design of a bacterial inoculum that is able to promote plant growth across diverse crop species, a broad range of soils, climatic conditions, and abiotic and biotic stresses (Compant et al., 2019). The development of inoculums consisting of multiple bacterial species that use a range of modes of action for biocontrol would reduce the risk of pathogens being able to evolve resistance to a single biocontrol strain (Backer et al., 2018). By incorporating bacteria with a suite of activities a bacterial consortium could be effective across more stages of the plant lifecycle, resulting in increased plant growth. Studies of bacterial consortia have shown that diverse bacterial inoculants show greater survival, reduced levels of plant disease and increased plant biomass (Hu et al., 2016, 2017). Harnessing rapidly generated information about bacterial gene function from transposon insertion sequencing studies has great potential for creating optimized bacterial consortia for crop inoculation across a range of conditions.

Alternative strategies to leverage the plant growth promoting effects of beneficial bacteria are using secreted bacterial metabolites or selecting desirable plant traits. The compounds that beneficial bacteria secrete can act as biostimulants and agricultural soils could be supplemented with these microbial metabolites to trigger plant responses and stimulate plant growth (Backer et al., 2018). The insights from transposon insertion sequencing could uncover ways to boost production of these bioactive secondary metabolites by plant-associated bacteria (Shelake et al., 2019). The findings from transposon insertion sequencing studies could also assist in the identification of desirable plant traits that support the attraction of beneficial bacteria or promote resistance to pathogen colonization. These traits could be selected in plant breeding programs (Syed Ab Rahman et al., 2018; Compant et al., 2019).

If there is a shift in public opinion about genetic engineering, then information gleaned from transposon insertion sequencing studies could be used to direct alterations of plant-associated microbes for the development of crop protection and crop enhancement products. Utilizing information from transposon insertion sequencing to identify targets for gene editing could lead to faster design and development of control methods for crop diseases (Miesel et al., 2003; Royet et al., 2019). Once high throughput analysis rapidly identifies a target gene or pathway, a pathogen could be genetically engineered to create an attenuated or avirulent isolate for use in biological control (Muñoz et al., 2019). These engineered strains can still colonize plants, compete with their pathogenic kin for space and nutrients, and may even induce plant defenses, but they themselves are no longer infectious (Couteaudier, 1992). For example, in glasshouse and field trials *hrpG* mutants of the tomato bacterial spot pathogen *Xanthomonas campestris* pv.

vesicatoria 75-3 provided up to 76% reduction in disease severity (Moss et al., 2007). Use of genetic engineering in this way could dramatically shorten the timeframe for the development of crop protection products.

Traditional methods of bacterial gene editing and the alteration of protein production, such as gene knockouts and heterologous expression, are now complemented by new gene editing techniques, such as CRISPR-Cas9. Researchers repurposed this bacterial and archaeal cell defense mechanism as a generic gene editing system when they realized that it could be used to recognize and cut specific DNA sequences (Barrangou and Doudna, 2016). As a gene editing mechanism CRISPR-Cas9 is faster and easier to carry out and less expensive than traditional methods, such as homologous recombination (Shelake et al., 2019). Further into the future, synthetic biology and synthetic genomics hold the promise of designing and building bespoke microorganisms capable of specific plant growth promoting functions. Transposon insertion sequencing will play an important role in identifying the genes and pathways that underpin plant growth promotion and will facilitate the design of such synthetic microbes. The performance of these synthetic beneficial microorganisms can be iteratively improved through the classic synthetic biology "design, build, test, learn" cycle (McArthur et al., 2015; Pouvreau et al., 2018).

CONCLUSION

We are currently in the midst of a paradigm shift about the role microbes play in our lives. We are transitioning from a simplistic view that microbes are largely disease-causing agents to a more sophisticated understanding of the ubiquitous nature of commensal microbes and the benefits microbes can provide for their hosts. Increasingly society is looking to microbes as a solution for complex problems. Just as research into the human microbiome is revolutionizing medicine and personalized medicine is on the horizon, the future of agriculture may come from rapid advances in our understanding of plant-associated bacterial gene functions and creating tailor made microbial solutions for particular environments, diseases and cropping regimes.

The end goal of these studies is not to have a one-size-fits-all plant growth promoting solution. Conducting transposon insertion studies on plant-associated bacteria will rapidly provide vast amounts of functional information about substantial numbers of agriculturally relevant bacterial genes. As the pressure to use agricultural land most effectively with the least chemical inputs continues to increase, this knowledge will inform agricultural practice so we can increase crop yields and allow us to tackle the challenges of overpopulation and the changing climate.

AUTHOR CONTRIBUTIONS

BF wrote the manuscript. ST and IP conceptualized the idea for this work and critically revised the manuscript. BF, ST, and IP approved the final version of the manuscript.

REFERENCES

Ali, S. Z., Sandhya, V., Grover, M., Linga, V. R., and Bandi, V. (2011). Effect of inoculation with a thermotolerant plant growth promoting *Pseudomonas putida* strain AKMP7 on growth of wheat (*Triticum* spp.) under heat stress. *J. Plant Interact.* 6, 239–246. doi: 10.1080/17429145.2010.545147

Arkhipova, T., Galimsyanova, N., Kuzmina, L., Vysotskaya, L., Sidorova, L., Gabbasova, I., et al. (2019). Effect of seed bacterization with plant growth-promoting bacteria on wheat productivity and phosphorus mobility in the rhizosphere. *Plant Soil Environ.* 65, 313–319. doi: 10.17221/752/2018-pse

Arnold, M. F. F., Shabab, M., Penterman, J., Boehme, K. L., Griffitts, J. S., and Walker, G. C. (2017). Genome-Wide sensitivity analysis of the microsymbiont *Sinorhizobium meliloti* to symbiotically important, defensin-like host peptides. *mBio* 8:e1060-17. doi: 10.1128/mBio.01060-17.

Backer, R., Rokem, J. S., Ilangumaran, G., Lamont, J., Praslickova, D., Ricci, E., et al. (2018). Plant growth-promoting rhizobacteria: context, mechanisms of action, and roadmap to commercialization of biostimulants for sustainable agriculture. *Front. Plant Sci.* 9:1473. doi: 10.3389/fpls.2018.01473

Barahona, E., Navazo, A., Yousef-Coronado, F., Aguirre de Cárcer, D., Martínez-Granero, F., Espinosa-Urgel, M., et al. (2010). Efficient rhizosphere colonization by *Pseudomonas fluorescens* f113 mutants unable to form biofilms on abiotic surfaces. *Environ. Microbiol.* 12, 3185–3195. doi: 10.1111/j.1462-2920.2010.02291.x

Bardin, M., Ajouz, S., Comby, M., Lopez-Ferber, M., Graillot, B., Siegwart, M., et al. (2015). Is the efficacy of biological control against plant diseases likely to be more durable than that of chemical pesticides? *Front. Plant Sci.* 6:566. doi: 10.3389/fpls.2015.00566

Barnawal, D., Bharti, N., Maji, D., Chanotiya, C. S., and Kalra, A. (2012). 1-Aminocyclopropane-1-carboxylic acid (ACC) deaminase-containing rhizobacteria protect *Ocimum sanctum* plants during waterlogging stress via reduced ethylene generation. *Plant Physiol. Biochem.* 58, 227–235. doi: 10.1016/j.plaphy.2012.07.008

Barquist, L., Boinett, C. J., and Cain, A. K. (2013). Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol.* 10, 1161–1169. doi: 10.4161/rna.24765

Barquist, L., Mayho, M., Cummins, C., Cain, A. K., Boinett, C. J., Page, A. J., et al. (2016). The TraDIS toolkit: Sequencing and analysis for dense transposon mutant libraries. *Bioinformatics* 32, 1109–1111. doi: 10.1093/bioinformatics/btw022

Barrangou, R., and Doudna, J. A. (2016). Applications of CRISPR technologies in research and beyond. *Nat. Biotechnol.* 34, 933–941. doi: 10.1038/nbt.3659

Bishop, A. H., Rachwal, P. A., and Vaid, A. (2014). Identification of genes required by *Bacillus thuringiensis* for survival in soil by transposon-directed insertion site sequencing. *Curr. Microbiol.* 68, 477–485. doi: 10.1007/s00284-013-0502-7

Borel, B. (2017). CRISPR, microbes and more are joining the war against crop killers. *Nature* 543, 302–304. doi: 10.1038/543302a

Buée, M., De Boer, W., Martin, F., van Overbeek, L., and Jurkevitch, E. (2009). The rhizosphere zoo: an overview of plant-associated communities of microorganisms, including phages, bacteria, archaea, and fungi, and of some of their structuring factors. *Plant Soil* 321, 189–212. doi: 10.1007/s11104-009-9991-3

Bünemann, E. K., Schwenke, G. D., and Van Zwieten, L. (2006). Impact of agricultural inputs on soil organisms - a review. *Soil Res.* 44, 379–406. doi: 10.1071/SR05125

Calero, P., Jensen, S. I., Bojanovič, K., Lennen, R., Koza, A., and Nielsen, A. T. (2017). Genome-wide identification of tolerance mechanisms towards p-coumaric acid in *Pseudomonas putida. Biotechnol. Bioeng.* 15, 762–774. doi: 10.1002/bit.26495

Chang, Y.-C., Hu, Z., Rachlin, J., Anton, B. P., Kasif, S., Roberts, R. J., et al. (2016). COMBREX-DB: an experiment centered database of protein function: knowledge, predictions and knowledge gaps. *Nucleic Acids Res.* 44, D330–D335. doi: 10.1093/nar/gkv1324

Chao, M. C., Abel, S., Davis, B. M., and Waldor, M. K. (2016). The design and analysis of transposon insertion sequencing experiments. *Nat. Rev. Microbiol.* 14, 119–128. doi: 10.1038/nrmicro.2015.7

Chaudhuri, R. R., Morgan, E., Peters, S. E., Pleasance, S. J., Hudson, D. L., Davies, H. M., et al. (2013). Comprehensive assignment of roles for *Salmonella typhimurium* genes in intestinal colonization of food-producing animals. *PLoS Genet.* 9:e1003456. doi: 10.1371/journal.pgen.1003456

Cole, B. J., Feltcher, M. E., Waters, R. J., Wetmore, K. M., Mucyn, T. S., Ryan, E. M., et al. (2017). Genome-wide identification of bacterial plant colonization genes. *PLoS Biol.* 15:e2002860. doi: 10.1371/journal.pbio.2002860

Compant, S., Samad, A., Faist, H., and Sessitsch, A. (2019). A review on the plant microbiome: ecology, functions, and emerging trends in microbial application. *J. Adv. Res.* 19, 29–37. doi: 10.1016/j.jare.2019.03.004

Cordell, D., Drangert, J. O., and White, S. (2009). The story of phosphorus: global food security and food for thought. *Glob. Environ. Chang.* 19, 292–305. doi: 10.1016/j.gloenvcha.2008.10.009

Correa-Garcia, S., Pande, P., Seguin, A., St-Arnaud, M., and Yergeau, E. (2018). Rhizoremediation of petroleum hydrocarbons: a model system for plant microbiome manipulation. *Microb. Biotechnol.* 11, 819–832. doi: 10.1111/1751-7915.13303

Couteaudier, Y. (1992). "Competition for carbon in soil and rhizosphere, a mechanism involved in biological control of fusarium wilts," in *Biological Control of Plant Diseases*, eds E. C. Tjamos, G. C. Papavizas, and R. J. Cook (Boston, MA: Springer), 99–104. doi: 10.1007/978-1-4757-9468-7_13

Cowley, L. A., Low, A. S., Pickard, D., Boinett, C. J., Dallman, T. J., Day, M., et al. (2018). Transposon insertion sequencing elucidates novel gene involvement in susceptibility and resistance to phages T4 and T7 in *Escherichia coli* O157. *mBio* 9:e00705-18. doi: 10.1128/mBio

de Moraes, M. H., Desai, P., Porwollik, S., Canals, R., Perez, D. R., Chu, W., et al. (2017). *Salmonella* persistence in tomatoes requires a distinct set of metabolic functions identified by transposon insertion sequencing. *Appl. Environ. Microbiol.* 83, 1–18. doi: 10.1128/AEM.03028-16

diCenzo, G. C., Benedict, A. B., Fondi, M., Walker, G. C., Finan, T. M., Mengoni, A., et al. (2018). Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium *Sinorhizobium meliloti. PLoS Genet.* 14:e1007357. doi: 10.1371/journal.pgen.1007357

diCenzo, G. C., Zamani, M., Checcucci, A., Fondi, M., Griffitts, J. S., Finan, T. M., et al. (2019). Multidisciplinary approaches for studying rhizobium-legume symbioses. *Can. J. Microbiol.* 65, 1–33. doi: 10.1139/cjm-2018-0377

Duong, D. A., Jensen, R. V., and Stevens, A. M. (2018). Discovery of *Pantoea stewartii* ssp. stewartii genes important for survival in corn xylem through a Tn-Seq analysis. *Mol. Plant Pathol.* 19, 1929–1941. doi: 10.1111/mpp.12669

Emmerson, M., Morales, M. B., Oñate, J. J., Batáry, P., Berendse, F., Liira, J., et al. (2016). How agricultural intensification affects biodiversity and ecosystem services. *Adv. Ecol. Res.* 55, 43–97. doi: 10.1016/bs.aecr.2016.08.005

FAO (2016). *The State of Food and Agriculture 2016: Climate Change, Agriculture and Food Security*. Rome: FAO.

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* 15, 579–590. doi: 10.1038/nrmicro.2017.87

Flores-Félix, J. D., Silva, L. R., Rivera, L. P., Marcos-García, M., García-Fraile, P., Martínez-Molina, E., et al. (2015). Plants probiotics as a tool to produce highly functional fruits: the case of *Phyllobacterium* and vitamin C in strawberries. *PLoS One* 10:e0122281. doi: 10.1371/journal.pone.0122281

Fungicide Resistance Action Committee [FRAC] (2018). *List of Plant Pathogenic Organisms Resistant to Disease Control Agents*. Available

online at: www.frac.info/docs/default-source/publications/list-of-resistant-plant-pathogens/list-of-resistant-plant-pathogenic-organisms_may-2018.pdf?sfvrsn=a2454b9a_2 (accessed September 21, 2019).

Gallagher, L. A., Shendure, J., and Manoil, C. (2011). Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* Using Tn-seq. *mBio* 2:e00315-10. doi: 10.1128/mBio.00315-10

Gallai, N., Salles, J. M., Settele, J., and Vaissière, B. E. (2009). Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecol. Econ.* 68, 810–821. doi: 10.1016/j.ecolecon.2008.06.014

Gawronski, J. D., Wong, S. M. S., Giannoukos, G., Ward, D. V., and Akerley, B. J. (2009). Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16422–16427. doi: 10.1073/pnas.0906627106

Geiger, F., Bengtsson, J., Berendse, F., Weisser, W. W., Emmerson, M., Morales, M. B., et al. (2010). Persistent negative effects of pesticides on biodiversity and biological control potential on European farmland. *Basic Appl. Ecol.* 11, 97–105. doi: 10.1016/j.baae.2009.12.001

Glick, B. R. (ed.) (2015). "Introduction to plant growth-promoting bacteria," in *Beneficial Plant-Bacterial Interactions* (Cham: Springer), 243.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2012). The challenge of food security. *Science* 327, 812–818. doi: 10.4337/9780857939388

Gonzalez-Mula, A., Lachat, J., Mathias, L., Naquin, D., Lamouche, F., Mergaert, P., et al. (2019). The biotroph *Agrobacterium tumefaciens* thrives in tumors by exploiting a wide spectrum of plant host metabolites. *New Phytol.* 222, 455–467. doi: 10.1111/nph.15598

Goodman, A. L., McNulty, N. P., Zhao, Y., Leip, D., Mitra, R. D., Lozupone, C. A., et al. (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* 6, 279–289. doi: 10.1016/j.chom.2009.08.003

Goodman, A. L., Wu, M., and Gordon, J. I. (2011). Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nat. Protoc.* 6, 1969–1980. doi: 10.1038/nprot.2011.417

Gosal, S. K., Gill, G. K., Sharma, S., and Walia, S. S. (2018). Soil nutrient status and yield of rice as affected by long-term integrated use of organic and inorganic fertilizers. *J. Plant Nutr.* 41, 539–544. doi: 10.1080/01904167.2017.1392570

Handelsman, J., and Stabb, E. V. (1996). Biocontrol of soilborne plant pathogens. *Plant Cell* 8, 1855–1869. doi: 10.1105/tpc.8.10.1855

Hassan, K. A., Cain, A. K., Huang, T., Liu, Q., Elbourne, L. D. H., Boinett, C. J., et al. (2016). Fluorescence-based flow sorting in parallel with transposon insertion site sequencing identifies multidrug efflux systems in *Acinetobacter baumannii*. *mBio* 7:e01200-16. doi: 10.1128/mBio.01200-16

Helmann, T. C., Deutschbauer, A. M., and Lindow, S. E. (2019a). Genome-wide identification of *Pseudomonas syringae* genes required for fitness during colonization of the leaf surface and apoplast. *Proc. Natl. Acad. Sci. U.S.A.* 116:201908858. doi: 10.1073/pnas.1908858116

Helmann, T. C., Ongsarte, C. L., Lam, J., Deutschbauer, A. M., and Lindow, S. E. (2019b). Genome-wide transposon screen of a *Pseudomonas syringae* mexB mutant reveals the substrates of efflux transporters. *bioRxiv* [Preprint]. doi: 10.1101/684605

Hershey, D. M., Fiebig, A., Crosson, S., Fuqua, C., and Laub, M. T. (2019). A genome-wide analysis of adhesion in *Caulobacter crescentus* identifies new regulatory and biosynthetic components for holdfast assembly. *mBio* 10:e02273-18. doi: 10.1128/mBio.02273-18

Hoffman, L. M., and Jendrisak, J. J. (2002). Transposomes: a system for identifying genes involved in bacterial pathogenesis. *Methods Enzymol.* 358, 128–140. doi: 10.1016/S0076-6879(02)58085-X

Hoffman, L. M., Jendrisak, J. J., Meis, R. J., Goryshin, I. Y., and Reznikoff, W. S. (2000). Transposome insertional mutagenesis and direct sequencing of microbial genomes. *Genetica* 108, 19–24. doi: 10.1023/A:1004083307819

Howell, C. R., and Stipanovic, R. D. (1980). Suppression of *Pythium ultimum*-induced damping-off of cotton seedlings by *Pseudomonas fluorescens* and its antibiotic, Pyoluteorin. *Phytopathology* 70, 712–715. doi: 10.1094/Phyto-70-712

Hu, J., Wei, Z., Friman, V.-P. P., Gu, S.-H. H., Wang, X.-F. F., Eisenhauer, N., et al. (2016). Probiotic diversity enhances rhizosphere microbiome function and plant disease suppression. *mBio* 7:e01790-16. doi: 10.1128/mBio.01790-16

Hu, J., Wei, Z., Weidner, S., Friman, V. P., Xu, Y. C., Shen, Q. R., et al. (2017). Probiotic *Pseudomonas* communities enhance plant growth and nutrient assimilation via diversity-mediated ecosystem functioning. *Soil Biol. Biochem.* 113, 122–129. doi: 10.1016/j.soilbio.2017.05.029

Huang, J., Wei, Z., Hu, J., Yang, C., Gu, Y., Mei, X., et al. (2017). *Chryseobacterium nankingense* sp. *nov.* WR21 effectively suppresses *Ralstonia solanacearum* growth via intensive root exudates competition. *BioControl* 62, 567–577. doi: 10.1007/s10526-017-9812-1

Jangid, K., Williams, M. A., Franzluebbers, A. J., Sanderlin, J. S., Reeves, J. H., Jenkins, M. B., et al. (2008). Relative impacts of land-use, management intensity and fertilization upon soil microbial community structure in agricultural systems. *Soil Biol. Biochem.* 40, 2843–2853. doi: 10.1016/j.soilbio.2008.07.030

Jarboe, L. R. (2018). Improving the success and impact of the metabolic engineering design, build, test, learn cycle by addressing proteins of unknown function. *Curr. Opin. Biotechnol.* 53, 93–98. doi: 10.1016/j.copbio.2017.12.017

Kakkanat, A., Phan, M. D., Lo, A. W., Beatson, S. A., and Schembri, M. A. (2017). Novel genes associated with enhanced motility of *Escherichia coli* ST131. *PLoS One* 12:e0176290. doi: 10.1371/journal.pone.0176290

Ke, X., Feng, S., Wang, J., Lu, W., Zhang, W., Chen, M., et al. (2019). Effect of inoculation with nitrogen-fixing bacterium *Pseudomonas stutzeri* A1501 on maize plant growth and the microbiome indigenous to the rhizosphere. *Syst. Appl. Microbiol.* 42, 248–260. doi: 10.1016/j.syapm.2018.10.010

Khan, A. L., Waqas, M., Kang, S.-M., Al-Harrasi, A., Hussain, J., Al-Rawahi, A., et al. (2014). Bacterial endophyte *Sphingomonas* sp. LK11 produces gibberellins and IAA and promotes tomato plant growth. *J. Microbiol.* 52, 689–695. doi: 10.1007/s12275-014-4002-7

Khan, M. S., Zaidi, A., and Wani, P. A. (2007). Role of phosphate-solubilizing microorganisms in sustainable agriculture — A review. *Agron. Sustain. Dev.* 27, 29–43. doi: 10.1051/agro:2006011

Kohl, J., Kolnaar, R., and Ravensberg, W. J. (2019). Mode of action of microbial biological control agents against plant diseases: relevance beyond efficacy. *Front. Plant Sci.* 10:845. doi: 10.3389/fpls.2019.00845

Lai, K. K., Davis-Richardson, A. G., Dias, R., and Triplett, E. W. (2016). Identification of the genes required for the culture of *Liberibacter crescens*, the closest cultured relative of the *Liberibacter* plant pathogens. *Front. Microbiol.* 7:547. doi: 10.3389/fmicb.2016.00547

Langridge, G. C., Phan, M. D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., et al. (2009). Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. *Genome Res.* 19, 2308–2316. doi: 10.1101/gr.097097.109

Lennen, R. M., and Herrgard, M. J. (2014). Combinatorial strategies for improving multiple-stress resistance in industrially relevant *Escherichia coli* strains. *Appl. Environ. Microbiol.* 80, 6223–6242. doi: 10.1128/AEM.01542-14

Levy, A., Conway, J. M., Dangl, J. L., and Woyke, T. (2018). Elucidating bacterial gene functions in the plant microbiome. *Cell Host Microbe* 24, 475–485. doi: 10.1016/j.chom.2018.09.005

Liu, Z., Beskrovnaya, P., Melnyk, R. A., Hossain, S. S., Khorasani, S., O'Sullivan, L. R., et al. (2018). A genome-wide screen identifies genes in rhizosphere-associated *pseudomonas* required to evade plant defenses. *mBio* 9:e00433-18. doi: 10.1128/mBio.00433-18

Lucas, J. A., Hawkins, N. J., and Fraaije, B. A. (2015). The evolution of fungicide resistance. *Adv. Appl. Microbiol.* 90, 29–92. doi: 10.1016/bs.aambs.2014.09.001

McArthur, G. H. IV, Nanjannavar, P. P., Miller, E. H., and Fong, S. S. (2015). Integrative metabolic engineering. *AIMS Bioeng.* 2, 93–103. doi: 10.3934/bioeng.2015.3.93

McDonald, M. C., Renkin, M., Spackman, M., Orchard, B., Croll, D., Solomon, P. S., et al. (2019). Rapid parallel evolution of azole fungicide resistance in Australian populations of the wheat pathogen *Zymoseptoria tritici*. *Appl. Environ. Microbiol.* 85, e1908–e1918. doi: 10.1128/AEM.01908-18

Meeske, A. J., Rodrigues, C. D. A., Brady, J., Lim, H. C., Bernhardt, T. G., and Rudner, D. Z. (2016). High-throughput genetic screens identify a large and diverse collection of new sporulation genes in *Bacillus subtilis*. *PLoS Biol.* 14:e1002341. doi: 10.1371/journal.pbio.1002341

Mesarich, C. H., Rees-George, J., Gardner, P. P., Ghomi, F. A., Gerth, M. L., Andersen, M. T., et al. (2017). Transposon insertion libraries for the characterization of mutants from the kiwifruit pathogen *Pseudomonas syringae* pv. *actinidiae*. *PLoS One* 12:e0172790. doi: 10.1371/journal.pone.0172790

Miesel, L., Greene, J., and Black, T. A. (2003). Genetic strategies for antibacterial drug discovery. *Nat. Rev. Genet.* 4, 442–456. doi: 10.1038/nrg1086

Moss, W. P., Byrne, J. M., Campbell, H. L., Ji, P., Bonas, U., Jones, J. B., et al. (2007). Biological control of bacterial spot of tomato using *hrp* mutants of *Xanthomonas campestris* pv. *vesicatoria*. *Biol. Control* 41, 199–206. doi: 10.1016/j.biocontrol.2007.01.008

Muñoz, I. V., Sarrocco, S., Malfatti, L., Baroncelli, R., and Vannacci, G. (2019). CRISPR-Cas for fungal genome editing: a new tool for the management of plant diseases. *Front. Plant Sci.* 10:135. doi: 10.3389/fpls.2019.00135

Mutalik, V. K., Novichkov, P. S., Price, M. N., Owens, T. K., Callaghan, M., Carim, S., et al. (2019). Dual-barcoded shotgun expression library sequencing for high-throughput characterization of functional traits in bacteria. *Nat. Commun.* 10:308. doi: 10.1038/s41467-018-08177-8

Naik, K., Mishra, S., Srichandan, H., Singh, P. K., and Sarangi, P. K. (2019). Plant growth promoting microbes: potential link to sustainable agriculture and environment. *Biocatal. Agric. Biotechnol.* 21:101326. doi: 10.1016/j.bcab.2019.101326

Nian, H., Zhang, J., Song, F., Fan, L., and Huang, D. (2007). Isolation of transposon mutants and characterization of genes involved in biofilm formation by *Pseudomonas fluorescens* TC222. *Arch. Microbiol.* 188, 205–213. doi: 10.1007/s00203-007-0235-8

Pandey, S., Ghosh, P. K., Ghosh, S., De, T. K., and Maiti, T. K. (2013). Role of heavy metal resistant *Ochrobactrum* sp. and *Bacillus* spp. strains in bioremediation of a rice cultivar and their PGPR like activities. *J. Microbiol.* 51, 11–17. doi: 10.1007/s12275-013-2330-7

Pardey, P. G., Beddow, J. M., Hurley, T. M., Beatty, T. K. M., and Eidman, V. R. (2014). A bounds analysis of world food futures: global agriculture through to 2050. *Aust. J. Agric. Resour. Econ.* 58, 571–589. doi: 10.1111/1467-8489.12072

Paulsen, I. T., Cain, A. K., and Hassan, K. A. (2017). Physical enrichment of transposon mutants from saturation mutant libraries using the TraDISort approach. *Mob. Genet. Elem.* 7:e1313805. doi: 10.1080/2159256X.2017.1313805

Perry, B. J., Akter, M. S., and Yost, C. K. (2016). The use of transposon insertion sequencing to interrogate the core functional genome of the legume symbiont *Rhizobium leguminosarum*. *Front. Microbiol.* 7:1873. doi: 10.3389/fmicb.2016.01873

Perry, B. J., and Yost, C. K. (2014). Construction of a *mariner*-based transposon vector for use in insertion sequence mutagenesis in selected members of the *Rhizobiaceae*. *BMC Microbiol.* 14:298. doi: 10.1186/s12866-014-0298-z

Phalan, B., Balmford, A., Green, R. E., and Scharlemann, J. P. W. (2011). Minimising the harm to biodiversity of producing more food globally. *Food Policy* 36, S62–S71. doi: 10.1016/j.foodpol.2010.11.008

Pinski, A., Betekhtin, A., Hupert-Kocurek, K., Mur, L. A. J., and Hasterok, R. (2019). Defining the genetic basis of plant–endophytic bacteria interactions. *Int. J. Mol. Sci.* 20:E1947. doi: 10.3390/ijms20081947

Poole, P., Ramachandran, V., and Terpolilli, J. (2018). Rhizobia: from saprophytes to endosymbionts. *Nat. Rev. Microbiol.* 16, 291–303. doi: 10.1038/nrmicro2017.171

Pouvreau, B., Vanhercke, T., and Singh, S. (2018). From plant metabolic engineering to plant synthetic biology: the evolution of the design/build/test/learn cycle. *Plant Sci.* 273, 3–12. doi: 10.1016/j.plantsci.2018.03.035

Price, M. N., Wetmore, K. M., Waters, R. J., Callaghan, M., Ray, J., Liu, H., et al. (2018a). Mutant phenotypes for thousands of bacterial genes of unknown function. *Nat. Chem. Biol.* 557, 503–509. doi: 10.1038/s41586-018-0124-0

Price, M. N., Zane, G. M., Kuehl, J. V., Melnyk, R. A., Wall, J. D., Deutschbauer, A. M., et al. (2018b). Filling gaps in bacterial amino acid biosynthesis pathways with high-throughput genetics. *PLoS Genet.* 14:e1007147. doi: 10.1371/journal.pgen.1007147

Ramsey, M. E., Hyde, J. A., Medina-Perez, D. N., Lin, T., Gao, L., Lundt, M. E., et al. (2017). A high-throughput genetic screen identifies previously uncharacterized *Borrelia burgdorferi* genes important for resistance against reactive oxygen and nitrogen species. *PLoS Pathog.* 13:e1006225. doi: 10.1371/journal.ppat.1006225

Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C., and Foley, J. A. (2012). Recent patterns of crop yield growth and stagnation. *Nat. Commun.* 3:1293. doi: 10.1038/ncomms2296

Reznikoff, W. S. (2008). Transposon Tn*5*. *Annu. Rev. Genet.* 42, 269–286. doi: 10.1146/annurev.genet.42.110807.091656

Rocha, I. D., Ma, Y., Souza-Alonso, P., Vosatka, M., Freitas, H., and Oliveira, R. S. (2019). Seed coating: a tool for delivering beneficial microbes to agricultural crops. *Front. Plant Sci.* 10:1357. doi: 10.3389/fpls.2019.01357

Rosconi, F., de Vries, S. P. W., Baig, A., Fabiano, E., and Grant, A. J. (2016). Essential Genes for *in vitro* Growth of the Endophyte *Herbaspirillum seropedicae* SmR1 as revealed by transposon insertion site sequencing. *Appl. Environ. Microbiol.* 82, 6664–6671. doi: 10.1128/aem.02281-16

Royet, K., Parisot, N., Rodrigue, A., Gueguen, E., and Condemine, G. (2019). Identification by Tn-seq of *Dickeya dadantii* genes required for survival in chicory plants. *Mol. Plant Pathol.* 20, 287–306. doi: 10.1111/mpp.12754

Rudrappa, T., Biedrzycki, M. L., and Bais, H. P. (2008). Causes and consequences of plant-associated biofilms. *FEMS Microbiol. Ecol.* 64, 153–166. doi: 10.1111/j.1574-6941.2008.00465.x

Santhanam, R., Luu, V. T., Weinhold, A., Goldberg, J., Oh, Y., and Baldwin, I. T. (2015). Native root-associated bacteria rescue a plant from a sudden-wilt disease that emerged during continuous cropping. *Proc. Natl. Acad. Sci. U.S.A.* 112, E5013–E5120. doi: 10.1073/pnas.1505765112

Santiago, M., Lee, W., Fayad, A. A., Coe, K. A., Rajagopal, M., Do, T., et al. (2018). Genome-wide mutant profiling predicts the mechanism of a Lipid II binding antibiotic. *Nat. Chem. Biol.* 14, 601–608. doi: 10.1038/s41589-018-0041-4

Scharf, B. E., Hynes, M. F., and Alexandre, G. M. (2016). Chemotaxis signaling systems in model beneficial plant–bacteria associations. *Plant Mol. Biol.* 90, 549–559. doi: 10.1007/s11103-016-0432-4

Serrania, J., Johner, T., Rupp, O., Goesmann, A., and Becker, A. (2017). Massive parallel insertion site sequencing of an arrayed *Sinorhizobium meliloti* signature-tagged mini-Tn*5* transposon mutant library. *J. Biotechnol.* 257, 9–12. doi: 10.1016/j.jbiotec.2017.02.019

Shelake, R. M., Pramanik, D., and Kim, J.-Y. (2019). Exploration of plant-microbe interactions for sustainable agriculture in CRISPR Era. *Microorganisms* 7:E269. doi: 10.3390/microorganisms7080269

Sivakumar, R., Ranjani, J., Vishnu, U. S., Jayashree, S., Lozano, G. L., Miles, J., et al. (2019). Evaluation of InSeq to identify genes essential for *Pseudomonas aeruginosa* PGPR2 corn root colonization. *G3* 9, 651–661. doi: 10.1534/g3.118.200928

Syed Ab Rahman, S. F., Singh, E., Pieterse, C. M. J., and Schenk, P. M. (2018). Emerging microbial biocontrol strategies for plant pathogens. *Plant Sci.* 267, 102–111. doi: 10.1016/j.plantsci.2017.11.012

Tscharntke, T., Clough, Y., Wanger, T. C., Jackson, L., Motzke, I., Perfecto, I., et al. (2012). Global food security, biodiversity conservation and the future of agricultural intensification. *Biol. Conserv.* 151, 53–59. doi: 10.1016/j.biocon.2012.01.068

van Opijnen, T., Bodi, K. L., and Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods* 6, 767–772. doi: 10.1038/nmeth.1377

Van Opijnen, T., and Camilli, A. (2012). A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res.* 22, 2541–2551. doi: 10.1101/gr.137430.112

van Opijnen, T., and Camilli, A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat. Rev. Microbiol.* 11, 435–442. doi: 10.1038/nrmicro3033

Vorholt, J. A., Vogel, C., Carlström, C. I., and Müller, D. B. (2017). Establishing causality: opportunities of synthetic communities for plant microbiome research. *Cell Host Microbe* 22, 142–155. doi: 10.1016/j.chom.2017.07.004

Welkie, D. G., Rubin, B. E., Chang, Y. G., Diamond, S., Rifkin, S. A., LiWang, A., et al. (2018). Genome-wide fitness assessment during diurnal growth reveals an expanded role of the cyanobacterial circadian clock protein KaiA. *Proc. Natl. Acad. Sci. U.S.A.* 115, E7174–E7183. doi: 10.1073/pnas.1802940115

Wetmore, K. M., Price, M. N., Waters, R. J., Lamson, J. S., He, J., Hoover, C. A., et al. (2015). Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons. *mBio* 6:e00306-15. doi: 10.1128/mBio.00306-15

Wheatley, R. M., Ramachandran, V. K., Geddes, B. A., Perry, B. J., Yost, C. K., and Poole, P. S. (2017). Role of O2 in the Growth of *Rhizobium leguminosarum* bv. *viciae* 3841 on Glucose and Succinate. *J. Bacteriol.* 199:e572-16. doi: 10.1128/JB.00572-16

Wu, F., Li, J., Chen, Y., Zhang, L., Zhang, Y., Wang, S., et al. (2019). Effects of phosphate solubilizing bacteria on the growth, photosynthesis, and nutrient uptake of *Camellia oleifera* abel. *Forests* 10:348. doi: 10.3390/f10040348

Wu, G., Liu, Y., Xu, Y., Zhang, G., Shen, Q., and Zhang, R. (2018). Exploring elicitors of the beneficial Rhizobacterium *Bacillus amyloliquefaciens* SQR9 to induce plant systemic resistance and their interactions with plant signaling pathways. *Mol. Plant Microbe Interact.* 31, 560–567. doi: 10.1094/MPMI-11-17-0273-R

Yang, G., Billings, G., Hubbard, T. P., and Park, J. S. (2017). Time-resolved transposon insertion sequencing reveals genome-wide fitness dynamics during infection. *mBio* 8:e01581-17. doi: 10.1128/mBio.01581-17

Zhang, R., Vivanco, J. M., and Shen, Q. (2017). The unseen rhizosphere root–soil–microbe interactions for crop production. *Curr. Opin. Microbiol.* 37, 8–14. doi: 10.1016/j.mib.2017.03.008

Tweaking Photosynthesis: FNR-TROL Interaction as Potential Target for Crop Fortification

Hrvoje Fulgosi* and Lea Vojta

Laboratory for Molecular Plant Biology and Biotechnology, Division of Molecular Biology, Institute Ruđer Bošković, Zagreb, Croatia

Photosynthesis not only supplies plants with needed energy for growth and nutrient storage, but links light-to-chemical energy conversion with redox regulatory networks of the entire cell. By modulating photosynthetic electron flow, plants can adapt to constantly changing light and environmental conditions. Different mechanisms direct electrons formed during light reactions to either energy-conserving or energy-dissipating pathways. Recently described dynamic interactions of the flavoenzyme ferredoxin:NADP$^+$ oxidoreductase (FNR) with protein TROL represent elegant molecular switch that can partition electrons downstream of photosystem I. FNR-TROL bifurcation can control energy transfer to either linear flow, which results in NADPH production, or to the rapid electron sink that efficiently prevents reactive oxygen species propagation. Plant genome editing of TROL represents an unexploited alley for improvement of plant stress and defense responses, productivity, and eventually agricultural yield.

The most well-known and established function of plant flavoenzyme ferredoxin: NADP(H) oxidoreductase (FNR) in photosynthetic energy conversion is the synthesis of NADPH (Forti and Bracale, 1984; Carillo and Ceccarelli, 2003; Shin, 2004; Mulo, 2011). This enzymatic reaction utilizes two molecules of reduced ferredoxin (Fd) to produce one molecule of NADPH (Arakaki et al., 1997; Medina and Gómez-Moreno, 2004). This conversion is known as the last step in the linear electron transfer (LET) chain (Allen, 2003; Rochaix, 2011). Generated NADPH is utilized in numerous downstream biochemical reactions, both in chloroplast stroma and in the rest of the cell (Rochaix, 2011; Scheibe and Dietz, 2012). Electron cycling from ferredoxin to NADPH only occurs in the light. In the dark, as well as in non-photosynthetic organisms, the FNR primarily works in reverse, utilizing NADPH to provide reduced ferredoxin for various metabolic pathways, such as oxidative stress response, nitrogen fixation, terpenoid biosynthesis, steroid metabolism, and iron–sulfur protein biogenesis (Aliverti et al., 2008). Thus, FNR links fundamental process of light-to-chemical energy conversion with general plant metabolism (Foyer and Noctor, 2005). The electron donor Fd is a small iron-sulfur protein that acts simultaneously as a hub, and as a bottle neck for the distribution of electrons supplied by photosystem I (PSI) (Paul and Foyer, 2001; Hanke et al., 2004; Balmer et al., 2006; Joliot and Johnson, 2011). FNR binding to thylakoid membranes of vascular plants has been proposed to take place via cytochrome b_6/f complex (Clark et al., 1984; Zhang et al., 2001), PSI-E subunit (Andersen et al., 1992), and in a complex with NADPH dehydrogenase (Quiles and Cuello, 1998). However, in all these studies the exact protein association domain responsible for FNR-binding has not been identified. Most recently, the FNR interaction with photosynthetic membranes has also been demonstrated in organisms other than vascular plants (Mosebach et al., 2017; Pini et al., 2019). In a number of essential publications, it has been demonstrated that the distribution of FNR between soluble- and thylakoid-bound states can have profound influence on FAD cofactor assembly (Miyake et al., 1998; Onda and Hase, 2004), Fd-FNR interactions, FNR catalytic properties, oxidative and photo-oxidative

stress tolerance, and the regulation of photosynthesis (Palatnik et al., 1997; Rodriguez et al., 2007; Hanke et al., 2008; Peng et al., 2008; Benz et al., 2009, 2010; Jurić et al., 2009; Alte et al., 2010; Yacoby et al., 2011; Twachtmann et al., 2012; Vojta and Fulgosi, 2012; Vojta et al., 2012; Goss and Hanke, 2014; Lintala et al., 2014). Particularly interesting, transgenic plants overexpressing chloroplastic FNR acquire increased tolerance to oxidative stress, while simultaneously displaying normal rates of photosynthesis (Rodriguez et al., 2007). Further, FNR knock-down plants, which have stunted growth and restricted photosynthetic activity, are particularly susceptible to photo-oxidative damage (Hajirezaei et al., 2002; Lintala et al., 2012). Perhaps the most intriguing finding was that the release of FNR from thylakoid membranes could be regulated by oxidative stress, in methyl viologen-dependent fashion (Palatnik et al., 1997). Lastly, in t*ic62trol* double mutants no binding of FNR to thylakoid proteins could be detected, whatsoever (Lintala et al., 2014).

The hypothesis describing additional functions and physiological roles of chloroplastic FNR in vascular plants has been proposed (Benz et al., 2010; Goss and Hanke, 2014). This concepts rest on a series of publications mostly describing binding and release of FNR from the soluble chloroplastic protein Tic62 (Benz et al., 2009; Lintala et al., 2014). Tic62 (62 kDa component of the translocon at the inner envelope of chloroplasts) has initially been discovered as a component of chloroplast inner envelope protein translocation machinery (Küchler et al., 2002), but has subsequently been associated with at least two other compartments, chloroplast stroma, and chloroplast photosynthetic membranes (Balsera et al., 2007). The essential protein region recognized in Tic62 as being responsible for the binding of FNR is present in multiple repeats in many, but not all, Tic62 family members studied (Balsera et al., 2007). The same motif was further identified at the C-terminus of the thylakoid rhodanase-like protein TROL (Jurić et al., 2009). This domain, dubbed membrane recruitment motif (MRM), binds FNR with high efficiency. In Tic62, the interaction was postulated to be light- and pH-dependent (Benz et al., 2009). In 2010, Benz et al. proposed that the majority of thylakoid-localized FNR is bound to the membrane via two interaction partners, Tic62 and TROL (Benz et al., 2009). They further postulate that soluble form of FNR is sufficient to sustain photosynthetic energy conversion, while the thylakoid pool largely performs other functions. Finally, they suggest that light- and redox-states regulate the distribution of FNR. In the view of that, FNR is stored at the membranes in the dark and during morning hours, thus allowing advantageous flow of electrons to various Fd-dependent pathways (Benz et al., 2009; Vojta et al., 2012). Such scenario could assure proper metabolic adjustments depending to environmental cues.

Soluble Tic62, however, does not contain MRM in all plant species (Balsera et al., 2007), and its attachment points at the thylakoids have yet to be determined. TROL, however, is *bona fide* integral membrane protein with dual localization (Jurić et al., 2009; Vojta et al., 2018). It is present in the mature form in thylakoid membranes, and in the precursor form at the inner chloroplast envelope. TROL is located near PSI, mostly in non-appressed thylakoid regions (Jurić et al., 2009). All known

vascular plant TROL sequences contain FNR MRM (Jurić et al., 2009). Additionally, a single MRM of TROL binds FNR with several fold higher affinity then the Tic62 MRM (Jurić et al., 2009). In Arabidopsis, TROL associated with FNR can be isolated in the dynamic supramolecular complex of ∼190 kDa (Jurić et al., 2009).

The unique hallmark of TROL is the lumen-located rhodanase-like domain (RHO), which is most likely inactive in a sense of sulfur detoxification, as it contains the aspartate residue instead of the conserved cysteine in the putative active site (Jurić et al., 2009). RHO domains are intriguing due to their ancient origin and structural identity with the catalytic domains of cell-division-cycle (CDC25) dual-specificity phosphatases. Inactive RHO domains are implicated in redox-sensing and were shown to interact with quinolinediones (Brisson et al., 2005). Plant sulfur-transferases and rhodanases have recently been reviewed in two comprehensive papers (Moseler et al., 2019; Selles et al., 2019).

Recently, alternative mechanism of FNR binding and release from TROL, involving redox sensing by the RHO domain, has been put forward (Vojta and Fulgosi, 2012, 2019). According to this model, certain (redox) signal(s) of lumenal origin can be sensed by the RHO and further transduced across thylakoid membrane, resulting in differential FNR binding on the stromal side (Vojta and Fulgosi, 2012). The role of the proline-rich region of TROL, dubbed PEPE, which precedes the MRM, has also been postulated (Jurić et al., 2009). Due to a repeating sequence of turn-inducing proline residues, PEPE might serve as molecular swivel, allowing free movement of bound FNR, or even its alternative associations with various supramolecular complexes, or membrane domains. Such dynamic FNR recruitment might be responsible for alternative partitioning of photosynthetic electrons, and/or prioritization of Fd-dependent pathways. Further, it has been demonstrated that chloroplasts of Arabidopsis *trol* mutants proliferate substantially less superoxide anion radicals then the WT (Vojta et al., 2015). This reduction can be recorded in *trol* chloroplast pre-acclimated to dark and growth-light conditions. Even more remarkable, *trol* chloroplast proliferate almost 40 percent less superoxide anion radicals even in the presence of ROS-generating methyl viologen (paraquat herbicide) (Vojta et al., 2015). These finding are in line with the previously published observations and suggest that FNR permanently detached from TROL can either efficiently scavenge superoxide anion, or that electrons are very rapidly partitioned into certain other pathways(s), different from the LET. In fact, results suggest that LET is preferential only when TROL-FNR association is established (Jurić et al., 2009; Vojta et al., 2015). Alternatively, in the absence of TROL, electrons rapidly flow to other sinks downstream of PSI donor site (Vojta et al., 2015). Finally, it has been demonstrated that in the absence of TROL light- and/or pH-dependent dynamic recruitment of FNR to thylakoids is entirely abolished (Vojta and Fulgosi, 2016). Apparently, TROL-FNR interaction might be the most prominent, if not exclusive, dynamic-type interaction of FNR with photosynthetic membranes of vascular plants.

TROL-FNR interaction could be an entrance to crop improvement *via* various modifications to the interaction itself,

or to upstream and/or downstream reactions and pathways. For example, TROL itself could be down- or up-regulated, or its FNR-binding or regulatory domains could be modified by genome engineering. FNR could also be a target of engineering, for example by exchange of FNR enzymes from C4 to C3 plants or *vice versa*. C4 FNR interacts with TROL ITEP domain with many fold higher affinity than the C3 FNR (Rac and Fulgosi, 2019). The increase of TROL levels by itself would not necessarily imply an increase in photosynthetic efficiency, since the improvement would depend on the availability of reduced ferredoxin and the ability of the system to provide the necessary reduction equivalents. However, alterations in the production of reduction equivalents could affect the redox homeostasis of chloroplasts and ultimately produce aberrations instead of plant benefit.

To conclude, we further iterate the concept of photosynthetic membrane recruitment of FNR and emphasize the importance of TROL-FNR dynamic interaction. We posit that TROL-FNR interaction is an important and overlooked mechanism for regulation and prioritization of energy-conserving and energy-dissipating pathways in vascular plant photosynthesis. We propose that TROL interaction with FNR is useful target for genome editing of agriculturally important species, potentially providing more stress-tolerant crops.

AUTHOR CONTRIBUTIONS

HF and LV wrote the manuscript.

FUNDING

REFERENCES

Aliverti, A., Pandini, V., Pennati, A., de Rosa, M., and Zanetti, G. (2008). Structural and functional diversity of ferredoxin-NADP$^+$ reductases. *Arch. Biochem. Biophys.* 474, 283–291. doi: 10.1016/j.abb.2008.02.014

Allen, J. F. (2003). Cyclic, pseudocyclic and noncyclic photophosphorylation: new links in the chain. *Trends Plant Sci.* 8, 15–19. doi: 10.1016/S1360-1385(02)00006-7

Alte, F., Stengel, A., Benz, J.-P., Petersen, E., Soll, J., Groll, M., et al. (2010). Ferredoxin:NADPH oxidoreductase is recruited to thylakoids by binding to polyproline type II helix in a pH dependent manner. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19260–19265. doi: 10.1073/pnas.1009124107

Andersen, B., Scheller, H. V., and Moller, B. L. (1992). The PSI E subunit of photosystem I binds ferredoxin:NADP$^+$ oxidoreductase. *FEBS Lett.* 311, 169–173. doi: 10.1016/0014-5793(92)81391-X

Arakaki, A. K., Ceccarelli, E. A., and Carrillo, N. (1997). Plant-type ferredoxin-NADP$^+$ reductases: a basal structural framework and a multiplicity of functions. *FASEB J.* 11, 133–140. doi: 10.1096/fasebj.11.2.9039955

Balmer, Y., Vensel, W. H., Cai, N., Manieri, W., Schürmann, P., Hurkman, W. J., et al. (2006). A complete ferredoxin/thioredoxin system regulates fundamental processes in amyloplasts. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2988–2993. doi: 10.1073/pnas.0511040103

Balsera, M., Stengel, A., Soll, J., and Bölter, B. (2007). Tic62: a protein family from metabolism to protein translocation. *BMC Evol Biol.* 7:43. doi: 10.1186/1471-2148-7-43

Benz, J.-P., Stengel, A., Lintala, M., Lee, Y.-H., Weber, A., Philippar, K., et al. (2009). Arabidopsis Tic62 and ferredoxin-NADP(H) oxidoreductase form light-regulated complexes that are integrated into the chloroplast redox poise. *Plant Cell* 21, 3965–3983. doi: 10.1105/tpc.109.069815

Benz, J. P., Lintala, M., Soll, J., Mulo, P., and Bölter, B. (2010). A new concept for ferredoxin-NADP(H) oxidoreductase binding to plant thylakoids. *Trends Plant Sci.* 15, 608–613. doi: 10.1016/j.tplants.2010.08.008

Brisson, M., Nguyen, T., Wipf, P., Joo, B., Day, B. W., Skoko, J. S., et al. (2005). Redox regulation of Cdc25B by cell-active quinolinediones. *Mol. Pharm.* 68, 1810–1820. doi: 10.1124/mol.105.016360

Carillo, N., and Ceccarelli, E. A. (2003). Open questions in ferredoxin-NADP$^+$ reductase catalytic mechanism. *Eur. J. Biochem.* 270, 1900–1915. doi: 10.1046/j.1432-1033.2003.03566.x

Clark, R. D., Hawkesford, M. J., Coughlan, S. J., Bennett, J., and Hind, G. (1984). Association of ferredoxin–NADP$^+$ oxidoreductase with the chloroplast cytochrome b–f complex. *FEBS Lett.* 174, 137–142. doi: 10.1016/0014-5793(84)81092-3

Forti, G., and Bracale, M. (1984). Ferredoxin–ferredoxin NADP reductase interaction. *FEBS Lett.* 166, 81–84. doi: 10.1016/0014-5793(84)80049-6

Foyer, C. H., and Noctor, G. (2005). Redox homeostasis and antioxidant signalling: a metabolic interface between stress perception and physiological responses. *Plant Cell* 17, 1866–1875. doi: 10.1105/tpc.105.033589

Goss, T., and Hanke, G. (2014). The end of the line: can ferredoxin and ferredoxin NADP(H) oxidoreductase determine the fate of photosynthetic electrons? *Curr. Prot. Pept. Sci.* 15, 385–393. doi: 10.2174/1389203715666140327113733

Hajirezaei, M. R., Peisker, M., Tschiersch, H., Palatnik, J. F., Valle, E. M., Carrillo, N., et al. (2002). Small changes in the activity of chloroplastic NADP$^{(+)}$-dependent ferredoxin oxidoreductase lead to impaired plant growth and restrict photosynthetic activity of transgenic tobacco plants. *Plant J.* 29, 281–293. doi: 10.1046/j.0960-7412.2001.01209.x

Hanke, G., Kimata-Ariga, Y., Taniguchi, I., and Hase, T. (2004). A post genomic characterization of Arabidopsis ferredoxins. *Plant Physiol.* 134, 255–264. doi: 10.1104/pp.103.032755

Hanke, G. T., Endo, T., Satoh, F., and Hase, T. (2008). Altered photosynthetic electron channelling into cyclic electron flow and nitrite assimilation in a mutant of ferredoxin: NADP(H) reductase. *Plant Cell Environ.* 31, 1017–1028. doi: 10.1111/j.1365-3040.2008.01814.x

Joliot, P., and Johnson, G. N. (2011). Regulation of cyclic and linear electron flow in higher plants. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13317–13322. doi: 10.1073/pnas.1110189108

Jurić, S., Hazler-Pilepić, K., Tomašić, A., Lepeduš, H., Jeličić, B., Puthiyaveetil, S., et al. (2009). Tethering of ferredoxin:NADP$^+$ oxidoreductase to thylakoid membranes is mediated by novel chloroplast protein TROL. *Plant J.* 60, 783–794. doi: 10.1111/j.1365-313X.2009.03999.x

Küchler, M., Decker, S., Hörmann, F., Soll, J., and Heins, L. (2002). Protein import into chloroplasts involves redox-regulated proteins. *EMBO J.* 21, 6136–6145. doi: 10.1093/emboj/cdf621

Lintala, M., Lehtimäki, N., Benz, J. P., Jungfer, A., Soll, J., Aro, E. M., et al. (2012). Depletion of leaf-type ferredoxin-NADP$^{(+)}$ oxidoreductase results in the permanent induction of photoprotective mechanisms in Arabidopsis chloroplasts. *Plant J.* 70, 809–817. doi: 10.1111/j.1365-313X.2012.04930.x

Lintala, M., Schuck, N., Thormählen, I., Jungfer, A., Weber, K.-L., Weber, A.-P., et al. (2014). Arabidopsis *tic62 trol* mutant lacking thylakoid bound ferredoxin-NADP$^+$ oxidoreductase shows distinct metabolic phenotype. *Mol. Plant.* 7, 45–57. doi: 10.1093/mp/sst129

Medina, M., and Gómez-Moreno, C. (2004). Interaction of ferredoxin-NADP$^{(+)}$ reductase with its substrates: optimal interaction for efficient electron transfer. *Photosynth. Res.* 79, 113–131. doi: 10.1023/B:PRES.0000015386.67746.2c

Miyake, C., Schreiber, U., Hormann, H., Sano, S., and Asada, K. (1998). The FAD-enzyme monodehydroascorbate radical reductase mediates photoreduction of superoxide radicals in spinach thylakoid membranes. *Plant Cell Physiol.* 39, 821–829. doi: 10.1093/oxfordjournals.pcp.a029440

Mosebach, L., Heilmann, C., Mutoh, R., Gäbelein, P., Steinbeck, J., Happe, T., et al. (2017). Association of Ferredoxin:NADP$^+$ oxidoreductase with

the photosynthetic apparatus modulates electron transfer in *Chlamydomonas reinhardtii*. *Photosynth. Res*. 134, 291–306. doi: 10.1007/s11120-017-0408-5

Moseler, A., Selles, B., Rouhier, N., and Couturier, J. (2019). Novel insights into the diversity of the sulfurtransferase family in photosynthetic organisms with emphasis on oak. *N. Phytol*. 1–11. doi: 10.1111/nph.15870

Mulo, P. (2011). Chloroplast-targeted ferredoxin-NADP$^+$ oxidoreductase (FNR): structure, function and location. *Biochim. Biophys. Acta* 1807, 927–934. doi: 10.1016/j.bbabio.2010.10.001

Onda, Y., and Hase, T. (2004). FAD assembly and thylakoid membrane binding of ferredoxin:NADP$^+$ oxidoreductase in chloroplasts. *FEBS Lett*. 564, 116–120. doi: 10.1016/S0014-5793(04)00325-4

Palatnik, J. F., Valle, E. M., and Carillo, N. (1997). Oxidative stress causes ferredoxin-NADP$^+$ reductase solubilization from the thylakoid membranes in methyl viologen-treated plants. *Plant Physiol*. 115, 1721–1727. doi: 10.1104/pp.115.4.1721

Paul, M. J., and Foyer, C. H. (2001). Sink regulation of photosynthesis. *J. Exp. Bot*. 52, 1383–1400. doi: 10.1093/jexbot/52.360.1383

Peng, L., Shimizu, H., and Shikanai, T. (2008). The chloroplast NAD(P)H dehydrogenase complex interacts with photosystem I in Arabidopsis. *J. Biol. Chem*. 283, 34873–34879. doi: 10.1074/jbc.M803207200

Pini, M., Elman, T., and Yacoby, I. (2019). Binding of ferredoxin NADP$^+$ oxidoreductase (FNR) to plant photosystem I. *BBA – Bioenergetics* 1860, 689–698. doi: 10.1016/j.bbabio.2019.07.007

Quiles, M. J., and Cuello, J. (1998). Association of ferredoxin–NADP oxidoreductase with the chloroplastic pyridine nucleotide dehydrogenase complex in barley leaves. *Plant Physiol*. 117, 235–244. doi: 10.1104/pp.117.1.235

Rac, A., and Fulgosi, H. (2019). Qualitative and quantitative dataset of TROL protein interaction with C3 and C4 ferredoxin: NADP$^+$ oxidoreductases. *Data Brief* 28:105038. doi: 10.1016/j.dib.2019.105038

Rochaix, J. D. (2011). Regulation of photosynthetic electron transport. *Biochim. Biophys. Acta* 1807, 375–383. doi: 10.1016/j.bbabio.2010.11.010

Rodriguez, R. E., Lodeyro, A., Poli, H. O., Zurbriggen, M., Peisker, M., Palatnik, J. F., et al. (2007). Transgenic tobacco plants overexpressing chloroplastic ferredoxin-NADP(H) reductase display normal rates of photosynthesis and increased tolerance to oxidative stress. *Plant Physiol*. 143, 639–649. doi: 10.1104/pp.106.090449

Scheibe, R., and Dietz, K.-J. (2012). Reduction-oxidation network for flexible adjustment of cellular metabolism in photoautotrophic cells. *Plant Cell Environ*. 35, 202–216. doi: 10.1111/j.1365-3040.2011.02319.x

Selles, B., Moseler, A., Rouhier, N., and Couturier, J. (2019). Rhodanese domain-containing sulfurtransferases: multifaceted proteins involved in sulfur trafficking in plants. *J. Exp. Bot*. 70, 4139–4154. doi: 10.1093/jxb/erz213

Shin, M. (2004). How is ferredoxin-NADP reductase involved in the NADP photoreduction of chloroplasts? *Photosynth. Res*. 80, 307–313. doi: 10.1023/B:PRES.0000030456.96329.f9

Twachtmann, M., Altmann, B., Muraki, N., Voss, I., Okutani, S., Kurisu, G., et al. (2012). N-terminal structure of maize ferredoxin:NADP$^+$ reductase determines recruitment into different thylakoid membrane complexes. *Plant Cell*. 24, 2979–2991. doi: 10.1105/tpc.111.094532

Vojta, L., Carić, D., Cesar, V., Antunović Dunić, J., Lepeduš, H., Kveder, M., et al. (2015). TROL-FNR interaction reveals alternative pathways of electron partitioning in photosynthesis. *Sci. Rep*. 5:10085. doi: 10.1038/srep10085

Vojta, L., Čuletić, A., and Fulgosi, H. (2018). Effects of TROL presequence mutagenesis on its import and dual localization in chloroplasts. *Int. J. Mol. Sci*. 19:569. doi: 10.3390/ijms19020569

Vojta, L., and Fulgosi, H. (2012). "Energy Conductance from Thylakoid Complexes to Stromal Reducing Equivalents," in *Advances in Photosynthesis - Fundamental Aspects*, ed M. M. Najafpour (Rijeka: InTech), 175–190. doi: 10.5772/26693

Vojta, L., and Fulgosi, H. (2016). Data supporting the absence of FNR dynamic photosynthetic membrane recruitment in trol mutants. *Data Brief* 7, 393–396. doi: 10.1016/j.dib.2016.02.044

Vojta, L., and Fulgosi, H. (2019). Topology of TROL protein in thylakoid membranes of *Arabidopsis thaliana*. *Physiol. Plantarum*. 166, 300–308. doi: 10.1111/ppl.12927

Vojta, L., Horvat, L., and Fulgosi, H. (2012). Balancing chloroplast redox status – regulation of FNR binding and release. *Period. Biol*. 114, 25–31.

Yacoby, I., Pochekailov, S., Toporik, H., Ghirardi, M. L., King, P. W., and Zhang, S. (2011). Photosynthetic electron partitioning between [FeFe]-hydrogenase and ferredoxin:NADP$^+$-oxidoreductase (FNR) enzymes *in vitro*. *Proc. Natl. Acad. Sci. U.S.A*. 108, 9396–9401. doi: 10.1073/pnas.1103659108

Zhang, H., Whitelegge, J. P., and Cramer, W. A. (2001). Ferredoxin: NADP$^+$ oxidoreductase is a subunit of the chloroplast cytochrome b6f complex. *J. Biol. Chem*. 276, 38159–38165. doi: 10.1074/jbc.M105454200

Cynara cardunculus L. as a Multipurpose Crop for Plant Secondary Metabolites Production in Marginal Stressed Lands

Helena Domenica Pappalardo, Valeria Toscano, Giuseppe Diego Puglia, Claudia Genovese and Salvatore Antonino Raccuia*

Consiglio Nazionale delle Ricerche-Istituto per i Sistemi Agricoli e Forestali del Mediterraneo, Catania, Italy

Cardoon (*Cynara cardunculus* L.) is a Mediterranean crop, member of the Asteraceae family, characterized by high production of biomass and secondary metabolites and by a good adaptation to climate change, usable in green chemistry, nutraceutical, and pharmaceutical sectors. Recent studies demonstrated the ability of cardoon to grow up in a stressful environment, which is associated with enhanced biosynthesis of biologically active compounds in these plants, and this effect is increased by abiotic stresses (salt, heat, pollution, and drought stress) that characterize many world marginal areas, affected by the climate changes. The plant response to these stresses consists in implementing different processes that modify some plant biological functions, such as alleviating both cellular hyperosmolarity and ion disequilibrium or synthesizing antioxidant molecules. The aim of this work was to investigate different cardoon response mechanisms to abiotic stresses and to evaluate their influence on the biologically active compounds biosynthesis. Following this purpose, we analyzed the ability of cardoon seeds to germinate under different salt stress conditions, and on the sprouts obtained, we measured the total phenol content and the antioxidant activity. Moreover, the growth of cardoon seedlings grown under heavy metals stress conditions was monitored, and the expression levels of heavy metal transport–associated genes were analyzed. The results showed the ability of cardoon plants to tolerate abiotic stress, thanks to different defense mechanisms and the possibility to obtain biomass with high content of biologically active molecules by exploiting the natural tolerance of this species for abiotic stresses. Moreover, we identified some important genes encoding for metal transportation that may be involved in arsenic and cadmium uptake and translocation in *C. cardunculus*. Then, this species can be considered as a promising crop for green chemistry and energy in marginal lands.

Keywords: heavy metals, salt, sprout, gene expression, antioxidant activity

INTRODUCTION

The climate, over the centuries, has always changed because of natural processes, but in the last 100 years, these changes have been much more severe and much faster than the changes that occurred in the past.

Climate change caused an increase in unfavorable or stressful environment. Abiotic stresses, such as drought, heat, cold, salt, or heavy metals like arsenic (As) and cadmium (Cd) in the soil, are exacerbated by climate change (Fedoroff et al., 2010). Drought and salinity are prime environmental stresses that influence the geographical distribution of plants in nature, limit their agriculture productivity, and threaten food security (Zhu, 2016). More than 6% of the world's total land and approximately 20% of irrigated land are affected by salt stress (Munns and Tester, 2008).

Human activities, industrialization, and modern agricultural practices are mainly responsible for the increase in environmental contamination by heavy metals (Kavamura and Esposito, 2010; Miransari, 2011; Singh et al., 2016). The use of pesticides, fertilizers, municipal and compost wastes, and also heavy metal release from smelting industries and metalliferous mines contributed to contaminate, in a decisive way, large areas of land with heavy metals (Yang et al., 2005; Singh et al., 2016). Either the presence of salt or heavy metals in soils affect the biological cycle of the plants. Although salt stress influences all growth stages of a plant, seed germination and seedling growth stages are known to be more sensitive for most plant species (Begum et al., 1992; Cuartero et al., 2006), and germination has been reported to decline with increasing salinity levels (Houle et al., 2001). Instead, the presence of heavy metals during seedling growth and plant establishment stage causes morphological abnormalities leading to yield reduction (Amari et al., 2017). Both these stresses give rise to the production of reactive oxygen species (ROS) compounds, such as O_2, H_2O_2, and OH^- (Mittler, 2002), which damage membranes and macromolecules. The plants, as a response to these stresses, have developed several strategies. One of these is the accumulation of compatible solutes in their organs in response to osmotic stress; the primary function of these solutes is to maintain cell turgor and thus the driving gradient for water uptake (Gupta and Goyal, 2017). Another strategy is the production of antioxidant compounds (ROS scavengers), such as polyphenols, which improve the antioxidant defense and can thus increase tolerance to different stress factors (Cushman and Bohnert, 2000).

The most commonly found heavy metals at contaminated sites are arsenic and Cd (Mulligan et al., 2001). These elements are non-essential metals with no known biological function in plants, so there is not a specific transporter system for them, but they can use the same transporters of the essential nutrient uptake (Verbruggen et al., 2009; Mendoza-Cózatl et al., 2011). Different gene families were proposed to be putatively involved in the uptake of As and Cd in plants (Fan et al., 2018). The Zinc Iron Protein (ZIP) family was mainly associated with the metal transportation from the soil (Guerinot, 2000). ZIP5 and ZIP6 obtained from *Thlaspi caerulescens* cloned in *Arabidopsis thaliana* indicated that both genes act in metal homeostasis (Wu et al., 2009). Natural Resistance of Macrophage (NRAMP) are

metal transporters, and in *A. thaliana*, a non-hyperaccumulator plant, NRAMPs have high affinity with Fe and Mn transporters (Thomine et al., 2003; Lanquar et al., 2010) and also retain heavy metal transport (Ni and Cd) ability (Thomine et al., 2003; Oomen et al., 2009). In rice, *OsNRAMP1* expression is induced during As stress at the same time of other stress-responsive genes, transporters, heat shock proteins, metallothioneins, and sulfate-metabolizing proteins (Tiwari et al., 2014). NRAMP3 and NRAMP4 are responsible for Cd^{2+} efflux from the vacuole (Verbruggen et al., 2009). The heavy metal ATPases (HMAs) operate in heavy metal transport and play a role in metal homeostasis and tolerance (Gupta et al., 2013). The HMA3 was proposed to be involved in the vacuolar storage of Cd in *A. thaliana* (Verbruggen et al., 2009). Among the phosphate transporters (PHT) family, in *Arabidopsis*, overexpression of PHT1 or PHT7 causes hypersensitivity to arsenate, due to increased As uptake, whereas As resistance is enhanced through YCF1-mediated vacuolar sequestration (LeBlanc et al., 2013). The large ATP-binding cassette (ABC) transporters are involved in the transfer of different substances, including carbohydrates, lipids, xenobiotics, ions, and heavy metals (Kim et al., 2007). The *AtABCC1* and *AtABCC2* are the major vacuolar transporters of peptide-chelating heavy metals (Song et al., 2010) mediating AsIII–PC complex transport to the vacuole in *Arabidopsis*.

The possibility of using some species for phytoremediation of soils has already been widely studied, but the identified hyperaccumulators are mainly herbaceous plants, which have some limits: metal selectivity, low biomass, shallow root systems, and slow growth rates (Krämer, 2010; Cun et al., 2014; Fan et al., 2018). Therefore, some high-biomass perennial plants have been studied recently as potential candidates for phytoremediation (Fan et al., 2018). Among these crops, *Cynara cardunculus* L. plays an increasingly important role; previous research (Llugany et al., 2012), in fact, demonstrated both the ability of this species to survive quite well in polluted and stressed soil and its beneficial properties linked to its nutritional and nutraceutical characteristics to protect the body from oxidative stress.

Cardoon (*C. cardunculus* L.) is a perennial species of the *Asteraceae* family with annual growth cycle. The plant is well adapted to Mediterranean climates characterized by hot and dry summers (Raccuia and Melilli, 2007; Toscano et al., 2016). It comprises three taxa, *C. cardunculus* L. subsp. *scolymus* (L.) Hegi = *C. cardunculus* L. var. *scolymus* (L.) Hayek (globe artichoke), *C. cardunculus* L. var. *altilis* DC. (leafy or domestic cardoon), and *C. cardunculus* L. var. *sylvestris* Lam. (wild cardoon), considered to be the wild ancestor of globe artichoke (Rottenberg and Zohary, 1996; Raccuia et al., 2004b). The wild cardoon is a robust thistle well-adapted to Mediterranean semiarid climate. The domestic one is well known for its high biomass and for its use as raw material in green chemistry. In fact, from this plant, it is possible to realize an innovative range of bioproducts (bioplastics, biolubricants, home and personal care items, food fragrances, plant protection additives, etc.), with a positive impact on the environment and farmers' income (Toscano et al., 2016). Moreover, the cardoon biomass, which contains cellulose, hemicellulose, and lignin, can be used to produce energy (Raccuia and Melilli, 2007;

Ierna et al., 2012; Ciancolini et al., 2013; Toscano et al., 2016; Ottaiano et al., 2017; Gominho et al., 2018; Petropoulos et al., 2019; Turco et al., 2019). Recent studies, on morphological and physiological characteristics and on seed germination process, showed intraspecific variability among different cardoon populations under salt and moisture stresses (Mauromicale and Ierna, 2004; Raccuia et al., 2004a; Benlloch-González et al., 2005; Pagnotta and Noorani, 2014; Toscano et al., 2016). The resistance developed by this plant to both salt and heavy metal stresses was shown to be associated with a greater synthesis and the accumulation of the secondary metabolites (Mauromicale and Licandro, 2002; Raccuia et al., 2004b; Argento et al., 2016; Leonardi et al., 2016; Pappalardo et al., 2016), consisting of high amounts of polyphenolic compounds and inulin conferring healthy properties to this plant (Raccuia et al., 2004a; Pandino et al., 2011; Genovese et al., 2016a,b).

The aim of this work was to investigate the cardoon response mechanisms to different abiotic stresses. To this end, we focused on the ability of the seed to germinate under saline stress and its effect on the synthesis of antioxidant compounds. Moreover, we monitored the seedling growth in presence of heavy metals (As and Cd) and the expression levels of ion transporter genes associated with their translocation.

MATERIALS AND METHODS

Plant Materials

For the different trials of this study, two genotypes were used: a wild cardoon genotype, "A14SR" (*C. cardunculus* L. var. *sylvestris*), and one domestic cardoon variety (*C. cardunculus* L. var. *altilis*). All the genotypes belong to the *cardoon* germplasm of the section of Catania of the National Research Council–Institute for Agricultural and Forest System in the Mediterranean (CNR-ISAFOM of Catania, Italy). The seeds were collected by hand from dried flower heads by shaking and lightly pounding them, making sure not to damage the seeds themselves during the

summertime. The seeds from domestic variety were used to carry out germination tests in salt stress conditions, whereas seeds of both genotypes were used to measure the expression levels of the genes associated with salt tolerance and heavy metal transport.

Germination Test With NaCl

To verify the effect of salt stress on the biosynthesis and the concentration of secondary metabolites, stress conditions were induced during the germination of domestic cardoon seeds by three different salt concentrations: 0, 60, and 120 mM of NaCl.

For this experiment, 10-cm diameter Petri dishes were used with a filter paper placed at the bottom. Three experimental replicas were prepared, consisting of eight Petri dishes each for every salt concentration (0, 60, and 120 mM). All dishes were filled with 30 seeds and 10 mL of saline solution; then, each plate was closed with parafilm and placed in the growth chamber with a 12-/12-h light-dark cycle photoperiod and a 15°C/25°C temperature cycle (**Figure 1**) (Raccuia et al., 2004a). The young sprouts obtained were grown for 20 days.

Total Phenol Content and Antioxidant Activity

After 20 days of growth, fresh and dry weight of 1,000 sprouts was recorded, and all the sprouts were stored at −80°C to be used for TPC (total phenol content) and the AA (antioxidant activity) measurements.

Phenolic compounds were extracted from the plant material with 80% methanol solution with a ratio of 1:10 (wt/vol); the mixture thus obtained was sonicated for 44 min in an ultrasonic tank. Afterward, the sample was centrifuged at 3000 g for 3 min, and the supernatant obtained filtered with a 0.20-μm filter. The obtained sample was used for the determination of TPC and AA.

The TPC was determined by the Folin–Ciocalteu method, as described by Dewanto et al. (2002). The results were expressed as milligrams of gallic acid equivalent (GAE)/g of sample. The calibration curve of gallic acid ranged from 25 to 200 μg/mL.

FIGURE 1 | Fresh weight **(A)** and dry weight **(B)** percentage of *altilis* sprouts grown at different salt concentrations (0, 60, and 120 mM). The error bars represent the error mean of three biological replicates. Different letters indicate differences at $P \leq 0.05$ (ANOVA).
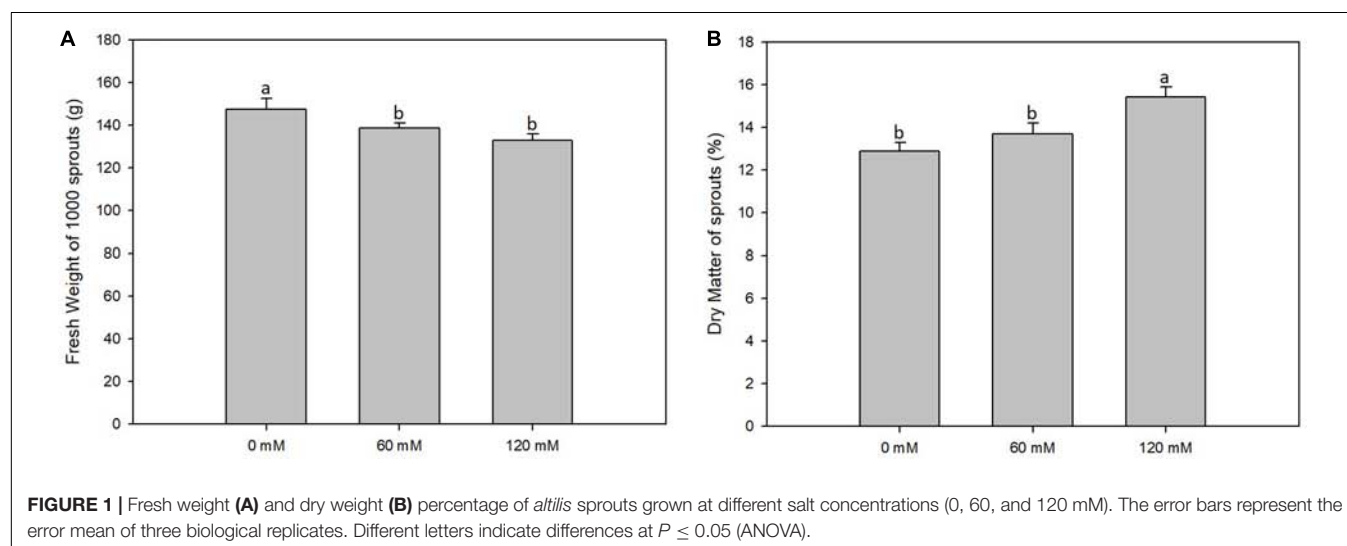
The AA was determined on cardoon sprouts using 2,2 diphenyl-1-picrylhydrazyl (DPPH) radical according to Brand-Williams et al. (1995). The antioxidant capacity was calculated by using a calibration curve obtained from known concentrations of Trolox in MeOH (10–200 μmol/L). The results were expressed as micromoles of Trolox equivalent (TE)/g of sample. All samples were analyzed in triplicates.

Seedlings Growth Analysis in Heavy Metal Stress Conditions

For this experiment, *altilis* and *sylvestris* seeds were sown in one-half MS medium with the addition of 0, 25, or 50 μM of cadmium sulfate hydrate and sodium arsenate dibasic heptahydrate, at 20°C/25°C and 12-h photoperiod. After 21 days from the seed sowing, the seedlings were collected and dissected into shoot and root portions measuring the length of the two organs. Shoot height was measured from root collar to the longest leaf extremity, and root length was measured from the root collar to the longest root apex.

Identification of Cardoon Genes Likely Associated With Plant Stress Response

To identify the cardoon orthologous genes involved in plant stress response, up to 100 mg of tissue was used to extract total RNA using protocols based on Chang et al. (1993). One microgram of total RNA, for each biological replicate, was reverse transcribed using ImProm-II[TM] Reverse Transcription System (Promega, Madison, WI, United States) according to the manufacturer's instructions.

To isolate cardoon genes involved in heavy metal uptake, we selected the NRAMP3, ZIP11, ABCC, HMA, and PHT *A. thaliana* genes sequences, and we searched orthologous sequences within the *C. cardunculus* var. *scolymus* genome v.1[1] and cardoon transcriptome (Puglia et al., 2019) using a local BLASTX and BLASTN analysis, respectively, with an *E*-value cutoff of 10^{-5} for both algorithms. The same procedure was carried out for the isolation of cardoon housekeeping genes, EF1 alpha, and GAPDH, which were used in the reverse transcriptase–quantitative polymerase chain reaction (RT-qPCR) analyses. The cloning primers were designed within the conserved domain regions using Primer3 software (Rozen and Skaletsky, 1999) (**Table 1**). The PCR was performed with PerfectTaq DNA polymerase (5 PRIME, Hilden, Germany), according to the manufacturer's instructions and cloned into the pJET vector (CloneJET PCR Cloning Kit; Thermo Scientific, Waltham, MA, USA). The DNA sequences were deposited to GenBank database (GenBank accession numbers from MN889990 to MN889996). To confirm the isolation of partial coding sequence of the genes, the obtained sequences were searched over the nucleotide collection (nt) database with BLASTN algorithm, and the first match was considered to confirm the correctness of isolated sequence. Moreover, to identify the protein region domain within the obtained contig sequences, we carried

[1]www.artichokegenome.unito.it

TABLE 1 | Cloning and RT-qPCR primer sequences designed for cardoon genes likely associated with heavy metal uptake and for housekeeping genes.

Gene name	Cloning forward primer sequence	Cloning reverse primer sequence	T_a	RT-qPCR forward primer sequence	RT-qPCR reverse primer sequence	Tm
NRAMP3	TTGATGCTACAATCAACCGA	GGGCAGAGTAGTACCATCAC	57°C	GGTGTAAGGAAGTTAGAGGCCC	AGCTTTGGAACCACGAGACC	60°C
ZIP 11	GAGATTGTCCATAGAGCTTTCCA	CCATGCCTCGTTTCCTTTTCT	48°C	TGCCTCGTTTCCTTTTCTTC	CTCGGGTGCTTCGTCGT	60°C
HMA	CGCACAGTCTCGTTCAGTTG	GGTTCCCACGTCCGGCAAG	50°C	CGGGCACGATTACTAGAGGA	CTAGCCTTGCTCTCGATGCT	60°C
ABCC	TGACCCCTTCAATGAGCACA	CGGTCACAATCGATGATGGT	54°C	TAAGTCTTTCGCGTGCATTG	TTCTTCGCGAATGGTCTTTT	60°C
PHT	GCATCTTCCTCATTCTCGCC	GTCATGGCCACCCTTTGTTT	48°C	AAGATTTCAGCAGGGACGAC	ACGACAACCGTCTTGGATTC	60°C
GAPDH	TGARTCHACYGGTGTCTTCA	TCRAYVACACGRGARCTGTA	53°C	AGTACGACAGTGTTCATGGCC	CTGAAGCCGAAAACAGCGAC	60°C
EF	TCCTTCTTGTCCACGCTCTT	AGTTGGCCGTGTTGAAACTG	53°C	TGACCCCAGTTTCAACACGG	AAGAGGCCATCAGACAAGCC	60°C
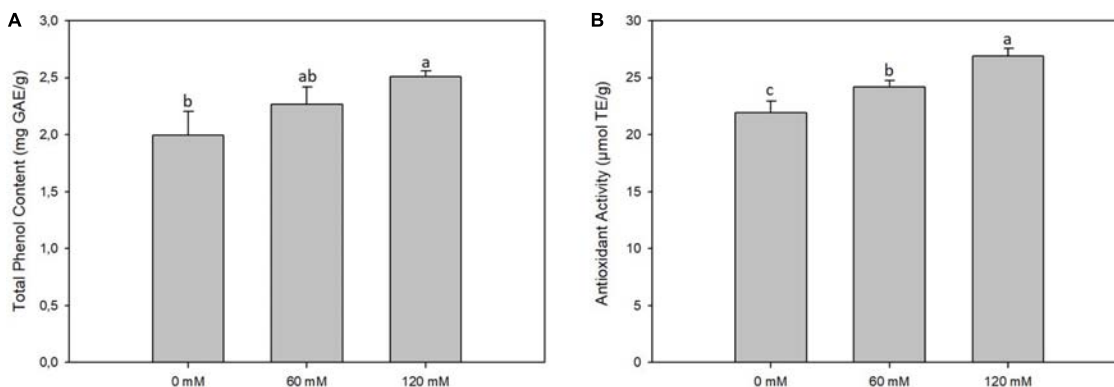
FIGURE 2 | Total phenol content **(A)** and antioxidant activity **(B)** of *altilis* sprouts grown in at different salt concentrations (0, 60, and 120 mM). The error bars represent the error mean of three biological replicates. Different letters indicate differences at $P \leq 0.05$ (ANOVA).

TABLE 2 | Analysis of variance of plant length, separated in shoot and root, measured after 21 days of treatment with As and Cd.

Source of variation	*sylvestris* Mean squares			*altilis* Mean squares		
	df	Absolute value	% of total	df	Absolute value	% of total
Organ (O)	1	5.84	*	13.94	1	5.14	**	6.83
Metal (M)	1	1.17	ns	2.80	1	30.25	***	40.24
Concentration (C)	2	15.13	***	36.11	2	29.40	***	39.11
O * M	1	0.06	ns	0.15	1	1.14	ns	1.51
O * C	2	16.97	***	40.48	2	1.33	ns	1.77
M * C	2	2.59	ns	6.18	2	7.62	***	10.13
O * M * C	2	0.15	ns	0.35	2	0.31	ns	0.41

*Partition of the treatment sum of squares into main effect and interaction. *Significant at 0.05 probability level. **Significant at 0.01 probability level. ***Significant at 0.001 probability level. ns, non-significant.*
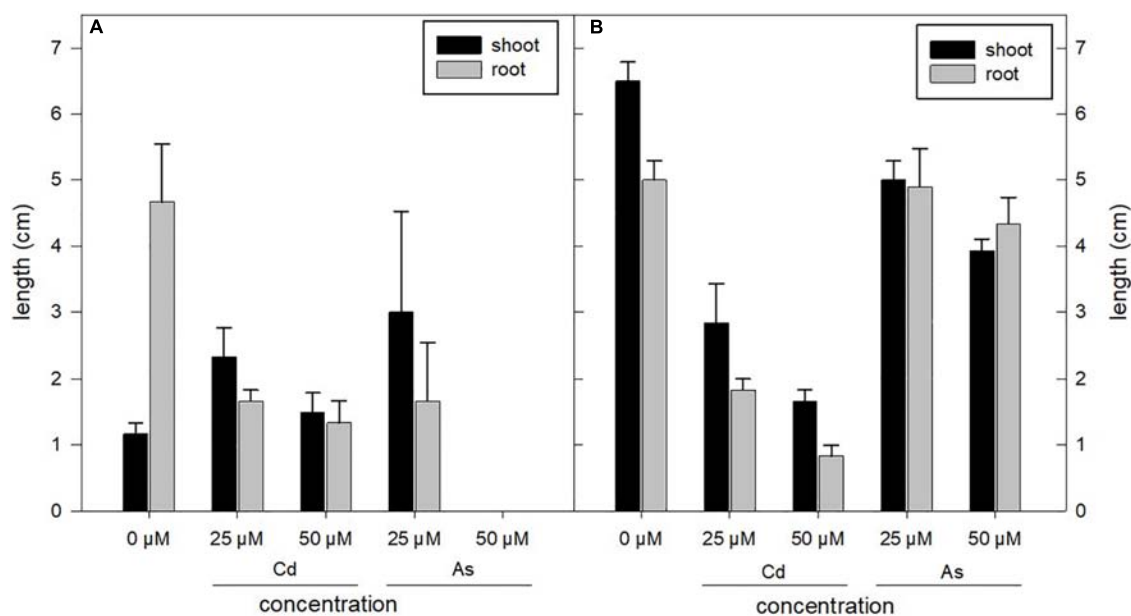
FIGURE 3 | Root and shoot lengths of *sylvestris* **(A)** and *altilis* **(B)** harvested after 21 days of treatment under different concentrations of As and Cd. The error bars indicate the standard deviation of three biological replicates.

out a BLASTP analysis of the translate sequences choosing nr (non-redundant protein sequences) database excluding *C. cardunculus* species (taxid 4265) or limited to *C. cardunculus* species (taxid 4265).

Transcriptional Levels of the Genes Likely Associated With Heavy Metal Stress Response

RNA was extracted from shoots and roots of plants grown for 2 and 3 weeks in presence of different concentrations of Cd or As. Reverse transcription reactions were performed by using ImProm-II[TM] Reverse Transcription System (Promega) according to the manufacturer's instructions. One microgram of total RNA was used for the cDNA synthesis. The RT-qPCR primers for all the heavy metal–associated cardoon genes were designed using primer3 website (Rozen and Skaletsky, 1999), setting annealing temperature at 60°C, and the gene expression levels were normalized using isolated housekeeping genes. The real-time PCR reactions were performed on a Rotorgene 6000 cycler (Qiagen, Hilden, Germany) with the QuantiNova SYBR Green Kit (Qiagen). At least three biological and three technical replicates per biological replicate were analyzed using real-time PCR analysis.

Data Analysis

All data were submitted to Bartlett test for the homogeneity of variance and then analyzed using analysis of variance (ANOVA) with CoStat program (CoHort Software, Monterey, CA, United States). Means were separated on the basis of the least significant difference, when the *F* test was significant at least at 0.05 probability level. Analysis of variance at three ways completely randomized was used to analyze the factors that mostly influenced the transcriptional levels of the considered genes, whereas ANOVA at one way was used to highlight

FIGURE 4 | The level of gene expression of NRAMP3 in seedlings of *altilis* and *sylvestris* grown up to 2 and 3 weeks under Cd **(A,B)** and As **(C,D)** at different concentrations. The fold change is the Ct value with respect to the housekeeping genes (EF1α and GAPDH) that is considered 1. The error bars represent the error mean of three biological replicates. Letters indicate only significantly different values ($P \leq 0.05$) between concentrations in each moment of sampling.

the factors' influence on the variable analyzed. In order to provide a more comprehensive analysis of the effect of the individual metal stress treatments, we carried out a principal component analysis (PCA) using the RT-qPCR expression fold-change variation of the five metal-response–associated genes along with the length of shoot and root measured at less than 25 and 50 µM of As and Cd.

RESULTS

NaCl Stress

The domestic cardoon germination percentage under NaCl stress was always greater than 87% at all concentrations considered (data not reported).

Fresh weight of 21-day-old sprouts was, on the average of the treatments, 138.1 g. The three different concentrations did not show much difference between them, although the highest weight (147.5 g) recorded for sprouts not subjected to saline treatment was reduced linearly as the salt concentration increased from 60 mM NaCl (138.5 g) to 120 mM (133.0 g) (**Figure 1**). Dry weight of the shoots showed no significant differences among the tests with different salt concentrations (data not shown). The data obtained led us to deduce that dry matter content found in the sprouts comes essentially from that one contained in the seed and therefore is not affected by salt treatment. This is in accordance with the data obtained concerning the percentage of dry matter, which was affected by salt treatments. In fact, the lowest value was recorded for the control (12.88%), and the highest value was found in the presence of NaCl maximum concentration (15.41%) (**Figure 1**).

On the average of the treatments, the TPC in sprouts was 2.26 mg GAE/g, and we observed an increase of up to 25% in sprouts subjected to a higher saline treatment (120 mM) (**Figure 2**). The AA of sprouts, on the average of the treatments, was 24.96 µmol TE/g, linearly varying between the value of

FIGURE 5 | The level of gene expression of ZIP11 in seedlings of *altilis* and *sylvestris* grown up to 2 and 3 weeks under Cd **(A,B)** and As **(C,D)** at different concentrations. The fold change is the Ct value with respect to the housekeeping genes (EF1α and GAPDH) that is considered 1. The error bars represent the error mean of three biological replicates. Letters indicate only significantly different values ($P \leq 0.05$) between concentrations in each moment of sampling.

21.92 μmol TE/g found for the extracts of control shoots and 26.87 μmol TE/g found in the shoots subjected to the higher salt treatment (120 mM) (**Figure 2**).

Seedlings Growth Analysis in Heavy Metal Stress Conditions

Length of seedlings grown under As and Cd stress conditions (0, 25, and 50 μM concentrations) measured after 3 weeks of treatments was differentially affected in the two genotypes analyzed (**Figure 3**) as the variance analysis confirmed (**Table 2**). In *sylvestris* genotype, organ–concentration interaction, with 16.97%, was the main cause of variation, differently with respect to *altilis* for which this interaction was not significant. On the contrary, in *altilis* genotype, the metal–concentration interaction was the main and only significant cause of variation

$(P < 0.001)$, whereas for *sylvestris*, this interaction was not significant.

Under As treatment, *altilis* genotype presented a growth rate similar to the control (0 μM), whereas under Cd treatment, a significant reduction in roots and shoots length was observed, showing a lower tolerance to the metal than that observed in the previous germination phase (**Figure 3**). However, shoots remained vital with no evidence of chlorosis. In *sylvestris* subjected to both metal stresses, root and shoot lengths decreased with the increase in heavy metal concentration, but at 25 μM, this was more evident in Cd than in As. Root length resulted more than twofold lower compared to untreated controls.

The root/seedling length ratio was also influenced by the metal and concentration used. In particular, for *sylvestris* in presence of Cd, the ratios were 80, 42, and 47% at 0, 25, and 50 μM, respectively, whereas they were 43, 39, and 33% in *altilis* at the same metal concentrations. As for As treatments, the ratios were

FIGURE 6 | The level of gene expression of HMA in seedlings of *altilis* and *sylvestris* grown up to 2 and 3 weeks under Cd **(A,B)** and As **(C,D)** at different concentrations. The fold change is the Ct value with respect to the housekeeping genes (EF1α and GAPDH) that is considered 1. The error bars represent the error mean of three biological replicates. Letters indicate only significantly different values ($P \leq 0.05$) between concentrations in each moment of sampling.

80 and 35% at 0 and 25 μM in *sylvestris*, and at 50 μM, seeds germinated, but their growth was arrested eventually. For *altilis* at 0, 25, and 50 μM of As, the root/seedling length ratios were 43, 49, and 52%, respectively, showing minor changes compared to control. Regarding the shoot/roots length ratio across the treatments, it varied very strongly in *sylvestris*, compared to both *altilis* and control.

Identification of Cardoon Genes Likely Associated With Plant Stress Response

The BLASTN and BLASTP analyses (**Supplementary Tables S1, S2**) of obtained cDNA contigs confirmed the identity and reliability of our sequences, not previously identified in cardoon. In fact, the nucleotide sequences present in cardoon database are just predicted from amino acid sequences.

In both genotypes, *altilis* and *sylvestris*, transcriptional levels can be influenced by the concentration of metals in the medium, type of metals, and time of growth. As regards the NRAMP3 gene, its expression level was strongly influenced by the genotype and the time of growth ($P < 0.05$) (**Figure 4**). The ANOVA at one way showed a high relation between concentration of metals and gene expression, more in *sylvestris* than in *altilis* ($P < 0.01$). For both shoots and roots, NRAMP3 was expressed higher in *sylvestris* and down-regulated in *altilis*. In particular, after 3 weeks, at least a twofold and fourfold increase in expression levels was observed in *sylvestris*, respectively, with 25 and 50 μM regardless of the type of metal used.

Similarly to NRAMP3, the transcriptional levels of ZIP11 were influenced by the genotype and time of growth ($P < 0.05$). In both genotypes, the type of metal did not influence the
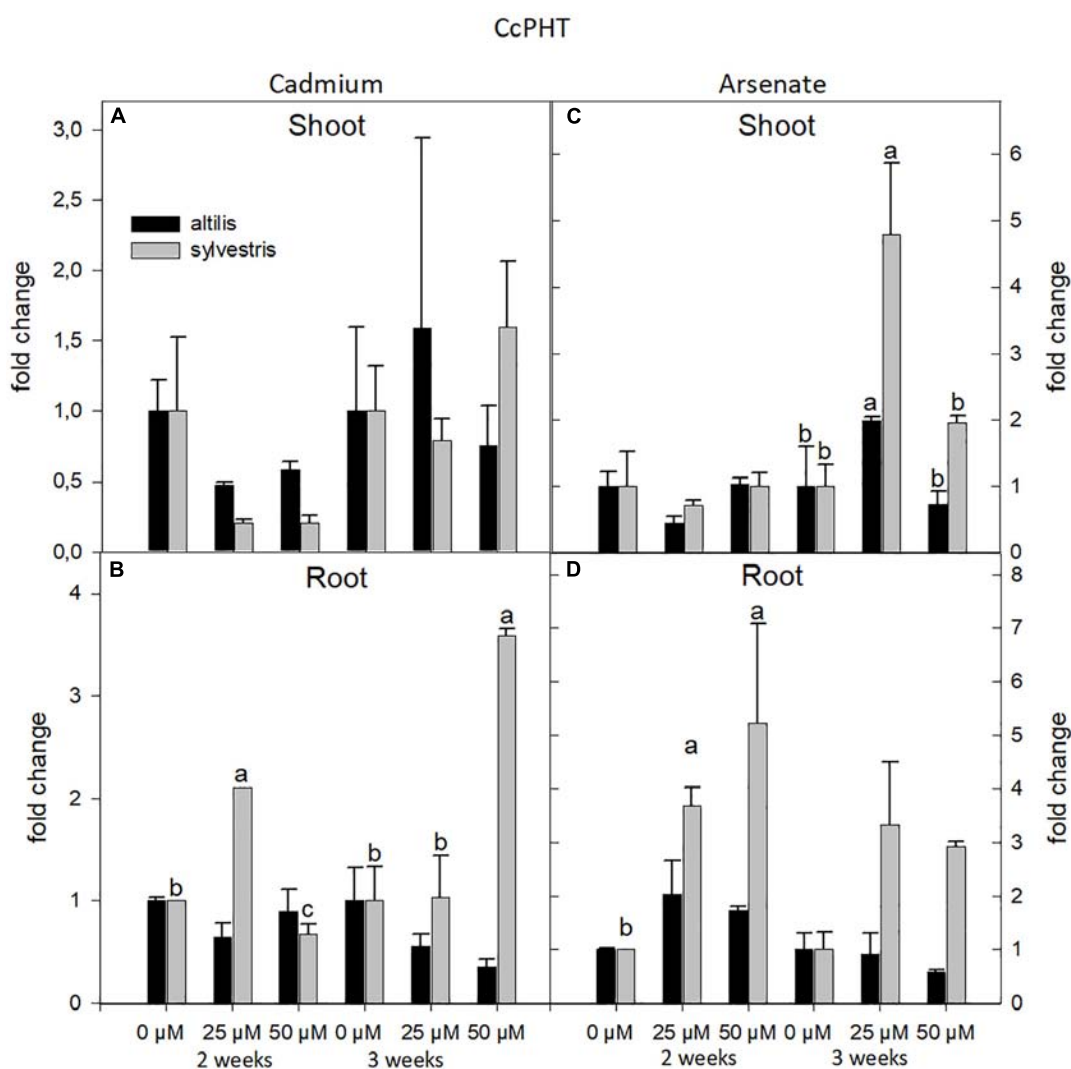
FIGURE 7 | The level of gene expression of PHT in seedlings of *altilis* and *sylvestris* grown up to 2 and 3 weeks under Cd **(A,B)** and As **(C,D)** at different concentrations. The fold change is the Ct value with respect to the housekeeping genes (EF1α and GAPDH) that is considered 1. The error bars represent the error mean of three biological replicates. Letters indicate only significantly different values ($P \leq 0.05$) between concentrations in each moment of sampling.

transcriptional levels. In roots of *sylvestris* after 3 weeks, the level of ZIP11 mRNAs significantly increased in presence of Cd and As ($P < 0.05$), whereas in *altilis*, the transcriptional levels decreased with increasing levels of both metals (**Figure 5**).

The expression of HMA was influenced by concentration and genotype. In particular, highly significant values, contributed by two of the three sources of variation (genotype, time, and concentration of metal), were observed in root treated with Cd ($P < 0.05$). Shoots and roots showed similar response to stresses, averaging the contribution of all the three sources of variation ($P < 0.05$), but the transcriptional level at 3 weeks in shoots of *altilis* with As was threefold increased than in *sylvestris* (**Figure 6**).

The transcriptional level of PHT was strongly influenced by the metal and genotype. In particular, highly significant values, contributed by the three sources of variation (genotype, time, and concentration of metal), were observed in shoot and root treated

with As ($P < 0.01$). In roots, but not in shoots, the expression of the PHT transcript, averaging the contribution of all the three sources of variation, increased compared to control, and this was more evident in *sylvestris* than *altilis* ($P < 0.01$). The type of metal affected the expression of PHT. In fact, in presence of As, the PHT transcriptional level showed a twofold increase with respect to Cd. In particular, after 2 weeks in the *sylvestris* roots, the expression increased linearly with the concentration of metal ranging from zerofold, threefold, and to fivefold at 0, 25, and 50 μM of Cd, respectively (**Figure 7**). Under As treatment, the genotype (*sylvestris*) was the factor that more markedly contributed to transcriptional increase in both roots and shoots after 3 weeks of treatment ($P < 0.001$).

Moreover, the ABCC transcriptional levels were mainly influenced by the treatments used (type of metal and its concentration). In particular, highly significant values, contributed by the three sources of variation (genotype, time,
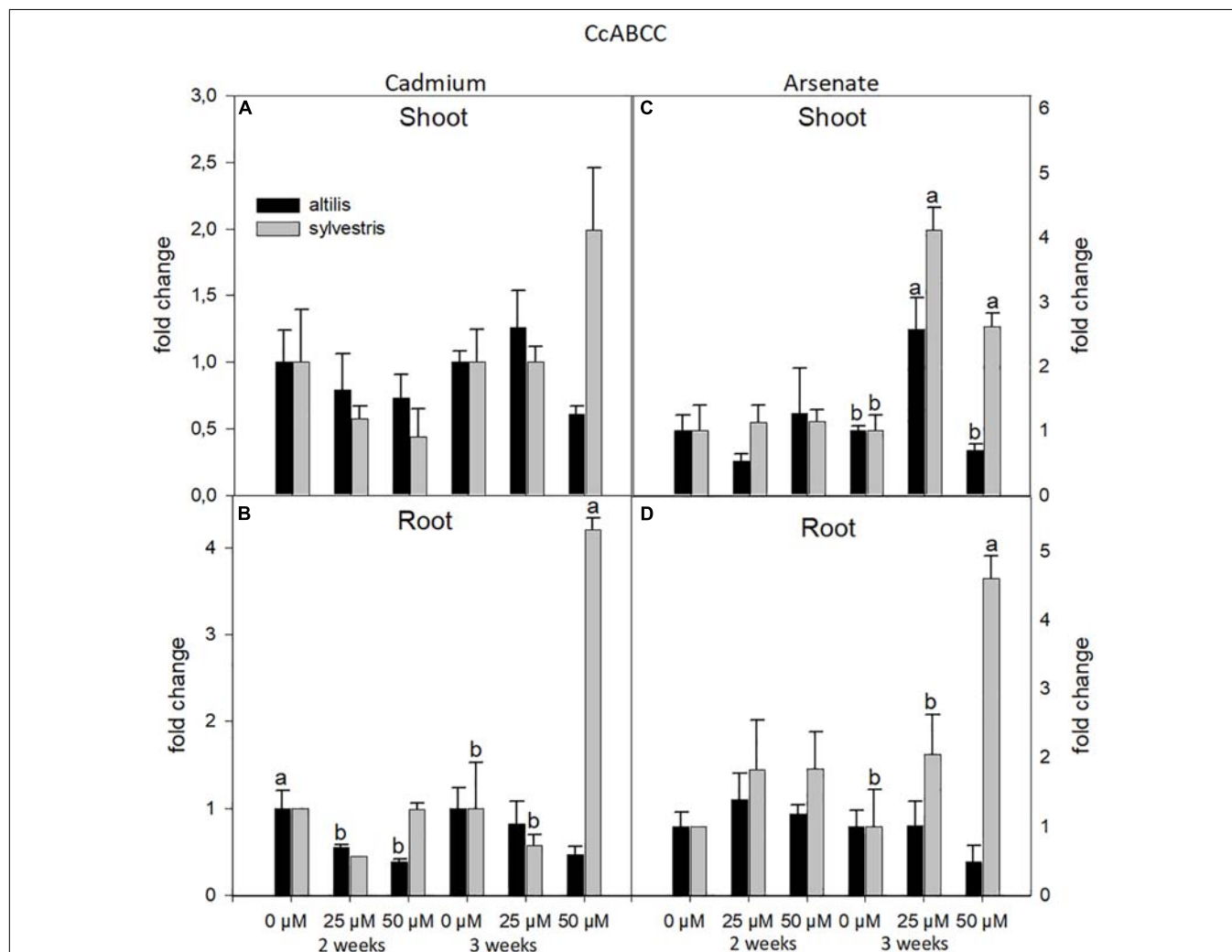
FIGURE 8 | The level of gene expression of ABCC in seedlings of *altilis* and *sylvestris* grown up to 2 and 3 weeks under Cd **(A,B)** and As **(C,D)** at different concentrations. The fold change is the Ct value with respect to the housekeeping genes (EF1α and GAPDH) that is considered 1. The error bars represent the error mean of three biological replicates. Letters indicate only significantly different values ($P \leq 0.05$) between concentrations in each moment of sampling.

and concentration of metal), were observed in root treated with As ($P < 0.001$).

On average, ABCC gene expression was not affected by the organ type (shoots or roots), but it was mainly influenced by the type of metal. In fact, with As, and not with Cd, the ABCC transcriptional levels increased by twofold in *sylvestris* compared to *altilis* (**Figure 8**). In particular, in wild cardoon, in shoots at 3 weeks, and in roots with As at 2 and 3 weeks, the ABCC mRNA was up-regulated, showing a clear involvement of this gene in As response ($P < 0.01$).

Principal Component Analysis

The PCA obtained using gene expression data and morphological characteristics in response to the presence of metals shows that for both varieties the values range from the first to the fourth quarter (**Figure 9**). However, *altilis* forms more compact clusters, which are more dispersed for sylvestris, instead. Moreover, the cluster localization is more influenced by metal type for *altilis*, whereas it is more dependent on the metal concentration for *sylvestris*.
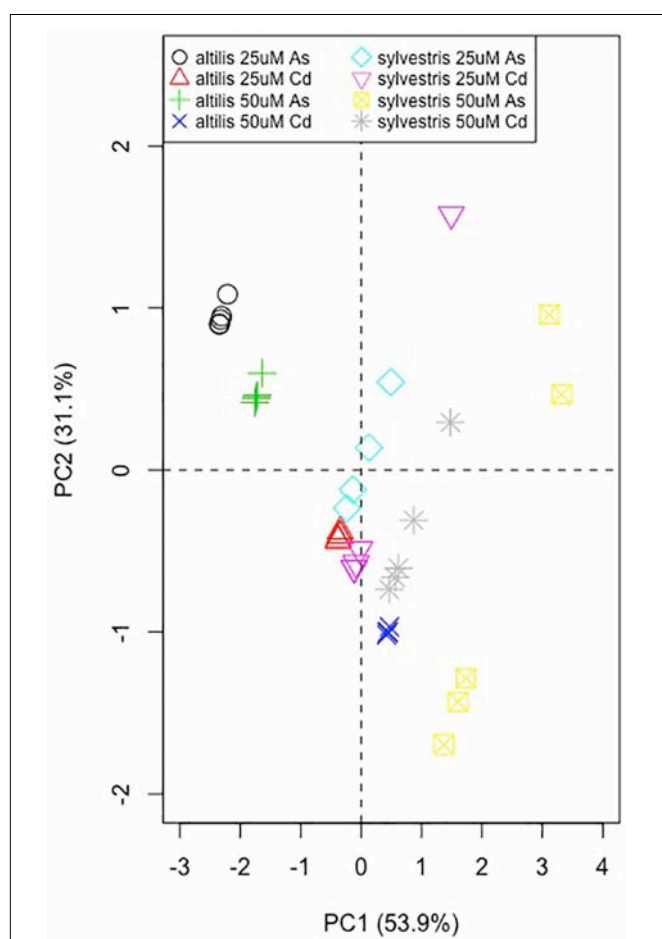
FIGURE 9 | Principal component analysis using the RT-qPCR fold-change values of the five metal-response–associated genes along with the shoot and root length.

DISCUSSION

In this study, we observed that salt stress did not reveal any significant effects on germination of domestic cardoon, while it linearly limited the development of fresh sprout biomass with the increase in NaCl concentrations. However, dry weight of sprouts did not show any significant variation between the salt and the control conditions, suggesting that the NaCl presence influences the water uptake capability of the plant decreasing water content of the seedlings. This salinity tolerance trait, in the early developmental stages, can be considered a plant adaptive strategy that allows us to consider *C. cardunculus* as a facultative halophyte species as reported by Benlloch-González et al. (2005).

As expected, salt stress induced the synthesis of phenolic substances in proportion to the increase in NaCl concentrations, confirming the important role of these molecules for the tolerance to stress conditions in plants. As well as for the total phenols, the trend of AA was directly proportional to salt concentration, demonstrating that the increase in the polyphenol content corresponds to a greater AA. These results reveal the possible use of salinity as an efficient technique for increasing the secondary metabolite content in plants grown for nutraceutical use, as documented for other elicitors or species (Bistgani et al., 2019; Hassini et al., 2019). In particular, in *C. cardunculus*, which is well-known for its high amount of polyphenolic compounds (Pandino et al., 2011), technical interventions based on the variation of NaCl concentrations in the germination solution could effectively modulate the bioactive molecules content in the sprouts.

In this study, for the first time, the influence of As and Cd on the growth of different genotypes of *C. cardunculus* sprouts was investigated. Their presence reduced the sprout growth in a significantly different way depending on the *C. cardunculus* genotypes. In particular, the *sylvestris* root and shoot lengths were mainly influenced by the concentration of metals; instead, the *altilis* growth was more affected by both the type of metal and its concentration. This finding is in accordance with growth rate documented for wheat and halophytes species in presence of heavy metals (Öncel et al., 2000; Leonardi et al., 2016; Amari et al., 2017).

Moreover, in the present study, we identified five *C. cardunculus* genes, NRAMP3, ZIP11, HMA, PHT, and ABCC, which seem to be involved in heavy metal stress response. Our results showed that in *sylvestris*, NRAMP3 expression is up-regulated in roots in presence of As or Cd with an increase in expression levels with longer treatments. This result agrees with Fallen et al. (2005), in which NRAMP3 and NRAMP4 were associated with Cd2^+ efflux from the vacuole. Their overexpression increased Cd sensitivity and determined the release of vacuolar Fe2^+ in *Arabidopsis*. Instead, no data are available in literature on NRAMP3 expression with respect to As.

ZIP proteins are generally responsible for the metal-ion homeostasis through the uptake of cations into the cytosol (Colangelo and Guerinot, 2006; Hou et al., 2017). Usually ZIP transporters are involved in the uptake and accumulation of Fe and Zn, but may also be responsible for Cd or other heavy metal transport (Guerinot, 2000). In *Solanum torvum* roots, IRT2 and

ZIP11 are associated with Zn transport (Xu et al., 2012). In the present study, expression of the ZIP11 transporter in wild cardoon was higher in shoot and roots subjected to Cd treatment. Similarly, ZIP11 mRNA levels increased after 3 weeks of exposure of the seedlings to As, and this is the first study to document the expression variation of this gene in association with As presence.

The uptake of As(V) in plants occurs via inorganic phosphate (Pi) system, because Pi transporters cannot distinguish between the similar electrochemical profiles of Pi and As(V) (Sánchez-Pardo et al., 2015). In our experiments, in *sylvestris* genotype, the phosphate transporter was up-regulated in roots under As treatment, expression of which was strongly influenced by the concentrations of metal. In *altilis*, the expression levels of PHT were elevated also in control; probably for this reason, we did not observe any significant variation in the gene expression levels across the treatments.

These results can be associated with what was observed by Di Tusa et al. (2016), who showed in *Pteris vittata* an increase in As accumulation when the plants expressed PvPht1;3. In *Arabidopsis*, the expression pattern of PHT1;1 in the presence of As(V) decreased significantly as compared to limiting Pi condition in the natural variants, whereas the expression of PHT1;4 was higher in presence of As(V) to limiting Pi condition (Shukla et al., 2015).

Despite the documented role of the HMAs in the heavy metal transport in other species (Verbruggen et al., 2009; Fan et al., 2018), the identified cardoon HMA did not show a significant difference in the expression levels in the presence of Cd or As as measured for the other genes. Our results indicate that its expression level was mainly influenced by concentration and genotype, with the *altilis* showing to be more sensitive with respect to *sylvestris*.

In cardoon, ABCC transcriptional levels, measured under As treatment, in roots of *sylvestris*, remained up-regulated compared to untreated sample. The increase in the expression was influenced by the time of exposure, with the highest level at 50 μM after 3 weeks. A similar response was observed in shoots after 3 weeks of As treatment. These results are in accordance with Song et al. (2010), who showed that *Arabidopsis* isoforms AtABCC1 and AtABCC2 mediate AsIII–PC complex transport to the vacuole, and overexpression of AtABCC1 increases As tolerance only when coexpressed with phytochelatins (PCS). In rice, a similar ABC transporter, OsABCC1, is critical for the vacuolar AsIII–PC sequestration and As detoxification, thus reducing As accumulation in rice grains. For this reason, knockout of OsABCC1 leads to the increase in As sensitivity (Song et al., 2014).

The results obtained in the present study lead us to conclude that cardoon seed germination and seedling establishment can take place in salt and heavy metal stress conditions. In addition, we documented the possibility to use abiotic stresses to improve the bioactive molecules content in cardoon sprouts. Moreover, we identified some important genes encoding for metal transportation that may be involved in the uptake and translocation of As and Cd in *C. cardunculus*.

These findings, in a context of climate change and environmental pollution, can be useful tools for the possibility of exploiting marginal lands for the cultivation of species, such as cardoon plants, able to develop in stressed environments and suitable for green chemistry and energetic purposes. This chance could represent an economically valid alternative for farmers and for the agriculture of the future.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

All authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis, approved the final version of the manuscript. SR, HP conceived and designed the study. HP, CG, VT, and GP contributed to production and assembly of data. HP, CG, VT, GP, and SR analyzed and interpreted the data. HP, CG, VT, and GP drafted the manuscript. SR critically revised the manuscript for important intellectual content. HP and GP contributed to the statistical analysis. SR supervised the study.

SUPPLEMENTARY MATERIAL

REFERENCES

Amari, T., Ghnaya, T., and Abdelly, C. (2017). Nickel, cadmium and lead phytotoxicity and potential of halophytic plants in heavy metal extraction. *S. Afr. J. Bot.* 111, 99–110. doi: 10.1016/j.sajb.2017.03.011

Argento, S., Puglia, G., Pappalardo, H., Pulvirenti, M., Melilli, M. G., and Raccuia, S. A. (2016). Seed germination responses to salt stress in wild and cultivated sicilian cardoon genotypes. *Acta Hortic.* 1147, 9–14. doi: 10.17660/actahortic.2016.1147.2

Begum, F., Karmoker, J. L., Fattah, Q. A., and Maniruzzaman, A. F. M. (1992). The effect of salinity on germination and its correlation with K+, Na+, Cl-

accumulation in germinating seeds of *Triticum aestivum* L. cv. *Akbar. Plant Cell Physiol.* 33, 1009–1014. doi: 10.1093/oxfordjournals.pcp.a078324

Benlloch-González, M., Fournier, J. M., Ramos, J., and Benlloch, M. (2005). Strategies underlying salt tolerance in halophytes are present in *Cynara cardunculus. Plant Sci.* 168, 653–659. doi: 10.1016/j.plantsci.2004.09.035

Bistgani, Z. E., Hashemi, M., DaCosta, M., Craker, L., Maggi, F., and Morshedloo, M. R. (2019). Effect of salinity stress on the physiological characteristics, phenolic compounds and antioxidant activity of *Thymus vulgaris* L. and *Thymus daenensis Celak. Ind. Crops Prod.* 135, 311–320. doi: 10.1016/j.indcrop.2019.04.055

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

Chang, S., Puryear, J., and Cairney, J. (1993). A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* 11, 113–116. doi: 10.1007/bf02670468

Ciancolini, A., Alignan, M., Pagnotta, M. A., Vilarem, G., and Crinò, P. (2013). Selection of italian cardoon genotypes as industrial crop for biomass and polyphenol production. *Ind. Crops Prod.* 51, 145–151. doi: 10.1016/j.indcrop.2013.08.069

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

Cuartero, J., Bolarín, M. C., Asíns, M. J., and Moreno, V. (2006). Increasing salt tolerance in the tomato. *J. Exp. Bot.* 57, 1045–1058. doi: 10.1093/jxb/erj102

Cun, P., Sarrobert, C., Richaud, P., Chevalier, A., Soreau, P., Auroy, P., et al. (2014). Modulation of Zn/Cd P1B2-ATPase activities in arabidopsis impacts differently on Zn and Cd contents in shoots and seeds. *Metallomics* 6, 2109–2116. doi: 10.1039/c4mt00182f

Cushman, J. C., and Bohnert, H. J. (2000). Genomic approaches to plant stress tolerance. *Curr. Opin. Plant Biol.* 3, 117–124. doi: 10.1016/s1369-5266(99)00052-7

Dewanto, V., Wu, X., Adom, K. K., and Liu, R. H. (2002). Thermal processing enhances the nutritional value of tomatoes by increasing total antioxidant activity. *J. Agric. Food Chem.* 50, 3010–3014. doi: 10.1021/jf0115589

Di Tusa, S. F., Fontenot, E. B., Wallace, R. W., Silvers, M. A., Steele, T. N., Elnagar, A. H., et al. (2016). A member of the phosphate transporter 1 (Pht1) family from the arsenic-hyperaccumulating fern *Pteris vittata* is a high-affinity arsenate transporter. *New Phytol.* 209, 762–772. doi: 10.1111/nph.13472

Fallen, P., Kienzler, K., and Krieger-Liszkay, A. (2005). Mechanism of Cd^{2+} toxicity: Cd^{2+} inhibits photoactivation of photosystem II by competitive binding to the essential Ca^{2+} site. *Biochim. Biophys. Acta Bioenerg.* 1706, 158–164. doi: 10.1016/j.bbabio.2004.10.005

Fan, W., Liu, C., Cao, B., Qin, M., Long, D., Xiang, Z., et al. (2018). Genome-wide identification and characterization of four gene families putatively involved in cadmium uptake, translocation and sequestration in mulberry. *Front. Plant Sci.* 9:879. doi: 10.3389/fpls.2018.00879

Fedoroff, N. V., Battisti, D. S., Beachy, R. N., Cooper, P. J. M., Fischhoff, D. A., Hodges, C. N., et al. (2010). Radically rethinking agriculture for the 21st century. *Science* 327, 833–834. doi: 10.1126/science.1186834

Genovese, C., Brundo, M. V., Toscano, V., Tibullo, D., Puglisi, F., and Raccuia, S. A. (2016a). Effect of *Cynara* extracts on multiple myeloma cell lines. *Acta Hortic.* 1147, 113–118. doi: 10.17660/actahortic.2016.1147.16

Genovese, C., Platania, C., Venticinque, M., Calderaro, P., Argento, S., Scandurra, S., et al. (2016b). Evaluation of cardoon seeds presscake for animal feeding. *Acta Hortic.* 1147, 323–328. doi: 10.17660/actahortic.2016.1147.45

Gominho, J., Curt, M. D., Lourenço, A., Fernández, J., and Pereira, H. (2018). *Cynara cardunculus* L. as a biomass and multi-purpose crop: a review of 30 years of research. *Biomass Bioenergy* 109, 257–275. doi: 10.1016/j.biombioe.2018.01.001

Guerinot, M. L. (2000). The ZIP family of metal transporters. *Biochim. Biophys. Acta Biomembr.* 1465, 190–198. doi: 10.1016/s0005-2736(00)00138-3

Gupta, D. K., Corpas, F. J., and Palma, J. M. (eds) (2013). *Heavy Metal Stress in Plants.* Heidelberg: Springer Berlin Heidelberg.

Gupta, S. K., and Goyal, M. R. (2017). *Soil Salinity Management in Agriculture.* New York, NY: Apple Academic Press.

Hassini, I., Rios, J. J., Garcia-Ibañez, P., Baenas, N., Carvajal, M., and Moreno, D. A. (2019). Comparative effect of elicitors on the physiology and secondary metabolites in broccoli plants. *J. Plant Physiol.* 239, 1–9. doi: 10.1016/j.jplph.2019.05.008

Hou, J., Liu, X., Cui, B., Bai, J., and Wang, X. (2017). Concentration-dependent alterations in gene expression induced by cadmium in *Solanum lycopersicum*. *Environ. Sci. Pollut. Res.* 24, 10528–10536. doi: 10.1007/s11356-017-8748-4

Houle, G., Morel, L., Reynolds, C. E., and Siégel, J. (2001). The effect of salinity on different developmental stages of an endemic annual plant, *Aster laurentianus* (Asteraceae). *Am. J. Bot.* 88, 62–67. doi: 10.2307/2657127

Ierna, A., Mauro, R. P., and Mauromicale, G. (2012). Biomass, grain and energy yield in *Cynara cardunculus* L. as affected by fertilization, genotype and harvest time. *Biomass Bioenergy* 36, 404–410. doi: 10.1016/j.biombioe.2011.11.013

Kavamura, V. N., and Esposito, E. (2010). Biotechnological strategies applied to the decontamination of soils polluted with heavy metals. *Biotechnol. Adv.* 28, 61–69. doi: 10.1016/j.biotechadv.2009.09.002

Kim, D. Y., Bovet, L., Maeshima, M., Martinoia, E., and Lee, Y. (2007). The ABC transporter AtPDR8 is a cadmium extrusion pump conferring heavy metal resistance. *Plant J.* 50, 207–218. doi: 10.1111/j.1365-313X.2007.03044.x

Krämer, U. (2010). Metal hyperaccumulation in Plants. *Annu. Rev. Plant Biol.* 61, 517–534. doi: 10.1146/annurev-arplant-042809-112156

Lanquar, V., Ramos, M. S., Lelièvre, F., Barbier-Brygoo, H., Krieger-Liszkay, A., Krämer, U., et al. (2010). Export of vacuolar manganese by *AtNRAMP3* and *AtNRAMP4* is required for optimal photosynthesis and growth under manganese deficiency. *Plant Physiol.* 152, 1986–1999. doi: 10.1104/pp.109.150946

LeBlanc, M. S., McKinney, E. C., Meagher, R. B., and Smith, A. P. (2013). Hijacking membrane transporters for arsenic phytoextraction. *J. Biotechnol.* 163, 1–9. doi: 10.1016/j.jbiotec.2012.10.013

Leonardi, C., Pappalardo, H., Genovese, C., Puglia, G., Bua, G. D., Dugo, G., et al. (2016). Mechanisms of phytoextraction in *Cynara cardunculus* L. growing under cadmium and arsenic stress. *Acta Hortic.* 1147, 139–144. doi: 10.17660/actahortic.2016.1147.19

Llugany, M., Miralles, R., Corrales, I., Barceló, J., and Poschenrieder, C. (2012). *Cynara cardunculus* a potentially useful plant for remediation of soils polluted with cadmium or arsenic. *J. Geochem. Explor.* 123, 122–127. doi: 10.1016/j.gexplo.2012.06.016

Mauromicale, G., and Licandro, P. (2002). Salinity and temperature effects on germination, emergence and seedling growth of globe artichoke. *Agronomie* 22, 443–450. doi: 10.1051/agro:2002011

Mauromicale, G., and Ierna, A. (2004). Biomass and grain yield in *Cynara cardunculus* l. genotypes grown in a permanent crop with low input. *Acta Hortic.* 660, 593–598. doi: 10.17660/actahortic.2004.660.91

Mendoza-Cózatl, D. G., Jobe, T. O., Hauser, F., and Schroeder, J. I. (2011). Long-distance transport, vacuolar sequestration, tolerance, and transcriptional responses induced by cadmium and arsenic. *Curr. Opin. Plant Biol.* 14, 554–562. doi: 10.1016/J.PBI.2011.07.004

Miransari, M. (2011). Hyperaccumulators, arbuscular mycorrhizal fungi and stress of heavy metals. *Biotechnol. Adv.* 29, 645–653. doi: 10.1016/j.biotechadv.2011.04.006

Mittler, R. (2002). Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci.* 7, 405–410. doi: 10.1016/s1360-1385(02)02312-9

Mulligan, C. N., Yong, R. N., and Gibbs, B. F. (2001). Remediation technologies for metal-contaminated soils and groundwater: an evaluation. *Eng. Geol.* 60, 193–207. doi: 10.1016/s0013-7952(00)00101-0

Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* 59, 651–681. doi: 10.1146/annurev.arplant.59.032607.092911

Öncel, I., Keleş, Y., and Üstün, A. S. (2000). Interactive effects of temperature and heavy metal stress on the growth and some biochemical compounds in wheat seedlings. *Environ. Pollut.* 107, 315–320. doi: 10.1016/s0269-7491(99)00177-3

Oomen, R. J. F. J., Wu, J., Lelièvre, F., Blanchet, S., Richaud, P., Barbier-Brygoo, H., et al. (2009). Functional characterization of NRAMP3 and NRAMP4 from the metal hyperaccumulator *Thlaspi caerulescens*. *New Phytol.* 181, 637–650. doi: 10.1111/j.1469-8137.2008.02694.x

Ottaiano, L., Di Mola, I., Impagliazzo, A., Cozzolino, E., Masucci, F., Mori, M., et al. (2017). Yields and quality of biomasses and grain in *Cynara cardunculus* L. grown in southern Italy, as affected by genotype and environmental conditions. *Ital. J. Agron.* 11, 375–382. doi: 10.4081/ija.2017.954

Pagnotta, M. A., and Noorani, A. (2014). "Genetic diversity assessment in European Cynara collections" in Genomics of Plant Genetic Resources, eds R., Tuberosa, A., Graner, E., Frison (Dordrecht: Springer), 559–584. doi: 10.1007/978-94-007-7572-5_23

Pandino, G., Lombardo, S., Mauromicale, G., and Williamson, G. (2011). Phenolic acids and flavonoids in leaf and floral stem of cultivated and wild *Cynara cardunculus* L. genotypes. *Food Chem.* 126, 417–422. doi: 10.1016/j.foodchem.2010.11.001

Pappalardo, H., Leonardi, C., Genovese, C., Toscano, V., Melilli, M. G., and Raccuia, S. A. (2016). Effects of heavy metals on seedlings germination and

growth in different cardoon genotypes. *Acta Hortic.* 1147, 281–288. doi: 10. 17660/actahortic.2016.1147.39

Petropoulos, S., Fernandes, Â, Stojkoviæ, D., Pereira, C., Taofiq, O., Di Gioia, F., et al. (2019). Cotton and cardoon byproducts as potential growing media components for *Cichorium spinosum* L. commercial cultivation. *J. Clean. Prod.* 240:118254. doi: 10.1016/j.jclepro.2019.118254

Puglia, G. D., Prjibelski, A. D., Vitale, D., Bushmanova, E., Schmid, K. J., and Raccuia, S. A. (2019). Hybrid transcriptome sequencing approach improved assembly and gene annotation in *Cynara cardunculus* (L.). *BMC Genomics*

Raccuia, S. A., Cavallaro, V., and Melilli, M. G. (2004a). Intraspecific variability in *Cynara cardunculus* L. var. sylvestris Lam. Sicilian populations: seed germination under salt and moisture stresses. *J. Arid Environ.* 56, 107–116. doi: 10.1016/s0140-1963(03)00006-5

Raccuia, S. A., Mainolfi, A., Mandolino, G., and Melilli, M. G. (2004b). Genetic diversity in *Cynara cardunculus* revealed by AFLP markers: comparison between cultivars and wild types from Sicily. *Plant Breed.* 123, 280–284. doi: 10.1111/j.1439-0523.2004.00983.x

Raccuia, S. A., and Melilli, M. G. (2007). Biomass and grain oil yields in *Cynara cardunculus* L. genotypes grown in a mediterranean environment. *Field Crop Res.* 101, 187–197. doi: 10.1016/j.fcr.2006.11.006

Rottenberg, A., and Zohary, D. (1996). The wild ancestry of the cultivated artichoke. *Genet. Resour. Crop Evol.* 43, 53–58. doi: 10.1007/bf00126940

Rozen, S., and Skaletsky, H. (1999). Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods protoc.* 132, 365–386. doi: 10.1385/1-59259-192-2:365

Sánchez-Pardo, B., Cantero, C., and Zornoza, P. (2015). Alleviation of arsenic stress in cardoon plants via the supply of a low cadmium concentration. *Environ. Exp. Bot.* 109, 229–234. doi: 10.1016/j.envexpbot.2014.07.004

Shukla, T., Kumar, S., Khare, R., Tripathi, R. D., and Trivedi, P. K. (2015). Natural variations in expression of regulatory and detoxification related genes under limiting phosphate and arsenate stress in *Arabidopsis thaliana*. *Front. Plant Sci.* 6:898. doi: 10.3389/fpls.2015.00898

Singh, S., Parihar, P., Singh, R., Singh, V. P., and Prasad, S. M. (2016). Heavy metal tolerance in plants: role of transcriptomics, proteomics, metabolomics, and ionomics. *Front. Plant Sci.* 6:1143. doi: 10.3389/fpls.2015. 01143

Song, W. Y., Park, J., Mendoza-Cózatl, D. G., Suter-Grotemeyer, M., Shima, D., Hörtensteiner, S., et al. (2010). Arsenic tolerance in arabidopsis is mediated by two ABCC-type phytochelatin transporters. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21187–21192. doi: 10.1073/pnas.1013964107

Song, W. Y., Yamaki, T., Yamaji, N., Ko, D., Jung, K. H., Fujii-Kashino, M., et al. (2014). A rice ABC transporter, OsABCC1, reduces arsenic accumulation in the grain. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15699–15704. doi: 10.1073/pnas. 1414968111

Thomine, S., Lelièvre, F., Debarbieux, E., Schroeder, J. I., and Barbier-Brygoo, H. (2003). AtNRAMP3, a multispecific vacuolar metal transporter involved in plant responses to iron deficiency. *Plant J.* 34, 685–695. doi: 10.1046/j.1365-313x.2003.01760.x

Tiwari, M., Sharma, D., Dwivedi, S., Singh, M., Tripathi, R. D., and Trivedi, P. K. (2014). Expression in arabidopsis and cellular localization reveal involvement of rice NRAMP, *OsNRAMP1*, in arsenic transport and tolerance. *Plant Cell Environ.* 37, 140–152. doi: 10.1111/pce.12138

Toscano, V., Sollima, L., Genovese, C., Melilli, M. G., and Raccuia, S. A. (2016). Pilot plant system for biodiesel and pellet production from cardoon: technical and economic feasibility. *Acta Hortic.* 1147, 429–442. doi: 10.17660/actahortic. 2016.1147.60

Turco, R., Tesser, R., Cucciolito, M. E., Fagnano, M., Ottaiano, L., Mallardo, S., et al. (2019). *Cynara cardunculus* biomass recovery: an eco-sustainable, nonedible resource of vegetable oil for the production of poly(lactic acid) bioplasticizers. *ACS Sustain. Chem. Eng.* 7, 4069–4077. doi: 10.1021/acssuschemeng.8b05519

Verbruggen, N., Hermans, C., and Schat, H. (2009). Mechanisms to cope with arsenic or cadmium excess in plants. *Curr. Opin. Plant Biol.* 12, 364–372. doi: 10.1016/j.pbi.2009.05.001

Wu, J., Zhao, F. J., Ghandilyan, A., Logoteta, B., Guzman, M. O., Schat, H., et al. (2009). Identification and functional analysis of two ZIP metal transporters of the hyperaccumulator *Thlaspi caerulescens*. *Plant Soil* 325, 79–95. doi: 10.1007/ s11104-009-0151-6

Xu, J., Sun, J., Du, L., and Liu, X. (2012). Comparative transcriptome analysis of cadmium responses in *Solanum nigrum* and *Solanum torvum*. *New Phytol.* 196, 110–124. doi: 10.1111/j.1469-8137.2012.04235.x

Yang, X. E., Ye, H. B., Long, X. X., He, B., He, Z. L., Stoffella, P. J., et al. (2005). Uptake and accumulation of cadmium and zinc by sedum alfredii hance at different Cd/Zn supply levels. *J. Plant Nutr.* 27, 1963–1977. doi: 10.1081/pln-200030082

Zhu, J.-K. (2016). abiotic stress signaling and responses in plants. *Cell* 167, 313–324. doi: 10.1016/j.cell.2016.08.029

Enhancing Biomass and Lutein Production From *Scenedesmus almeriensis*: Effect of Carbon Dioxide Concentration and Culture Medium Reuse

Antonio Molino[1], Sanjeet Mehariya[1,2], Angela Iovine[1,2], Patrizia Casella[1], Tiziana Marino[2], Despina Karatza[2], Simeone Chianese[2]* and Dino Musmarra[2]

[1] Department of Sustainability-CR Portici, ENEA Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Portici, Italy, [2] Department of Engineering, University of Campania "Luigi Vanvitelli", Aversa, Italy

The main purpose of this study is to investigate the effects of operative parameters and bioprocess strategies on the photo-autotrophic cultivation of the microalgae *Scenedesmus almeriensis* for lutein production. *S. almeriensis* was cultivated in a vertical bubble column photobioreactor (VBC-PBR) in batch mode and the bioactive compounds were extracted by accelerated solvent extraction with ethanol at 67°C and 10 MPa. The cultivation with a volume fraction of CO_2 in the range 0–3.0%v/v showed that the highest biomass and lutein concentrations – 3.7 g/L and 5.71 mg/g, respectively – were measured at the highest CO_2 concentration and using fresh growth medium. Recycling the cultivation medium from harvested microalgae resulted in decreased biomass and lutein content. The nutrient chemical composition analysis showed the highest consumption rates for nitrogen and phosphorus, with values higher than 80%, while sulfate and chloride were less consumed.

Keywords: microalgae, photo-autotrophic cultivation, *Scenedesmus almeriensis*, lutein, ASE

INTRODUCTION

Lutein is a carotenoid found in plants and phototrophic microorganisms. It is classified as a primary xanthophyll because of the presence of two hydroxyl functional groups in its chemical structure (Chen et al., 2017; Xie et al., 2017; Bowen et al., 2018). Lutein acts as a light-harvesting pigment that improves photosynthesis efficiency and prevents photo-damage in plant cells (Ruban et al., 2007). For its outstanding antioxidant, anti-inflammatory and colorant properties, lutein is widely used as a nutraceutical for human health (Fernández-Sevilla et al., 2010). Lutein is used in the treatment of age-related macular degeneration, and for the prevention of cardiovascular diseases and some types of cancers (Coleman and Chew, 2007; Cerón et al., 2008). The existing commercial source of natural lutein is marigold (*Tagetes erecta L.*) petals (Lin et al., 2015b). However, the lutein content of marigold petals is around 0.03% (based on dry biomass weight), and their harvesting and separation carry a high cost (Lin et al., 2015a). Therefore, there is a need for an alternative source for cost-effective production.

Microalgae have been considered as a potential source of natural lutein as their pigment content (0.5-1.2% based on dry biomass weight) is far higher than that of conventional sources (Cerón et al., 2008; Sánchez et al., 2008b; Lin et al., 2015a,b; Chen et al., 2017). Additionally, producing lutein from microalgae offers several environmental benefits, such as a carbon dioxide mitigation as their photosynthesis efficiency is 10-15 times higher than that of terrestrial plants (Lam et al., 2012; Yen et al., 2013; Xie et al., 2014). The microalgae growth rate is 5–10 times higher than that of higher plants. They can be grown in seawater and brackish water, as well as on non-arable land, therefore they do not compete with conventional agriculture crops for resources (Del Campo et al., 2001; Utomo et al., 2013; Ho et al., 2014; Lin et al., 2015a). Scenedesmus almeriensis is recognized as a rich source of lutein, containing up to 4.5 mg/g (dry weight), when grown in outdoor culture conditions (Del Campo et al., 2007); moreover, lutein content can be increased by manipulating growth conditions, such as light intensity and temperature, till to 5.4 mg/g (dry weight) (Sánchez et al., 2008b).

Christian et al. (2018) repot that the astaxanthin content of Haematococcus pluvialis using high concentrations of CO_2 (15%v/v) as the carbon source can achieve the value of about 36 mg/g (dry weight). Cheng et al. (2016) observed the highest biomass (0.65 g/L) and astaxanthin (45 mg/L) concentrations in Haematococcus pluvialis grown in 6% CO_2.

Microalgae cultivation requires large amounts of water and nutrients, which reduces the cost-effectiveness of the entire bio-compound extraction process (Hadj-Romdhane et al., 2013). The re-use of culture media could be a solution for the development of large-scale cultures to minimize water use and nutrient consumption (Fret et al., 2017). For the cultivation of Scenedesmus obliquus, Livansky et al. (1996) found that water use and nutrient consumption could be reduced of about 64 and 16%, respectively, by recycling the growth medium. For Chlorella vulgaris Hadj-Romdhane et al. (2012) reported that the use of optimized culture conditions during medium recycling could decrease water use and nutrient consumption of 75 and 62%, respectively.

As CO_2 is used as the carbon source, microalgae cultivation also contributes to CO_2 sequestration, which makes it one of the most promising approaches to deal with global warming (Salih, 2011; Chen and Liu, 2018). Not only does the higher CO_2 concentration increase the carbon source available for the growth of microalgae, but it also improves the assimilation of nutrients in their biomass (de Assis et al., 2019). The microalgae can efficiently convert atmospheric CO_2 into organic biomass via carbon fixation. Several inexpensive sources of CO_2 can be explored, such as CO_2 -rich industrial exhaust gases and fermentation effluent gases that could make the microalgae cultivation process cost-effective and eco-friendly (Lam et al., 2012; Xie et al., 2014; Chen and Liu, 2018). Patel et al. (2019) cultivated Chlorella protothecoides in 5% of CO_2 and obtained 4.12 g/L of biomass. Xie et al. (2017) cultivated Desmodesmus sp. F51 and evaluated the effect of CO_2 concentration on microalgae biomass and lutein production. Six different CO_2 concentrations (0.03, 2.5, 5.0, 7.5, 10.0, and 12.5%) were used. The biomass productivity and the specific growth rate were higher when

CO_2 concentration increased from 0.03 to 2.5% and decreased when CO_2 concentration was further increased to 12.5% (Xie et al., 2017). On the basis of the above-mentioned studies 3.0% was chosen as maximum value to evaluate the effect of CO_2 concentration in this study.

Moheimani (2016) showed that Tetraselmis suecica can be grown in CO_2 from a coal-fired power plant flue gas and by reusing the growth medium. CO_2 biofixation with nutrient recycling and the addition of monoethanolamine were tested on Spirulina sp. cultivation by da Rosa et al. (2015). They highlighted that Spirulina can be produced using recycled medium, in spite of a reduce protein and lipid content. Cui et al. (2019) showed that using flue gas from a biomass plant and recycling the growth to cultivate Spirulina sp. a 42% lower nutrient consumption was achieved, with no significant differences between fresh medium and recycled medium in terms of protein and phycocyanin contents.

To the best of the authors knowledge, the combined effect of CO_2 concentration and medium recycling on the S. almeriensis growth and lutein production has not been investigated anywhere before. The present study aims at this dual purpose: developing a lab-scale methodology to recycle the supernatant/filtrate growth medium obtained from the harvesting of the microalgae biomass, and assessing how different CO_2 concentrations (0–3%v/v) and fresh and recycled growth media influence biomass and lutein production. The biomass was harvested by filtration, then lutein was obtained by accelerated solvent extraction (ASE) at 67°C and 10 MPa. Ethanol, a green solvent belonging to the class of the Generally Recognized as Safe (GRAS) solvents, was used during the extraction step. The lutein content was measured by u-HPLC.

MATERIALS AND METHODS

Microalgae and Growth Medium

Seed culture of S. almeriensis was provided by AlgaRes Srl (Rome, Italy), and used for the cultivation under laboratory conditions. Microalgae cells were cultivated in a modified Mann & Myers medium (Mann and Myers, 1968; Barceló-Villalobos et al., 2019), consisting of $NaNO_3$ (1.0 g/L), K_2HPO_4 (0.1 g/L), $MgSO_4*7H_2O$ (1.2 g/L), and $CaCl_2$ (0.3 g/L).

Moreover, 10 mL of a solution of micronutrients, containing Na_2EDTA (0.001 mg/L), $MnCl_2$ (1.4 mg/L), $ZnSO_4*7H_2O$ (0.33 mg/L), $FeSO_4*2H_2O$ (2 mg/L), $CuSO_4*5H_2O$ (0.002 mg/L), and $Co(NO_3)_2*6H_2O$ (0.007 mg/L), were added to 990 mL of the growth medium.

Photo-Bioreactor

Scenedesmus almeriensis was cultivated in a vertical bubble column photo-bioreactor (VBC-PBR), made of plexiglass, with a working volume of 1.25 L (effective height = 680 mm; external diameter = 60 mm; thickness = 10 mm) and with a volume to surface ratio (V/S) of 11.5 L/m². The VBC-PBR was fed with a gaseous mixture ($N_2/O_2/CO_2$) from tanks and was equipped to a monitor and control system which allowed fine tuning of the gaseous mixture flow rate, the temperature, the pH and the light intensity. The bottom of the reactor was equipped with three

TABLE 1 | Scheme of the inoculum and the medium reuse during the cultivation of *Scenedesmus almeriensis* cultures under batch mode at different cultivation conditions.

Experiment	Experimental setup	CO$_2$ concentration (%v/v)	Source of inoculum	Growth phase - composed by		
				Inoculum volume	Fresh medium volume	Reuse medium volume
A	Fresh medium	Air	Algares	200	1000	0
B	Fresh medium	0.5	Culture transfer after 10 days of growth with air*	1200	0	0
C	Reuse medium	0.5	Culture from 0.5%v/v CO$_2$ with fresh medium**	270	550	390
D	Fresh medium	1.5	Culture from 3.0% v/v CO$_2$ with reuse medium***	40	1160	0
E	Reuse medium	1.5	Culture from 1.5% v/v CO$_2$ with fresh medium****	280	550	370
F	Fresh medium	3	Culture from 0.5% v/v CO$_2$ with reuse medium*****	70	1130	0
G	Reuse medium	3	Culture from 3.0% v/v CO$_2$ with fresh medium******	230	660	310

*Inoculum derived from sample A; **inoculum derived from sample B; ***inoculum derived from sample G; ****inoculum derived from sample D; *****inoculum derived from sample C; ******inoculum derived from sample F.

FIGURE 1 | Experimental set-up schematization.
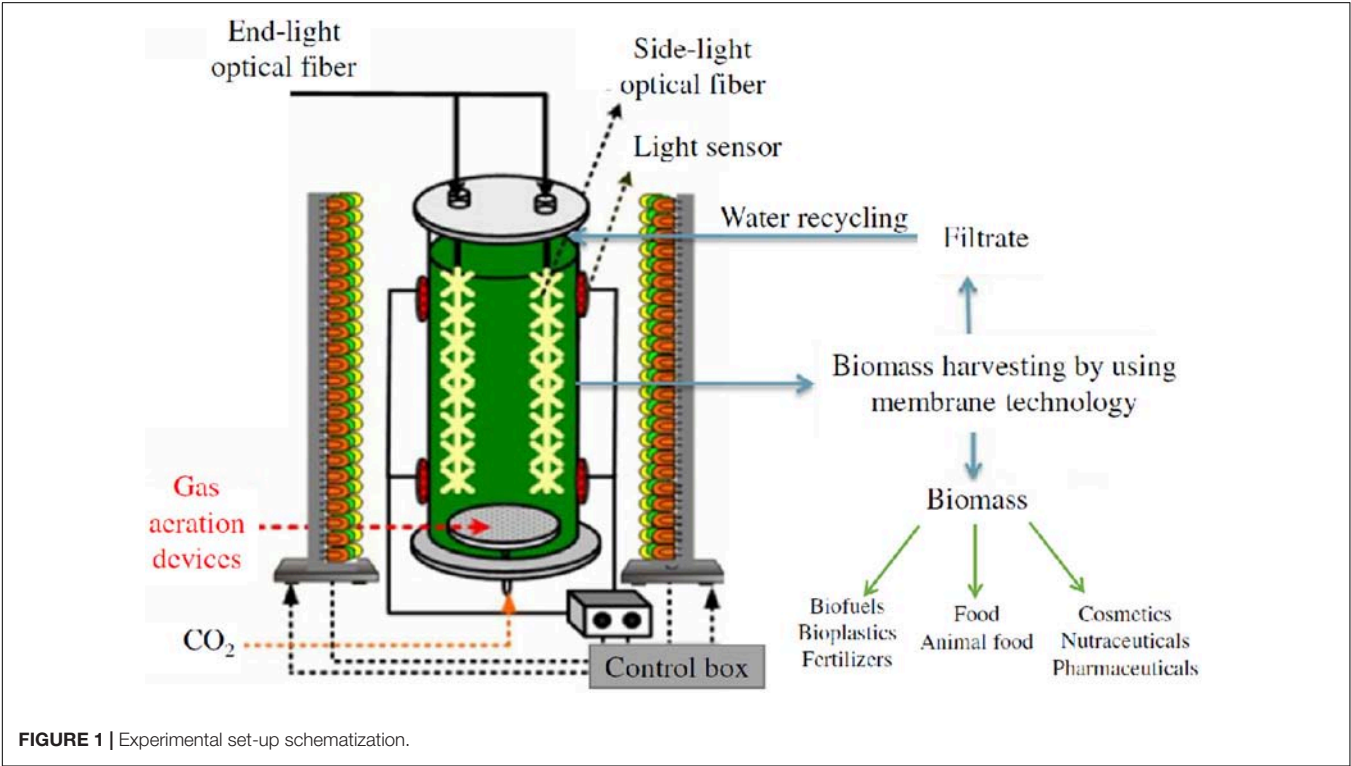
sintered steel spargers, installed through 3 fileted holes (1/8″), to feed the gaseous mixture into the reactor. The gaseous mixture flow rate was regulated by Bronkhorst gas flow controllers^{TM}, with a flow control accuracy of 0.5%. The top of the reactor was equipped with a temperature sensor (thermocouple) and with a pH sensor. The temperature control system consisted of an AISI 316L coaxial pipe (diameter = 60.3 mm; thickness = 1 mm) in which water was used as cooling fluid. Thanks to a heat pump, the temperature control system allowed to regulate the temperature inside the reactor with a precision of ±1°C in the range 15–35°C.

The lighting system consisted of a semi-cylinder structure, located at a distance of 100 mm from the VBC-PBR, with blue, white and red lights from a selective LED system (only blue/only white/only red or a mix of them). The lighting system was controlled and regulated by a SCADA (Supervisory Control and Data Acquisition) system, equipped with a touch-screen, a custom software and a PC to collect and to record experimental data of temperature, gas flow rate, pH, and light intensity. A schematization of the experimental set-up is sketched in **Figure 1**.

Growth Conditions

The microalgae were grown under white light with a lux intensity of 4000 lux on the surface of the VBC-PBR, and with a gaseous mixture flow rate (N$_2$/O$_2$/CO$_2$) of 50 mL/min, in which the CO$_2$ content was varied in the range 0–3.0%v/v (O$_2$ = 21%v/v). The temperature was kept constant at 28°C. The pH of the culture medium changed in the range 7.5–8.5 due to the addition of CO$_2$.
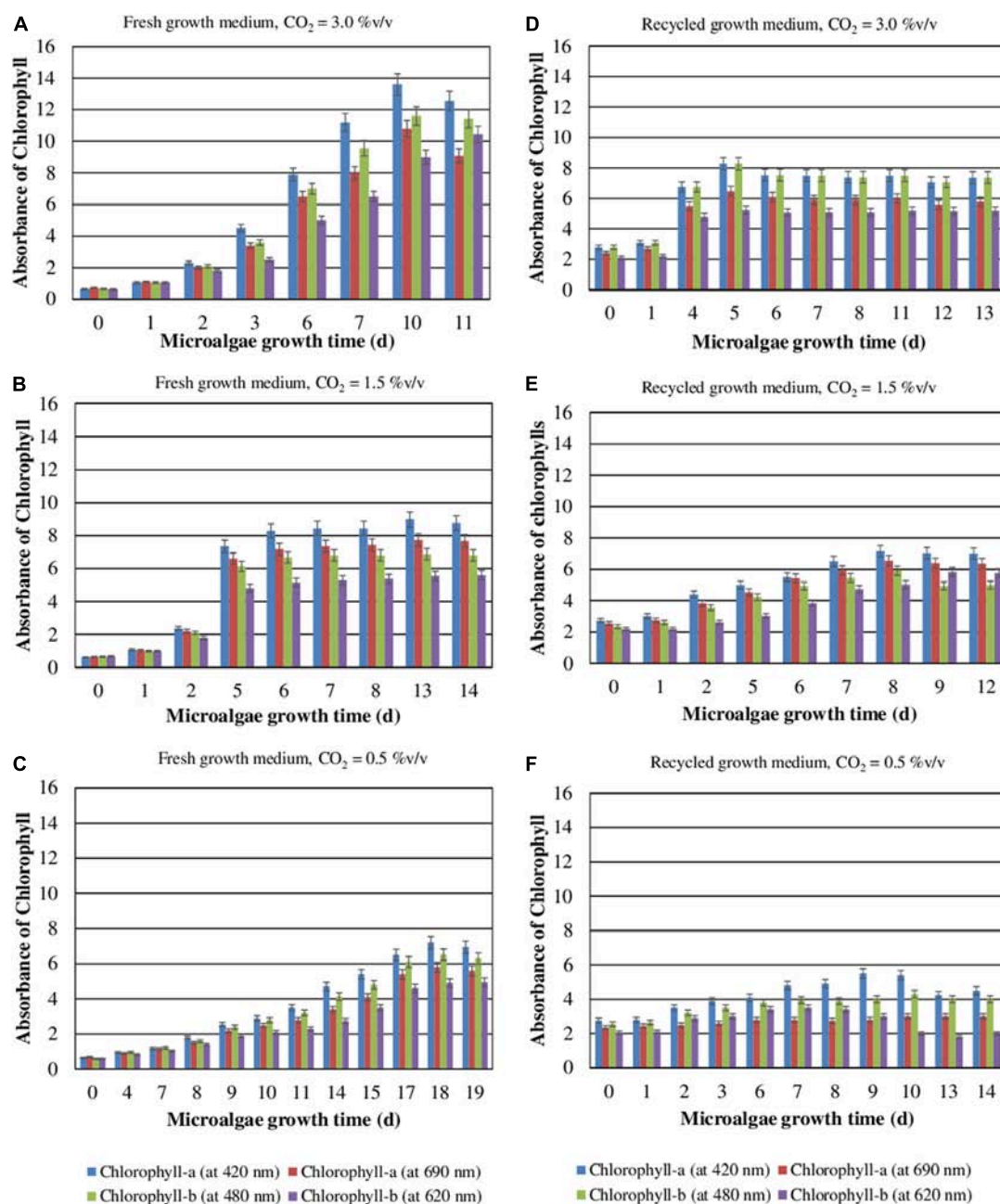
FIGURE 2 | Contents of chlorophyll-a and chlorophyll-b as a function of the *S. almeriensis* growth time at different CO_2 contents with fresh **(A–C)** and recycled growth medium **(D–F)**. Standard deviation was calculated on three biological replications.

Two cultivation conditions were investigated. The microalgae were first cultured in the fresh medium from Mann and Myers with the composition described herein (see section "Microalgae and growth medium"). Then around 300 mL of the culture medium were stored to be used as an inoculum in further experiments and the remaining culture was filtered to measure the biomass and lutein content. In the second step, the growth medium recovered by filtration was mixed with the fresh medium and with a given amount of inoculum to achieve an optical density of 0.6–0.7 at 420 nm (see section "Microalgae and growth medium"). In both cases, CO_2 varied in the range 0–3.0%v/v. Each experimental step was carried out with a working volume of 1.2 L and the compositions of the medium reused are reported in **Table 1**. Each experimental condition for microalgae growth (expressed as chlorophyll content and biomass concentration), including, nutrient chemical analysis, extraction yield and lutein content measurements, was investigated in replicates, and for each condition, the standard deviation (SD) value was calculated.

FIGURE 3 | *Scenedesmus almeriensis* concentration (based on dry biomass weight) as a function of growth time at different CO₂ contents with fresh **(A)** and recycled **(B)** growth medium. Standard deviation was calculated on three biological replications.

Microalgae Growth Assessment

The *S. almeriensis* cell growth was monitored by determining the absorbance of the samples at 420 nm (Chlorophyll-a), 480 nm (Chlorophyll-b), 690 nm (Chlorophyll-a), and 620 nm (Chlorophyll-b) with a UV/Visible spectrophotometer (Multiskan, Thermo Fisher Scientific, United States), as reported in the literature (Lu et al., 2017; Almomani and Örmeci, 2018; Chirivella-Martorell et al., 2018). The biomass dry weight (BDW) was calculated using the absorbance values at different biomass concentrations measured during the growth phase in the recycled medium. The following

calibration curve between absorbance and concentration was obtained:

$$DBC = (0.0867 * A) - 0.1868$$

where DBC is the concentration of biomass on dry weight (g/L) and A is the total absorbance obtained from the sum of the absorbance values at the four chlorophyll wavelengths (**Supplementary Figure 1**).

For the final dry weight determination, cell cultures were dewatered by vacuum filtration using a vacuum filter with a pore size of 0.45 μm (Sigma-Aldrich, United States) and the

pellets were lyophilized for 24 h. Three biological replications were carried out.

Accelerated Solvent Extraction

Lutein was extracted from mechanically pre-treated *S. almeriensis* cells by the accelerated solvent extraction method, using the Dionex-ASE 200 extractor (Salt Lake City, UT, United States). The pre-treatment was performed via ball milling according to the procedure described elsewhere (Mehariya et al., 2019). Four consecutive extraction cycles were performed using ethanol, a green solvent belonging to the class of the Generally Recognized as Safe (GRAS) solvents, at 67°C and 10 MPa for the complete biomass discoloring. The extraction conditions were optimized elsewhere (Molino et al., 2018c). At the end of each extraction run (20 min), the extracts were collected in 40 mL amber glass vials, by flushing the system with 6.6 mL of fresh solvent, and the system was purged for 1 min with nitrogen (Purity ≥ 99.999%). Five technical replications were carried out.

Growth Medium Characterization

The chemical analysis of the nutrient concentrations (initial and final) was carried out using an ion Chromatograph (Dionex ICS-1100, Thermo Fisher Scientific, Massachusetts, United States). The Dionex ICS-1100 is an integrated ion chromatography system equipped with a pump, an injection valve, and a conductivity detector. Several nutrients, such as Mg^{2+}, SO_4^{2-}, Na^+, NO_3^-, NO_2^-, Ca_2^+, Cl^-, K^+, and PO_4^{3-} were analyzed.

The extract obtained after each extraction cycle was divided in equal parts and placed in two different vials adding

FIGURE 4 | *Scenedesmus almeriensis* productivity (based on dry biomass weight) at different CO$_2$ contents with fresh and with recycled growth medium. Standard deviation was calculated on three biological replications.

TABLE 2 | *Scenedesmus almeriensis* productivity (based on dry biomass weight) at different CO$_2$ contents with fresh and with recycled growth medium: ANOVA (one-way; α = 0.05) results.

CO$_2$ concentration (%v/v)	*p*-value
0.5	0.00049
1.5	0.0017
3	0.088

BHT at 0.1wt% as an antioxidant for saponification and gravimetric analysis.

Lutein Measurement

The total lutein content was gravimetrically quantified after the complete removal of the solvent using a Zymark TurboVap evaporator (Zymark, Hopkinton, MA, United States). Before measuring lutein, the saponification of the samples was carried out in order to remove lipids and chlorophyll, for avoiding the overlap of the spectra with the species present in the carotenoid family (Vechpanich and Shotipruk, 2011; Molino et al., 2018a,b; Sanzo et al., 2018). In particular, the saponification was carried out adding 1 mL of a NaOH solution in methanol (0.05 M) to 5 mL of the extract. This solution was left in the dark in an inert atmosphere for 7 h. Once this step was completed, the sample was neutralized using 3 mL of a NH$_4$Cl solution in methanol (0.05 M). After saponification, lutein was measured using a u-HPLC Agilent 1290 Infinity II with Zorbax reverse phase C18 column with the methanol-water (95:5, v/v) mixture as the mobile phase solvent. Flow rate and column temperature were kept constant at 0.4 mL/min and 28°C, respectively. Five technical replications were carried out.

Statistical Analysis

ANOVA analysis (one-way; α = 0.05) was carried out to compare the results of the effect of the growth medium (fresh and recycled) at different CO$_2$ on the biomass productivity, on the nutrient consumption, and on the extraction yield and lutein content.

RESULTS AND DISCUSSION

CO$_2$ Content and Recycled Medium Effects on Chlorophyll Content

During their autotrophic growth, the microalgae perform photosynthesis using CO$_2$ as the inorganic source of carbon. The effects of CO$_2$ concentration and medium reuse on the chlorophyll content during the autotrophic cultivation of *S. almeriensis* were evaluated (**Figure 2**). The results show that the accumulation of chlorophyll increases when CO$_2$ content raises and decreases when the medium is recycled. The results also demonstrate that the higher the CO$_2$ concentration the higher the chlorophyll content and the lower the time required to reach the peak. The maximum concentrations of chlorophyll-a at 420 nm and chlorophyll-b at 480 nm, 13.6 and 11.6 respectively, were achieved in 10 days with a CO$_2$ content of 3.0%v/v and the fresh medium. In addition, the recycled growth medium exhibited a lower photosynthetic efficiency, in terms of chlorophyll content reduction (**Figure 2**), which could be due to a lower amount of nutrients available. The amount of chlorophyll needed for an efficient light absorption in autotrophic cultivation, where light is the only source of energy, may explain these results. Moreover, the toxic compounds accumulating in the medium during the first culture step may inhibit the photosynthesis efficiency, which may lead to a decrease in the chlorophyll content.

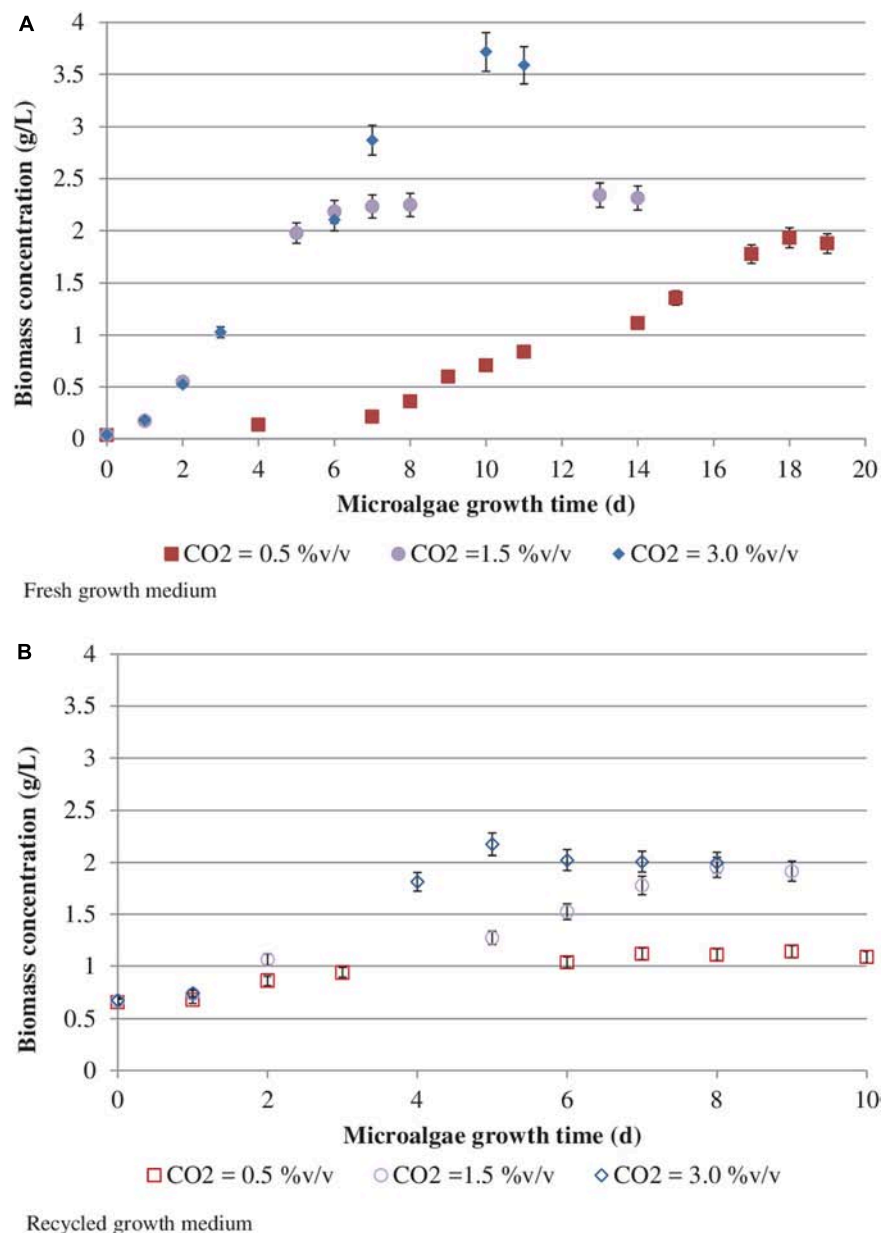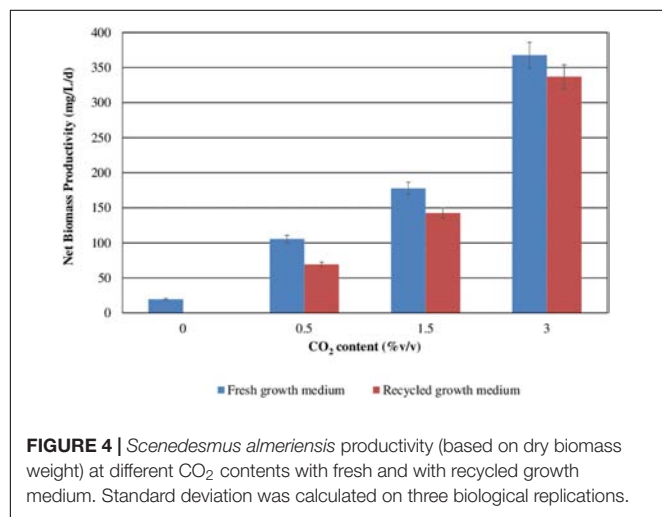## CO$_2$ Content and Recycled Medium Effects on Biomass Concentration

Figure 3 reports *S. almeriensis* concentration (based on dry biomass weight) as a function of growth time for different CO$_2$ contents with fresh and recycled growth medium. **Figure 4** reports *S. almeriensis* productivity (based on dry biomass weight) for different CO$_2$ contents with fresh and with recycled growth medium. Results in **Figure 3** show that with the fresh growth medium, the higher the CO$_2$ content the higher the microalgae concentration and the lower the cultivation time. The highest biomass concentration was equal to 3.7 g/L and was achieved by aeration with 3.0%v/v of CO$_2$ in 10 days. With the CO$_2$ contents of 1.5%v/v and of 0.5%v/v, the biomass concentration was about 2.3 g/L (cultivation time = 14 days) and about 1.9 g/L (cultivation time 18 days), respectively. The biomass concentration decreased when the recycled growth medium was used. After 6 days of cultivation, the reused medium caused saturation in cell growth for all the investigated CO$_2$ concentrations. Results in **Figure 4** show that *S. almeriensis* productivity markedly increases as CO$_2$ concentrations is increased, and evidence that the productivity is higher by using fresh medium than recycled medium. These results agree with those found for continuous cultures of *C. vulgaris*, grown in a recycled medium (Hadj-Romdhane et al., 2012). The recovered medium might lead to an osmotic stress with adverse effects on biomass production and quality (Hadj-Romdhane et al., 2012). Rodolfi et al. (2003) reported that the cultivation of *Nannochloropsis* sp. in fed-batch mode with the recycled medium came to a halt when cell concentration reached about 3 g/L (growth time of 16 days). When the fresh medium was used, the biomass concentration increased above 5 g/L (growth time of 28 days). According to Rodolfi et al. (2003), *Nannochloropsis* sp. released soluble inhibitors and particulate organic matter in the supernatant that seemed to inhibit the cell growth, which could justify the finding of this study. Nevertheless, a statistical analysis (one-way ANOVA with α = 0.05) on the outcomes shown in **Figure 4** was performed and the results are reported in **Table 2**. Significant differences in

biomass productivity were observed between fresh and recycled growth media at 0.5% and 1.5% v/v CO$_2$ contents. No statistically significant differences were observed at 3% v/v CO$_2$ content.

Both in the presence of the fresh medium and of the recycled medium, the biomass productivity increases concomitantly with the CO$_2$ content, passing from about 100 to about 360 mg/L/day with the fresh medium, and from about 75 mg/L/day to about 340 mg/L/day with the recycled medium.

Dineshkumar et al. (2015a) reported that, by increasing CO$_2$ concentration from 0.8 to 2.5%, the photosynthesis efficiency improved, which may improve the biomass productivity of *Chlorella minutissima* MCC-27, while a further increase in CO$_2$ (>5%) negatively influenced the productivity of microalgae. The optimal value of CO$_2$ content, able to enhance the growth rate and the product accumulation in *C. minutissima* MCC- 27, of 3.5%v/v was found by the particle swarm optimization technique. However, it should be noted that the optimal CO$_2$ content for microalgae growth usually differs from species to species and depends on the culture medium and the growth conditions. It was found that *Chlorella* sp. and *Nannochloropsis oculata* exhibited an optimal growth with a CO$_2$ concentration of 2%v/v and the growth of microalgae was completely inhibited when the CO$_2$ was higher than 5%v/v (Chiu et al., 2008, 2009). Liu et al. (2019) found that the dynamics of biomass concentration, and the biomass productivity, was higher in *Arthrospira platensis* cultures aerated with a CO$_2$ content of 0.5%, while the growth decreased at lower or higher CO$_2$ contents than 0.5%v/v. When CO$_2$ contents were higher than 10%v/v, the algal cells only exhibited a slight increase at initial hours to days and finally bleached by the end of experiments. In summary, the optimum CO$_2$ level should be identified for each microalgae species to attain the highest biomass productivity (Santos et al., 2018).

CO$_2$ Content and Recycled Medium Effects on Nutrient Consumption

Microalgae growth requires an adequate amount of several nutrients. Among them, nitrogen (N) and phosphorus (P) are

TABLE 3 | Initial and final nutrient concentration during the cultivation of *Scenedesmus almeriensis* cultures under batch mode at different cultivation conditions.

	Concentration (mg/L)											
	0.5%v/v CO$_2$				1.5%v/v CO$_2$				3.0%v/v CO$_2$			
	Fresh medium		Reused medium		Fresh medium		Reused medium		Fresh medium		Reused medium	
Nutrients	IC	FC	IC	FC	IC	FC	IC	FC	IC	FC	IC	FC
Mg$_2$$^+$	90.25	56.26	80.21	68.88	115.22	58.88	88.41	72.10	111.06	50.55	90.56	70.78
SO$_4$$^{2-}$	435.21	355.76	440.12	400.56	406.28	378.61	410.58	398.25	448.87	400.25	425.21	411.72
Na$^+$	245.23	158.25	167.85	147.25	291.80	150.25	236.85	180.63	270.73	120.35	217.23	160.10
NO$_3$$^-$	520.38	9.85	333.97	10.52	722.34	10.70	356.25	70.58	630.89	0.00	426.60	11.77
Ca$_2$$^+$	81.46	57.89	85.07	65.86	101.45	28.74	66.25	45.15	94.18	23.59	81.79	47.10
Cl$^-$	147.85	92.20	136.55	119.95	183.84	144.30	160.58	152.32	182.53	157.25	125.25	123.67
K$^+$	35.81	22.65	35.22	30.83	42.41	24.12	38.98	33.52	39.42	20.79	41.03	34.35
PO$_4$$^{3-}$	17.81	0.00	23.65	0.00	52.79	0.00	22.52	0.00	40.18	0.00	25.49	0.00

Standard deviation was less than 5% in all operative conditions. **Where- IC**, *initial concentration;* **FC**, *final concentration.*
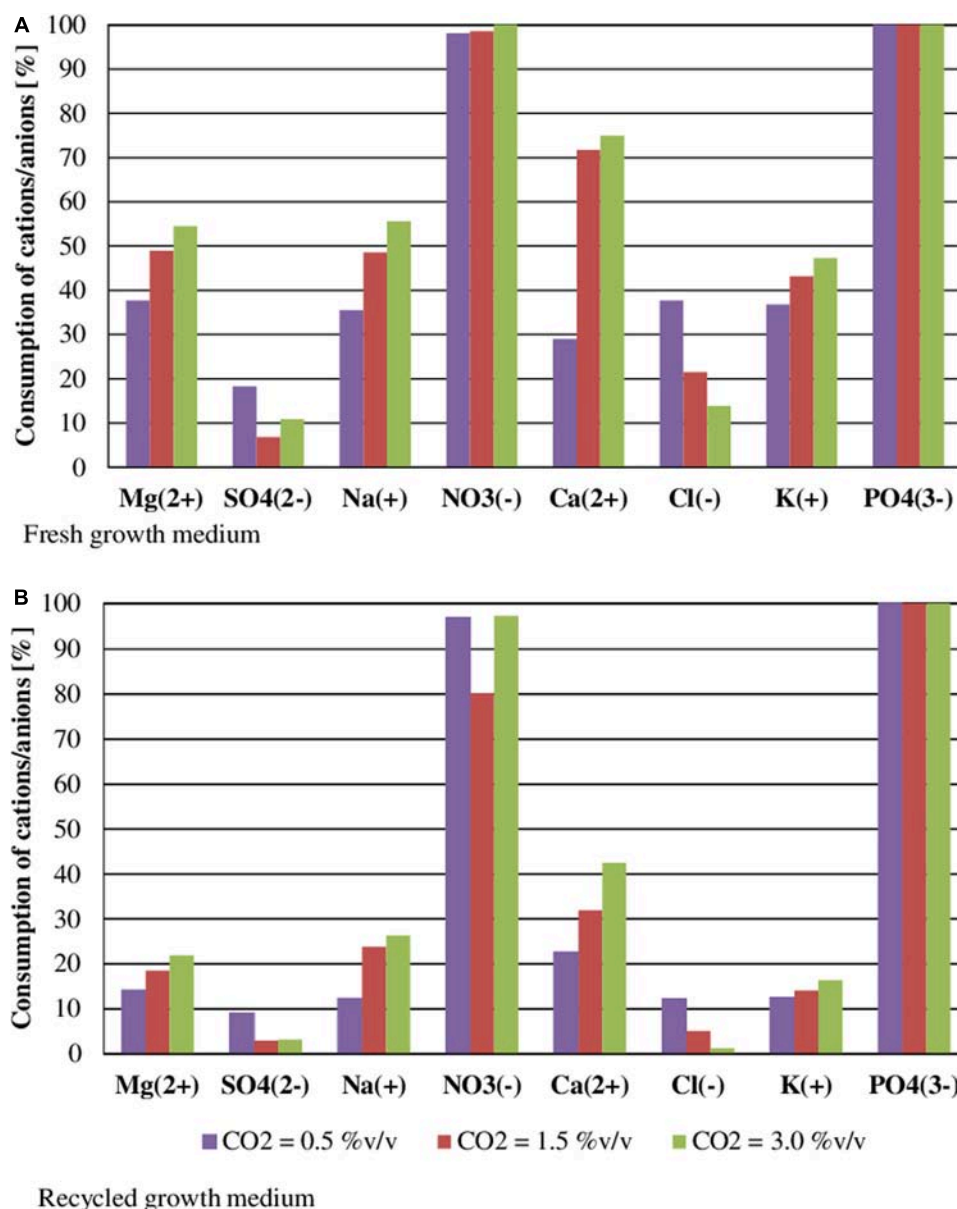
FIGURE 5 | Effect of CO$_2$ content on nutrient consumption during the cultivation of *S. almeriensis* with fresh growth medium **(A)** and with recycled growth medium **(B)**. Standard deviation was calculated on five technical replications and it was less than 5% in all operative conditions.

essential for the development of cells and their metabolic activity. They are usually used as buffer agents (Choi et al., 2017). In this study, initial and final concentrations of nutrients were measured (**Table 3**). The chemical analysis shows that the highest nutrient concentrations were found in the fresh medium. The initial concentration of nutrients was slightly different due to the varied nutrient concentration in the inoculum and filtrate.

Figure 5 shows nutrient consumption during the cultivation of *S. almeriensis* with fresh and reused growth medium in different CO$_2$ contents. N and P were highly consumed in all experimental conditions, as they are the main contributors to sustain microalgae growth (Lam et al., 2012; Choi et al., 2017).

At the end of the experiments PO$_4^{3-}$ was totally used up in every CO$_2$ concentration with both fresh and reused growth media, even if its initial concentration was varied in the range 17.81–52.79 mg/L (**Table 3**). This may suggest that PO$_4^{3-}$ is necessary for *S. almeriensis* growth, which confirms the results by Hadj-Romdhane et al. (2012). A similar consideration can be put forward for NO$_3^-$: 98–100% consumption rate with the fresh medium and a slightly lower rate (80–97%) with the recycled medium.

Results in **Table 3** indicate that the higher the CO$_2$ content the higher the consumption of nutrients, except for Cl$^-$ and SO$_4^{2-}$. The consumption of Cl$^-$ and of SO$_4^{2-}$ was lower than
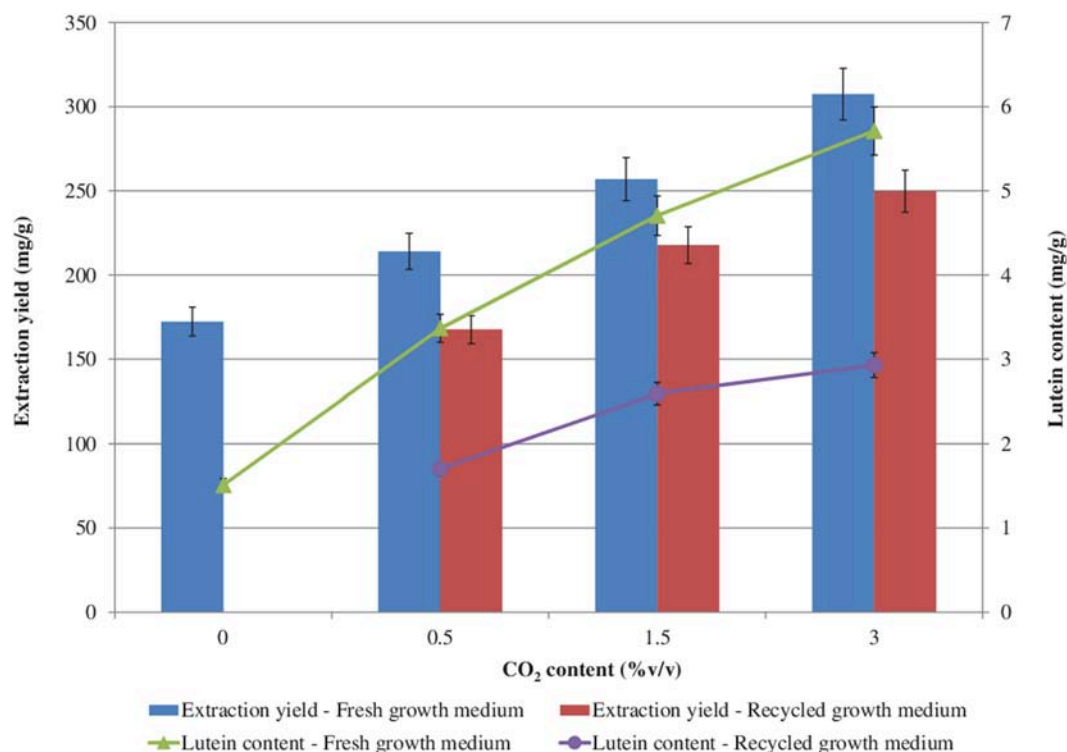
FIGURE 6 | Extraction yield and lutein content (based on dry biomass weight) from *S. almeriensis* at different CO_2 contents with fresh growth medium and with recycled growth medium. Standard deviation was calculated on five technical replications.

37 and 18%, respectively, with both the fresh and the recycled growth medium for all the investigated CO_2 contents, this results suggests that by growing *S. almeriensis* in the presence of CO_2 the biomass produced has a different composition and in particular a lower amount of Cl^- and SO_4^{2-}. Similar results were obtained in tests of *Chlorella vulgaris* cultivation by growth medium recycling (Hadj-Romdhane et al., 2013). The consumption of Ca_2^+, Cl^-, K^+, Mg_2^+, Na^+, NO_3^-, and SO_4^{2-} was reduced up to 40% using the recycled medium. Remarkably, during the growth with a CO_2 content of 0.5%v/v, the consumption of Cl^- was around 37 and 12% with the fresh and the reused growth

medium, respectively. During the growth with the recycled medium, an increase in osmolality can be found, leading to a decrease of the consumption of nutrients (Hadj-Romdhane et al., 2012). Therefore, the recycled growth medium could limit the growth of *S. almeriensis*, by decreasing the microalgae nutrient utilization.

Results of statistical analysis are reported in **Table 4**. Significant differences in nutrient consumption were observed at all CO_2 contents and for all nutrients, except NO_3^-, between fresh and recycled growth media. No statistically significant differences were observed in NO_3^- consumption at 0.5 and 3% v/v CO_2 contents.

CO_2 Content and Recycled Medium Effects on Extraction Yield and Lutein Production

The effects of CO_2 and of the recycled growth medium on the extraction yield and lutein content are shown in **Figure 6**. The higher the CO_2 concentration the higher the extraction yield and the lutein content, both with the fresh and with the recycled growth medium. However, with the recycled growth medium, slightly lower extraction yield and lutein content values were found with respect to the growth with the fresh one. The highest extraction yield (307.44 mg/g) and the highest lutein content (5.71 mg/g) were achieved with a CO_2 content of 3.0%v/v with the fresh growth medium.

TABLE 4 | Nutrient consumption at different CO_2 contents with fresh and with recycled growth medium: ANOVA (one-way; $\alpha = 0.05$) results.

	p-value		
	CO_2 concentration (%v/v)		
Nutrients	**0.5**	**1.5**	**3**
Mg^{2+}	0.000022	0.000022	0.000026
SO_4^{2-}	0.000065	0.000035	0.000012
Na^+	0.000019	0.000047	0.000041
NO_3^-	0.90	0.002564	0.37
Ca^{2+}	0.002159	0.000033	0.000099
Cl^-	0.000014	0.000008	0.000004
K^+	0.000016	0.000015	0.000018

TABLE 5 | Extraction yield and lutein extraction at different CO$_2$ contents with fresh and with recycled growth medium: ANOVA (one-way; $\alpha = 0.05$) results.

CO$_2$ concentration (%v/v)	p-value	
	Extraction yield	Lutein extraction
0.5	0.0042	0.00010
1.5	0.0151	0.00017
3	0.0073	0.00013

With the recycled growth medium, a 1.3-fold decrease in the extraction yield and a 2-fold decrease in the lutein content (maximum reductions) were measured, which might be attributed to the lower photosynthetic activity observed during the cultivation with the recovered growth medium (Hadj-Romdhane et al., 2012, 2013).

A slightly lower lutein content (5.56 mg/g) was obtained with 150 mg/L ammonium-N from *Desmodesmus* sp. F51 (Xie et al., 2017). Chen and Liu (2018) evaluated the effect of medium replacement during a two-stage cultivation process of *Chlorella sorokiniana* MB-1. The 60% replacement of the effluent from the first stage to the second stage with the fresh medium resulted in a lutein content below 3.5 mg/g, while the 80% replacement of the effluent showed a lutein concentration lower than 2.5 mg/g (Chen and Liu, 2018).

Results of statistical analysis are reported in **Table 5**. Significant differences in extraction yield and lutein extraction were observed between fresh and recycled growth media at all CO$_2$ contents.

The lutein content obtained in this study, equal to 5.71 mg/g, was lower than the value of 8.54 mg/g reported by Molino et al. (2019) who investigated the performance of the same microalgal species using a vertical bubble column photo-bioreactor with a volume to surface ratio (V/S) of 56.6 L/m^2. The different lutein content can be explained by the different photo-bioreactor configuration in the two research works: V/S ratio = 11.5 L/m^2 (this work) and = 56.6 L/m^2 (Molino et al., 2019). On the other hand, the value reported in this work is slightly higher than those reported in microalgae-based lutein studies using batch phototrophic conditions (Sánchez et al., 2008b; Ho et al., 2014; Xie et al., 2014, 2017), while it is comparable with the values reported by Sánchez et al. (2008a,b) who studied *S. almeriensis* growth in a continuous mode. Ho et al. (2014) optimized the effect of light intensity founding that 75 μmol/m^2/s was the optimum choice to attain higher lutein contents with continuous aeration of 2.5%v/v CO$_2$; the maximum lutein content of 5.52 mg/g from *Scenedesmus obliquus* FSP-3 was produced by supplying 8.0 mM calcium nitrate as the nitrogen source during cultivation (Ho et al., 2014). Dineshkumar et al. (2015a) cultivated *Chlorella minutissima* MCC-27 in a 2-L airlift photo-bioreactor and achieved a maximum lutein productivity of 3.45 mg/L/d in batch system. Moreover, lutein productivity of *Chlorella minutissima* MCC-27 could be further enhanced to 4.32 mg/L/d by optimization of the operative parameters such as light intensity, CO$_2$ concentration and gaseous flow rate (Dineshkumar et al., 2015a). These observations highlight the

TABLE 6 | Lutein content: microalgae comparison (adapted from Molino et al., 2019).

Microalgal strain	Lutein content (mg/g dry biomass)	References
Chlorella sorokiniana	3.0	Cordero et al., 2011
Scenedesmus obliquus FSP-3	4.52	Ho et al., 2014
Chlorella zofingiensis	7.2	Del Campo et al., 2004
Scenedesmus almeriensis	5.5	Sánchez et al., 2008a
Scenedesmus almeriensis	5.3	Sánchez et al., 2008b
Desmodesmus sp. F51	5.5	Xie et al., 2013
Coccomyxa onubensis	6.2	Vaquero et al., 2012
Muriellopsis sp.	4.3	Del Campo et al., 2007
Chlorella zofingiensis	3.4	Del Campo et al., 2007
Chlorella minutissima	6.37	Dineshkumar et al., 2015a
Chlorella minutissima MCC-27	6.05	Dineshkumar et al., 2015b
Scenedesmus almeriensis	8.54	Molino et al., 2019
Chlorella sorokiniana MB-1	3.5 (60% recycled medium)	Chen and Liu, 2018
	2.5 (80% recycled medium)	
Scenedesmus almeriensis	5.71 (fresh medium)	This study
	2.93 (recycled medium)	

importance of the photo-bioreactor configuration choice for the selected microalgae strain.

A comparison between the lutein content extracted in this study and the literature results is reported in the following table (**Table 6**):

CONCLUSION

The effects of the fresh and the recycled growth medium, at different CO$_2$ concentrations on the productivity and the concentration of *S. almeriensis*, and on the lutein production were investigated. The recycled medium was found to limit the growth of *S. almeriensis*, as a decreased amount of available nutrients was observed, which corresponded to a slight decrease in the biomass productivity and the lutein content. On the other hand, *S. almeriensis* resulted to be potentially effective for CO$_2$ bio-fixation with promising performance rates; in particular, the higher the CO$_2$ content the higher the biomass productivity, as well as the microalgae concentration and the lutein content. *S. almeriensis* could be an ideal candidate for the commercial production of lutein from microalgae and for bio-fixation of CO$_2$ at the same time.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

DM and AM: conceptualization. SM, AI, and DM: data curation. AI: formal analysis. SM: investigation. AI, PC, DK, and SC: methodology. SM and TM: writing – original draft. SM: writing – review draft. AM: project administration and resources. DM and AM: supervision.

REFERENCES

Almomani, F. A., and Örmeci, B. (2018). Monitoring and measurement of microalgae using the first derivative of absorbance and comparison with chlorophyll extraction method. *Environ. Monit. Assess.* 190:90. doi: 10.1007/s10661-018-6468-y

Barceló-Villalobos, M., Serrano, C. G., Zurano, A. S., García, L. A., Maldonado, S. E., Peña, J., et al. (2019). Variations of culture parameters in a pilot-scale thin-layer reactor and their influence on the performance of *Scenedesmus almeriensis* culture. *Bioresour. Technol. Rep.* 6, 190–197. doi: 10.1016/j.biteb.2019.03.007

Bowen, P. E., Herbst-Espinosa, S. M., Hussain, E. A., and Stacewicz-Sapuntzakis, M. (2018). Esterification does not impair lutein bioavailability in humans. *J. Nutr.* 132, 3668–3673. doi: 10.1093/jn/132.12.3668

Cerón, M. C., Campos, I., Sánchez, J. F., Acién, F. G., Molina, E., and Fernández-Sevilla, J. M. (2008). Recovery of lutein from microalgae biomass: development of a process for *Scenedesmus almeriensis* biomass. *J. Agric. Food Chem.* 56, 11761–11766. doi: 10.1021/jf8025875

Chen, C. Y., Ho, S. H., Liu, C. C., and Chang, J. S. (2017). Enhancing lutein production with *Chlorella sorokiniana* Mb-1 by optimizing acetate and nitrate concentrations under mixotrophic growth. *J. Taiwan Inst. Chem. Eng.* 79, 88–96. doi: 10.1016/j.jtice.2017.04.020

Chen, C. Y., and Liu, C. C. (2018). Optimization of lutein production with a two-stage mixotrophic cultivation system with *Chlorella sorokiniana* MB-1. *Bioresour. Technol.* 262, 74–79. doi: 10.1016/j.biortech.2018.04.024

Cheng, J., Li, K., Yang, Z., Zhou, J., and Cen, K. (2016). Enhancing the growth rate and astaxanthin yield of *Haematococcus pluvialis* by nuclear irradiation and high concentration of carbon dioxide stress. *Bioresour. Technol.* 204, 49–54. doi: 10.1016/j.biortech.2015.12.076

Chirivella-Martorell, J., Briz-Redón, Á, and Serrano-Aroca, Á (2018). Modelling of biomass concentration, multi-wavelength absorption and discrimination method for seven important marine microalgae species. *Energies* 11:1089. doi: 10.3390/en11051089

Chiu, S. Y., Kao, C. Y., Chen, C. H., Kuan, T. C., Ong, S. C., and Lin, C. S. (2008). Reduction of CO2 by a high-density culture of *Chlorella* sp. in a semicontinuous photobioreactor. *Bioresour. Technol.* 99, 3389–3396. doi: 10.1016/j.biortech.2007.08.013

Chiu, S.-Y., Kao, C.-Y., Tsai, M.-T., Ong, S.-C., Chen, C.-H., and Lin, C.-S. (2009). Lipid accumulation and CO2 utilization of *Nannochloropsis oculata* in response to CO2 aeration. *Bioresour. Technol.* 100, 833–838. doi: 10.1016/j.biortech.2008.06.061

Choi, Y. Y., Joun, J. M., Lee, J., Hong, M. E., Pham, H. M., Chang, W. S., et al. (2017). Development of large-scale and economic pH control system for outdoor cultivation of microalgae *Haematococcus pluvialis* using industrial flue gas. *Bioresour. Technol.* 244, 1235–1244. doi: 10.1016/j.biortech.2017.05.147

Christian, D., Zhang, J., Sawdon, A. J., and Peng, C. A. (2018). Enhanced astaxanthin accumulation in *Haematococcus pluvialis* using high carbon dioxide concentration and light illumination. *Bioresour. Technol.* 256, 548–551. doi: 10.1016/j.biortech.2018.02.074

Coleman, H., and Chew, E. (2007). Nutritional supplementation in age-related macular degeneration. *Curr. Opin. Ophthalmol.* 18, 220–223. doi: 10.1097/ICU.0b013e32814a586b

Cordero, B. F., Obraztsova, I., Couso, I., Leon, R., Vargas, M. A., and Rodriguez, H. (2011). Enhancement of lutein production in *Chlorella sorokiniana* (chorophyta) by improvement of culture conditions and random mutagenesis. *Mar. Drugs* 9, 1607–1624. doi: 10.3390/md9091607

Cui, H., Yang, Z., Lu, Z., Wang, Q., Liu, J., and Song, L. (2019). Combination of utilization of CO2 from flue gas of biomass power plant and medium recycling to enhance cost-effective Spirulina production. *J. Appl. Phycol.* 31, 2175–2185. doi: 10.1007/s10811-019-1736-y

da Rosa, G. M., Moraes, L., Cardias, B. B., Souza, M., de Souze, R. A. Z., and Costa, J. A. V. (2015). Chemical absorption and CO2 biofixation via the cultivation of *Spirulina* in semicontinuous mode with nutrient recycle. *Bioresour. Technol.* 192, 321–327. doi: 10.1016/j.biortech.2015.05.020

de Assis, T. C., Calijuri, M. L., Assemany, P. P., de PaulaPereira, A. S. A., and Martins, M. A. (2019). Using atmospheric emissions as CO2 source in the cultivation of microalgae: productivity and economic viability. *J. Clean. Prod.* 215, 1160–1169. doi: 10.1016/j.jclepro.2019.01.093

Del Campo, J. A., García-González, M., and Guerrero, M. G. (2007). Outdoor cultivation of microalgae for carotenoid production: current state and perspectives. *Appl. Microbiol. Biotechnol.* 74, 1163–1174. doi: 10.1007/s00253-007-0844-9

Del Campo, J. A., Rodríguez, H., Moreno, J., Vargas, M. A., Rivas, J., and Guerrero, M. G. (2001). Lutein production by *Muriellopsis* sp. in an outdoor tubular photobioreactor. *J. Biotechnol.* 85, 289–295. doi: 10.1016/S0168-1656(00)00380-1

Del Campo, J. A., Rodríguez, H., Moreno, J., Vargas, M. Á, Rivas, J., and Guerrero, M. G. (2004). Accumulation of astaxanthin and lutein in *Chlorella zofingiensis* (Chlorophyta). *Appl. Microbiol. Biotechnol.* 64, 848–854. doi: 10.1007/s00253-003-1510-5

Dineshkumar, R., Dash, S. K., and Sen, R. (2015a). Process integration for microalgal lutein and biodiesel production with concomitant flue gas CO2 sequestration: a biorefinery model for healthcare, energy and environment. *RSC Adv.* 5, 73381–73394. doi: 10.1039/c5ra09306f

Dineshkumar, R., Dhanarajan, G., Dash, S. K., and Sen, R. (2015b). An advanced hybrid medium optimization strategy for the enhanced productivity of lutein in *Chlorella minutissima*. *Algal Res.* 7, 24–32. doi: 10.1016/j.algal.2014.11.010

Fernández-Sevilla, J. M., Acién Fernández, F. G., and Molina Grima, E. (2010). Biotechnological production of lutein and its applications. *Appl. Microbiol. Biotechnol.* 86, 27–40. doi: 10.1007/s00253-009-2420-y

Fret, J., Roef, L., Blust, R., Diels, L., Tavernier, S., Vyverman, W., et al. (2017). Reuse of rejuvenated media during laboratory and pilot scale cultivation of *Nannochloropsis* sp. *Algal Res.* 27, 265–273. doi: 10.1016/j.algal.2017.09.018

Hadj-Romdhane, F., Jaouen, P., Pruvost, J., Grizeau, D., Van Vooren, G., and Bourseau, P. (2012). Development and validation of a minimal growth medium for recycling *Chlorella vulgaris* culture. *Bioresour. Technol.* 123, 366–374. doi: 10.1016/j.biortech.2012.07.085

Hadj-Romdhane, F., Zheng, X., Jaouen, P., Pruvost, J., Grizeau, D., Croué, J. P., et al. (2013). The culture of *Chlorella vulgaris* in a recycled supernatant: effects

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00415/full#supplementary-material

on biomass production and medium quality. *Bioresour. Technol.* 132, 285–292. doi: 10.1016/j.biortech.2013.01.025

Ho, S. H., Chan, M. C., Liu, C. C., Chen, C. Y., Lee, W. L., Lee, D. J., et al. (2014). Enhancing lutein productivity of an indigenous microalga *Scenedesmus obliquus* FSP-3 using light-related strategies. *Bioresour. Technol.* 152, 275–282. doi: 10.1016/j.biortech.2013.11.031

Lam, M. K., Lee, K. T., and Mohamed, A. R. (2012). Current status and challenges on microalgae-based carbon capture. *Int. J. Greenh. Gas Control* 10, 456–469. doi: 10.1016/j.ijggc.2012.07.010

Lin, J. H., Lee, D. J., and Chang, J. S. (2015a). Lutein in specific marigold flowers and microalgae. *J. Taiwan Inst. Chem. Eng.* 49, 90–94. doi: 10.1016/j.jtice.2014.11.031

Lin, J. H., Lee, D. J., and Chang, J. S. (2015b). Lutein production from biomass: marigold flowers versus microalgae. *Bioresour. Technol.* 184, 421–428. doi: 10.1016/j.biortech.2014.09.099

Liu, W., Wang, J., and Liu, T. (2019). Low pH rather than high CO$_2$ concentration itself inhibits growth of Arthrospira. *Sci. Total Environ.* 666, 572–580. doi: 10.1016/j.scitotenv.2019.02.312

Livansky, K., Dedic, K., Binova, J., Tichy, V., Novotny, P., and Doucha, J. (1996). Influence of the nutrient solution recycling on the productivity of *Scenedesmus obliquus*, utilization of nutrients and water in outdoor cultures. *Arch. Hydrobiol.* 114, 105–113. doi: 10.1127/algol_stud/81/1996/105

Lu, L., Yang, G., Zhu, B., and Pan, K. (2017). A comparative study on three quantitating methods of microalgal biomass. *Indian J. Geo Mar. Sci.* 46, 2265–2272.

Mann, J. E., and Myers, J. (1968). On pigments, growth, and photosynthesis of phaeodactylum tricornutum 2. *J. Phycol.* 4, 349–355. doi: 10.1111/j.1529-8817.1968.tb04707.x

Mehariya, S., Iovine, A., Di Sanzo, G., Larocca, V., Martino, M., Leone, G., et al. (2019). Supercritical fluid extraction of lutein from *Scenedesmus almeriensis*. *Molecules* 24:1324. doi: 10.3390/molecules24071324

Moheimani, N. R. (2016). Tetraselmis suecica culture for CO$_2$ bioremediation of untreated flue gas from a coal-fired power station. *J. Appl. Phycol.* 28, 2139–2146. doi: 10.1007/s10811-015-0782-3

Molino, A., Iovine, A., Casella, P., Mehariya, S., Chianese, S., Cerbone, A., et al. (2018a). Microalgae characterization for consolidated and new application in human food, animal feed and nutraceuticals. *Int. J. Environ. Res. Public Health* 15:2436. doi: 10.3390/ijerph15112436

Molino, A., Mehariya, S., Iovine, A., Larocca, V., Di Sanzo, G., Martino, M., et al. (2018b). Extraction of astaxanthin and lutein from microalga *Haematococcus pluvialis* in the red phase using CO$_2$ supercritical fluid extraction technology with ethanol as co-solvent. *Mar. Drugs* 16:432. doi: 10.3390/md16110432

Molino, A., Rimauro, J., Casella, P., Cerbone, A., Larocca, V., Chianese, S., et al. (2018c). Extraction of astaxanthin from microalga *Haematococcus pluvialis* in red phase by using generally recognized as safe solvents and accelerated extraction. *J. Biotechnol.* 283, 51–61. doi: 10.1016/j.jbiotec.2018.07.010

Molino, A., Mehariya, S., Karatza, D., Chianese, S., Iovine, A., Casella, P., et al. (2019). Bench-scale cultivation of microalgae *Scenedesmus almeriensis* for CO$_2$ capture and lutein production. *Energies* 12:2806. doi: 10.3390/en12142806

Patel, A. K., Joun, J. M., Hong, M. E., and Sim, S. J. (2019). Effect of light conditions on mixotrophic cultivation of green microalgae. *Bioresour. Technol.* 282, 245–253. doi: 10.1016/j.biortech.2019.03.024

Rodolfi, L., Zittelli, G. C., Barsanti, L., Rosati, G., and Tredici, M. R. (2003). Growth medium recycling in *Nannochloropsis* sp. mass cultivation. *Biomol. Eng.* 20, 243–248. doi: 10.1016/s1389-0344(03)00063-7

Ruban, A. V., Berera, R., Ilioaia, C., Van Stokkum, I. H. M., Kennis, J. T. M., Pascal, A. A., et al. (2007). Identification of a mechanism of photoprotective energy dissipation in higher plants. *Nature* 450, 575–578. doi: 10.1038/nature06262

Salih, F. M. (2011). Microalgae tolerance to high concentrations of carbon dioxide: a review. *J. Environ. Prot.* 02, 648–654. doi: 10.4236/jep.2011.25074

Sánchez, J. F., Fernández-Sevilla, J. M., Acién, F. G., Cerón, M. C., Pérez-Parra, J., and Molina-Grima, E. (2008a). Biomass and lutein productivity of *Scenedesmus almeriensis*: influence of irradiance, dilution rate and temperature. *Appl. Microbiol. Biotechnol.* 79, 719–729. doi: 10.1007/s00253-008-1494-2

Sánchez, J. F., Fernández, J. M., Acién, F. G., Rueda, A., Pérez-Parra, J., and Molina, E. (2008b). Influence of culture conditions on the productivity and lutein content of the new strain Scenedesmus almeriensis. *Process Biochem.* 43, 398–405. doi: 10.1016/j.procbio.2008.01.004

Santos, L. O., De, M. G., Grima, E. M., Alberto, J., Costa, V., and Aci, F. G. (2018). Engineering strategies for the enhancement of *Nannochloropsis gaditana* outdoor production: influence of the CO2 flow rate on the culture performance in tubular photobioreactors. *Process Biochem.* 76, 171–177. doi: 10.1016/j.procbio.2018.10.010

Sanzo, G., Mehariya, S., Martino, M., Larocca, V., Casella, P., Chianese, S., et al. (2018). Supercritical carbon dioxide extraction of astaxanthin, lutein, and fatty acids from *Haematococcus pluvialis* microalgae. *Mar. Drugs* 16:334. doi: 10.3390/md16090334

Utomo, R. P., Chang, Y. R., Lee, D. J., and Chang, J. S. (2013). Lutein recovery from *Chlorella* sp. *ESP*-6 with coagulants. *Bioresour. Technol.* 139, 176–180. doi: 10.1016/j.biortech.2013.04.025

Vaquero, I., Ruiz-Domínguez, M. C., Márquez, M., and Vílchez, C. (2012). Cu-mediated biomass productivity enhancement and lutein enrichment of the novel microalga *Coccomyxa onubensis*. *Process Biochem.* 47, 694–700. doi: 10.1016/j.procbio.2012.01.016

Vechpanich, J., and Shotipruk, A. (2011). Recovery of free lutein from tagetes erecta: determination of suitable saponification and crystallization conditions. *Sep. Sci. Technol.* 46, 265–271. doi: 10.1080/01496395.2010.506904

Xie, Y., Ho, S. H., Chen, C. N. N., Chen, C. Y., Ng, I. S., Jing, K. J., et al. (2013). Phototrophic cultivation of a thermo-tolerant *Desmodesmus* sp. for lutein production: effects of nitrate concentration, light intensity and fed-batch operation. *Bioresour. Technol.* 144, 435–444. doi: 10.1016/j.biortech.2013.06.064

Xie, Y., Zhao, X., Chen, J., Yang, X., Ho, S. H., Wang, B., et al. (2017). Enhancing cell growth and lutein productivity of *Desmodesmus* sp. *F*51 by optimal utilization of inorganic carbon sources and ammonium salt. *Bioresour. Technol.* 244, 664–671. doi: 10.1016/j.biortech.2017.08.022

Xie, Y. P., Ho, S. H., Chen, C. Y., Chen, C. N. N., Liu, C. C., Ng, I. S., et al. (2014). Simultaneous enhancement of CO$_2$ fixation and lutein production with thermo-tolerant *Desmodesmus* sp. *F*51 using a repeated fed-batch cultivation strategy. *Biochem. Eng. J.* 86, 33–40. doi: 10.1016/j.bej.2014.02.015

Yen, H. W., Hu, I. C., Chen, C. Y., Ho, S. H., Lee, D. J., and Chang, J. S. (2013). Microalgae-based biorefinery - From biofuels to natural products. *Bioresour. Technol.* 135, 166–174. doi: 10.1016/j.biortech.2012.10.099

Genetic Control of Reproductive Traits in Tomatoes Under High Temperature

Maria José Gonzalo[1], Yi-Cheng Li[2], Kai-Yi Chen[2], David Gil[3], Teresa Montoro[3], Inmaculada Nájera[4], Carlos Baixauli[4], Antonio Granell[1] and Antonio José Monforte[1]*

[1] Instituto de Biología Molecular y Celular de Plantas, Universitat Politècnica de València-Consejo Superior de Investigaciones Científicas, Valencia, Spain, [2] Department of Agronomy, National Taiwan University, Taipei, Taiwan, [3] Enza Zaden Centro de Investigación S.L., Almería, Spain, [4] Centro de Experiencias de Cajamar en Paiporta, Paiporta, Spain

Global climate change is increasing the range of temperatures that crop plants must face during their life cycle, giving negative effects to yields. In this changing scenario, understanding the genetic control of plant responses to a range of increasing temperature conditions is a prerequisite to developing cultivars with increased resilience. The current work reports the identification of Quantitative Trait Loci (QTL) involved in reproductive traits affected by temperature, such as the flower number (FLN) and fruit number (FRN) per truss and percentage of fruit set (FRS), stigma exsertion (SE), pollen viability (PV) and the incidence of the physiological disorder tipburn (TB). These traits were investigated in 168 Recombinant Inbred Lines (RIL) and 52 Introgression Lines (IL) derived from the cross between *Solanum lycopersicum* var. "MoneyMaker" and *S. pimpinellifolium* accession TO-937. Mapping populations were cultivated under increased temperature regimen conditions: T1 (25°C day/21°C night), T2 (30°C day/25°C night) and T3 (35°C day/30°C night). The increase in temperature drastically affected several reproductive traits, for example, FRS in Moneymaker was reduced between 75 and 87% at T2 and T3 when compared to T1, while several RILs showed a reduction of less than 50%. QTL analysis allowed the identification of genomic regions affecting these traits at different temperatures regimes. A total of 22 QTLs involved in reproductive traits at different temperatures were identified by multi-environmental QTL analysis and eight involved in pollen viability traits. Most QTLs were temperature specific, except QTLs on chromosomes 1, 2, 4, 6, and 12. Moreover, a QTL located in chromosome 7 was identified for low incidence of TP in the RIL population, which was confirmed in ILs with introgressions on chromosome 7. Furthermore, ILs with introgressions in chromosomes 1 and 12 had good FRN and FRS in T3 in replicated trials. These results represent a catalog of QTLs and pre-breeding materials that could be used as the starting point for deciphering the genetic control of the genetic response of reproductive traits at different temperatures and paving the road for developing new cultivars adapted to climate change.

Keywords: pollen viability, fruit set, QTL, introgression line, tipburn, abiotic stress

INTRODUCTION

Under the current scenario of global warming, temperature projections estimate a 2–5°C increase in temperature by the end of the twenty-first century (IPCC, 2014). Agriculture production will be greatly affected by this temperature rise, as high temperatures have a negative impact on crops, causing an array of morpho-anatomical, physiological and biochemical changes, which negatively affect plant growth and development and may lead to a drastic reduction in yield (Wahid et al., 2007; Bita and Gerads, 2013). The major losses due to heat stress are expected to occur at low latitude regions (temperate and tropical areas). In fact, yield reduction due to heat stress has been documented in many crops such as wheat, rice, barley, sorghum, maize, chickpea, canola, and more (Hasanuzzaman et al., 2013; Challinor et al., 2014; Alsamir et al., 2019). In the case of tomatoes, a 28% yield reduction due to high temperatures has already been reported (Alsamir et al., 2019).

Tomatoes are one of the most important horticultural crops worldwide and they are currently cultivated in a wide range of agroclimatic regions, either in open fields or under greenhouse conditions. Vegetative growth in tomatoes is well adapted to high temperatures. However, high temperature affects growth and development of different tomato plant organs or structures. For instance, a decrease in the flower number with increased temperatures has been observed (Charles and Harris, 1972) although the effect on fruit set has more dramatic consequences for the yield. Optimal temperatures for setting fruit in field conditions are between 21 and 24°C (Geisenberg and Stewart, 1986) temperatures exceeding 32°C during the day, and/or not decreasing below 21°C during the night (Moore and Thomas, 1952) has dramatic consequences for fruit set and total yield (El Ahmadi and Stevens, 1997). The decrease in tomato fruit set under long-term mildly elevated temperatures has been shown to correlate with a decrease in pollen viability (Dane et al., 1991; Peet et al., 1998; Pressman et al., 2002, 2006; Xu et al., 2017b). An inserted stigma is also an important trait to ensure self-pollination in cultivated tomatoes (Rick and Dempsey, 1969; Chen and Tanksley, 2004). Growth at high temperatures may lead to the protrusion of the style above the anther cone (exsertion, Levy et al., 1978; Sato et al., 2006) negatively affecting flower pollination (Charles and Harris, 1972; Dane et al., 1991; Xu et al., 2017b). High temperatures also increase the incidence of tipburn, necrosis of the apical vegetative and reproductive tissues that have been related to insufficient water absorption and nutritional unbalance (Starck et al., 1994; Chung et al., 2010).

Screening for heat tolerance among tomato cultivars has been carried out in a limited number of cultivars, uncovering just a few "thermotolerant" genotypes (Dane et al., 1991; Opena et al., 1992; Abdul-Baki and Stommel, 1995; Grilli et al., 2007; Kugblenu et al., 2013; Xu et al., 2017b; Ruggieri et al., 2019). Most of the screenings were based on the capacity of plants to set fruit at high temperatures. Given the complexity of this trait, which is also affected by many other factors (Charles and Harris, 1972; Abdul-Baki, 1991; Dane et al., 1991; Adams et al., 2001; Alsamir et al., 2019) some authors have based their screenings on pollen viability as well (Dane et al., 1991;

Paupière et al., 2017; Driedekons et al., 2018) as this trait appears to be the most sensitive to high temperatures and under the hypothesis that the dissection of the processes from flower to fruit set is an effective strategy to determine the genetic basis of thermotolerance in tomatoes.

Accessions from wild species of *Solanum* spp. such as *S. pimpinellifolium*, L., *S. pennellii* Correll, *S. habrochaites* S. Knapp and D. M Spooner, *S. chmielewskii* C. M. Rick et al. and *S. cheesmaniae* (L. Riley) Fosberg have been found to be tolerant to high temperatures (Alam et al., 2010; Nahar and Ullah, 2011, 2012; Golam et al., 2012; Paupière et al., 2017; Driedekons et al., 2018). However, our understanding of the genetic control of heat tolerance in tomatoes is still very limited. Up to the present, very few reports have addressed this issue, resulting in very few QTLs involved in heat tolerance identified in tomatoes (Grilli et al., 2007; Lin et al., 2010; Xu et al., 2017b; Ruggieri et al., 2019; Wen et al., 2019). In general, these studies have been performed using simple mapping populations (F_2), with a limited sample size or with low dense marker coverage. As a consequence, the number of reported QTLs involved in heat tolerance is relatively low and the stability of their effects still needs to be verified.

There are many reasons that may explain the lack of information on the genetic control of heat tolerance in tomatoes. The choice of the proper trait to evaluate this tolerance is a crucial matter. Reproductive traits have been used extensively, including flowering traits, pollen viability and/or fruit set (for example Xu et al., 2017a). On the other hand, the heat injury index and physiological traits have been investigated less frequently (Wen et al., 2019). The relationship between both types of traits has not been studied, although it would be critical for understanding the global response to abiotic stress. Only a tiny fraction of the tomato germplasm has been screened for heat tolerance, more screening efforts would help to identify tolerance sources and to develop mapping populations suitable to investigate the different factors involved in this tolerance. Lastly, the lack of powerful mapping populations such as recombinant inbred lines (RILs), introgression lines (ILs) or multiparent advanced generation intercross (MAGIC) designed for the study of heat tolerance hampers the identification of QTLs.

In this manuscript, we take advantage of the availability of two populations derived from the cross between *S. lycopersicum* cv. 'Moneymaker' and *S. pimpinellifolium* accession TO-937: a set of 168 RILs (Alba et al., 2009) and a set of 52 ILs (Barrantes et al., 2014). Even though TO-937 has not been cataloged as a heat tolerant genotype, transgressive segregants are often obtained when crossing cultivated and exotic or wild genotypes (de Vicente and Tanksley, 1993; Diaz et al., 2014). In fact, transgressive QTLs have been identified in the TO-937-x-MoneyMaker RIL and IL populations for other traits previously, for example QTLs involved in disease resistance, vegetative growth and fruit quality (Alba et al., 2009; Powell et al., 2012; Salinas et al., 2013; Capel et al., 2015, 2017; Barrantes et al., 2016). In the current work, a first evaluation of both populations for reproductive traits in different temperature regimens was conducted, finding a promising segregation of heat tolerance in both of them. Therefore, we decided to study them in further replicated experiments to obtain insights on the genetic control

of the heat tolerance segregation found in these populations. The joint analysis of those populations could also be a powerful approach for such complex traits. For example, Rambla et al. (2017) successfully studied the genetic control of fruit volatile composition with the joint analysis of both populations. On one hand, RIL genetic architecture is appropriate for the study of traits under complex genetic control and to map QTLs with a relatively good resolution, whereas IL genetic architecture allows the accurate estimation of the effects of single QTL in a desired genetic background, facilitating QTL cloning and integration of QTL in applied marker assisted selection breeding.

This report describes the identification of candidate QTLs involved in the capacity of set fruits under different temperature regimens that can be used to improve heat tolerance in tomatoes.

MATERIALS AND METHODS

Plant Material and Growing Conditions

Two mapping population derived from the cross between *S. lycopersicum* cv. "MoneyMaker" (MM) and the *S. pimpinellifolium* accession TO-937: 167 Recombinant Inbred Lines population (RILs) (Alba et al., 2009) and 56 Introgression Lines (ILs), each one carrying a different introgression from TO-937 into MM genetic background (Barrantes et al., 2014), were investigated in the current research.

Mapping populations were cultivated in greenhouses under controlled temperature conditions in two facilities: Centro de Experiencias Cajamar belonging to Fundación Cajamar Comunidad Valenciana (FCCV, Paiporta, Spain) and National University of Taiwan (NTW, Taipei, Taiwan).

The 167 RILs were assayed for two consecutive years, 2016 and 2017, at FCCV, with three plants per RIL. The three plants were planted in the same bag, thus for statistical analysis, the RIL value was the mean of the three plants. Also, 25 replicates of MM and three of TO-937 were assayed in 2016. In 2017, due to limitations of seed availability for TO-937, only eight replicates of MM were assayed. Plants were grown under a stepwise temperature increase (T1: 25°C day/20°C night; T2: 30°C day/25°C night; T3: 35°C day/30°C night) as follows: each temperature regimen was established for 4 weeks. For 2 weeks, the plants were allowed to flower without any restriction. In the third week, the number of flowers was recorded and in the fourth week the fruit set was recorded as the number of observed developing fruits. After recording the fruit set, all the flowers and fruits were pruned from the plant, in order to avoid the physiological effects of previous fruit load in the new inflorescences, and the temperature was increased to the next regimen.

In addition, the RIL population was also analyzed in the year 2018 in greenhouses under controlled conditions at the National University of Taiwan (NTW). RILs were grown under one of the following temperatures conditions: T2 (30°C day/25°C night), T3 (35°C day/30°C night). The plants were not subjected to stepwise temperature increase; rather they were cultivated in different greenhouses for each temperature treatment.

A preliminary analysis of the 56 ILs was carried out in 2016 at FCCV with the same experimental set up as the RILs, with one replicate of three plants per IL and 25 replicates for MM, evenly distributed among the ILs. A selection of the 12 more promising ILs (SP_1-3, SP_1-4, SP_2-2, SP_5-5, SP_6-3, SP_7-4, SP_11-4, SP_12-1, SP_12-2, SP_12-3, SP_12-4, SP_12-5) were re-evaluated in 2017 at FCCV using the same temperature regimens. In 2019, the experiment was replicated again - three ILs (SP_12-1, SP_12-2 and SP_12-4) with low FRS in 2017 were replaced with ILs with introgressions from chromosome 2 and 7 (SP_2-4, SP_2-5, SP_7-3) as QTLs were previously identified in those genomic regions using the RILs. For all these experiments, ILs were grown following a completely randomized design with five and three replicates (with three plants each) of each genotype in 2017 and 2019, respectively, and 6 replicates of MM with three plants per replicate. Additionally, ILs SP_7-1, SP_7-2, SP_7-3 and SP_7-4, selected to verify tolerance to tipburn (TB, see below), were included in the 2017 experiment at FCCV with a completely randomized design with four replicates of three plants.

Phenotyping

On the third week of each temperature treatment the number of flowers (FLN), the degree of stigma exsertion (SE, scored as 0: not exserted, 1: slightly exserted and 2: very exserted) were recorded in the second and third truss. The number of fruits set (FRN), and fruit set proportion (FRS = 100 × FRN/FLN) and the incidence of tipburn (TB) on apical tissues was recorded on the fourth week as presence/absence (**Table 1**).

Several traits related to pollen viability were measured for the 167 RILs in 2016 and 2017. For pollen tube germination (TG), recorded in 2016 and 2017, pollen was collected directly from fresh flowers and cultivated for 16 h at room temperature in an 18% sucrose medium. TG was measured as the number of germinated grains (scored as 0: no germination, 1: between 1 and 25% germination, 2: between 26 and 75% germination and 3: more than 75% of the pollen grains germinated) counted with a stereomicroscope with epi-illumination. Pollen viability was assessed in 2017 with two techniques: Aniline Blue staining (AB) and flux cytometry of three flowers from each plant. The aniline blue stain was used to identify and evaluate pollen grains by visualization with a Nikon Eclipse E600 microscope, the grain number was scored as 0: no pollen, 1: 25% of the pollen stained, 2: between 26 and 75% stained and 3: more than 75% of the pollen grains stained. The images from the microscope were analyzed with Image J software[1] to calculate the pollen number (AB) (number of grains in 200 μM). Cytometry analysis was performed in the Enza-Zaden España S.L. facilities with an AmphaTMZ32 flow cytometer (Amphasys AG, Switzerland) to measure the percentage of viable pollen (VP).

Statistical Analysis

The basic statistics (mean, standard deviation, maximum and minimum values), trait distribution and the Pearson correlations were calculated among traits, years and treatments. IL means were contrasted with the recurrent control MM mean with a Dunnett's test ($p < 0.05$) in 2017 and 2018 experiments.

---

[1]https://imagej.nih.gov/ij/

TABLE 1 | Description of the reproductive traits studied in the current report.

Trait name	Trait abbreviation	Trait description
Flower number	FLN	Mean of flower number in the second and third truss for each temperature
Fruit number	FRN	Mean of fruit set in the second and third truss for each temperature
Fruit Set Proportion	FRS	$100 \times$ (Ratio FRN/FLN)
Stigma Exsertion	SE	0: no exserted, 1: slightly exserted and 2: very exserted
Tipburn	TB	Apical necrosis. 0: no TB and 1: TB incidence
Pollen tube germination	TG	Number of germinated pollen grains
Aniline Blue	AB	Pollen viability based on aniline blue staining
Viable pollen	VP	Percentage of viable pollen based on flow cytometry

The analyses were implemented with JMP (2019) software (version 12.1.0).

QTL Analysis

The map used for the QTL analysis was previously generated and contained 4,932 Single Nucleotide Markers (SNP) from the 8K SNP SOLCAP Infinium chip (Sim et al., 2012). The map was condensed to 1,279 SNPs with QTL IciMapping (Meng et al., 2015) to facilitate the computational analysis. Multi-environmental QTL analysis was performed with IciMapping. LOD threshold for a significance level $p < 0.05$ was obtained by a permutation test with 1,000 resamplings. Additionally, composite interval mapping (CIM, Zeng, 1994) was performed for each independent experiment with Windows QTL Cartographer 2.5[2] (Wang et al., 2007). Multi-environment QTLs were named with an abbreviation of the trait, followed by the chromosome number, the number of the QTL within chromosome, the temperature regimen (T1, T2 or T3) and the suffix _2E (indicating two environments). QTLs for single experiments were named accordingly, adding a suffix with the experiment year (i.e., _16 and _17). As TP was scored as presence/absence, a χ^2 test was performed for each marker.

RESULTS

Phenotypic Variation for Reproductive Traits at Different Temperatures in the RIL Population

Table 2 depicts the mean for the reproductive traits among the parent genotypes TO-937 and MM in the 2016 and 2017 experiments (TO-937 could not be assayed in 2017). MM maintained flower production in the three temperature regimens, but fructification decreased drastically (between 75 and 85%) at T2 and T3 in both years as a consequence of the heat stress. TO-937 did not set fruit at T2 and T3, or flower at T3, showing even more sensitivity to heat stress.

The RIL population MM x-TO-937 was evaluated for reproductive traits at the different temperature regimens in a preliminary experiment in 2016. A transgressive segregation was observed at T2 and T3, with a number of RILs showing

a high fruit set at high temperatures, suggesting the presence of genetic variability for heat tolerance in the current RIL population (**Figure 1** and **Supplementary Table S1**). Therefore, RILs were evaluated in two additional experiments (2017, 2018). The distributions of the reproductive traits FLN, FRN and FRS among different experiments and temperature regimens are shown in **Figure 1** and **Supplementary Table S1**. At T1 (optimal temperature), trait distributions generally fitted into a normal distribution. The range of the distributions were wide, transgressive segregants and were observed for all traits in both directions (very low and very high values). However, at T2 the shape of the distributions changed significantly (**Figure 1**), skewing toward lower values. FRN and FRS dropped more drastically than FLN, although transgressive segregants toward higher values were obtained. At T3, the skew toward low values was even more dramatic with an important proportion of RIL displaying very low values, although a transgressive segregation to higher values was also observed. In the case of FRN and FRS, the median corresponded to values equal or very close to 0. The effect of the thmperature increase was even more dramatic in the 2018 experiment performed by NTW. The different facilities used in NTW (a medium sized greenhouse compared with a large greenhouse in the FCCV facilities), external environmental effects (higher humidity in NTW) and the different method used to impose the temperature regimen (stepwise in FCCV vs. direct in NTW) may explain the differences in trait distribution. Nevertheless, these results showed that the reduced ability to produce fruit at high temperature was mostly due to the reduction in the plant's ability to set fruits rather than due to the reduction in the number of flowers. The broad range of the distributions and the observation of transgressive segregants in different experiments for heat tolerance demonstrated the presence of genetic variability for heat tolerance in this population, highlighting that several RILs were capable of setting fruit even at extremely high temperatures.

The consistency and robustness of the results was tested by correlation analysis for each trait-temperature combination over the different experiments. Correlations among years for the FLN were positive and highly significant at every temperature regimen (**Table 3**), although correlations were more robust among the 2016 and 2017 experiments than with the 2018 experiment. FRN was also highly correlated between the 2016 and 2017 experiments at the three temperature treatments. Correlations of FRN and FRS between both 2016 and 2017

---

[2]http://statgen.ncsu.edu/qtlcart/WQTLCart.htm

TABLE 2 | Means and standard deviation for reproductive traits flower number (FLN), fruit number (FRN) and fruit set proportion (FRS) among the parent genotypes TO-937 and MoneyMaker in the three temperature regimens (T1: 25°C day/20°C night; T2: 30°C day/25°C night; T3: 35°C day/30°C night) from the 2016 and 2017 recombinant inbred line experiments.

	2016		2017	
	TO-937	**MoneyMaker**	**TO-937**	**MoneyMaker**
FLN_T1	12.17 ± 0.57	8.05 ± 1.12	nd	8.81 ± 1.71
FLN_T2	6.83 ± 1.25	8.22 ± 2.08	nd	5.88 ± 1.55
FLN_T3	0	7.10 ± 16.03	nd	4.50 ± 1.45
FRN_T1	10.33 ± 1.04	5.70 ± 2.19	nd	8.71 ± 1.76
FRN_T2	0	0.73 ± 2.04	nd	1.83 ± 1.21
FRN_T3	0	1.00 ± 9.46	nd	0.81 ± 0.68
FRS_T1	84.93 ± 12.86	70.81 ± 3.54	nd	98.73 ± 2.4
FRS_T2	0	8.83 ± 1.99	nd	32.36 ± 18.82
FRS_T3	0	14.08 ± 20.50	nd	18.15 ± 13.39

nd: no data.

with 2018 were not significant indicating that non-controlled factors affected that experiment as discussed above. Regarding the correlations between traits within temperature regimens, high correlations were observed between FLN and FRN in 2016 and 2017 experiments at all temperatures. Correlations between FRN and FRS were also generally significant and high in the all experiments and temperatures. On the other hand, FLN and FRS showed variable correlations, but these were mainly non-significant or negative.

Identification of QTL Involved in Reproductive Traits in the RIL Population

Due to the low correlation of the 2018 experiment results with 2016 and 2017 data (**Table 3**), a multi-environment QTL analysis was performed only with 2016 and 2017 data. A total of 22QTL were detected across traits and temperature regimens (**Table 4**). QTL analysis performed independently in each experiment showed a total of 46 QTLs, 10 of them from the 2018 experiment (**Supplementary Table S2**) and 34 of them were detected in two experiments or at least in two temperature conditions. QTLs detected in just one experiment or temperature regimen were not considered as reliable QTLs, so they were not taken into account for further discussion. Most of the QTLs detected by multi-environment analysis were also detected by single environment analysis (**Table 4** and **Supplementary Table S2**).

A total of five QTLs were detected for FLN by multi-environmental analysis across temperatures (**Table 4**). Among them, the QTLs with more stable effects among temperatures and years was localized to chromosome 2 in the 105 cM positon (*fln2.1_T1-2E, fln2.1_T2_2E*). The phenotypic variance explained by this QTL ranged between 8 and 10% with low QTL-x-environment (QTL-x-E) interaction and with the TO-937 allele increasing FLN. Other QTLs detected in two temperature regimens mapped onto chromosomes 4 and 6 (**Table 4**). Additionally, a QTL was detected in the 2018 experiment on chromosome 11 at T3 (*fln11.1_T3_18*, **Supplementary Table S2**), which could correspond to *fln11.2_T2* (**Table 4**), although they were detected at different temperature regimens. Interestingly, a number of QTLs on chromosomes 2, 4, 6, and 11 showed effects

on different temperatures, with MM or TO-937 alleles, depending on the QTL, increasing FLN.

Eight multi-environment QTLs were identified for FRN (**Table 4**). QTLs on chromosome 4, 6, and 7 were detected at T2, with low QTL-x-E interaction and the TO-937 allele increasing FRN for *frn7.1_T2_2E*, whereas the MM allele increased the trait for the other two (*frn6.2_T1_2E* and *frn6.2_T2_2E*). These QTLs were also detected by single environment analysis (**Supplementary Table S2**). QTLs on chromosome 2 were detected in the three temperature conditions (*frn2.1_T1_16, frn2.1_T1_17, frn2.1_T2_16, frn2.1_T2_17, frn2.2_T3_17*, **Supplementary Table S2**) and located in the same region as *frn2.2_T1_2E* (100–130 cM, **Table 4**). QTLs on chromosome 12 (*frn12.1T2_18* and *frn12.1_T3_18*) were also detected in the 2018 experiment (**Supplementary Table S2**).

Two multi-environment QTLs for FRS on chromosomes 6 and 7 were mapped in T2, overlapping with FRN QTLs in the same chromosome region (**Table 4**). As for FRN, the MM allele of *frs6.1_T2_2E* increased FRS, whereas the TO-937 allele increased the trait for *frs7.1_T2_2E*. *frs6.1_T2_2E* displayed low QTL-x-E interaction, whereas it was more important for *frs7.1_T2_2E*. Additionally, QTLs on chromosome 12 *frs12.1_T2_18* and *frs12.1_T3_18* were detected in the 2018 experiment in the two high temperature regimens, with TO-937 alleles increasing FRS (**Supplementary Table S2**). Overall, the percentage of variance explained by the QTLs detected by multi-environment analysis or in two temperatures/experiments by single environmental analysis were relatively modest (5–12%, **Table 4** and **Supplementary Table S2**).

One multi-environment QTL for SE was detected on chromosome 2 at T1 and T2 (**Table 4**). These QTLs explained between 9.5 and 15% of the phenotypic variance and showed low QTL-x-E interaction. QTLs in the same genome region were also detected in the 2018 experiment, with the TO-937 alleles increasing SE (**Supplementary Table S2**).

Analysis of Reproductive Traits in Introgression Lines

The complete set of 56 ILs was analyzed for the 2016 experiment. At T1 and T2, the FLN means were similar in MM and IL
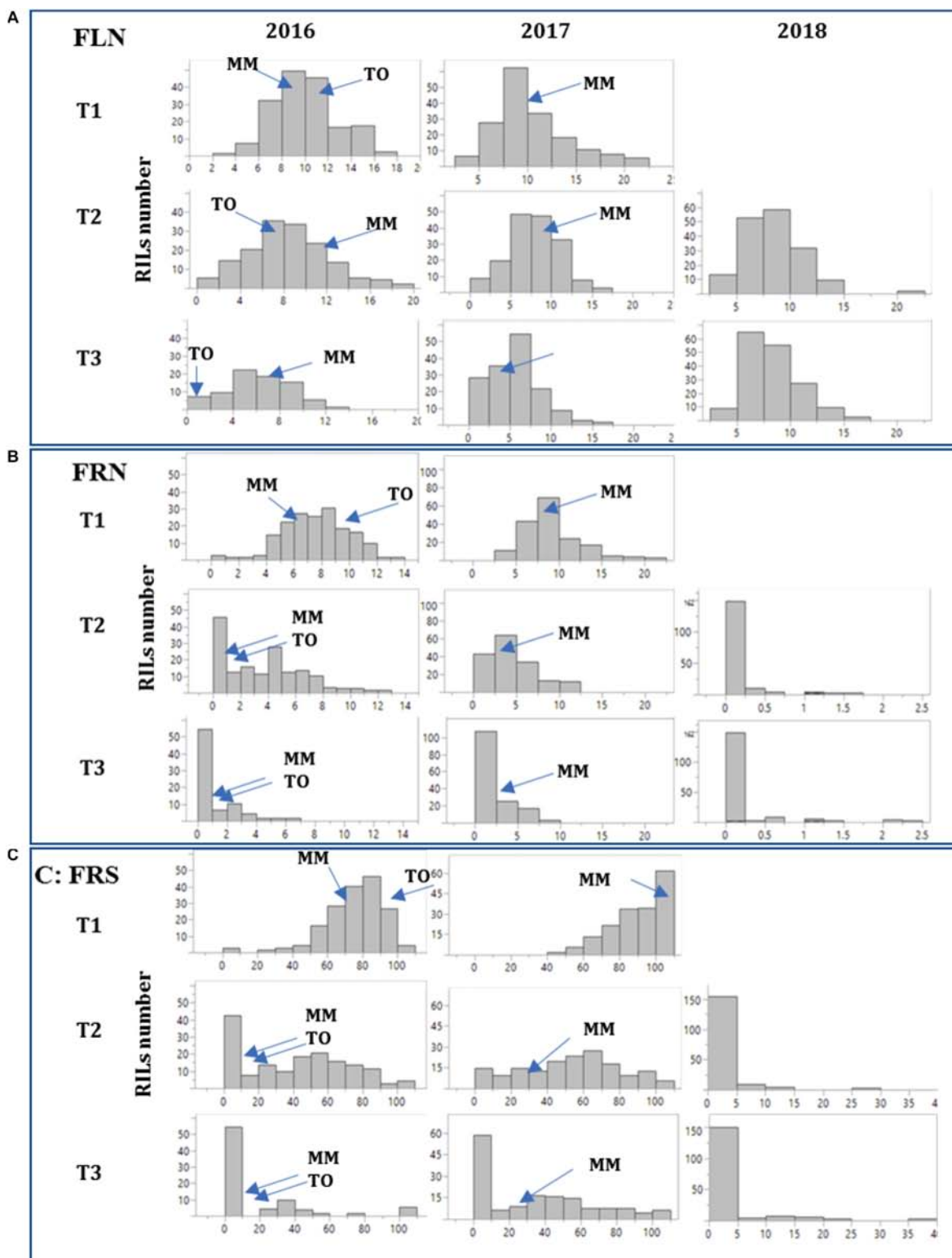
FIGURE 1 | Histograms depicting the distribution of reproductive traits in the RIL population. Reproductive traits flower number (FLN, **A**), number of fruits per inflorescence (FRN, **B**) and percentage of fruit set (FRS, **C**) were studied in three experiments (2016, 2017, and 2018) at three temperature regimens (T1: 25°C day/20°C night; T2: 30°C day/25°C night; T3: 35°C day/30°C night) except in 2018, where the temperature T1 was not assayed. The frequency histograms for each trait in the "MoneyMaker"-x-TO-937 Recombinant Inbred Lines (RIL) across the years are depicted. Means of the parents "Moneymaker" (MM) and TO-937 (TO) are indicated with arrows for the 2016 experiment.

TABLE 3 | Linear correlations among the three temperature treatments (T1: 25°C day/20°C night; T2: 30°C day/25°C night; T3: 35°C day/30°C night) and experiments (2016, 2017, and 2018) for flower number (FLN), fruit number (FRN) and fruit set proportion (FRS) in the MM-xTO-937 recombinant inbred population. Correlations among years are highlighted in yellow.

T1 (25°C/20°C; day/night)

	FLN16_T1	FLN17_T1	FRN16_T1	FRN17_T1	FRS16_T1			
FLN17_T1	0.72**	–	–	–	–	–	–	–
FRN16_T1	0.74**	0.48**	–	–	–	–	–	–
FRN17_T1	0.64**	0.88**	0.47**	–	–	–	–	–
FRS16_T1	−0.17*"	−0.18*	0.49**	−0.11	–	–	–	–
FRS17_T1	−0.17*	−0.31**	−0.02	0.14	0.19*	–	–	–

T2 (30°C/25°C; day/night)

	FLN16_T2	FLN17_T2	FLN18_T2	FRN16_T2	FRN17_T2	FRN18_T2	FRS16_T2	FRS17_T2
FLN17_T2	0.58**	–	–	–	–	–	–	–
FLN18_T2	0.29**	0.28**	–	–	–	–	–	–
FRN16_T2	0.57**	0.38**	0.20*	–	–	–	–	–
FRN17_T2	0.52**	0.74**	0.06	0.42**	–	–	–	–
FRN18_T2	−0.04	0.03	0.01	0.06	0.07	–	–	–
FRS16_T2	0.13	0.16*	0.04	0.79**	0.22**	0.10	–	–
FRS17_T2	0.30**	0.28**	−0.14	0.30**	0.78**	0.01	0.24**	–
FRS18_T2	−0.05	0.01	−0.04	0.07	0.06	0.97**	0.11	0.01

T3 (35°C/30°C; day/night)

	FLN16_T3	FLN17_T3	FLN18_T3	FRN16_T3	FRN17_T3	FRN18_T3	FRS16_T3	FRS17_T3
FLN17_T3	0.49**	–	–	–	–	–	–	–
FLN18_T3	0.28*	0.34**	–	–	–	–	–	–
FRN16_T3	0.28*	0.251*	0.18	–	–	–	–	–
FRN17_T3	0.16	0.49**	0.274**	0.44**	–	–	–	–
FRN18_T3	−0.14	−0.06	0.05	−0.06	0.11	–	–	–
FRS16_T3	−0.14	0.10	0.10	0.74**	0.292*	−0.07	–	–
FRS17_T3	−0.09	−0.01	0.08	0.30**	0.70**	0.13	0.24*	–
FRS18_T3	−0.09	−0.06	−0.04	−0.03	0.10	0.94**	−0.043	0.12

*Correlations among traits within experiment are highlighted in green (2016), blue (2017) and orange (2018) at *p < 0.05 and **p < 0.01.*

population for both temperatures, whereas FLN clearly decreased at T3 (**Supplementary Table S3** and **Figure S1**). In the case of FRN and FRS, a drastic decrease was already observed at T2 (**Supplementary Table S3** and **Figure S1**). Mean differences between ILs and MM were not important for FLN at T1 and T2, and no IL overcame MM at T3 (**Supplementary Table S3**). Regarding FRN, ILs with higher FRN than MM were observed at T2 and T3 (**Supplementary Table S3**), suggesting that some ILs may carry heat tolerance genes. Lastly, few ILs displayed higher FRS than MM at T1, 19 ILs at T2 and 12 ILs at T3 (**Figure 2**). Interestingly, ILs with introgression in chromosomes 1, 2, and 12 showed high FRS, overlapping with some of the reproductive trait QTLs described above in the RIL population.

In order to verify the heat tolerance of candidate ILs, an assay with five replicates per IL was carried out in 2017. Twelve ILs with better performance under high temperatures in the previous year's assay were analyzed at the same three temperatures regimens. No significant differences were found between MM and ILs for FLN at all three temperatures

(**Supplementary Table S3**). The same result was found for FRN and FRS at T2 (**Supplementary Table S3**). At T3, SP_12-5 displayed higher FRN and FRS than MM, whereas SP_1-4 showed higher FRS (**Supplementary Table S3**). In 2019, a similar experiment was implemented with 9 of the 12 previously selected ILs and three additional ones selected for carrying an introgression in the same position of a QTL detected in the RIL population. SP_1-4 and SP_12-2 showed higher FRN and FRS than MM at T3 (**Supplementary Table S3**). SP_1-4 also showed a high FRS at T3 in all three experiments. Thus, the QTLs on chromosome 12, *frn12.1_T3_18* and *frs12.1_T3_18* were confirmed with SP_12-2. On the other hand, QTLs for FRN on chromosome 1 were detected in the RILs at T1 (*frn.1_T1_2E*) and T2 (*frn1.1_T2_16*), although not at T3, neither for FRS, whereas SP1-4 showed effects on FRN and FRS at T3.

Pollen Viability

Pollen viability was determined by three different approaches: pollen tube germination (TG), aniline blue staining (AB)

TABLE 4 | QTLs detected for the reproductive traits flower number (FLN), fruit number (FRN) and fruit set proportion (FRS) and stigma exsertion (SE) at different temperature regimens (T1: 25°C day/20°C night; T2: 30°C day/25°C night; T3: 35°C day/30°C night) in two experiments (2016 and 2017) by multi-environment QTL analysis with ICIMapping.

Trait	Temp.	QTL	Chr	Peak marker	position (cM)	LOD	LOD (A)	LOD (AbyE)	PVE	PVE (A)	PVE (AbyE)	Add
FLN	T1	fln2.1_T1_2E	2	solcap_snp_sl_35759	39	5.84	4.63	1.21	9.14	5.18	3.96	0.68
		fln2.2_T1_2E	2	solcap_snp_sl_20052	108	12.38	5.39	7.00	10.53	6.27	4.26	−0.75
		fln6.1_T1_2E	6	solcap_snp_sl_57398	67	6.60	5.49	1.10	6.25	6.20	0.05	0.75
		fln10.1_T1_2E	10	solcap_snp_sl_8835	80	7.02	6.90	0.11	10.47	8.09	2.38	0.84
	T2	fln2.2_T2_2E	2	solcap_snp_sl_36189	105	8.67	7.01	1.66	8.02	7.87	0.15	−1.03
		fln4.1_T2_2E	4	solcap_snp_sl_41710	42	4.14	3.42	0.72	3.80	3.77	0.03	0.76
		fln6.1_T2_2E	6	solcap_snp_sl_57327	71	4.31	3.06	1.25	3.60	3.37	0.23	0.69
		fln11.1_T2_2E	11	solcap_snp_sl_21007	29	5.28	2.64	2.64	3.99	2.93	1.07	−1.01
	T3	fln2.1_T3_2E	2	solcap_snp_sl_49669	58	5.54	5.54	0.00	9.52	7.09	2.42	0.67
		fln4.1_T3_2E	4	solcap_snp_sl_41710	42	4.40	4.40	0.00	7.38	5.51	1.87	0.63
FRN	T1	frn1.1_T1_2E	1	SL20127_1427	59	4.79	3.60	1.20	4.71	4.54	0.17	0.53
		frn2.2_T1_2E	2	solcap_snp_sl_36189	105	8.16	3.55	4.61	7.01	4.49	2.52	−0.53
		frn6.1_T1_2E	6	solcap_snp_sl_41840	44	5.33	3.23	2.10	4.82	4.14	0.68	0.51
		frn6.2_T1_2E	6	solcap_snp_sl_57327	72	4.24	4.06	0.18	7.14	5.22	1.92	0.58
		frn10.1_T1_2E	10	solcap_snp_sl_59416	65	5.51	4.62	0.89	10.16	6.00	4.16	0.61
	T2	frn4.1_T2_2E	4	solcap_snp_sl_41710	42	4.64	4.59	0.05	6.15	6.01	0.14	0.70
		frn6.2_T2_2E	6	SGN-U579138_snp745	82	4.59	4.21	0.38	6.22	5.52	0.69	0.64
		frn7.1_T1_2E	7	solcap_snp_sl_70700	84	6.16	4.70	1.46	8.71	6.22	2.49	−0.67
FRS	T2	frs6.2_T2_2E	6	SGN-U579138_snp745	82	4.94	4.65	0.29	7.39	6.83	0.56	6.49
		frs7.2_T2_2E	7	solcap_snp_sl_37060	82	7.50	5.38	2.13	11.79	7.91	3.88	−6.85
SE	T1	se2.1_T1_2E	2	solcap_snp_sl_50058	79	10.35	9.88	0.47	12.83	12.75	0.08	−0.27
	T2	se2.1_T1_2E	2	solcap_snp_sl_50058	82	5.63	5.55	0.08	9.51	8.46	1.05	−0.16

QTLs are named with the trait abbreviation, followed by the chromosome number, the number of QTLs within the chromosome, temperature regimen and the suffix _2E. The marker closest to the maximum LOD score and its genetic position are also shown. The statistical estimates for each QTL include: Maximum LOD score for genetic effects (LOD), LOD score of additive effects [LOD (A)], LOD score for the interaction of additive effects with environment [LOD (AbyE)], percentage of phenotypic variance explained by the QTL (PVE), by the additive effects (A), interaction additive by environment (AbyE) and the additive value (Add), being negative when TO-937 alleles increased the trait and positive when MM increased it.
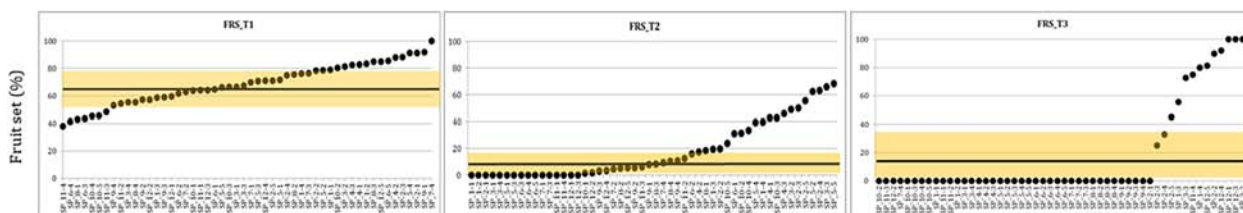
FIGURE 2 | Overview of the fruit set percentage among ILs in 2016 experiment. Means of percentage of fruit set (FRS) in 2016 experiment at three temperature regimens: (T1: 25°C day/20°C night; T2: 30°C day/25°C night; T3: 35°C day/30°C night) among TO-937 introgression lines (ILs). ILs are on the X axis in order according their FRS value. The yellow line indicates the "MoneyMaker" mean and the shadowed area the standard deviation.

and cytometry (VP) for the 2017 experiment (at all three temperatures), whereas in 2016 only TG was analyzed at T2 and T3. The distribution of the traits showed a decrease of all pollen viability traits as temperature increased, mostly in T3 (**Figure 3**). A low correlation was found for TG between 2016 and 2017 (**Supplementary Table S4**). On the other hand, correlations between AB and VP were positive and highly significant at all three temperature regimens (**Supplementary Table S4**).

However, TG did not show a significant correlation with the previous traits. Regarding to the correlation between pollen and the other reproductive traits, no significant correlations were found between pollen viability traits and FLN, FRN or FRS (**Supplementary Table S4**).

In 2016, no QTL was identified for TG. In 2017, 4 QTLs for TG, 2 QTLs for AB and 2 QTLs for VP were identified (**Table 5**). A region at the end of chromosome 3 contained QTLs involved

FIGURE 3 | Histograms of the distribution of pollen viability traits in the RIL population. Histograms at the three temperature regimens: (T1: 25°C day/20°C night; T2: 30°C day/25°C night; T3: 35°C day/30°C night) for the pollen viability traits studied in 2016 and 2017 experiments among MoneyMaker"-x-TO-937 Recombinant Inbred Lines (RIL). The mean for "Moneymaker" (MM) is indicated with an arrow. In 2016, pollen tube growth was not assayed at T1.

in all three pollen viability traits at T1, with the TO-937 allele increasing viability, which could reflect the presence of a QTL involved in pollen viability at normal temperatures. Interestingly, the QTL for TG and AB were detected at the top of chromosome 7 at T2, which could be a candidate for a pollen viability QTL under mild heat stress. Nevertheless, as these QTLs were studied in only one experiment, their effects should be verified with additional experiments.

Tipburn Incidence

Tipburn is a physiological disorder that usually occurs as a consequence of heat stress (Jenni et al., 2013) and it is characterized by the necrosis of the youngest leaves and inflorescences at the tip of the apex of the plant, bringing production of new vegetative and reproductive structures to a halt. A segregation of tipburn incidence (TB) was observed initially in the RIL population during the 2016 experiment, and it was also recorded in the 2017 experiment. The incidence of TB depended on the temperature regimen: at optimal temperatures TB was present in less than one-third of the plants, while at T3, over half of the RILs showed TB (**Figure 4**).

Several markers showed an association with TP incidence at chromosome 7 at T1 in 2016 and at T3 in 2017 experiments. The marker with the strongest association was located in position 58 cM (**Table 6**), although a large region of chromosome 7 also showed association with TP (**Supplementary Figure S2**), what could indicate the presence of a linked QTL. In order to verify this QTL, the incidence of TB was evaluated in ILs with introgressions in chromosome 7. ILs SP_7-3 and SP_7-4 did not show incidence of TP (**Figure 5**). The introgressions carried by these ILs only overlapped in the interval between markers solcap_snp_sl_70992

and solcap_snp_sl_70912, so it is reasonable to suggest that the QTL is within that marker interval. Nevertheless, the possibility of two linked genes located in the non-overlapping regions of the two ILs cannot be ruled out, and further fine mapping would be necessary to discern between these two hypotheses. SNP genotype data and genetic for RILs can be found in **Supplementary Table S5**. RIL and IL phenotypic data for all experiments are included in **Supplementary Tables S6, S7**.

DISCUSSION

Heat stress affects both vegetative growth and reproduction of plants, and in both cases, the plant response is complex and controlled by multiple genes. The choice of the specific vegetative or reproductive trait for studying heat tolerance will determine the genetic mechanism that can be identified. The genetic mechanism may be common for vegetative and reproductive traits or specific, i. e., depending on the trait a component of the tolerance would be studied. The effects of heat stress on vegetative traits are evident at high temperatures (i.e., 40°C, Wen et al., 2019), whereas reproductive traits are already affected by mild heat stress when minimum temperatures are above 25°C (Xu et al., 2017b). Currently, night temperatures over 25°C are not unusual during summer at regions where tomatoes are cultivated, such as the Mediterranean basin, so we decided to focus our research on identifying sources of heat tolerance on reproductive traits when temperatures increase above the minimum of 25°C.

Reproductive traits can be measured in different ways including the number of flowers, fruits per inflorescence and pollen viability. The last one has been proposed as an adequate

TABLE 5 | QTLs detected for the pollen viability traits pollen tube growth (TG), aniline blue staining (AB) and viable pollen by cytometry (VP) at different temperature regimens (T1: 25°C day/20°C night; T2: 30°C day/25°C night; T3: 35°C day/30°C night) in the 2017 experiment with the Moneymaker-x-TO-937 recombinant inbred lines.

Trait	Temp.	QTL	Chr	Peak marker	position (cM)	LOD	(%) Phenotipic variance	a
TG	T1	**TG3.1_T1**	3	solcap_snp_sl_20802	120.0	3.3	7.49	−0.30
		TG8.1_T1	8	solcap_snp_sl_64604	88.07	3.6	5.86	0.27
	T2	TG7.1_T2	7	solcap_snp_sl_62904	1.0	3.8	9.15	−0.23
	T3	**TG3.1_T3**	3	solcap_snp_sl_61460	134.32	4.4	1.85	0.14
AB	T1	AB3.1_T1	3	CL016802-0675	95.0	3.1	9.76	−0.28
	T2	AB7.1_T2	7	solcap_snp_sl_62904	1.0	2.9	6.42	−0.26
VP	T1	VP3.1_T1	3	solcap_snp_sl_58584	92.14	3.1	9.77	−4.29
		VP4.1_T1	4	solcap_snp_sl_47457	122.32	29.5	61.52	32.41

QTLs are named with the trait abbreviation, followed by the chromosome number, the number of QTLs within the chromosome and the temperature regimen. The additive value (A) is negative when TO-937 alleles increase the trait and positive when "MoneyMarker" alleles increase it. QTLs detected in at least two temperatures or experiments are highlighted in bold.
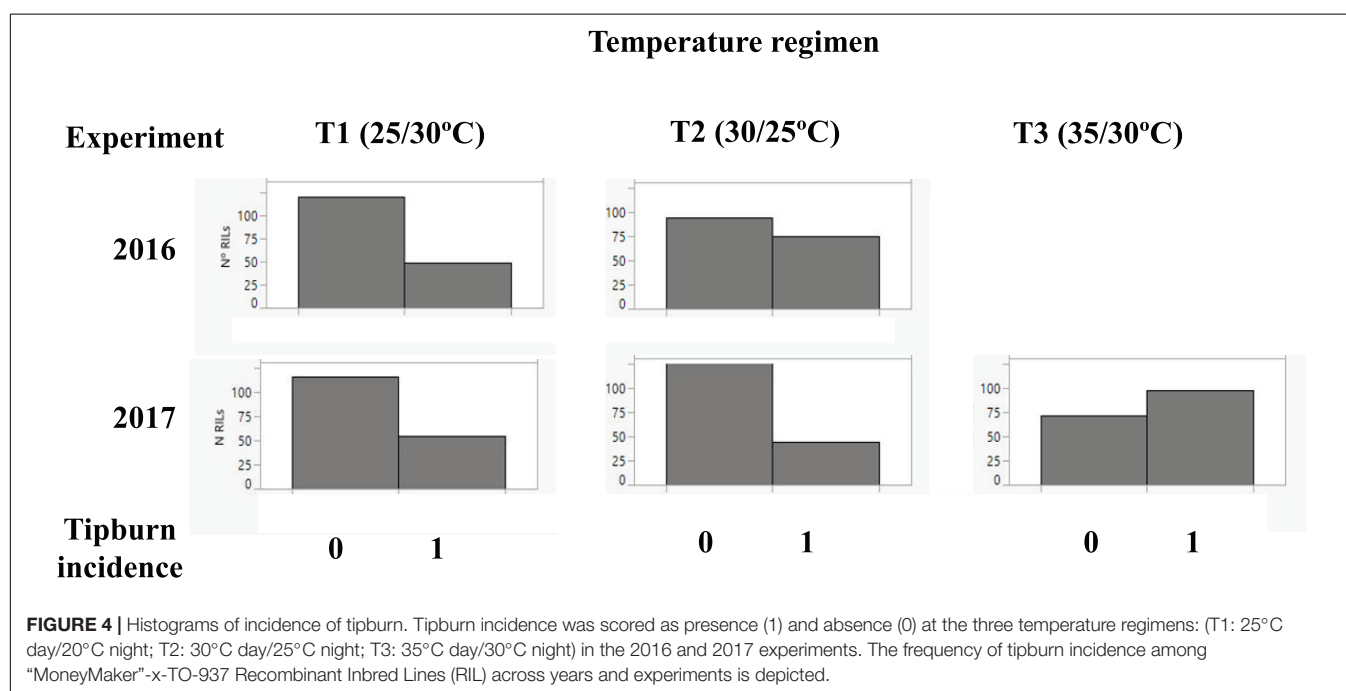
FIGURE 4 | Histograms of incidence of tipburn. Tipburn incidence was scored as presence (1) and absence (0) at the three temperature regimens: (T1: 25°C day/20°C night; T2: 30°C day/25°C night; T3: 35°C day/30°C night) in the 2016 and 2017 experiments. The frequency of tipburn incidence among "MoneyMaker"-x-TO-937 Recombinant Inbred Lines (RIL) across years and experiments is depicted.

TABLE 6 | Incidence of Tipburn (TB) among recombinant inbred lines (RILs) in the 2016 experiment at T1 (25°C day/20°C night) (16T1) and the 2017 experiment at T3 (35°C day/30°C night) (T3) and association with marker solcap_snp_sl_53335 (located in position 58 cM of chromosome 7).

Genotype	Number of RILs	Experiment			
		16T1		17T3	
		RILs with TB	RILs No TB	RILs with TB	RILs No TB
EE	58	28	30	42	16
PP	78	57	21	44	34
	p	0.0002		0.0011	

Genotype indicates homozygous "MoneyMarker" (MM) and homozygous TO-937 (PP) for the marker, followed by the total number of RILs with the genotypes and number of RILs within each genotypic class displaying or not displaying TB. In the lower part of the table, the significance of the χ^2 test for marker/TB association is shown.

FIGURE 5 | Validation of QTL for tipburn incidence on chromosome 7. The condensed genetic map with SOLCAP markers calculated in the MoneyMaker"-x-TO-937 Recombinant Inbred Lines (RIL) is depicted in the top of the figure. The extent of the TO-937 introgressions for the selected introgression lines (ILs) is indicated with a blue bar below the genetic map. The table on the left depicts the tipburn means for the ILs and "Moneymaker" (MM) in the 2017 experiment. Dunnet's test column indicates the statistically significant mean differences at $p < 0.05$, means with the different letters indicates significant mean differences.

indicator of heat tolerance due to its correlation with fruit set and a likely simpler genetic control mechanism (Xu et al., 2017b; Driedekons et al., 2018). In the current report, we have evaluated pollen viability using three different methods: TG, AB and VP. AB and VP showed strong correlations at all temperatures, but correlations with TG were not significant. On the other hand, QTLs involved in TG, AB and VP were detected in the same region of chromosome 3 at T1 with the TO-937 allele increasing pollen viability, although the effects of the QTLs were relatively modest (explaining less than 10% of phenotypic variance). In this same region in chromosome 3, a MetaQTL for pollen viability was detected from the joint analysis of four different experiments (Ayenan et al., 2019). Similarly, QTLs in the same region of chromosome 7 were found for TG and AB in T2, also with modest effect (also less than 10% of phenotypic variance). These results suggested that, even though TG is in general controlled by a different mechanism than AB and VP, some common mechanisms may exist. Xu et al. (2017b) detected a QTL for pollen viability on chromosome 11, but we did not find QTLs for any of the assayed pollen viability traits in that chromosome, which can be explained by the different germplasms used in the two reports. We also did not find correlation between pollen viability traits and fruit set, although the lack of biological replications in the RIL population might limit conclusions from this observation. Xu et al. (2017a) found a significant correlation at high temperature, but this was not significant at the control temperature, suggesting that the correlation may be dependent on the experimental conditions. Given the low correlation between pollen viability traits and the modest effects of the QTLs detected in the current report, the genetic variability for pollen variability traits in the current RIL seems to be low and likely not sufficient for accurately studying their genetic control.

The high and positive correlation observed in the reproductive traits FLN, FRN, and FRS between the 2016 and 2017 experiments indicated a significant hereditability at all the assayed temperatures. The low correlation observed with the 2018 experiment may be due to the environmental differences between NTW and FCCV greenhouse facilities and/or developmental stages of the plants when they were subjected to high temperatures in the NTW facilities (Charles and Harris,

1972). The limitation of biological replicates does not allow us to infer sturdy conclusions from these observations. Also, several QTLs and ILs displayed consistent effects at different temperature regimens and across years, indicating that the current mapping populations harbored sufficient genetic variability for studying the genetic control of these reproductive traits. Xu et al. (2017a) did not find a correlation between pollen viability and female fertility, so it is likely that the fruit set variability in the current experiment could be related to female fertility instead of pollen variability, which would explain the lack of correlation between pollen viability and fruit set.

Therefore, the discussion will be focused on FLN, FRN, and FRS. At mild heat stress (T2), FLN only showed a slight reduction, with this being more drastic at T3. On the other hand, detrimental effects of increased temperature were apparent for mild heat stress on FRN and FRS. Thus, even though, MM and TO-937 showed a drastic reduction of FRN and FRS at T2, and more drastic at T3, transgressive segregations were observed in the RIL population. The transgressive phenotypes appeared due to allelic combinations from the different parents, with a portion of RILs setting fruit at both high temperatures. Transgressive segregation is commonly reported when crossing exotic germplasms with cultivated tomatoes (de Vicente and Tanksley, 1993; Monforte et al., 2001; Shivaprasad et al., 2012; Capel et al., 2015). The fact that transgressive segregation for heat tolerance in reproductive traits has been found in a recombinant population derived from non-heat-tolerant parents reinforces the power of interspecific crosses for uncovering hidden genetic diversity. In some cases both TO-937 and MM alleles are able to increase reproductive traits at a high temperature in different QTLs. Xu et al. (2017a) in a cross between the heat-tolerant cultivar "Nagcarlang" and heat-susceptible NCHS-1, and Wen et al. (2019) in a cross between the heat-tolerant LA2093 (*S. pimpinellifolium*) and heat-susceptible LA1698 (cultivated tomato), also found that alleles from both parents were associated to tolerance. Thus, the tolerance determined in parents does not seem to be a perfect indicator of the genetic potential of the alleles carried by them, which increases the potential of the tomato germplasm for developing new heat tolerant cultivars by generating new genetic combinations.

The number of QTLs detected at each temperature regime/experiment combination was relatively low. Likely, a large number of biological replication could help to detect a larger number of QTLs. This low QTL detection was also observed in the previous works by Xu et al. (2017a); Ruggieri et al. (2019) and Wen et al. (2019). In fact, the meta-QTL analysis carried out by Ayenan et al. (2019) only found one meta-QTL in chromosome 1 for the number of flowers per inflorescence (Ayenan et al., 2019). The low QTL detection probably reflects the genetic complexity of the studied heat tolerance traits, resulting in only a fraction of the genes involved being detected by using common QTL mapping approaches. In the current report, the QTLs displaying more consistent effects across temperatures and experiments were involved in FLN and FRN. Among them, the QTL with the most consistent effect is located on the distal region of chromosome 2. Interestingly, the association of fruit numbers at high temperature with makers located in chromosome 2 was also reported by Ruggieri et al. (2019). Another region which can be highlighted is the FLN QTL on chromosome 11 that may correspond with the QTL qFP11 involved in flowers per inflorescence as reported by Xu et al. (2017a). While it is true that the number of QTL experiments dissecting the genetic control of heat tolerance in tomatoes is still very limited, the identification of similar genomic regions involved in heat tolerance in this and previous reports is encouraging and provides additional proof that we are on the right track.

The analysis of the IL population was initially intended to verify the QTLs detected in the RIL population, but also to identify new ones. Our results show that there was not a very high correspondence between the QTLs identified in both populations. The lack of verification may be due to the complex genetic control of the traits, i.e., the effect of the QTL may greatly depend on the genetic background, so specific multi-loci combinations would be necessary to express the heat tolerance. These multi-loci combinations may occur in a number of RILs, but they disappeared in ILs. Nevertheless, some interesting results were found among ILs. For instance, IL SP12-2 showed higher FRN and FRS at T3 than MM in the 2019 experiment which could indicate the effects of QTLs *frn12.1_T3_18*, *frs12.1_T3_18* detected in the RIL population. SP_1-4 displayed higher FRN and FRS than MM at T3, whereas the QTLs detected in this region with the RILs were involved only in FRN at T1 and T2 temperatures (*frn11.1_T1_16*, *frn1.1_T2_16*). It is likely that the QTL for FRN has stronger effects in the RILs at T1 and T2, whereas in the ILs the effects become more evident at T3 due to the differences in interactions with the genetic background as a result of the different genetic structure of both populations. Nevertheless, we cannot rule out that those are different QTLs. The changes in the genetic background during IL development can also induce new phenotypes that were not observed previously in early segregating populations. Unexpected phenotypes are commonly reported in IL populations, such as the increased intensity of the internal red color in tomatoes observed in a green-fruited *S. habrochaites* IL (Monforte et al., 2001) increased fruit weight from *S. pimpinellifolium* (Barrantes et al., 2016) fruits of a melon IL showing climacteric ripening from non-climacteric parents

(Vegas et al., 2013) or production of round fruits from parents that produce oval or elongated fruits (Diaz et al., 2014). As each IL carries a single introgression on the otherwise MM genetic background, the genetic complexity of the traits was reduced to a single locus, which will facilitate their future use in breeding programs and at the same time to identify the causal genes of the heat tolerance.

In addition to the reproductive traits, heat stress affects other factors such as stigma exsertion, increasing the protrusion of the style out of the anther. QTLs with strong effects on SE across experiment and temperature regimens were located on chromosome 2 in the same region as *se2.1* that was map-based cloned by Chen and Tanksley (2004). Xu et al. (2017a) also detected SE QTLs in that region under heat stress, supporting the finding that *se2.1* control of stigma exsertion is mostly independent on temperature conditions. The non-significant correlation between SE and FRS also reinforces the idea that SE is not an appropriate trait for evaluating the tolerance to high temperature in tomatoes (Lohar and Peat, 1997).

Lastly, TB has been studied thoroughly in leafy vegetables (such as lettuce) as the presence of this physiological disorder has a direct impact on the market value. In the case of tomato, the incidence of TB affects young developing tips and inflorescences, dramatically reducing tomato yield. We report a consistent QTL for TB on chromosome 7 successfully validated in ILs SP_7-3 and SP_7-4. According to the RIL experiment, the QTL would map in the central region of the chromosome (around 58 cM), whereas the IL analysis suggested that the QTL would map on the distal regions (around 95 cM), assuming that the QTL is located in the region where the SP_7-3 and SP_7-4 introgressions overlap. A closer look to the association of markers with TB incidence across chromosome 7 showed that markers located in the region between 50 and 100 cM were also associated with TB tolerance (**Supplementary Figure S2**), which may be interpreted as the presence of multiple linked QTLs. Therefore, we cannot rule out that SP_7-3 and SP_7-4 tolerance to TB is due to the same QTL or different linked QTLs on chromosome 7. Anyway, both ILs showed a very mild, negligible incidence of TB, whereas other ILs and MM were severely affected. This result suggested that TB can be prevented with just one locus (i. e., even though the ILs may carry different QTLs, the presence of only one of them is sufficient for tolerance). To the best of our knowledge, this is the first report of the mapping of TB resistance in tomatoes. Major QTLs for TB have also been recently reported in lettuce (Jenni et al., 2013; Macias Gonzalez et al., 2019) however, no candidate gene has been suggested yet. The precise location of QTL or QTLs on chromosome 7 will allow an assessment of whether the genetic control on TB resistance could be similar in both species, as soon as candidate genes can be defined in posterior research studies.

CONCLUSIONS

The genetic control of heat stress tolerance in tomatoes is largely unknown. In our experimental design, no true biological replicates were studied in the RIL population, which limited

some of the conclusions about the general genetic control of heat tolerance in the RIL population. Nevertheless, as geneticists, our interest was focused in the cosegregation of markers and traits to identify genomic regions associated with the tolerance. We identified new sources of variability for heat tolerance and dissected the tolerance to heat stress for reproductive traits using previously developed mapping populations. The strategy was successful, as heat tolerance associated RILs, ILs and markers were identified. The heat tolerance was mainly associated with female fertility rather than pollen viability, although some QTLs involved in pollen viability were also found. Some ILs showed tolerance in several replicated experiments, making them the proper choice for increasing the heat tolerance of tomato cultivars or to define candidate or casual genes. Some QTLs detected in the RIL population could not be verified with the ILs, indicating that the effects of these QTLs are dependent on the genetic background. For those cases, a single QTL selection approach (as the IL strategy) would not be the right approach and other strategies such as genomic selection may be more appropriate (Voss-Fels et al., 2019) for managing them in breeding programs. The study of two mapping populations with different genetic structures has allowed us to obtain some insights into different heat tolerance mechanisms. Also, for the first time, a locus for TB tolerance has been mapped on chromosome 7, and its transfer to applied or basic research will be straightforward. In summary, we report a large catalog of QTLs involved in tomato reproductive traits at different temperatures. Most of these QTLs are not involved in pollen viability traits, but they increase the fruit set at high temperatures. Therefore, it would be expected that in combination with the pollen viability QTL from other sources this could speed up the development of heat tolerant tomato cultivars. Nevertheless, fruit set is one of the components of global heat tolerance. Research on other traits such as vegetative development, fruit quality and postharvest behavior at high temperatures would complement the current study to design a holistic strategy to develop heat tolerant cultivars.

DATA AVAILABILITY STATEMENT

AUTHOR CONTRIBUTIONS

MG carried out the experiments, analyzed the data, and drafted the manuscript. Y-CL, K-YC, DG, and TM carried out experiments. IN and CB supervised greenhouse experiments. AG and AM designed the study and wrote the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

REFERENCES

Abdul-Baki, A. A. (1991). Tolerance of tomato cultivars and selected germplasm to heat stress. *J. Amer. Soc. Hort. Sci.* 116, 1113–1116.

Abdul-Baki, A. A., and Stommel, J. R. (1995). Pollen viability and fruit set of tomato genotypes under optimum and high-temperature regimes. *HortScience* 30, 115–117.

Adams, S. R., Cockshull, K. E., and Cave, C. R. J. (2001). Effect of temperatures on the growth and development of tomato fruits. *Ann. Bot.* 88, 869–877. doi: 10.1006/anbo.2001.1524

Alam, M., Sultana, N., Ahmad, S., Hossain, M., and Islam, A. (2010). Performance of heat tolerant tomato hybrid lines under hot, humid conditions. *Bangladesh J. Agr. Res.* 35, 367–373. doi: 10.3329/bjar.v35i3.6442

Alba, J. M., Montserrat, M., and Fernández-Muñoz, R. (2009). Resistance to the two-spotted spider mite (Tetranychusurticae) by acylsucroses of wild tomato (*Solanum pimpinellifolium*) trichomes studied in a recombinant inbred line population. *Exp. Appl. Acarol.* 47, 35–47.

Alsamir, M., Ahmand, N., Ariel, V., Mahmood, T., and Trethiwan, R. (2019). Phenotypic diversity and marker-trait association under heat stress in tomato (*Solanum lycopersicum* L.). *Aust. J. Crop Sci.* 13, 578–587. doi: 10.21475/ajcs.19.13.04.p1581

Ayenan, M. A. T., Danquah, A., Hanson, P., Ampomah-Dwamena, C., Sodedji, F. A. K., Asante, I. K., et al. (2019). Accelerating vreeding for heat tolerance in tomato (*Solanum lycopersicum* L.): an integrated approach. *Agronomy* 9:720.

Barrantes, W., Fernández-del-Carmen, A., López-Casado, G., Gonzalez-Sanchez, M. A., Fernandez-Muñoz, R., Granell, A., et al. (2014). Highly efficient genomics-assisted development of a library of introgression lines of *Solanum pimpinellifolium*. *Mol. Breed.* 34, 1817–1831.

Barrantes, W., Lopez-Casado, G., Garcia-Martinez, S., Alonso, A., Rubio, F., Ruiz, J. J., et al. (2016). Exploring new alleles involved in tomato fruit quality in an introgression line library of *Solanum pimpinellifolium*. *Front. Plant Sci.* 7:1172. doi: 10.3389/fpls.2016.01172

Bita, C. E., and Gerads, T. (2013). Plant tolerance to high temperatures in a changing environment: scientific fundamentals and production of heat stress-tolerant crops. *Front. Plant Sci.* 4:273. doi: 10.3389/fpls.2013.00273

Capel, C., Fernández del Carmen, A., Alba, J. M., Lima-Silva, V., Hernández-Gras, F., Salinas, M., et al. (2015). Wide-genome QTL mapping of fruit quality traits in a tomato RIL population derived from the wild-relative species *Solanum pimpinellifolium* L. *Theor. Appl. Genet.* 128, 2019–2035. doi: 10.1007/s00122-015-2563-4

Capel, C., Yuste-Lisbona, F. J., López-Casado, G., Angosto, T., and Cuartero, J. (2017). Multi-environment QTL mapping reveals genetic architecture of

fruit cracking in a tomato RIL *Solanum lycopersicum* × *S. pimpinellifolium* population. *Theor. Appl. Genet.* 130:213. doi: 10.1007/s00122-016-2809-9

Challinor, A. J., Watson, J., Lobell, D. B., Howden, S. M., Smith, D. R., and Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nat. Clim. Change* 4:287. doi: 10.1038/nclimate2153

Charles, W. B., and Harris, R. E. (1972). Tomato fruit-set at high and low temperatures. *Can. J. Plant Sci.* 52, 497–506.

Chen, K. Y., and Tanksley, S. D. (2004). High-resolution mapping and functional analysis of se2.1: a major stigma exsertion quantitative trait locus associated with the evolution from allogamy to autogamy in the genus *Lycopersicon*. *Genetics* 168, 1563–1573. doi: 10.1534/genetics.103.022558

Chung, M. Y., Vrebalov, J., Alba, R., Lee, J., McQuinn, R., Chung, J. D., et al. (2010). A tomato (*Solanum lycopersicum*) APETALA2/ERF gene, SlAP2a, is a negative regulator of fruit ripening. *Plant J.* 64, 936–947. doi: 10.1111/j.1365-313X.2010.04384.x

Dane, F., Hunter, A. G., and Chambliss, O. L. (1991). Fruit set, pollen fertility and combining ability of selected tomato genotypes under high-temperature field conditions. *J. Amer. Soc. Hort. Sci.* 116, 906–910.

de Vicente, M. C., and Tanksley, S. D. (1993). QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics* 134, 585–596.

Diaz, A., Zarouri, B., Fergany, M., Eduardo, I., Álvarez, J. M., Pico, B., et al. (2014). Mapping and introgression of QTL involved in fruit shape transgressive segregation in 'Piel de Sapo' melon (*Cucumis melo* L.). *PLoS One* 9:e104188. doi: 10.1371/journal.pone.0104188

Driedekons, N., Wolter-Arts, M., Huber, H., de Boer, G.-J., Vriezen, W., Mariani, C., et al. (2018). Exploring the natural variation for reproductive thermotolerance in wild tomato species. *Euphytica* 214:67.

El Ahmadi, A. B., and Stevens, M. A. (1997). Reproductive responses of heat-tolerant tomatoes to high temperatures. *J. Am. Soc. Hort. Sci.* 104, 686–691.

Geisenberg, C., and Stewart, K. (1986). "Field crop management," in *The Tomato Crop: a Scientific Basis for Improvement*, eds J. G. Atherton and J. Rudich (London: Chapman & Hall), 511–557.

Golam, F., Phorhan, Z. H., Nezhadahmani, A., and Rahman, M. (2012). Heat tolerance in tomato. *Life Sci. J.* 9:4.

Grilli, G. V., Braz, G. L. T., and Lemos, E. G. M. (2007). QTL identification for tolerance to fruit set in tomato by fAFLP markers. *Crop Breed. Appl. Biot.* 7, 234–241.

Hasanuzzaman, M., Nahar, K., Alam, M., Roychowdhury, R., and Fujita, M. (2013). Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. *Int. J. Mol. Sci.* 14, 9643–9684. doi: 10.3390/ijms14059643

IPCC (2014). *Climate Change: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, eds O. Edenhofer, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, et al. (Cambridge: Cambridge University Press).

Jenni, S., Truco, M. J., and Michelmore, R. W. (2013). Quantitative trait loci associated with tipburn, heat stress-induced physiological disorders, and maturity traits in crisphead lettuce. *Theor. Appl. Genet.* 126, 3065–3079. doi: 10.1007/s00122-013-2193-7

JMP (2019). *JSL Syntax Reference*. Cary, NC: SAS Institute Inc.

Kugblenu, Y. O., Oppong Danso, E., Ofori, K., Andersen, M., Abenney, M., Stephen, S., et al. (2013). Screening tomato genotypes for adaptation to high temperature in West Africa. *Acta Agri. Scand. B Soil Plant Sci.* 63, 516–522. doi: 10.1080/09064710.2013.813062

Levy, A., Rabinowitch, H. D., and Kedar, N. (1978). Morphological and physiological characters affecting flower drop and fruit set of tomatoes at high temperatures. *Euphytica* 27, 211–218. doi: 10.1007/BF00039137

Lin, K. H., Yeh, W. L., Chen, H. M., and Lo, H. F. (2010). Quantitative trait loci influencing fruit-related characteristics of tomato grown in high-temperature conditions. *Euphytica* 174, 119–135. doi: 10.1007/s10681-010-0147-6

Lohar, D. P., and Peat, W. E. (1997). Floral characteristics of heat-tolerant and heat-sensitive tomato (*Lycopersicon esculentum* Mill.) cultivars at high temperature. *Sci. Hort.* 73, 53–60.

Macias Gonzalez, M., Truco, M. J., Bertier, L. D., Jenni, S., Simko, I., Hayes, R. J., et al. (2019). Genetic architecture of tipburn resistance in lettuce. *Theor. Appl. Genet.* 132, 2209–2222. doi: 10.1007/s00122-019-03349-6

Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in bi-parental populations. *Crop J.* 3, 269–283.

Monforte, A., Friedman, E., Zamir, D., and Tanksley, S. D. (2001). Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: deductions about natural variation and implications for germplasm utilization. *Theor. Appl. Genet.* 102, 572–590. doi: 10.1007/s001220051684

Moore, E., and Thomas, W. (1952). Some effects of shading and paracolorphenoxy acetic acid on fruitfulness of tomatoes. *Am. Soc. Hort. Sci. Proc.* 60, 289–294.

Nahar, K., and Ullah, S. M. (2011). Effect of water stress on moisture content distribution in soil and morphological characters of two tomato (*Lycopersicon esculentum* Mill) Cultivars. *J. Sci. Res.* 3, 677–682. doi: 10.3329/jsr.v3i3.7000

Nahar, K., and Ullah, S. M. (2012). Morphological and physiological characters of tomato (*Lycopersicon esculentum* Mill) cultivars under water stress. *Bangladesh J. Agri. Res.* 37, 355–360.

Opena, R. T., Chen, J. T., Kuo, C. G., and Chen, H. M. (1992). "Genetic and physiological aspects of tropical adaptation in tomato. Adaptation of food crops to temperature and water stress," in *Proceedings of an International Symposium, 3–18 August*, (Shanhua: Asian Vegetable Research and Development Centre).

Paupière, M. J., van Haperen, P., Rieu, I., Visser, R. G. F., Tikunov, Y. M., and Bovy, A. G. (2017). Screening for pollen tolerance to high temperatures in tomato. *Euphytica* 213:130. doi: 10.1007/s10681-017-1927-z

Peet, M. M., Sato, S., and Gardner, R. G. (1998). Comparing heat stress effects on male fertile and male sterile tomatoes. *Plant. Cell Environ.* 21, 225–231. doi: 10.1046/j.1365-3040.1998.00281.x

Powell, A. L. T., Nguyen, C. V., Hill, T., Cheng, K. L., Figueroa-Balderas, R., Aktas, K., et al. (2012). Uniform ripening encodes a golden 2-like transcription factor regulating tomato fruit chloroplast development. *Science* 336, 1711–1715. doi: 10.1126/science.1222218

Pressman, E., Harel, D., Zamski, E., Shaked, R., Altahan, L., Rosenfeld, K., et al. (2006). The effect of high temperatures on the expression and activity of sucrose cleaving enzymes during tomato (*Lycopersicon esculentum*) anther development. *J. Hort. Sci. Biotech.* 81, 341–348. doi: 10.1080/14620316.2006.11512071

Pressman, E., Peet, M. M., and Pharr, D. M. (2002). The effect of heat stress on tomato pollen characteristics is associated with changes in carbohydrate concentration in the developing anthers. *Ann. Bot.* 90, 631–636. doi: 10.1093/aob/mcf240

Rambla, J. L., Medina, A., Fernández del Carmen, A., Barrantes, W., Grandillo, S., Cammareri, M., et al. (2017). Identification, introgression, and validation of fruit volatile QTLs from a red-fruited wild tomato species. *J. Exp. Bot.* 68, 429–442. doi: 10.1093/jxb/erw455

Rick, C. M., and Dempsey, W. H. (1969). Position of the stigma in relation to fruit setting of the tomato. *Bot. Gazette* 130, 180–186.

Ruggieri, V., Calafiore, R., Schettini, C., Rigano, M. M., Olivieri, F., Frusciante, L., et al. (2019). Exploiting genetic and genomic resources to enhance heat-tolerance in tomatoes. *Agronomy* 9:22. doi: 10.3390/agronomy9010022

Salinas, M., Capel, C., Alba, J. M., Mora, B., Cuartero, J., Fernández-Muñoz, R., et al. (2013). Genetic mapping of two QTL from the wild tomato *Solanum pimpinellifolium* L. controlling resistance against two-spotted spider mite (*Tetranychus urticae* Koch). *Theor. Appl. Genet.* 126, 83–92.

Sato, S., Kamiyama, M., Iwata, T., Makita, T., Furukawa, H., and Ikeda, H. (2006). Moderate increase of mean daily temperature adversely affects fruit set of *Lycopersicon esculentum* by disrupting specific physiological processes in male reproductive development. *Ann. Bot.* 97, 731–738. doi: 10.1093/aob/mcl037

Shivaprasad, P. V., Dunn, R. M., Santos, B. A., Bassett, A., and Baulcombe, D. C. (2012). Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. *EMBO J.* 31, 257–266. doi: 10.1038/emboj.2011.458

Sim, S. C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M. W., Van Deynze, A., et al. (2012). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS One* 7:e0040563. doi: 10.1371/journal.pone.0040563

Starck, Z., Siwiec, A., and Chotuj, D. (1994). Distribution of calcium in tomato plants in response to heat strees and plant growth regulators. *Plant Soil* 167, 143–148. doi: 10.1007/BF01587609

Vegas, J., Garcia-Mas, J., and Monforte, A. J. (2013). Interaction between QTLs induces an advance in ethylene biosynthesis during melon fruit ripening. *Theoretic. Appl. Genet.* 126, 1531–1544. doi:10.1007/s00122-013-2071-3

Voss-Fels, K. P., Cooper, M., and Hayes, B. (2019). Accelerating crop genetic gains with genomic selection. *Theoretic. Appl. Genet.* 132, 669–686. doi: 10.1007/s00122-018-3270-8

Wahid, A., Geleani, S., Ashraf, M., and Foolad, M. R. (2007). Heat tolerance in plants: an overview. *Env. Exp. Bot.* 61, 199–223. doi: 10.1016/j.envexpbot.2007.05.011

Wang, S., Basten, C. J., and Zeng, Z. B. (2007). *Windows QTL Cartographer 2.5.* Raleigh, NC: Department of Statistics, North Carolina State University.

Wen, J., Jiang, F., Weng, Y., Sun, M., Shi, X., Zhou, Y., et al. (2019). Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping, QTL-seq and RNA-seq in tomato. *BMC Plant Biol.* 19:398. doi: 10.1186/s12870-019-2008-3

Xu, J., Driedekons, N., Rutten, M. J. M., Vriezen, W. H., de Boer, G.-J., and Rieu, I. (2017a). Mapping quantitative trait loci for heat tolerance of reproductive traits in tomato (*Solanum lycopersicum*). *Mol. Breed.* 37:58. doi: 10.1007/s11032-017-0664-2

Xu, J., Wolters-Ars, M., Mariani, C., Huber, H., and Rieu, I. (2017b). Heat stress affects vegetative and reproductive performance and trait correlations in tomato (*Solanum lycopersicum*). *Euphytica* 213:156.

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136, 1457–1468.

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.

The reviewer GR declared a past co-authorship with several of the authors AM and MG to the handling Editor.

frontiers
in Plant Science

Field Resistance to *Phakopsora pachyrhizi* and *Colletotrichum truncatum* of Transgenic Soybean Expressing the *NmDef02* Plant Defensin Gene

Natacha Soto[1]*, Yuniet Hernández[1], Celia Delgado[1], Yamilka Rosabal[1], Rodobaldo Ortiz[2], Laura Valencia[1], Orlando Borrás-Hidalgo[1,3], Merardo Pujol[1] and Gil A. Enríquez[1]

[1] Soybean Biotechnology Laboratory, Plant Biotechnology Department, Center for Genetic Engineering and Biotechnology, Havana, Cuba, [2] National Institute of Agricultural Sciences, San José de las Lajas, Cuba, [3] Shandong Provincial Key Laboratory of Microbial Engineering, School of Biotechnology, Qilu University of Technology, Jinan, China

Fungal diseases lead to significant losses in soybean yields and a decline in seed quality; such is the case of the Asian soybean rust and anthracnose caused by *Phakopsora pachyrhizi* and *Colletotrichum truncatum*, respectively. Currently, the development of transgenic plants carrying antifungal defensins offers an alternative for plant protection against pathogens. This paper shows the production of transgenic soybean plants expressing the *NmDef02* defensin gene using the biolistic delivery system, in an attempt to improve resistance against diseases and reduce the need for chemicals. Transgenic lines were assessed in field conditions under the natural infections of *P. pachyrhizi* and *C. truncatum*. The constitutive expression of the *NmDef02* gene in transgenic soybean plants was shown to enhance resistance against these important plant pathogens. The quantification of the *P. pachyrhizi* biomass in infected soybean leaves revealed significant differences between transgenic lines and the non-transgenic control. In certain transgenic lines there was a strong reduction of fungal biomass, revealing a less severe disease. Integration and expression of the transgenes were confirmed by PCR, Southern blot, and qRT-PCR, where the Def1 line showed a higher relative expression of defensin. It was also found that the expression of the *NmDef02* defensin gene in plants of the Def1 line did not have a negative effect on the nodulation induced by *Bradyrhizobium japonicum*. These results indicate that transgenic soybean plants expressing the *NmDef02* defensin gene have a substantially enhanced resistance to economically important diseases, providing a sound environmental approach for decreasing yield losses and lowering the burden of chemicals in agriculture.

Keywords: *NmDef02* defensin gene, fungal resistance, *Colletotrichum truncatum*, *Phakopsora pachyrhizi*, soybean, *Bradyrhizobium japonicum*

INTRODUCTION

Although soybean [*Glycine max* (L.) Merryll] is one of the most economically important crops worldwide (Hartman et al., 2011; Rosa et al., 2015), its outstanding role in feeding the world through its contribution of both protein meal and vegetable oil, is jeopardized by the attack of fungal diseases at all growth stages, producing a considerable reduction in yields (Hartman et al., 2015).

Asian soybean rust caused by *Phakopsora pachyrhizi* (Sydow & Sydow) is the most destructive disease in soybean, causing early defoliation while affecting the weight and quality of the seeds (Hartman et al., 2005). *P. pachyrhizi* can reduce yields by over 80% when environmental conditions are favorable for disease development, and it can affect yields with a disease incidence of just 0.05% (Tremblay et al., 2012). It is found in the soybean-producing countries of South America (Yorinori et al., 2005), the United States (Schneider et al., 2005), and Mexico (Cárcamo Rodríguez et al., 2006). Asian soybean rust has also been reported in other Latin America countries (Yorinori et al., 2005) including Cuba (Pérez-Vicente et al., 2010).

Another important disease that affects soybean is anthracnose produced by *Colletotrichum truncatum*. The soybean plant can be infected at any stage of development (Yang and Hartman, 2015). It is prevalent in tropical and subtropical countries, causing severe effects on grains that lead to seedling loses (Yang and Hartman, 2015; Marmat and Ratnaparkhe, 2017). Furthermore, *C. truncatum* can systemically infect mature plants, and damages are greater under heavy rains with a high plant population (Yang and Hartman, 2015; Pawlowski and Hartman, 2016).

There are now no commercially available soybean cultivars with good agronomic characteristics that are resistant to these diseases. Moreover, modern fungicides cannot effectively control these pathogens, while increasing production costs, and having a strong negative impact on the environment through the use of chemicals (Abdallah et al., 2010; Godoy et al., 2016; Kawashima et al., 2016).

Certain Asian soybean rust resistance genes (*Rpp1-Rpp7*) have been identified in the soybean genome (Lemos et al., 2011; Yamanaka et al., 2013; King et al., 2017; Childs et al., 2018; Hossain and Yamanaka, 2019). However, these genes only confer pathotype-specific resistance, controlled by the interaction of the R genes in soybean with the virulence genes in pathotypes of *P. pachyrhizi* (Vittal et al., 2014; Yamanaka et al., 2015; Langenbach et al., 2016). The effectiveness of specific pathotype resistance genes, whether resistance is complete or incomplete, is usually short-lived, especially when it is evaluated against obligate pathogens such as *P. pachyrhizi* with a high variability and virulence (Yamanaka et al., 2013, 2016). Nevertheless, Kawashima et al. (2016) identified and cloned a gene (*CcRpp1*) from *Cajanus cajan* that confers resistance to different isolates of *P. pachyrhizi* when expressed in soybean.

To counteract fungal infection, plants develop innate immune systems that recognize the presence of pathogens and start effective defense responses (Lay and Anderson, 2005; van Loon et al., 2006). Plants produce pathogenesis-related protein-like defensins (Van Loon, 1997). Defensins are small antimicrobial peptides that play a fundamental role in the innate immunity of plants (Thomma et al., 2002, 2003; van der Weerden et al., 2013; Vriens et al., 2014; Moosa et al., 2018). Their biological activities consist of the inhibition of proteases, blocking of ionic channels, and inhibition of protein synthesis, among others (Graham et al., 2008). Defensins may inhibit the growth of a wide range of microorganisms and phytopathogenic insects; they may also be involved in abiotic stress adaptation (Tavares et al., 2008). This means that not only do defensins produce a defense against

plant pathogens, but they also generate adaptations to difficult conditions, a characteristic that makes them even more attractive for modern agriculture. The structure stabilized by disulfide bridges and cationic charge presented by defensins makes them very stable molecules, which is essential for the development of biotechnological products based on them (Tavares et al., 2008).

Growth inhibition by plant defensins of a wide range of pathogenic fungi is not associated with toxicity in mammalian or plant cells (Thomma et al., 2002). Studies of the biological activity, stability, and range of toxicity of an isolated chickpea defensin (Ca-AFP) revealed that there are no risks in using the gene in the production of transgenic crops (Islam, 2008). However, to date, three defensin-related proteins have been described as allergens (Singh et al., 2006; Petersen et al., 2015). It is therefore essential to assess the allergenicity and toxicity of genetically modified crops carrying defensin before they become a product for human and animal use. Consequently, plant defensins can be used to produce transgenic crops with improved resistance to pathogens.

Several genes encoding defensins have been successfully transferred to important plant species such as tobacco (Portieles et al., 2010; Lee et al., 2018), tomato (Abdallah et al., 2010), potato (Gao et al., 2000; Portieles et al., 2010; Kumar and Chakrabarti, 2018), rice (Kanzaki et al., 2002; Jha and Chattoo, 2010), and beans (Espinosa-Huerta et al., 2013), among others, producing resistance to different pathogens. It has been demonstrated that the expression of the *NmDef02* defensin gene in tobacco and potato transgenic plants produced a strong resistance against *Phytophthora infestans* under greenhouse and field conditions (Portieles et al., 2010).

Therefore, our objective in this study was to determine whether transgenic soybean plants expressing the *NmDef02* defensin gene are better equipped to overcome infection by *P. pachyrhizi* and *C. truncatum* under field conditions. The efficiency of the symbiosis of *Bradyrhizobium japonicum* with transgenic soybean plants carrying defensin was also evaluated, since the association with this bacterium is essential for atmospheric nitrogen fixation in soybean plants, eliminating the need for chemical nitrogen fertilization. This strategy is in line with the goals of decreasing yield losses, decreasing the use of chemicals, and contributing to an increase of 100% in yields that are required for sustaining a world population of nearly 10 billion people in 2050, all of which are challenges acknowledged by Next Generation Agriculture.

EXPERIMENTAL PROCEDURES

Plant Material

The soybean [*Glycine max* (L.)] was of the variety DT-84 from Vietnam. Embryonic axes of mature seeds were used as explants for their bombardment-mediated transformation, according to Soto et al. (2017).

Vector Construction

The pCP4EPSPS-DEF vector carrying the *cp4epsps* gene and the *NmDef02* defensin gene isolated from *Nicotiana megalosiphon*

was the vector system used for the transformation. The cassette with p35S/*NmDef02*/*tnos* obtained by Portieles et al. (2010) was cloned into the pCP4EPSPS binary vector (Soto et al., 2017) to generate the pCP4EPSPS-DEF vector (**Figure 1A**). This was done at the soybean biotechnology laboratory at CIGB, Havana.

Transformation, Selection, and Plant Regeneration

A total of 150 explants were bombarded with the pCP4EPSPS-DEF vector and selected in a MSB5 medium with 20 mg/L of glyphosate. The controls used were explants derived from cultivar DT-84, which were cultured under the same conditions and without selective agent. The regenerated shoots were excised and transferred to the same medium without selecting for rooting, as described by Soto et al. (2017). Plantlets were transferred to pots containing a mixture of organic material and zeolite (50/50) in an acclimatized greenhouse at 26–27°C to produce seeds. All seeds collected from each of the R0 generation transgenic lines were germinated under greenhouse conditions to obtain the T_1 generation and the following generations.

Analysis of the Integration of the *NmDef02* Gene Using Polymerase Chain Reaction (PCR)

Total genomic DNA was isolated from young leaves of glyphosate-resistant and control plants using the CTAB protocol (Doyle and Doyle, 1987). PCR was used to screen for transformants (T_1) carrying the *NmDef02* gene. Each reaction was performed in a total volume of 25 μl, and the PCR mixture consisted of 10 mM buffer Go Taq Green 5x, 10 mM dNTP; 20 pmol/μl of each primer, 1 unit of GoTaq DNA polymerase (Promega, United States), and 400 ng of the genomic DNA. The primers used were forward 5′-GCTGGCTTATGCTTCCTCTTCTTG-3′ and reverse 5′-TCACAGACTTGGACGCAGTTCG-3′. The reaction started with an initial denaturing step at 95°C for 3 min, followed by 30 cycles of the following profile: denaturing at 95°C for 1 min, annealing at 64°C for 1 min, synthesis at 72°C for 1 min followed by an extension at 72°C for 10 min. The PCR products were loaded onto 2% electrophoresis agarose gel and visualized using ethidium bromide.

Relative Expression of the *NmDef02* Gene Using qRT-PCR

Total RNA for the qPCR analysis was extracted from frozen leaf tissues of six soybean transgenic lines and non-transgenic plants using Tri-Reagent (Sigma-Aldrich, United States) according to the manufacturer's protocol. The RNA was sequentially treated with DNase I (Promega, United States) at 37°C for 15 min to remove the remaining genomic DNA. The integrity and yield of RNA were evaluated using agarose gel electrophoresis and a NanoDrop Spectrophotometer (Thermo Scientific), respectively. The cDNA was synthesized from 1 μg of total RNA using an oligo-(dT) primer and the Super-Script III reverse transcriptase kit (Invitrogen, United States) according to the manufacturer's instructions. qPCR reactions were carried out in a final volume of 15 μl containing 0.2 mM of each primer, 10 μl SYBR (QuantiTect SYBR Green PCR kit; Qiagen, Germany), and a dilution of the 25x cDNA. The soybean β-*actin* gene was selected as the housekeeping gene and used for normalizing the data. The primer sequences to amplify the *NmDef02* gene were forward 5′-AAGCTTATGCGTGAGTGCAAGGCTC-3′ and reverse 5′-CTGCAGTTAGCACTCGAATATAC-3′. The primer sequences from β-actin gene were forward 5′-GTGTCAGCCATACTGTCCCCATTT-3′ and reverse 5′-GTTTCAAGCTCTTGCTCGTAATCA-3′. The amplification conditions included: an initial 95°C denaturation step for 15 min, followed by denaturation for 15 s at 95°C, annealing for 30 s at 60°C, and extension for 30 s at 72°C for 40 cycles. Quantitative PCR was conducted using a Rotor-Gene 3000 PCR machine (Corbett, Sydney, NSW, Australia). The efficiency of the primers was determined by using serial dilutions of a mixture of different cDNAs (from each sample) with concentrations of 5x, 25x, 125x, and 625x. Further analysis of the dissociation temperature of the PCR products was performed to determine their specificity. The dissociation analysis and the Ct values were used by the Rotor-Gene equipment program (version 6.1) to determine the efficiency of the qPCR reactions.

The q-gene method was used to obtain the relative expression of the qPCR values, and they were analyzed with the Q-Gene 96 program (Muller et al., 2002). The results represent the mean of three biological and technical replicates on each transgenic line and the non-transgenic control. The amplified products were sequenced to verify their identity.

Southern Blot Analysis

Southern blot and hybridization were performed by following the protocol described by Sambrook et al. (1989). Genomic DNA (15 μg) from soybean plants (T4 generation) selected with glyphosate and evaluated in the field was digested with *Eco*RV. The digested DNA was electrophoresed on a 0.8% agarose gel and blotted onto a nylon membrane (Hybond N, Amersham Biosciences). Hybridization was carried out with a-[32P]-dATP-labeled *cp4epsps* gene as the probe, using the DNA random primer labeling kit (Promega, United States). The probe was obtained by PCR with *cp4epsps*-gene-specific primers to generate the 887 bp fragment. It was isolated from a 1% agarose gel and purified using the SV Gel Wizard Clean-Up System (Promega, United States).

Symbiosis of *Bradyrhizobium japonicum* With Transgenic Soybean Plants Expressing the *NmDef02* Defensin

Because of the essential role of *B. japonicum* in atmospheric nitrogen fixation in soybean, we determined the efficiency of its symbiosis with transgenic soybean plants carrying defensin. The test was performed with 30 transgenic plant seeds and 20 non-transgenic seeds, which were used as the control. The Semia 5080 strain of *B. japonicum* was used for inoculations. Seeds were planted in pots with zeolite and placed in plastic trays with water to maintain humidity within a greenhouse. A week after seed germination, seedlings were inoculated with 1 mL of the
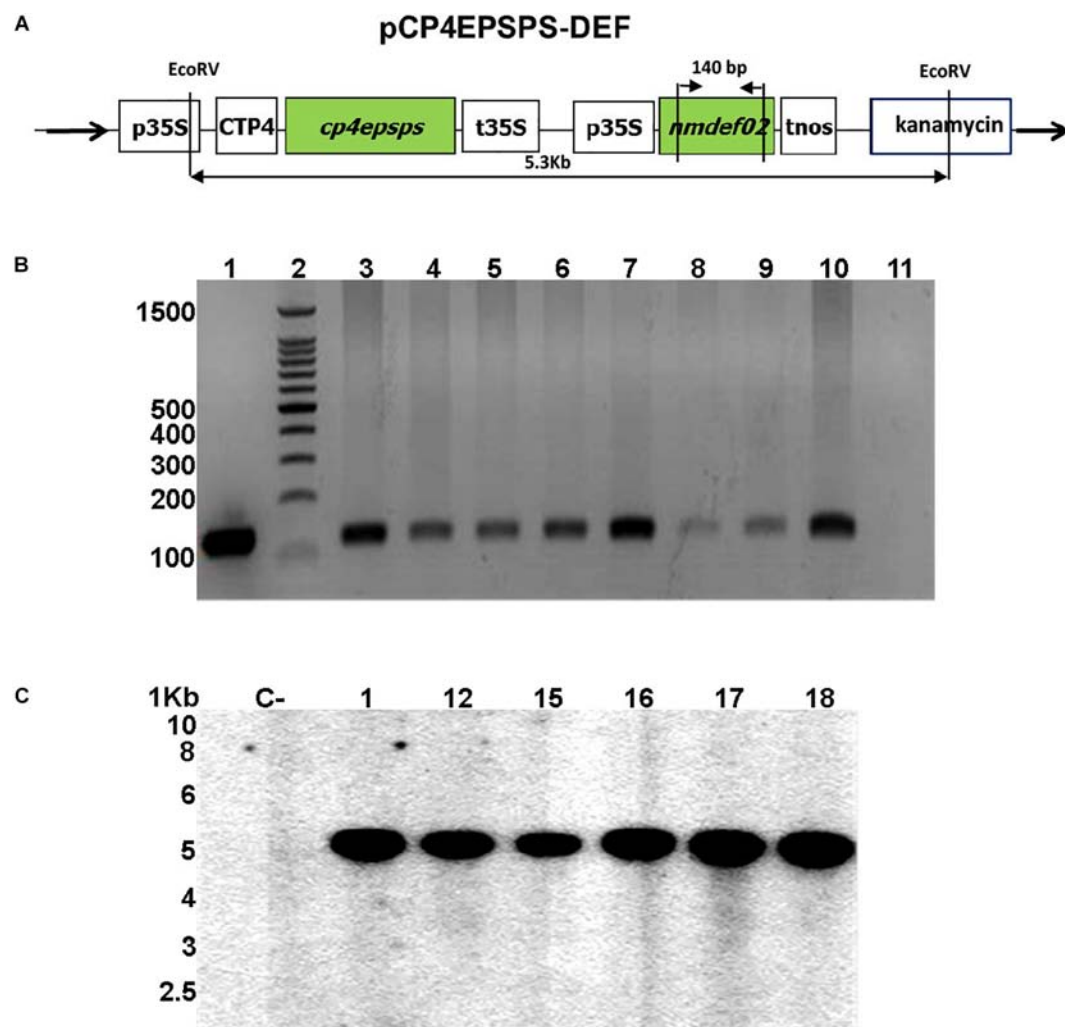
FIGURE 1 | Molecular analyses of transgenic soybean lines. **(A)** Schematic map of the pCP4EPSPS-DEF plasmid used for soybean transformation. The cassette contains the p35S: 35S promoter of the cauliflower mosaic virus, CTP: chloroplast transport peptide, *cp4epsps*: gene that encodes the CP4EPSPS protein, t35S: 35S terminator, *nmdef02*: defensin gene *NmDef02*, tnos: terminator of the nopaline synthase, and two restriction sites of the *Eco*RV enzyme (5.3 Kb). **(B)** PCR reaction products from genomic DNA of soybean plants transformed with the plasmid pCP4EPSPS-DEF. Amplification of a 140 bp sequence corresponding to the *NmDef02* gene. Lane 1: Positive control band of plasmid pCP4EPSPS-DEF; lane 2: 100 bp molecular weight marker (Promega); lanes 3–10: transgenic soybean lines; lane 11: non-transgenic plant (NT). **(C)** Genomic Southern blot analysis of transgenic soybean plants carrying the *Nmdef02* gene. Lane C: non-transgenic control. Lanes 1–18: selected transgenic lines.

diluted bacterial culture at 2×10^6 viable cells/pot. Uninoculated transformed and non-transformed seedlings were used for the control of the assay. The plants were collected during flowering and the following symbiotic efficiency indicators were quantified: number of nodules per plant, fresh weight of nodules per plant (g), fresh leaf weight per plant (g), and dry leaf weight per plant (g). There were three replicates of the experiment.

Phakopsora pachyrhizi Field Trials

Thirteen transgenic lines obtained by self-pollination (T_3) and non-transgenic plants (variety DT-84) were grown on an experimental area in Havana during the winter (November–March). The field experiment was authorized by the National Center for Biological Safety of Cuba with the license: LH47-L

(95) 13. Seeds were inoculated with *B. japonicum* and planted in a field near soybean plants affected by *P. pachyrhizi*. A randomized block design was used, with three blocks/line and 450 seeds/line. Plants were not treated with fungicidal products. The experiment was assessed daily. After the outbreak of rust symptoms, affected leaves were analyzed using an Envirologix QuickStix kit (Envirologix, United States) to confirm the presence of *P. pachyrhizi*.

The incidence of *P. pachyrhizi* was calculated by dividing the number of plants showing symptoms by the total number of plants in the experiment and multiplying the resulting value by 100. When the first symptoms of Asian soybean rust appeared, the severity of the disease (% of the area of the leaf affected by rust) was calculated according to the protocol proposed by

Ploper et al. (2006), through which the central folioles of the lower, middle, and upper parts of the plants were sampled. The following scale was used to calculate severity: Grade 1 (0%); Grade 1.5 (0.6–1%); Grade 2 (1–5%); Grade 3 (6–25%); Grade 4 (26–50%); Grade 5 (>50%). The second evaluation took place 10 days after the outbreak, and 20–30 plants were analyzed. The percentage of defoliated plants was calculated at 36 and 60 days after the start of symptoms. Plants were harvested and the following morpho-agronomic parameters of 30 plants for each line were evaluated: height of the plant (cm), height of the 1st pod (cm), number of branches, number of pods, number of seeds, and weight of seeds/plant (g).

Phakopsora pachyrhizi Biomass

Quantitative PCR was used to measure the fungal biomass of *P. pachyrhizi* in leaves as described by Lamour et al. (2006). Two folioles (from the upper and lower parts of the plants) were collected from each transgenic and non-transgenic plant 10 days after Asian soybean rust symptoms were observed. Thirty plants from each transgenic line and the non-transgenic control were used in this analysis. The collected plant material was frozen at −80°C. The leaves from each line and from the control were pooled separately, macerated in liquid nitrogen, and homogenized to use 1 g of tissue.

Genomic DNA was extracted from frozen leaf tissues of transgenic and non-transgenic plants using a modified CTAB protocol (Doyle and Doyle, 1987). The integrity and yield of DNA were evaluated using agarose gel electrophoresis and a NanoDrop Spectrophotometer (Thermo Scientific), respectively. The qPCR reactions were carried out in a final volume of 20 µl containing 200 ng DNA, 0.8 mM of each primer, and 10 µl SYBR (QuantiTect SYBR Green PCR kit; Qiagen, Germany) by using a Rotor-Gene 3000 PCR machine (Corbett, Sydney, NSW, Australia). The specific primers to amplify an ITS sequence of *P. pachyrhizi* were Ppm1 5′-GCAGAATTCAGTGAATCATCAAG-3′ forward and Ppa4 5′-TCAAAATCCAACAATTTCCC-3′ reverse (Frederick, 2006). The amplification conditions used included: an initial 95°C denaturation step for 15 min, followed by denaturation for 15 s at 95°C, annealing for 30 s at 50°C, and extension for 30 s at 72°C for 40 cycles.

For the quantification of biomass, a standard curve (1/10, 1/100, 1/1000) was made with DNA isolated from pustules of the fungus of highly infested plants. Data were analyzed in Rotor-Gene 3000 software (Corbett). The amplified products were sequenced to verify their identity.

Colletotrichum truncatum Field Trials

Three transgenic lines obtained by self-pollination (T_4) that were selected for resistance to Asian soybean rust and non-transgenic plants (susceptible variety DT-84) were grown on an experimental area of the National Institute of Agricultural Sciences (INCA), Mayabeque province, during the winter (November–March). Seeds were inoculated with *B. japonicum* and planted in soil with a history of a high incidence of anthracnose caused by *C. truncatum*. A total of 360 seeds from each line and 200 seeds from the non-transgenic plants were used

in this study. A randomized block design was used, with three blocks per line. Plants were not treated with fungicidal products, and they were evaluated weekly. After the outbreak of symptoms, infected pod samples were collected and the fungus was isolated and identified (Chen et al., 2006). The incidence of *C. truncatum* in the experiment was calculated by dividing the number of plants with symptoms by the total number of plants in the experiment and multiplying the result by 100. Plants were harvested and the morpho-agronomic parameters of 30 plants of each line were evaluated, specifically the height of the plant (cm), height of the 1st pod (cm), number of branches, number of pods, number of seeds, and weight of seeds/plant (g). Soybean plants transformed with the pCP4EPSPS-DEF (**Figure 1A**) and pCP4EPSPS (Soto et al., 2017) plasmids were also grown in disease-free soil, using non-transgenic plants (DT-84) as the control. Plants were harvested and the morpho-agronomic parameters of 30 plants were evaluated.

Statistical Analysis

Data were statistically analyzed by IBM SPSS Statistics 25 using ANOVA at the $P \leq 0.05$ level. The means of the experimental replicates were plotted, and the standard deviations are shown as error bars.

RESULTS

Transformation and Plant Regeneration

Particle acceleration-mediated transformation was carried out using a pCP4EPSPS-DEF vector carrying the glyphosate resistance gene and the *NmDef02* defensin gene under the control of the cauliflower mosaic virus 35S promoter (**Figure 1A**). The first herbicide-resistant shoots from explants via direct organogenesis were observed after 15 days in the selection medium. Data obtained in the transformation experiment showed that 19 out of 150 bombarded explants developed shoots in the selection medium with glyphosate. All transgenic lines showed similar growth to the non-transformed control and were transferred to greenhouse conditions until the T_2 seeds were harvested. After having developed their second trifoliate leaf, the plants were sprayed with a concentration of 360 g/L of glyphosate for resistant plant selection. In addition, the expression of the CP4 EPSPS protein was demonstrated in 22 rooted lines (T_0) using the Roundup Ready immunodetection kit.

Integration of Transgene in Soybean Plants

In order to analyze the stability of transgene integration in the T_1 generation, DNA of glyphosate-resistant lines underwent PCR analysis. This analysis detected the presence of the expected 140 bp fragment (**Figure 1B**), indicating the presence of the *NmDef02* gene in the transgenic soybean plants, while it was not detected in non-transformed plants.

Six transgenic lines selected in glyphosate and showing resistance in the field were screened by Southern blot analysis in generation T_3. Signals corresponding to the region of the

plasmid between the two sites that were recognized by the *Eco*RV enzyme were detected (**Figure 1A**). The signals showed a stable integration of the segment of the plasmid containing the *cp4epsps* and *NmDef02* genes in the genome of the transformed plants of a size of 5.3 Kb (**Figure 1C**). DNA isolated from non-transformed plants did not show any hybridization signal (**Figure 1C**).

Relative Expression of the *NmDef02* Defensin Gene in Transgenic Plants

The relative expression of the *NmDef02* gene in six transgenic soybean lines was evaluated by quantitative RT-PCR (**Figure 2**). The transgenic lines differed in defensin expression level. Although lines Def1, 17, and 18 showed significant difference ($p \leq 0.001$) compared to the non-transgenic control, Def1 showed the highest accumulation of defensin NmDef02.

Efficient Nodulation by *Bradyrhizobium japonicum* in Soybean Plants That Express the *NmDef02* Defensin

This test was carried out to evaluate the efficiency of the symbiosis of this bacterium with transgenic soybean plants carrying defensin. Taking into account that the Def1 transgenic line showed fungal resistance in the experiments performed under natural infection conditions, nodulation in these plants when inoculated with *B. japonicum* was evaluated. In this experiment, the transgenic plants showed a similar phenotypic development to the non-transformed plants used as a control, which was observed in the growth of the stem and color of the leaves, as well as in flowering under greenhouse conditions (**Figure 3A**). Inoculation produced nodulation in all transgenic plants and the control (**Figure 3B**). Nodes of different sizes were obtained in all of the inoculated plants (**Figure 3C**), and the highest number of

FIGURE 3 | Symbiosis of *Bradyrhizobium japonicum* with soybean plants carrying the *NmDef02* defense gene. **(A)** Transgenic plants in pots with zeolite and inoculated with *B. japonicum*. **(B)** Transgenic plant with nodules at the base of the stem, induced by inoculation (left) and plants without inoculation (right). **(C)** Nodules produced in inoculated transgenic plants. **(D)** Nodular mass with an internal red coloration indicating the production of the leg hemoglobin required for the fixation of atmospheric nitrogen.

nodules was observed in the plants harboring defensin, but this did not lead to marked differences in the average fresh weight of the nodules (**Table 1**). In all cases, the nodules were widely distributed in the root neck region (**Figure 3B**) and showed an internal red coloring (**Figure 3D**) due to the presence of the leg-hemoglobin protein. The number of functional nodules with a red coloring in the plants (**Table 1**) is an indirect indicator of the occurrence of the process of atmospheric nitrogen fixation by the bacteroide. The inoculated plants maintained an intense green color during the experiment and showed higher values of fresh weight and dry weight of the aerial parts than the non-inoculated plants. It is thus shown that the *NmDef02* defensin does not affect *B. japonicum* nodulation.

Field Resistance to *Phakopsora pachyrhizi* in Transgenic Lines

After the beginning of Asian soybean rust symptoms in the pod and grain formation stages, the presence of *P. pachyrhizi* was confirmed by the Envirologix QuickStix kit (**Figure 4A**), and reproductive structures were observed by a stereoscope (**Figures 4B,C**). Non-transgenic plants (DT-84) used as the control had a 100% incidence of rust (percentage of plants with pustules), demonstrating the high susceptibility of this cultivar to *P. pachyrhizi* (**Figure 4D**). In parallel, all transgenic lines showed signs of rust; in this case, the pustules were present in the leaves that were closest to the soil, although the percentage of affected plants in transgenic lines was lower than in non-transformed plants (**Figure 4D**). In this study, some transgenic plants presented a high incidence (**Figure 4D**) and low severity (**Figure 4E**) of Asian soybean rust. This was evident in line 18. This is so because, in this study, the incidence only showed the
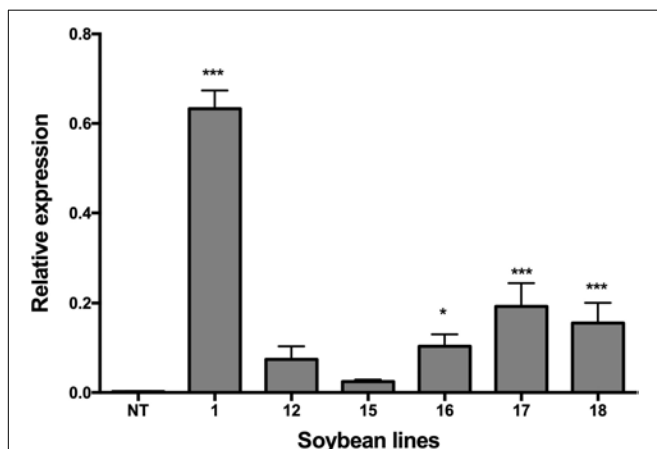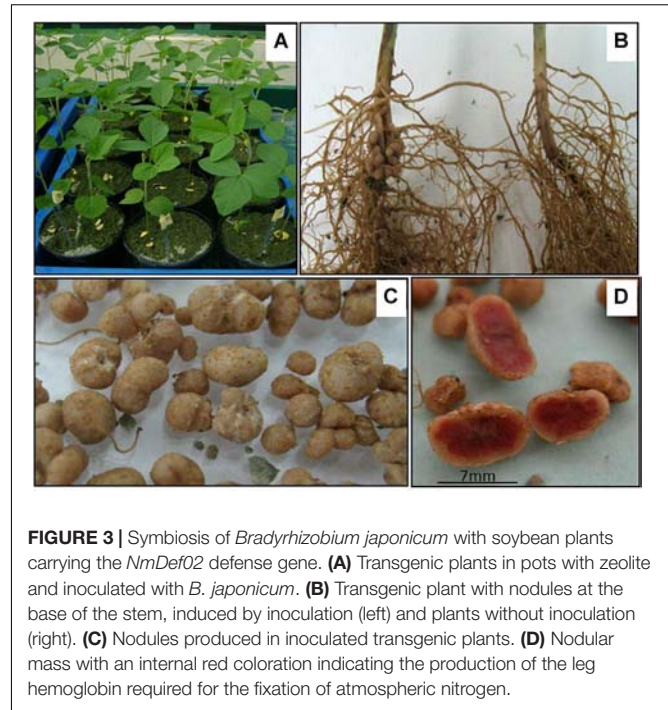

FIGURE 2 | qRT-PCR analyses of transgenic lines showing the transcript levels of the *NmDef02* gene in transgenic plants, compared to the constitutive expression of the endogenous β-*actin* of soybean. 1–18: Transgenic lines (Def1, 12, 15, 16, 17, and 18). NT: non-transgenic control. Bars represent mean ($n > 9$) and standard error of the results obtained. Asterisks indicate significant differences using ANOVA by Tukey's multiple range test with respect to the control (*$p \leq 0.05$, ***$p \leq 0.001$).

TABLE 1 | Nodulation in transgenic and non-transgenic soy plants (cultivate DT-84) inoculated with *Bradyrhizobium japonicum*.

Treatment	Number of nodules	Weight of fresh nodules (g)	Fresh weight of the aerial parts (g)	Dry weight of the aerial parts (g)
NT	0 ± 0	0 ± 0	5.5 ± 0.9^b	1.8 ± 0.6^b
NT + *B. japonicum*	8.9 ± 5.8^a	0.2 ± 0.1	8.2 ± 1.5^a	2.2 ± 0.5^a
Def 1	0 ± 0	0 ± 0	6.3 ± 1.1^b	1.9 ± 0.4^b
Def 1 + *B. japonicum*	12.4 ± 7^a	0.2 ± 0.1	7.8 ± 1.7^a	2.2 ± 0.5^a

Def1 (transgenic line) n = 30; NT (non-transgenic control DT-84) n = 20. The values represent the means ± standard deviation corresponding to three replicates. Different letters show significant differences of p < 0.05 according to Tukey's multiple range test.

dispersion of fungi in the experimental area. All plants with any symptoms at all were counted, even those with few pustules at the lowest part of the plant, as occurred in lines Def1, 12, 17, and 18. The severity of soybean leaf rust was estimated through the visual observation of plants using the scale proposed by Ploper et al. (2006). The results are shown in **Figure 4E**. According to this study, severity was higher in the lower parts of the plants, mainly in older trifoliate leaves. All transgenic lines had a significantly lower severity than the control DT-84 ($p < 0.05$), where more than 40% of the leaves involved were from the lower parts of the plant (**Figure 4E**). Transgenic lines Def1, 12, 16, 17, and 18 showed severe damage by rust; lines Def1 and Def12 had less than 8% of the leaves affected in the lower plant parts, with less than 5% in the middle and upper parts (**Figure 4E**). Resistance to *P. pachyrhizi* was determined by visual description of signs and symptoms observed on the soybean leaves in response to infection by Asian soybean rust. The presence of reddish-brown (RB) pustules with different levels of sporulation and without sporulation was observed in the leaves of these transgenic lines. Contrarily, in non-transgenic plants, abundant sporulation in uredinias was observed, which shows the high susceptibility of the DT-84 cultivar to *P. pachyrhizi*. The presence of dark brown pustules with limited sporulation was also evident in transgenic and non-transgenic plants. In our study, transgenic lines expressing the *NmDef02* gene showed different levels of resistance to *P. pachyrhizi*, displaying complete and incomplete resistance in lines Def1, 12, 17, and 18. Line Def12 also showed light brown lesions with reduced sporulation in some plants, suggesting a partial resistance. Some plants of the Def3, 4, 6, 10, and 14 lines also showed light brown lesions with reduced sporulation.

Quantitative PCR analysis confirmed the presence of *P. pachyrhizi*. The quantification of fungal biomass in plants made it possible to verify that all transgenic lines had a significantly lower amount of fungal biomass than the non-transgenic plants ($p < 0.05$), as shown in **Figure 5**. Transgenic lines Def1, 17, and 18 revealed the lowest amount of fungal biomass compared to the other transgenic lines and the non-transgenic control (**Figure 5**). The differences in fungal colonization between the transgenic soybean lines were expressed through the quantification of the fungal biomass, even when the visual differences in the signs and symptoms were not evident, as in the cases of lines Def3, 4, 6, and 14.

Leaves of transgenic plants remained green even when they were affected by *P. pachyrhizi* (**Figure 6A**), thus contrasting with non-transgenic plants, which showed intense chlorosis in their leaves (**Figure 6A**) followed by premature defoliation (**Figure 6B**). The highest percentage of defoliation was observed in non-transgenic plants at 60 days after rust infection. At that time, transgenic lines Def1, 5, 12, 17, and 18 showed defoliation of less than 10% (**Figure 6C**).

The incidence of rust at this stage caused early defoliation, which had a negative impact on all parameters. The disease affected the number of branches, pods and seeds, as well as the weight of seeds per plant in the non-transgenic control. In contrast, most of the transgenic lines were significantly superior to the control in all parameters related to yield. The results are summarized in **Table 2**.

Resistance to *Colletotrichum truncatum* in Transgenic Lines

A total of 360 transgenic plants carrying the *NmDef02* defensin gene and 200 non-transgenic plants (DT-84) representing the controls were used in the experiment under conditions favoring the incidence of Anthracnose. Symptoms of irregular brown-shaped spots on pods, petioles and stems, similar to those described for anthracnose of soybeans, were observed in non-transgenic plants in the grain formation phase (**Figure 7A**). Tissues from the affected pods and stems were observed under an optical microscope to examine the structure of the fungus. The presence of *C. truncatum* was confirmed by PCR (data not shown).

In this study, line Def1 showed a high resistance to this pathogen (**Figure 7B**) because only some plants (10%) (**Figure 7C**) presented pods with spots at the basal zone of the plant at the end of the cycle. The Def12 and Def16 transgenic lines also showed symptoms of anthracnose, but their incidence was less than in the non-transgenic control (**Figure 7C**). These transgenic lines showed irregular spots on the pods of some plants, but their leaves remained green until the final stage of plant development, as occurred in line Def1 (**Figure 7B**). In contrast, 100% of DT-84 plants were affected (**Figure 7C**) and showed irregular brown spots on pods, as well as leaf chlorosis and high premature defoliation (**Figure 7D**). On the other hand, the Def12 and Def16 transgenic lines showed a certain reduction in seed quality in some plants. The seeds of line Def1, however, remained healthy (**Figure 7E**). Seeds and pods of the non-transgenic control were highly affected by the fungus (**Figure 7E**), showing wrinkling, mold and, in some cases, turning dark brown, which is similar to the symptoms reported for anthracnose associated with Colletotrichum in soybean. After plants were
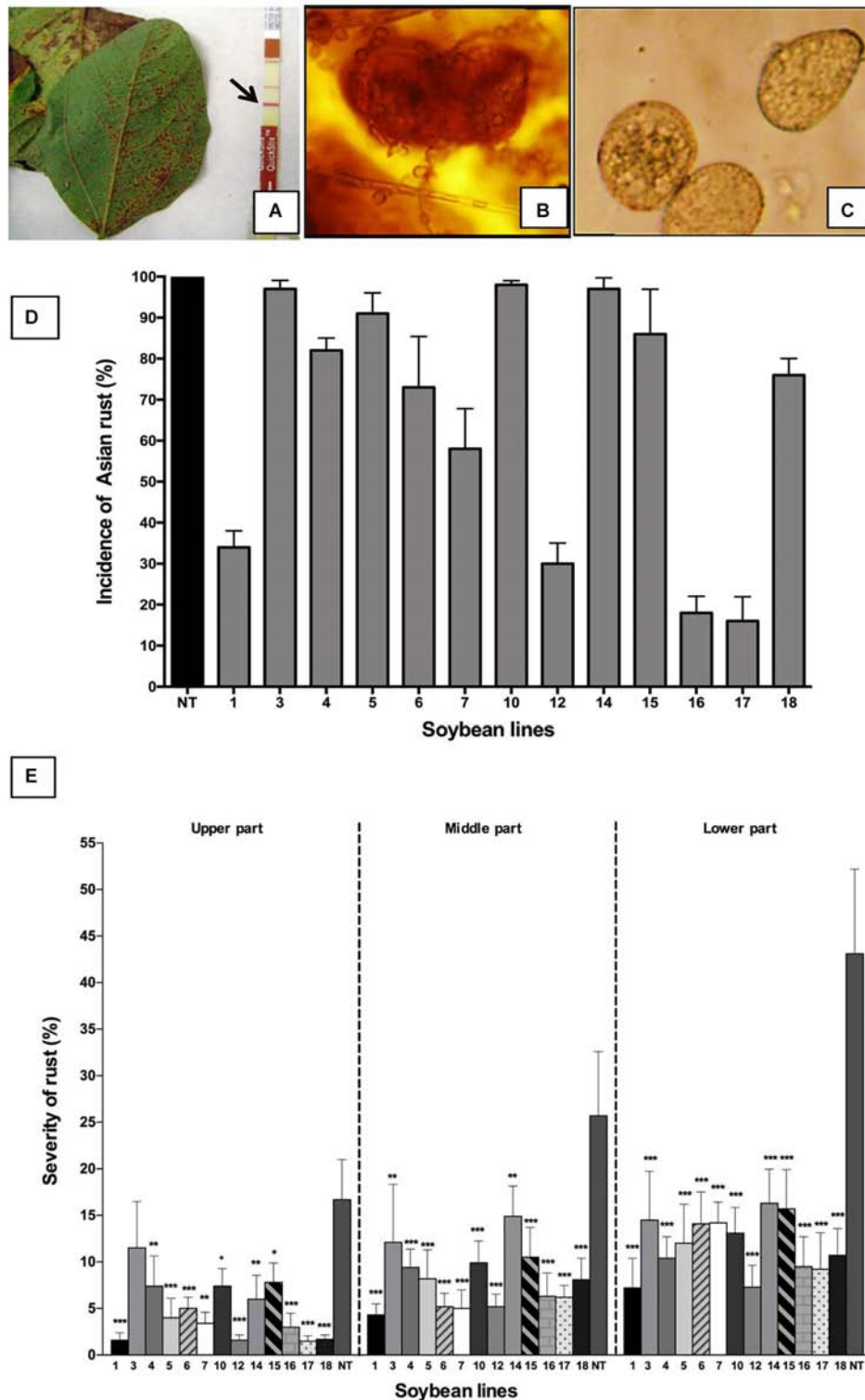
FIGURE 4 | Impact of Asian soybean rust on soybean plants. **(A)** Symptoms of soybean rust in soybean plants. Detection of *P. pachyrhizi* using the Envirologix QuickStix immunodetection strips. Arrow indicates a positive line. **(B)** Urediosoros with uredospore visualized under an optical microscope (20X). **(C)** Uredospore observed under an optical microscope (40X). **(D)** Incidence of soybean rust in soybean plants. The percentages of plants affected by the *P. pachyrhizi* fungus in the transgenic lines (1–18) and in the non-transgenic control (NT) are shown. Bars represent the deviation of the means (n = 3). **(E)** Evaluation of soybean rust severity in transgenic soybean plants. The data show the average of two experiments. Severity was determined in different parts of the plant (upper, middle, and lower parts) (n = 20). Asterisks indicate significant differences using ANOVA by Tukey's multiple range test compared to the control (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$).
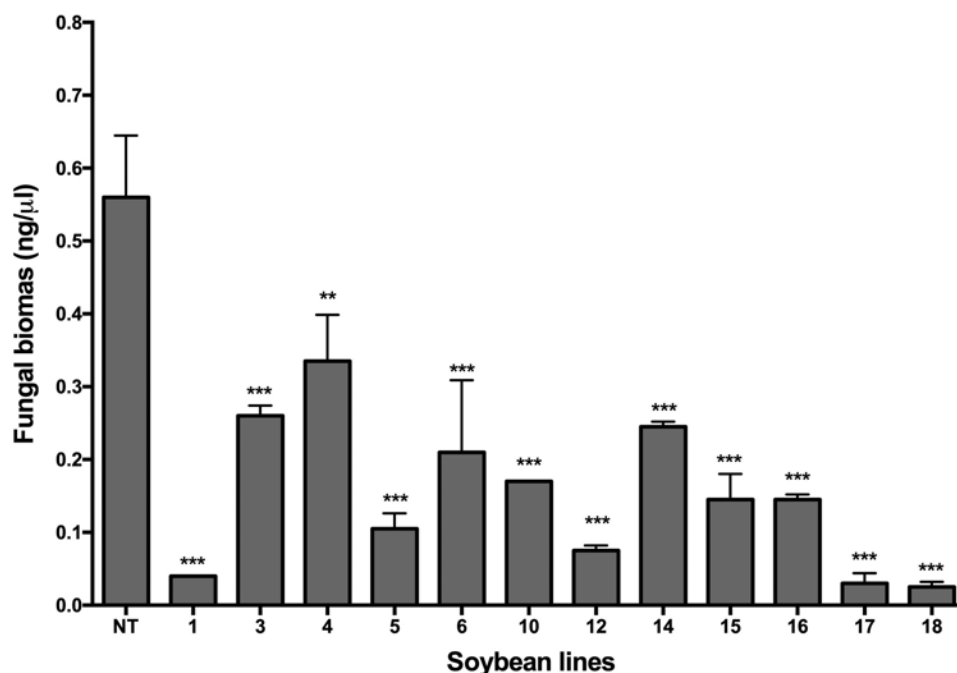
FIGURE 5 | Analysis by qPCR for the quantification of *P. pachyrhizi* DNA in transgenic soybean plants. 1–18: Transgenic lines, and NT: non-transgenic control, DT-84. Bars represent the deviation of the means (*n* = 3). Asterisks indicate significant differences using ANOVA by Tukey's multiple range test compared to the control (**$p \leq 0.01$, ***$p \leq 0.001$).

harvested, the results showed a statistically better performance for line Def1 compared to lines Def12 and Def16 and the non-transgenic control in all parameters evaluated (**Table 3**).

Transgenic soybean lines did not present any detrimental agronomic features compared to the non-transgenic control DT-84 when plants were grown in disease-free soil (**Table 4**). All soybean plants showed a similar vegetative development. The parameters evaluated show differential behavior, both in transgenic plants carrying the *NmDef02* gene and the *cp4epsps* gene and in non-transgenic plants, as shown in **Table 4**. These results demonstrate that the overexpression of defensin did not have a negative impact on parameters related to yield in soybean plants.

DISCUSSION

The production of transgenic plants expressing antimicrobial genes is able to provide broad resistance against different pathogens while reducing the use of chemical pesticides. In the current study, we obtained the first evidence of resistance to the hemibiotrophic fungus *C. truncatum* and the biotrophic fungus *P. pachyrhizi* in soybean plants transformed with the *NmDef02* defensin gene under the 35S constitutive promoter.

Defensin has antifungal activity and produces membrane disruption by pore formation in the cell membrane (Thevissen et al., 2007). A high concentration of defensins produces severe membrane permeabilization, which leads to fungal death (Thevissen et al., 2003; Seo et al., 2014). Previous studies have

shown that plant defensins are accumulated in the peripheral cell layers of cotyledons, hypocotyls, endosperm, tubers, fruits, and floral organs, including style, ovary, filaments of stamen, and anthers (Thevissen et al., 2003). These defensin locations are consistent with their role in the first line of defense against potential pathogens (Thevissen et al., 2003; Lay and Anderson, 2005; De Coninck et al., 2013). Plant defensins can also be found in stomatal cells and cell walls of the sub-stomatic cavity; these are involved in plant protection against pathogens that penetrate the stomata (Prema and Pruthvi, 2012).

In this study, we have observed high protection against the pathogenic fungi *P. pachyrhizi* and *C. truncatum* in soybean leaves and pods, which may be favored by the constitutive overexpression of *NmDef02* defensin in the membranes. Previously, Portieles et al. (2010) showed that the constitutive expression of the *NmDef02* defensin gene provided strong resistance to *P. infestans* in transgenic potato plants under greenhouse and field conditions.

The inhibitory activity of defensins on the growth of a wide range of hemibiotrophic and necrotrophic fungi has been observed through *in vitro* studies at micromolar concentrations (Portieles et al., 2010; Lacerda et al., 2016). Nevertheless, studies to determine the antifungal activity of plant defensins against biotrophic fungi are much more difficult, since they are difficult to cultivate *in vitro* according to Kaur et al. (2011). These fungi establish a long-term feeding relationship with the living host cells. Studies recently published by Lacerda et al. (2016), reported that the Drr230a defensin expressed in yeast affected the *in vitro* germination of the spores of the *P. pachyrhizi* fungus.
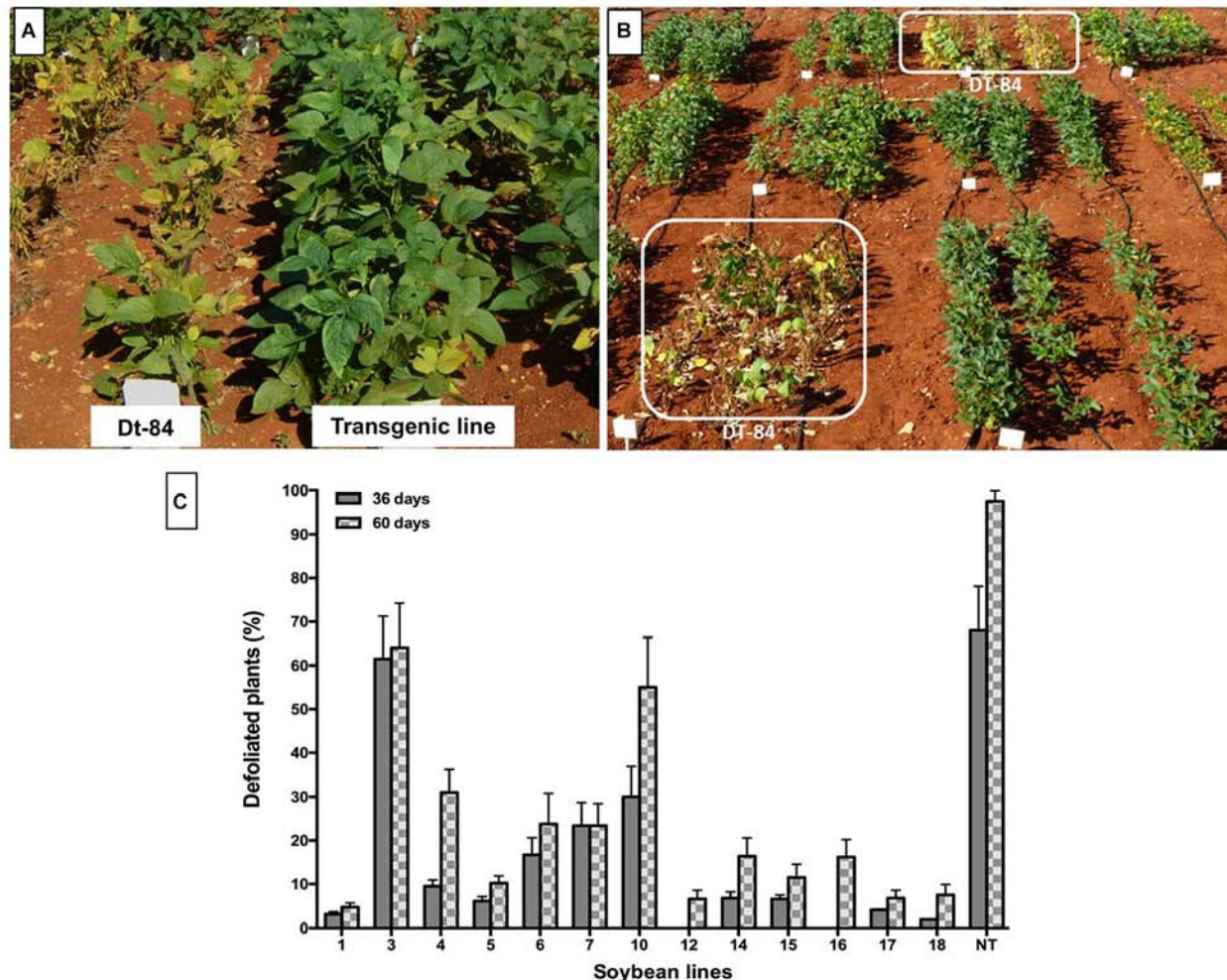
FIGURE 6 | Soybean plants growing in the field and naturally infested with *Phakopsora pachyrhizi* during winter. **(A)** Symptoms of chlorosis by Asian soybean rust in non-transgenic plants (DT-84). **(B)** Symptoms of defoliation in non-transgenic plants due to early maturation. **(C)** Defoliation in soybean plants due to the incidence of Asian soybean rust. Evaluation at 36 and 60 days after the appearance of soybean rust symptoms; 1–18: Transgenic lines, NT: non-transgenic control. Bars represent the deviation of the means (*n* = 3).

They observed less severity of rust caused by *P. pachyrhizi* in leaves that were artificially inoculated with the fungus and the defensin. Also, the results obtained in our study showed that the transgenic lines expressing the *NmDef02* defensin gene are able to inhibit the development and sporulation of the *P. pachyrhizi* fungus under natural infection conditions. Similar results were reported with the *CcRpp1* gene isolate from *C. cajan* and cloned in soybean, where it conferred specific resistance to *P. pachyrhizi* (Kawashima et al., 2016).

The complex interactions occurring between the pathogen, its host, and the environment are expressed as the incidence or severity of a disease. In this study, the average severity of rust on older leaves (in the lower third of the plants) was statistically higher than on younger leaves (in the upper part of the plants). Similar results were reported by Xavier et al. (2017), who showed that younger soybean plants are more susceptible to Asian soybean rust than older plants but that older trifoliate leaves

had the highest disease severity. Although all transgenic lines expressed signs of Asian soybean rust, they were of low severity compared with the non-transgenic control. Studies by Vittal et al. (2014) showed that the presence of reddish-brown pustules (RB) without sporulation indicated complete resistance and that RB lesions with different levels of sporulation meant that there was an incomplete resistance. Both types of lesions were found in the transgenic plants that remained with green leaves, unlike the non-transgenic plants, which showed chlorosis accompanied by the appearance of pustules. On the other hand, the presence of dark brown pustules with limited sporulation in certain transgenic plants shows partial resistance, as described by Pham et al. (2009). Our data support the idea that the constitutive expression of the *NmDef02* gene produced a decrease in the number of pustules in the transgenic lines, and this led to a lower fungal biomass. In contrast, the leaves of the control plants (DT-84) colonized by the pathogen showed a lot of uredias with abundant sporulation

TABLE 2 | Agronomic field test with transgenic soybean lines affected by Asian soybean rust.

Line	Size (cm)	Height of 1st pod (cm)	Number of branches	Number of pods	Number of seeds	Weight of seeds/plant (g)
Def1	35,8[bc]	8,4[ab]	2,9[abc]	45,5[ab]	70,7[ab]	9,5[abcd]
Def3	30.1[def]	6.8[cd]	1.7[d]	29.1[ef]	50.8[bcd]	7.1[def]
Def4	34.3[cd]	7.0[bcd]	2.8[abc]	41.0[abcd]	69.0[abc]	10.3[abc]
Def5	22.8[g]	6.3[cd]	2.3[cd]	28.6[ef]	51.4[bcd]	7.7[cde]
Def6	25.9[fg]	6.6[cd]	2.3[cd]	29.9[def]	48.5[cd]	6.0[ef]
Def7	30.1[def]	6.5[cd]	2.5[bcd]	35.0[bcde]	65.0[abc]	7.2[def]
Def10	31.8[cde]	6.2[d]	2.5[bcd]	38.9[abcde]	64.1[abc]	8.6[bcde]
Def12	40.9[ab]	7.8[abc]	3.6[a]	39.2[abcde]	73.8[a]	8.1[bcde]
Def14	32.2[cde]	6.3[cd]	2.6[bcd]	38.0[abcde]	63.1[abc]	8.3[bcde]
Def15	27.4[efg]	6.7[cd]	2.6[bcd]	32.4[cdef]	51.6[bcd]	7.6[cdef]
Def16	42.8[a]	9.3[a]	3.5[ab]	38.0[abcde]	60.5[abc]	10.8[ab]
Def17	40.8[ab]	7.7[bcd]	3.4[ab]	47.2[a]	81.5[a]	11.7[a]
Def18	35.1[dc]	7.7[bcd]	3.1[abc]	43.5[abc]	79.2[a]	11.0[ab]
NT	27.6[efg]	6.6[cd]	0.5[e]	22.0[f]	31.2[d]	4.7[f]
SD	8.49	1.97	1.38	14.84	27.61	3.84
SE	0.41	0.10	0.07	0.72	1.35	0.19

A simple classification ANOVA was used. Averages with different letters indicate significant differences of $p < 0.05$ according to Tukey's multiple range test. NT, non-transgenic control; SD, standard deviation; SE, standard error of the means. These data correspond to the average of 30 plants for each line and the non-transgenic control.

FIGURE 7 | Evaluation of transgenic soybean lines affected by anthracnose (*Colletotrichum truncatum*) in the field experiment. **(A)** Anthracnose symptoms observed in non-transgenic plants cv. DT-84. **(B)** Healthy plants of the Def1 transgenic line. **(C)** Evaluation of the number of plants affected by anthracnose in the field experiment. Transgenic lines (1, 12, and 16). Non-transgenic control (NT). Bars represent the deviation of the means (*n* = 3). **(D)** Early maturity and defoliation in soybean plants. Left: Def1 line, right: non-transgenic control. **(E)** Anthracnose symptoms observed in pods and seeds of non-transgenic plants cv. DT-84. Top: Def1 transgenic line, bottom: non-transgenic control.

and a high amount of *P. pachyrhizi* biomass, demonstrating a compatibility response with the pathogen, as described by some authors (Yamanaka et al., 2010; Vittal et al., 2014). The reduced

sporulation of the fungus observed in some plants from the transgenic lines is evidence of resistance to *P. pachyrhizi* in this experiment under natural infection conditions. This effect of the

TABLE 3 | Agronomic field trial with transgenic soybean lines affected by anthracnose.

Line	Size (cm)	Height of 1st pod (cm)	Number of branches	Number of pods	Number of seeds	Weight of seeds/plant (g)
Def1	42.28[a]	7.21[a]	3.27[a]	69.57[a]	148.83[a]	29.53[a]
Def12	32.60[b]	6.65[ab]	2.67[ab]	39.43[bc]	69.23[bc]	11.60[b]
Def16	26.71[c]	6.79[ab]	2.47[b]	40.80[b]	75.83[b]	14.00[b]
NT	28.44[c]	6.04[b]	0.53[c]	30.97[c]	55.67[c]	10.25[b]
SD	7.31	1.26	1.39	20.35	45.74	10.07
SE	0.67	0.12	0.13	1.86	4.18	0.92

Averages with different letters in the same column indicate significant statistical differences of p < 0.05 according to the Tukey's multiple range test. NT, non-transgenic control. SD, standard deviation. SE, standard error. These data correspond to the average of 30 plants for each line.

TABLE 4 | Agronomic field trial with transgenic soybean lines under field conditions without fungal diseases.

Lines	Size (cm)	Height of 1st pod (cm)	Number of branches	Number of pods	Number of seeds	Weight of seeds/plant (g)
1	40.5[ab]	10.5[b]	2.5[ab]	55.0[a]	126.0[a]	26.6[a]
17	40.2[ab]	11.2[b]	3.4[a]	42.7[bc]	99.8[b]	20.1[b]
18	40.0[ab]	10.6[b]	2.5[abc]	39.1[bc]	87.1[bc]	20.0[b]
4	42.9[a]	14.2[a]	1.9[bcd]	38.2[bc]	79.1[bc]	20.5[b]
9	37.5[bc]	10.2[b]	1.2[d]	32.8[c]	63.7[c]	12.9[c]
12	32.7[c]	9.7[b]	2.1[bcd]	40.3[bc]	83.1[bc]	18.3[b]
NT	43.3[a]	9.6[b]	1.4[cd]	47.9[ab]	94.5[b]	22.1[ab]

A simple classification ANOVA was performed. Averages with different letters indicate significant differences of p < 0.05 according to Tukey's multiple range test. These data correspond to the average of 30 plants for each transgenic line carrying the NmDef02 gene (1, 17, 18), the cp4epsps gene (4, 9, 12), and the non-transgenic control (NT).

inhibition of fungal germination and growth of spores by plant defensins was also reported in transgenic bean plants that carry the *pdf1.2* defensin gene against *Colletotrichum lindemuthianum* (Espinosa-Huerta et al., 2013).

Plant-pathogen interaction studies have shown that RB lesions can vary in color from light to dark red (Rosa et al., 2015). Because of this, the authors consider that the color of the lesions is not a reliable indicator of resistance or susceptibility to *P. pachyrhizi*. Lesion color is not always a reliable indicator, because it is influenced by environment (Yamanaka et al., 2010). Studies conducted by Miles et al. (2011) also showed that in some cases the severity of the rust is not related to lesion type. However, they found that the number of uredias per leaf area is inversely related to yield (Miles et al., 2011). Similar results were obtained in this study, with different types of lesions on the leaves of transgenic and non-transgenic plants. In addition, the high fungal biomass detected in the plants was related to the large number of uredias present on the leaves, regardless of the type of lesion. It was also consistent with the inverse relationship between severity and yield parameters, where the non-transgenic control plants affected by the fungus had a high percentage of uredias on the leaves and a small number of pods and seeds, as shown in **Table 2**.

Rust infestation was more severe in certain lines, which also had more fungal biomass on leaves. They were also affected by early defoliation, suggesting that although they were less affected than the DT-84 control, defensin expression was not enough to avoid fungal damage. Studies conducted by Ntui et al. (2010) with the wasabi defensin gene showed that fungal resistance is associated with the level of expression of the protein. In our study, we found a high correlation between

the relative expression of *NmDef02*, as determined by qRT-PCR, and the high resistance against *P. pachyrhizi* in transgenic lines Def1, 17, and 18. Interestingly, lines Def12 and 16, which had relatively low expression of defensin, also showed resistance against *P. pachyrhizi* but developed symptoms due to *C. truncatum* infection.

The constitutive expression of *NmDef02* also influenced the proliferation of C. *truncatum*, because there was a decrease in the number of lesions on the transgenic plants, in which line Def1 showed increased resistance to this pathogen. In contrast, DT-84 plants used as controls showed a high susceptibility to *C. truncatum*, with abundant lesions in stems and pods. Similar results were observed in common bean (*Phaseolus vulgaris* L.) carrying the *pdf1.2* defensin gene, where the authors achieved a significant reduction in the formation of lesions in transgenic lines infected with *Colletotrichum* sp. (Espinosa-Huerta et al., 2013).

Some authors state that plants with a short life cycle are able to avoid yield reduction due to Asian soybean rust. This type of mechanism could be a form of horizontal resistance based on an escape mechanism or an unfavorable environment for disease development (Santos et al., 2018). This, however, depends on the susceptibility of the variety and development stage of the plants at the time of the appearance of the pathogen. In the present study, the short cycle cultivar DT-84 used as a non-transgenic control was highly susceptible to *P. pachyrhizi* and *C. truncatum* under the conditions of natural infection, and the plants underwent complete defoliation before concluding their maturation cycle. The early defoliation of these soybean plants reduced productivity by interfering with their physiological processes, thus resulting in less normal pods, fewer seeds per

pod and lower grain weight. Disease progression during the pod formation and pod-filling periods is most detrimental to yield (Kawuki et al., 2004). The negative effect of defoliation on crop yields was also observed in Brazilian soybean cultivars affected by Asian soybean rust (da Silva et al., 2015; Childs et al., 2018). This defoliation can affect the natural mechanisms of resistance, making them less active and increasing the susceptibility of soybean to end-of-cycle diseases.

To conclude, in experiments where no chemical fungicides were applied, transgenic plants showed increased resistance to *P. pachyrhizi* and *C. truncatum* until the end-of-cycle stage, where other pathogens normally appear. Molecular analyses showed the presence of the transgene in the progeny of these lines, where transgenic line Def1 accumulated the highest transcript levels and displayed the highest degree of resistance to both diseases. This could explain why severity and incidence of and defoliation by Asian soybean rust in plants of this transgenic line were lower. Similarly, the reduced fungal biomass present in the transgenic plants coincides with a reduced sporulation of the pathogen, which demonstrates the antifungal effect exerted by defensin *NmDef02* on *P. pachyrhizi*.

The antifungal effect of the *NmDef02* defensin had been previously demonstrated by Portieles et al. (2010). However, it is not obvious that there is resistance to a biotrophic fungus such as *P. pachyrhizi*, which is difficult to control, or against *C. truncatum*, because resistance to fungal pathogens is not only obtained by introducing this defensin into a culture. An example of this is the susceptibility to these fungi observed in some transgenic lines evaluated in the field.

Transgenic soybean plants had a similar development to non-transgenic plants when inoculated with *B. japonicum*. The use of rhizospheric microorganisms in the preparation of inoculants for soybeans was very important in maintaining high productivity with a lower environmental impact, as demonstrated by some authors (Menéndez et al., 2014; Nápoles García et al., 2014). This study also showed that the expression of the *NmDef02* defensin gene in soybean plants had no negative effect on the nodulation induced by *B. japonicum*, a bacterium that plays an essential role in the technology of this crop in Cuba. Kaur et al. (2017) studied the symbiosis of mycorrhizae with transgenic wheat plants carrying MtDef4.2 defensin. This study also showed that the expression of that defensin in apoplast can provide resistance to leaf rust, without having a negative effect on the symbiosis with that beneficial fungus.

To the best of our knowledge, this is the first report on the transformation of soybean with a defensin gene for resistance to fungal pathogens. We demonstrated that the overexpression of the *NmDef02* gene resulted in a delay in progression of the fungi. However, more evaluations of these transgenic lines against these pathogens in different weather conditions are necessary, taking into account other parameters that were not taken into account in this study, to confirm resistance. A complete resistance to *P. pachyrhizi* was not found in the transgenic lines, and the application of fungicide is needed to completely control the pathogen. Evidently, the use of transgenic plants expressing this defensin would reduce the number of chemical fungicide applications in the field for an integral pest management in soybean with a minimal environmental impact. These results provide a sound environmental approach to decrease yield losses and to lower the burden of chemicals, both goals targeted by Next Generation Agriculture.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

NS and GE conceived and designed the research work. YR performed the construction of pCP4EPSPS-DEF plasmid. NS, YH, and CD performed the soybean transformation experiments. NS, YH, CD, LV, RO, and GE performed the field experiments. NS, YH, OB-H, and GE conducted molecular analysis and analyzed the data. NS wrote the manuscript. GE, MP, and OB-H reviewed the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Abdallah, N. A., Shah, D., Abbas, D., and Madkour, M. (2010). Stable integration and expression of a plant defensin in tomato confers resistance to fusarium wilt. *GM Crops* 1, 344–350. doi: 10.4161/gmcr.1.5.15091

Cárcamo Rodríguez, A., Ríos, J. A., and Hernández, J. (2006). First report of Asian soybean rust caused by *Phakopsora pachyrhizi* from Mexico. *Plant Dis.* 90, 1260–1260. doi: 10.1094/PD-90-1260B

Chen, L., Chu, C., Liu, C., Chen, R., and Tsay, J. (2006). PCR-based detection and differentiation of anthracnose pathogens, colletotrichum gloeosporioides and *C. truncatum*, from vegetable soybean in Taiwan. *J. Phytopathol.* 154, 654–662.

Childs, S. P., Buck, J. W., and Li, Z. (2018). Breeding soybeans with resistance to soybean rust (*Phakopsora pachyrhizi*). *Plant Breed.* 137, 250–261.

da Silva, A. F., Sediyama, T., dos Santos Silva, F. C., Bezerra, A. R. G., Rosa, D. P., Cruz, C. D., et al. (2015). Effect of defoliation from the bottom to the top on the yield of Brazilian soybean cultivars. *Afr. J. Agric. Res.* 10, 3296–3304.

De Coninck, B., Cammue, B. P. A., and Thevissen, K. (2013). Modes of antifungal action and in planta functions of plant defensins and defensin-like peptides. *Fungal Biol. Rev.* 26, 109–120.

Doyle, J., and Doyle, J. (1987). CTAB DNA extraction in plants. *Phytochem. Bull.* 19, 11–15.

Espinosa-Huerta, E., Quintero-Jiménez, A., Sánchez-García, B. M., Acosta-Gallegos, J. A., and Mora-Avilés, M. A. (2013). Resistencia a *Colletotrichum lindemuthianum* en frijol común transgénico, expresando el gen defensina de *Arabidopsis thaliana. Rev. Mex. Cienc. Agríc.* 4, 1027–1042.

Frederick, R. D. (2006). *PCR Methods for the Identification and Detection of the Soybean Rust Pathogen Phakopsora Pachyrhizi.* Washington, DC: US Department of Agriculture. .

Gao, A.-G., Hakimi, S. M., Mittanck, C. A., Wu, Y., Woerner, B. M., Stark, D. M., et al. (2000). Fungal pathogen protection in potato by expression of a plant defensin peptide. *Nat. Biotechnol.* 18, 1307–1310. doi: 10.1038/82436

Godoy, C. V., Seixas, C. D. S., Soares, R. M., Marcelino-Guimarães, F. C., Meyer, M. C., and Costamilan, L. M. (2016). Asian soybean rust in Brazil: past, present, and future. *Pesqui. Agropecu. Brasileira* 51, 407–421. doi: 10.1094/PDIS-07-14-0772-SR

Graham, M., Silverstein, K., and VandenBosch, K. (2008). Defensin-like genes: genomic perspectives on a diverse superfamily in plants. *Plant Genome* 48(Suppl. 1), S3–S11.

Hartman, G. L., Miles, M. R., and Frederick, R. D. (2005). Breeding for resistance to soybean rust. *Plant Dis.* 89, 664–666. doi: 10.1094/PD-89-0664

Hartman, G. L., West, E. D., and Herman, T. K. (2011). Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. *Food Security* 3, 5–17.

Hartman, G. L., Rupe, J. C., Sikora, E. J., Domier, L. L., Davis, J. A., and Steffey, K. L. eds. (2015). *Compendium of Soybean Diseases and Pests.* St. Paul: Am Phytopath Society, 4–16.

Hossain, M. M., and Yamanaka, N. (2019). Pathogenic variation of Asian soybean rust pathogen in Bangladesh. *J. Gen. Plant Pathol.* 85, 90–100.

Islam, A. (2019). Preliminary risk assessment of a novel antifungal defensin peptide from chickpea (*Cicer arietinum* L.). *Appl. Biosaf.* 13, 222–230.

Jha, S., and Chattoo, B. B. (2010). Expression of a plant defensin in rice confers resistance to fungal phytopathogens. *Trans. Res.* 19, 373–384. doi: 10.1007/s11248-009-9315-7

Kanzaki, H., Nirasawa, S., Saitoh, H., Ito, M., Nishihara, M., Terauchi, R., et al. (2002). Overexpression of the wasabi defensin gene confers enhanced resistance to blast fungus (*Magnaporthe grisea*) in transgenic rice. *Theor. Appl. Genet.* 105, 809–814. doi: 10.1007/s00122-001-0817-9

Kaur, J., Fellers, J., Adholeya, A., Velivelli, S. L., El-Mounadi, K., Nersesian, N., et al. (2017). Expression of apoplast-targeted plant defensin MtDef4. 2 confers resistance to leaf rust pathogen *Puccinia triticina* but does not affect mycorrhizal symbiosis in transgenic wheat. *Trans. Res.* 26, 37–49. doi: 10.1007/s11248-016-9978-9

Kaur, J., Sagaram, U. S., and Shah, D. (2011). Can plant defensins be used to engineer durable commercially useful fungal resistance in crop plants? *Fungal Biol. Rev.* 25, 128–135.

Kawashima, C. G., Guimarães, G. A., Nogueira, S. R., MacLean, D., Cook, D. R., Steuernagel, B., et al. (2016). A pigeonpea gene confers resistance to Asian soybean rust in soybean. *Nat. Biotechnol.* 34:661. doi: 10.1038/nbt.3554

Kawuki, R., Tukamuhabwa, P., and Adipala, E. (2004). Soybean rust severity, rate of rust development, and tolerance as influenced by maturity period and season. *Crop Protect.* 23, 447–455.

King, Z. R., Childs, S. P., Harris, D. K., Pedley, K. F., Buck, J. W., Boerma, H. R., et al. (2017). A new soybean rust resistance allele from PI 423972 at the Rpp4 locus. *Mol. Breed.* 37:62.

Kumar, M., and Chakrabarti, S. K. (2018). Expression of β-defensin gene in potato confers enhanced resistance to *Ralstonia Solanacearum* L. *Defence Life Sci. J.* 3, 15–23.

Lacerda, A. F., Del Sarto, R. P., Silva, M. S., de Vasconcelos, E. A. R., Coelho, R. R., dos Santos, V. O., et al. (2016). The recombinant pea defensin Drr230a is active against impacting soybean and cotton pathogenic fungi from the genera Fusarium. Colletotrichum and Phakopsora. *3 Biotech* 6, 1–10. doi: 10.1007/s13205-015-0320-7

Lamour, K. H., Finley, L., Snover-Clift, K. L., Stack, J. P., Pierzynski, J., Hammerschmidt, R., et al. (2006). Early detection of Asian soybean rust using PCR. *Plant Health Prog.* 7:20.

Langenbach, C., Schultheiss, H., Rosendahl, M., Tresch, N., Conrath, U., and Goellner, K. (2016). Interspecies gene transfer provides soybean resistance to a fungal pathogen. *Plant Biotechnol. J.* 14, 699–708. doi: 10.1111/pbi.12418

Lay, F. T., and Anderson, M. A. (2005). Defensins-components of the innate immune system in plants. *Curr. Protein Peptide Sci.* 6, 85–101.

Lee, H.-H., Kim, J.-S., Hoang, Q. T., Kim, J.-I., and Kim, Y. S. (2018). Root-specific expression of defensin in transgenic tobacco results in enhanced resistance against *Phytophtora parasitica* var. nicotianae. *Eur. J. Plant Pathol.* 151, 811–823.

Lemos, N. G., Braccini, A. D. L., Abdelnoor, R. V., de Oliveira, M. C. N., Suenaga, K., and Yamanaka, N. (2011). Characterization of genes *Rpp*2. *Rpp*4, and *Rpp*5 for resistance to soybean rust. *Euphytica* 182:53. doi: 10.1007/s00122-017-2983-4

Marmat, N., and Ratnaparkhe, M. (2017). Molecular and phylogenetic studies of *Colletotrichum truncatum* associated with soybean anthracnose in India. *Plant Pathol. Q.* 7, 146–152.

Menéndez, C., Trujillo, L. E., Ramírez, R., González-Peña, D., Espinosa, D., Enriquez, G. A., et al. (2014). Producción de un inoculante líquido de *Bradyrhizobium japonicum* con alto impacto en la siembra mecanizada de la soya en Cuba. *Biotecnol. Apl.* 31, 116–120.

Miles, M., Bonde, M., Nester, S., Berner, D., Frederick, R., and Hartman, G. (2011). Characterizing resistance to *Phakopsora pachyrhizi* in soybean. *Plant Dis.* 95, 577–581.

Moosa, A., Farzand, A., Sahi, S. T., and Khan, S. A. (2018). Transgenic expression of antifungal pathogenesis-related proteins against phytopathogenic fungi – 15 years of success. *Israel J. Plant Sci.* 65:38. doi: 10.1080/07929978.2017.1288407

Muller, P. Y., Janovjak, H., Miserez, A. R., and Dobbie, Z. (2002). Short technical report processing of gene expression data generated by quantitative real-time RT-PCR. *Biotechniques* 32, 1372–1379.

Nápoles García, M. C., González-Anta, G., Ferreira, A., Rossi, A., Hernández Forte, I., and Costales Menéndez, D. (2014). Efecto de diferentes inoculantes sobre la nodulación de la soya cultivada en condiciones de estrés. *Cult. Trop.* 35, 45–51.

Ntui, V. O., Thirukkumaran, G., Azadi, P., Khan, R. S., Nakamura, I., and Mii, M. (2010). Stable integration and expression of wasabi defensin gene in "Egusi" melon (*Colocynthis citrullus* L.) confers resistance to Fusarium wilt and Alternaria leaf spot. *Plant Cell Rep.* 29, 943–954. doi: 10.1007/s00299-010-0880-2

Pawlowski, M. L., and Hartman, G. L. (2016). "Infection Mechanisms and Colonization Patterns of Fungi Associated with Soybean," in *Fungal Pathogenicity*, ed. S. Sultan (London: IntechOpen), 25.

Pérez-Vicente, L., Martínez-de la Parte, E., Pérez-Miranda, M., Martín-Triana, E., Borrás-Hidalgo, O., and Hernández-Estévez, I. (2010). First report of Asian rust of soybean caused by *Phakopsora pachyrhizi* in Cuba. *Plant Pathol.* 59, 803–803.

Petersen, A., Kull, S., Rennert, S., Becker, W.-M., Krause, S., Ernst, M., et al. (2015). Peanut defensins: novel allergens isolated from lipophilic peanut extract. *J. Allergy Clin. Immunol.* 136, 1295.–1301. doi: 10.1016/j.jaci.2015.04.010

Pham, T., Miles, M., Frederick, R., Hill, C., and Hartman, G. (2009). Differential responses of resistant soybean entries to isolates of *Phakopsora pachyrhizi. Plant Dis.* 93, 224–228. doi: 10.1094/PDIS-93-3-0224

Ploper, L. D., Escobar, D., Ivancovich, A., Diaz, C. G., Sillon, M., Formento, N., et al. (2006). Propuesta de protocolo para muestreo y evaluación de la roya asiática de la soja en Argentina. *Ava. Agroind. Estación Exp. Agro Ind. Obispo Colombres* 27, 35–37.

Portieles, R., Ayra, C., Gonzalez, E., Gallo, A., Rodriguez, R., Chacón, O., et al. (2010). NmDef02, a novel antimicrobial gene isolated from *Nicotiana megalosiphon* confers high-level pathogen resistance under greenhouse and field conditions. *Plant Biotechnol. J.* 8, 678–690. doi: 10.1111/j.1467-7652.2010.00501.x

Prema, G. U., and Pruthvi, T. P. M. (2012). Antifungal plant defensins. *Curr. Biot.* 6, 254–270.

Rosa, C., Spehar, C., and Liu, J. (2015). Asian soybean rust resistance: an overview. *J. Plant Pathol. Microb.* 6:2.

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor.

Santos, J. V. M. D., Yamanaka, N., Marcelino-Guimarães, F. C., de Toledo, J. F. F., Arias, C. A. A., and Abdelnoor, R. V. (2018). Molecular mapping of quantitative trait loci for agronomical traits in soybean under Asian soybean rust infection. *Crop Breed. App. Biotechnol.* 18, 390–398.

Schneider, R., Hollier, C., Whitam, H., Palm, M., McKemy, J., Hernandez, J., et al. (2005). First report of soybean rust caused by *Phakopsora pachyrhizi* in the continental United States. *Plant Dis.* 89, 774–774. doi: 10.1094/PD-89-0774A

Seo, H.-H., Park, S., Park, S., Oh, B.-J., Back, K., Han, O., et al. (2014). Overexpression of a defensin enhances resistance to a fruit-specific anthracnose fungus in pepper. *PLoS One* 9:e97936. doi: 10.1371/journal.pone.0097936

Singh, A., Mehta, A., Sridhara, S., Gaur, S., Singh, B., Sarma, P., et al. (2006). Allergenicity assessment of transgenic mustard (*Brassica juncea*) expressing bacterial codA gene. *Allergy* 61, 491–497. doi: 10.1111/j.1398-9995.2006. 01049.x

Soto, N., Delgado, C., Hernández, Y., Rosabal, Y., Ferreira, A., Pujol, M., et al. (2017). Efficient particle bombardment-mediated transformation of Cuban soybean (INCASoy-36) using glyphosate as a selective agent. *Plant Cell Tissue Organ Cult.* 128, 187–196.

Tavares, L. S., Santos, M. D. O., Viccini, L. F., Moreira, J. S., Miller, R. N., and Franco, O. L. (2008). Biotechnological potential of antimicrobial peptides from flowers. *Peptides* 29, 1842–1851. doi: 10.1016/j.peptides.2008.06.003

Thevissen, K., Ferket, K. K., François, I. E., and Cammue, B. P. (2003). Interactions of antifungal plant defensins with fungal membrane components. *Peptides* 24, 1705–1712. doi: 10.1016/j.peptides.2003.09.014

Thevissen, K., Kristensen, H. H., Thomma, B. P., Cammue, B. P., and Francois, I. E. (2007). Therapeutic potential of antifungal plant and insect defensins. *Drug Discov. Today* 12, 966–971. doi: 10.1016/j.drudis.2007.07.016

Thomma, B. P., Cammue, B. P., and Thevissen, K. (2002). Plant defensins. *Planta* 216, 193–202.

Thomma, B. P., Cammue, B. P., and Thevissen, K. (2003). Mode of action of plant defensins suggests therapeutic potential. *Curr. Drug Targets Infect. Disord.* 3, 1–8. doi: 10.2174/1568005033342000

Tremblay, A., Hosseini, P., Li, S., Alkharouf, N. W., and Matthews, B. F. (2012). Identification of genes expressed by *Phakopsora pachyrhizi*, the pathogen causing soybean rust, at a late stage of infection of susceptible soybean leaves. *Plant Pathol.* 61, 773–786.

van der Weerden, N. L., Bleackley, M. R., and Anderson, M. A. (2013). Properties and mechanisms of action of naturally occurring antifungal peptides. *Cell. Mole. Life Sci.* 70, 3545–3570. doi: 10.1007/s00018-013-1260-1

Van Loon, L. (1997). Induced resistance in plants and the role of pathogenesis-related proteins. *Eur. J. Plant Pathol.* 103, 753–765.

van Loon, L. C., Rep, M., and Pieterse, C. M. (2006). Significance of inducible defense-related proteins in infected plants. *Annu. Rev. Phytopathol.* 44, 135–162. doi: 10.1146/annurev.phyto.44.070505.143425

Vittal, R., Paul, C., Hill, C., and Hartman, G. (2014). Characterization and quantification of fungal colonization of *Phakopsora pachyrhizi* in soybean genotypes. *Phytopathology* 104, 86–94. doi: 10.1094/PHYTO-12-12-0334-R

Vriens, K., Cammue, B., and Thevissen, K. (2014). Antifungal plant defensins: mechanisms of action and production. *Molecules* 19, 12280–12303. doi: 10. 3390/molecules190812280

Xavier, S. A., Martins, D. C., Fantin, L. H., and Canteri, M. G. (2017). Older leaf tissues in younger plants are more susceptible to soybean rust. *Acta Scientiarum Agron.* 39, 17–24.

Yamanaka, N., Lemos, N. G., Uno, M., Akamatsu, H., Yamaoka, Y., Abdelnoor, R. V., et al. (2013). Resistance to Asian soybean rust in soybean lines with the pyramided three Rpp genes. *Crop Breedi. Appl. Biotechnol.* 13, 75–82.

Yamanaka, N., Morishita, M., Mori, T., Lemos, N. G., Hossain, M. M., Akamatsu, H., et al. (2015). Multiple Rpp-gene pyramiding confers resistance to Asian soybean rust isolates that are virulent on each of the pyramided genes. *Trop. Plant Pathol.* 40, 283–290.

Yamanaka, N., Morishita, M., Mori, T., Muraki, Y., Hasegawa, M., Hossain, M., et al. (2016). The locus for resistance to Asian soybean rust in PI 587855. *Plant Breed.* 135, 621–626.

Yamanaka, N., Yamaoka, Y., Kato, M., Lemos, N. G., Passianotto, A. L. D. L., dos Santos, J. V., et al. (2010). Development of classification criteria for resistance to soybean rust and differences in virulence among Japanese and Brazilian rust populations. *Trop. Plant Pathol.* 35, 153–162.

Yang, H.-C., and Hartman, G. L. (2015). Methods and evaluation of soybean genotypes for resistance to *Colletotrichum truncatum*. *Plant Dis.* 99, 143–148. doi: 10.1094/PDIS-03-14-0228-RE

Yorinori, J., Paiva, W., Frederick, R., Costamilan, L., Bertagnolli, P., Hartman, G., et al. (2005). Epidemics of soybean rust (*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001 to 2003. *Plant Dis.* 89, 675–677. doi: 10.1094/PD-89-0675

Molecular Mechanisms of the Floral Biology of *Jatropha curcas*: Opportunities and Challenges as an Energy Crop

Manali Gangwar and Jata Shankar*†

Genomic Laboratory, Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, India

Fossil fuel sources are a limited resource and could eventually be depleted. Biofuels have emerged as a renewable alternative to fossil fuels. *Jatropha* has grown in significance as a potential bioenergy crop due to its high content of seed oil. However, *Jatropha*'s lack of high-yielding seed genotypes limits its potential use for biofuel production. The main cause of lower seed yield is the low female to male flower ratio (1:25–10), which affects the total amount of seeds produced per plant. Here, we review the genetic factors responsible for floral transitions, floral organ development, and regulated gene products in *Jatropha*. We also summarize potential gene targets to increase seed production and discuss challenges ahead.

Keywords: *Jatropha curcas*, energy crop, transcriptome, biofuel, ABCDE model

INTRODUCTION

About 11 billion tons of oil is consumed worldwide each year for fuel. With this rate of oil consumption, we may soon exhaust the oil reservoir (Shafiee and Topal, 2009)[2]. Climate change is also greatly influenced by fossil fuel combustion. Therefore, sustainable and environmentally friendly alternative energy sources are needed. *Jatropha curcas* L (Euphorbiaceae) is a plant with potential for biodiesel production due to its high seed oil content (around 45–50%) (Achten et al., 2008). Compared with other oil plants, *Jatropha* has its own merits, including an outstanding adaptability to varied environments, smooth propagation, and greater fruit and seed size. Furthermore, *Jatropha* grows well in the desert, adapts to drought conditions, has a short gestation period, and assists in soil conservation. Despite its advantageous properties for biodiesel production, *Jatropha* has some limitations that restrict its commercialization as an energy crop, such as low seed yield, inconsistent flowering and fruiting, and relatively expensive plantation maintenance. The significant factors influencing its potential as biofuel feedstock are the oil content in seeds, the number of seeds per tree, the number of fruits on each branch of the plant, and the number of branches per plant. Seed yield at each inflorescence is largely dependent on the number of female flowers. *Jatropha*'s female to male flower ratio is quite small (1:25 to 1:30), which means that each inflorescence contains only about 10 to 12 female flowers (out of 300) that yield just 8 to 10 fruits. Therefore, a relatively small number of fruits are produced as compared to the total number of flowers (Kumar and Sharma, 2008). One way to increase the total seed yield in *Jatropha* would be to increase the number of female flowers per plant. In this context, we have discussed the genetic factors involved in the floral transformation, determination of sex, and floral growth of *Jatropha curcas*.

---

[1] https://www.ecotricity.co.uk/our-green-energy/energy-independence/the-end-of-fossil-fuels

BIOLOGY OF SEX DETERMINATION IN *JATROPHA CURCAS*

Sex determination processes allow floral organ development in plants. The two processes for forming a unisexual flower are (i) emergence of only one type of sex organ (unisexual tissues) and (ii) initiation of stamen and pistil followed by an arrest or abortion of one sex organ, which results in the functional immaturity of either stamens or carpels. The developmental arrest step occurs at an immature stage well before sexual maturity is reached (Ainsworth, 1999, 2000; Kater et al., 2001). There are two modes of sex determination and development in *Jatropha*. One mode is the development of male flowers with early adolescence without any female primordia. The other mode is by aborting male tissues, which results in female flowers developing (Li and Li, 2009; Wu et al., 2011). The male flower is unisexual right from the start, whereas the female flower is bisexual until its sixth developmental phase. Because of this, an inflorescence has three types of flowering sites; (i) female flowering site, (ii) male flowering site, and (iii) middle flowering site where both males and females may develop. Though male tissue abortion occurs in female flowers during sexual differentiation, traces of male tissue may be found in mature females. However, when abortion of male tissues fails in a female flower, it develops as a male at the female flowering site. Such inflorescence is known as middle type inflorescence. Due to the number of female flowers formed at middle type inflorescence, variation in the total number of female flowers occurs. An inflorescence statistical analysis found ∼75 percent of middle-type inflorescence and 0.09 percent of female flowers (Luo et al., 2007; Wu et al., 2011). For 18 female locations, Wu et al. (2011) found only seven female flowers. The female flowering sites and the sites occupied by middle-type inflorescence are important in increasing the number of female flowers. The presence of hermaphroditic flowers has also been recorded in *Jatropha*, showing structural similarity with female flowers but diffused stamens (Abdelgadir et al., 2010; Wu et al., 2011; Adriano-Anaya et al., 2016). A recent population analysis on *Jatropha*'s floral diversity and sex expression has grouped accessions into gynoecious (having only females), androecious (having only males flowers), and andromonoecious (having both bisexual and male flowers) plants showing no correlation with their geographic location (Adriano-Anaya et al., 2016). Of the 103 accessions from 33 sites in southern Mexico, 93.2 percent were monoecious, while others were androgynomonoecious, androecious, or gynoecious (**Figures 1A,B**). It has been hypothesized that male development commences through suppression of females, which might be the result of male sterility mutation in gynomonoecious plants (Salvador-Figueroa et al., 2015; Adriano-Anaya et al., 2016). No gynomonoecious plants of *Jatropha* have been found. The possible explanation, according to Adriano-Anaya et al. (2016), is that gynoecious *Jatropha* plants derive from hermaphrodite ancestors through a one-step mutation.
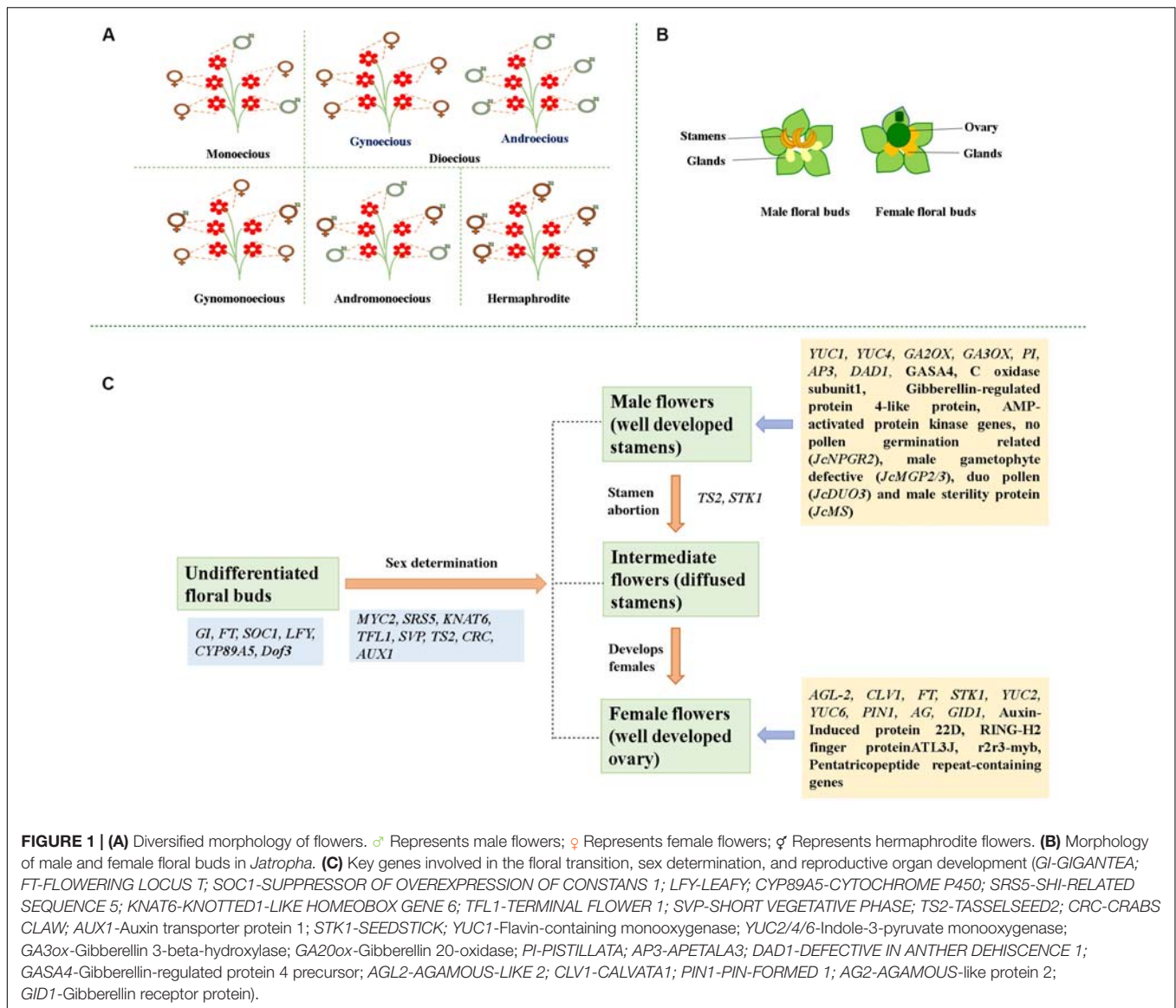

Abbreviations: BRs, brassinosteroids; CTK, cytokinin; GAs, gibberellic acids; JAs, jasmonic acid; *Jatropha, Jatropha curcas* L.

GENETIC FACTORS FOR VEGETATIVE TO A REPRODUCTIVE PHASE TRANSITION

In floral initiation, the apical shoot meristem differentiates into an inflorescence. The induction of floral signaling is genetically controlled by floral integrator genes, such as *FT* (*FLOWERING LOCUS T*), *FLC* (*FLOWERING LOCUS C*) and *SOC1* (*SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1*). Ye et al. (2014) reported that *JcFT* (*Jatropha Flowering locus T*) overexpression caused early flowering by shortening the bolting time. Li et al. (2014) characterized *FT* in *Jatropha*, and data from its spatial expression showed higher expression in reproductive phases. The *LFY* gene has recently been identified and overexpressed in both *Arabidopsis thaliana* and *Jatropha* (Tang et al., 2016). During the early stages of flowering, they observed a higher expression of *JcLFY* (*Jatropha LEAFY*). Transgenics with *JcLFY* overexpression showed early flowering and increased transcript levels of floral meristem identity genes, such as *JcAP1*, *JcAP3*, *JcSEP1*, *JcSEP3*, and *JcAG*. In addition, co-suppression of *LFY* in *Jatropha* resulted in delayed flowering, abnormal floral flowers replaced by sepaloid organs, and an increased rate of floral abortion (Tang et al., 2016). Recently, the role of *TFL1* homologs has been studied through the transgenic method, and their overexpression has resulted in delayed flowering due to reduced *AP1* and *FT* gene expression (Karlgren et al., 2013; Li et al., 2017). In contrast, Li et al. (2014) reported higher expression of *FT* in *Jatropha*'s reproductive phases and fruits. Circadian rhythms play an important role in the initiation of flowering. JcDof3, a plant-specific transcription factor with a conserved zinc finger (ZF) DNA-binding domain, is a circadian clock regulated gene. The C-terminal conserved region of Dof3 interacts with the F-box protein forming Dof3-Fbox complex regulating the expression of *CO*, a circadian clock regulating flowering gene (Yang et al., 2011). Foliar cytokinin (CTK) treatment upregulates genes *GI*, *SOC1,* and *LFY*, and inactivates genes *COP1* and *TFL1b* that maintain a flowering signal which promotes flowering (Chen et al., 2014; Pan et al., 2014). Thus, the interplay between the circadian rhythm and hormones control flowering genes and phase transition to inflorescence meristem in *Jatropha*.

MOLECULAR BASIS OF SEX DETERMINATION

Jatropha is a monoecious plant in which female flowers are formed due to stamen abortion/suppression. Remains of female tissues are not observed in male flowers, though remains of aborted stamens (male tissues) are present at the base in female flowers. By analyzing *Jatropha* floral buds for gene expression, the *SUPERMAN* gene was observed to suppress male tissue and promote the development of female tissue (Gangwar et al., 2018). A recent study suggested that, in *Arabidopsis*, the *SUPERMAN* gene not only bridges floral organogenesis and floral meristem but also regulates

FIGURE 1 | (A) Diversified morphology of flowers. ♂ Represents male flowers; ♀ Represents female flowers; ⚥ Represents hermaphrodite flowers. (B) Morphology of male and female floral buds in *Jatropha*. (C) Key genes involved in the floral transition, sex determination, and reproductive organ development (*GI-GIGANTEA; FT-FLOWERING LOCUS T; SOC1-SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1; LFY-LEAFY; CYP89A5-CYTOCHROME P450; SRS5-SHI-RELATED SEQUENCE 5; KNAT6-KNOTTED1-LIKE HOMEOBOX GENE 6; TFL1-TERMINAL FLOWER 1; SVP-SHORT VEGETATIVE PHASE; TS2-TASSELSEED2; CRC-CRABS CLAW; AUX1*-Auxin transporter protein 1; *STK1-SEEDSTICK; YUC1*-Flavin-containing monooxygenase; *YUC2/4/6*-Indole-3-pyruvate monooxygenase; *GA3ox*-Gibberellin 3-beta-hydroxylase; *GA20ox*-Gibberellin 20-oxidase; *PI-PISTILLATA; AP3-APETALA3; DAD1-DEFECTIVE IN ANTHER DEHISCENCE 1; GASA4*-Gibberellin-regulated protein 4 precursor; *AGL2-AGAMOUS-LIKE 2; CLV1-CALVATA1; PIN1-PIN-FORMED 1; AG2-AGAMOUS*-like protein 2; *GID1*-Gibberellin receptor protein).

auxin biosynthesis (Xu et al., 2018). Transcriptome analysis of *Jatropha*'s floral buds showed reduced expression of the stamen development gene *TASSELSEED 2* (*TS2*) that facilitated the growth of carpels (Chen et al., 2014). Transcriptomic analysis of different stages of male and female flower buds of *Jatropha* showed upregulation of *CRABS CLAW* (*CRC*) during development stages of female flowers. *CRC*, a C2C2-YABBY zinc finger protein, is involved in the regulation of carpel fusion and growth, nectary formation, and floral meristem termination (Xu et al., 2016; Gross et al., 2018). Genes encoding for inorganic phosphate transporter and ubiquitin carboxyl-terminal hydrolase were upregulated during female flower development and may contribute to embryo sac development (Xu et al., 2016). Further, upregulation of genes encoding for chlorophyll A/B-binding protein during initiation of carpel primordia may facilitate carpel differentiation. Genes encoding for Gibberellin-regulated protein 4-like protein,

cytochrome c-oxidase subunit 1 (mitochondrial gene), and AMP-activated protein kinase, however, were upregulated during stamen development. Upregulation of genes encoding for RING-H2 finger protein ATL3J (E3 ubiquitin ligases), *CLAVATA1* (receptor-like kinase), auxin-induced protein 22D, transcription factor R2R3-myb (regulating cell cycle genes and cytokinin signaling), and *AGAMOUS-LIKE-20* (MADS-box genes) have been identified during the late stage of female flower development, which may facilitate the maturation of female flower (Alvarez and Smyth, 1999; Pelaz et al., 2000; Makkena et al., 2012; Xu et al., 2016). In both male and female flower buds, genes such as ARP1 (Auxin repressed protein), X10A (auxin-induced protein), and GID1 (gibberellin receptor protein) were upregulated (Xu et al., 2016). The role of *JcFT*, a florigen and a key flowering pathway regulator in *Jatropha* showed significantly high transcript levels in female flowers (Li et al., 2014). Another transcriptomic study

identified *MYC2*, *TS2*, *KNAT6*, *SVP*, *TFL1*, and *SRS5* as sex determination regulators in *J. curcas* (Chen et al., 2017). The suppression of nodulin *MtN3* or *LESS ADHERENT POLLEN* (*LAP3*) resulted in small anthers, sterile pollens, and abortion of female flowers in *Oryza sativa*, *Vitis vinifera,* and *Medicago truncatula* (Chu et al., 2006; Ramos et al., 2014). In *Pisum sativum L*, carpel senescence has been induced as a result of increased lipoxygenase gene expression (Rodríguez-Concepción and Beltrán, 1995). Pentatricopeptide repeat-containing gene is expressed in the female embryo sac and restores the cytoplasmic male sterility in *Jatropha* (Bentolila et al., 2002; Xu et al., 2016). Key genes involved in the floral transition, sex determination and development of reproductive organs are shown in **Figure 1C** and **Table 1**. These studies shed light on how sex determination and differentiation occur in monoecious plants and how some of the genes expressed during floral differentiation suppress male flowering.

ABCDE MODEL FOR SEX DIFFERENTIATION

The ABCDE model is a scientific model that specifies the role of homeotic genes in the development of floral organs. Genes of the A class specify sepal development. The development of petals occurs by the combined effect of genes from the A and B classes. Both the B- and C-class genes are important for stamen growth. The carpel development and activity of ovules are determined by C-class and D-class genes, respectively. Recently, E-class genes were discovered to play a role in the development of carpel and ovary (Pelaz et al., 2000; Honma and Goto, 2001). A-, B-, C-, D-, and E-class genes are transcription factors with conserved DNA binding domains known as the MADS-box family and are involved in floral organogenesis regulation (Parěnicová et al., 2003; Chen et al., 2019). *PERIANTHA* (*PAN*), a bZIP transcription factor, activates *AG*, a C-class MADS-box protein that regulates floral organ numbers and whorl patterning (Maier et al., 2009). In *Elaeis guineensis*, the mutants *AP3* and *PISTILLATA* (*PI*) inhibited male tissues. *AG2* has a mixed C/D function gene, and its expression has been observed in ovule primordia and carpel of *Arabidopsis* and *Elaeis guineensis*, respectively (Favaro et al., 2003; Adam et al., 2007). FLORAL BINDING PROTEIN 11 (FBP11), a D-class gene, determines the formation of ovules in cucumbers (Favaro et al., 2003). An increase in the C-class gene transcription level arrests the development of sexual organs in monoecious plants, such as *Liquidambar styraciflua L* and *Rumex acetosa L* (Ainsworth, 2000). B-and C-class genes are regulated at a sex locus by a genetic switch that further controls the development of male or female flowers in *Populus trichocarpa* (Leseberg et al., 2006). B-class genes *PI* and *AP3* have been identified in the formation of stamen in *Jatropha*. A- and C-class gene *AG* and D-class gene *SEEDSTICK1* (*STKI*) have been reported for carpel development and maturation (Hui et al., 2017). Thus, the ABCDE model helps to understand the floral differentiation in *Jatropha*.

ROLE OF HORMONES IN SEX DETERMINATION

The process of flower development and sex determination is regulated by the interplay of endogenous hormones (auxins, cytokinins, gibberellins, abscisic acids, etc.). Auxin regulates sex determination in *Jatropha*. IAA enhanced female to male ratio from 1:27 to 1:23, and it also increased seed weight 3-fold (Joshi et al., 2011). Auxin biosynthesis and signaling are associated with genes such as *ARFs*, *AUX1*, and Transport inhibitor response 1 (*TIR1*). Transcriptome analysis of *Jatropha* suggested that *AUX1* is responsible for sex determination. The main source of auxin production is through Trp-dependent auxin biosynthesis, which participates in embryo patterning and reproductive organ development (Chen et al., 2017). In this pathway, IAA is produced from indole-3-pyruvic acid by *YUCCA* (*YUC*), a flavin-dependent monooxygenase (Stepanova et al., 2008). During stamen primordia formation, auxin is produced locally by *YUC1* and *YUC4* followed by *YUC2* and *YUC6* genes at late stages of stamen development (Cheng et al., 2006; Cecchetti et al., 2008). In mature gynoecia, *YUC4* and *YUC8* genes were expressed in the style, whereas *YUC2* was expressed in carpel valve tissues (Martínez-Fernández et al., 2014). Increased expression of *ARF 10/16/17/18* leads to abnormalities in females and abortion of organs, resulting in fewer seed sets (Huang et al., 2016).

Gibberellic acids also contribute to the development of the stamens in monoecious plants. Exogenous application of GA on the inflorescences of *Jatropha* resulted in a 2-fold increase in female flowering. However, inflorescence branches were not affected. Hui et al. (2018) reported the altered endogenous CTK (increased) and GA (decreased) ratio due to exogenous GA application, which resulted in an increased proportion of female flowers. However, a higher concentration of GA caused withering of floral buds. Hu et al. (2017) isolated the *JcGA2ox6* (Gibberellin oxidase) gene, which reduces the amount of endogenous GA4 (active gibberellin). They overexpressed *JcGA2ox6* gene in *Jatropha*, which led to decreased inflorescence size, decreased male and female flowers, and decreased seed length in transgenic plants. There was a significant decrease in both seed weight and oil content. *GA20ox* and *GA3ox* have been observed in other studies to enhance the development of stamen, whereas the exogenous application of GA3 led to a restricted development of pistils, thus enabling the male to expand. GA treatment enhanced the development of stamens in monoecious females, and it resulted in bisexual flowers in monoecious plants. GASA4 protein functions in stamen differentiation. *GID1*, a positive GA signaling pathway regulator, controls *Jatropha*'s female flowering (Roxrud et al., 2007; Hu et al., 2017). GA deficiency results in male sterility in plants. Therefore, GA allows the stamens to develop without affecting female flowers.

Paclobutrazol foliar application inhibits GA biosynthesis and promotes female flowering by suppressing no related pollen germination (*JcNPGR2*), male defective gametophyte (*JcMGP2*/3), duo pollen (*JcDUO3*), and male sterility protein (JcMS) genes, thus allowing female flowers to develop in *Jatropha* (Seesangboon et al., 2018).

TABLE 1 | Key genes involved in the floral transition, sex determination, and reproductive organ development.

	Vegetative to reproductive stage		
Gene name	**Plant spp.**	**Pathway/Association**	**References**
1-aminocyclopropane-1-carboxylate synthase (ACS1)	*Cucumis sativus*	Ethylene biosynthesis	Switzenberg et al., 2014
1-aminocyclopropane-1-carboxylate synthase 7 (ACS7)	*Cucumis sativus*	Ethylene biosynthesis	Switzenberg et al., 2014
Agamous (AG)	*Populus trichocarpa*	MADS-box regulators of differentiation, Homeotic genes	Brunner et al., 2000
Apetala 1/3 (AP1/3)	*Arabidopsis thaliana*	Floral meristem identity genes	Karlgren et al., 2013
Dof3	*Arabidopsis thaliana*	F-box protein regulates flowering time	Imaizumi, 2010
Flowering locus C (FLC)	*Arabidopsis thaliana*	Florigen signaling	Li et al., 2014
Flowering Locus T (FT)	*Arabidopsis thaliana*	Florigen signaling	Li et al., 2014
GIGANTEA (GI)	*Arabidopsis thaliana*	Circadian clock control and photoperiodism	Mizoguchi et al., 2005
LFY	*Arabidopsis thaliana*	Florigen signaling	Tang et al., 2016
Sepallata (SEP1/3)	*Arabidopsis thaliana*	Floral meristem identity genes	Pelaz et al., 2000
Suppressor of constans overexpression 1 (SOC1)	*Arabidopsis thaliana*	MADS-box protein	Chen et al., 2014
	Genes associated with flowering sex determination		
Aborted microspores (AMS)	*Arabidopsis thaliana; Capsicum annuum* L.	Pollen and anther development	Ye et al., 2010; Guo et al., 2018
Agamous (AG2)	*Elaeis guineensis*	Ovule primordia and carpel	Favaro et al., 2003; Adam et al., 2007
Agamous-like-2 (AGL-2)	*Arabidopsis thaliana*	Induces microsporogenesis, Embryo sac development	Pelaz et al., 2000
Apetala 3 (AP3)	*Elaeis guineensis*	Development of male tissues	Favaro et al., 2003; Adam et al., 2007
ARF 10/16/17/18	*Arabidopsis thaliana*	Female organ abortion	Huang et al., 2016
ATL3J	*Zea mays*	Embryo sac development	Xu et al., 2016
Auxin induced protein (*X10A*)	*Arabidopsis thaliana*	Stamen differentiation and embryo sac development	Xu et al., 2016
Auxin repressed protein (*ARP1*)	*Nicotiana tabacum*	Pollen maturation	Nakamura et al., 2004; Xu et al., 2016
Clavata1 (*CLV1*)	*Arabidopsis thaliana*	Peptide-receptor signaling	Alvarez and Smyth, 1999
CRABS CLAW (*CRC*)	*Arabidopsis thaliana; Oryza sativa*	Carpel fusion and growth, forming nectary	Gross et al., 2018
Cup-shaped cotyledon 2 (CUC2)	*Arabidopsis thaliana; Silene latifolia*	Forms boundary between the organs and separates organs with meristem	Li et al., 2010
Defective in Tapetal development and function 1 (TDF1)	*Arabidopsis thaliana*	Pollen and anther development	Zhu et al., 2008
Duo pollen (*DUO3*)	*Arabidopsis thaliana*	Regulator of Male Germline and embryogenesis	Brownfield et al., 2009
Floral binding protein 11 (*FBP11*)	*Cucumis sativus*	Ovule formation	Favaro et al., 2003
Gibberellin receptor protein (*GID1*)	*Oryza sativa*	Stamen differentiation and embryo sac development	Ueguchi-Tanaka et al., 2007; Xu et al., 2016
Less adherent pollen (LAP3)	*Oryza sativa, Vitis vinifera, Medicago truncatula*	Pollen development	Chu et al., 2006; Ramos et al., 2014
Lonely guy (LOG)	*Oryza sativa*	Maintains floral meristem activity and ovule development	Yamaki et al., 2011
Male gametophyte defective (MGP2/3)	*Arabidopsis thaliana*	Pollen tube growth and pollen germination	Deng et al., 2010
No pollen germination related (NPGR2)	*Arabidopsis thaliana*	A calmodulin-binding protein regulated pollen germination	Golovkin and Reddy, 2003
Pistillata (PI)	*Elaeis guineensis*	Development of male tissues	Favaro et al., 2003; Adam et al., 2007
Seedstick1 (STKI)	*Arabidopsis thaliana*	Carpel development	Pinyopich et al., 2003
Sporocyteless/nozzle (SPL/NZZ)	*Arabidopsis thaliana*	Regulates anther cell differentiation	Liu et al., 2009
Superman (SUP)	*Arabidopsis thaliana*	Suppresses stamen development	Prunet et al., 2017
Tasselseed 2 (TS2)	*Zea mays*	Stamen development by Pistil abortion	Acosta et al., 2009
YUCCA (YUC1/2/4/6)	*Arabidopsis thaliana*	Stamen development	Stepanova et al., 2011
	ABCDE Model Genes		
Pistillata (PI)	*Elaeis guineensis; Jatropha curcas*	B-class, stamen development	Favaro et al., 2003; Adam et al., 2007; Hui et al., 2017
Apetala 3 (AP3)	*Elaeis guineensis; Jatropha curcas*	B-class, stamen development	Favaro et al., 2003; Adam et al., 2007; Hui et al., 2017

(Continued)

TABLE 1 | Continued

	ABCDE Model Genes		
Gene name	Plant spp.	Pathway/Association	References
Agamous (*AG*)	*Jatropha curcas; Populus trichocarpa*	C-class, carpel differentiation	Brunner et al., 2000; Hui et al., 2017
SEEDSTICK1 (*STKI*)	*Arabidopsis thaliana; Jatropha curcas*	D-class, carpel maturation	Pinyopich et al., 2003; Hui et al., 2017
Sepallata (*SEP*)	*Arabidopsis thaliana; Jatropha curcas*	E-class, male floral initiation	Pelaz et al., 2000; Chen et al., 2019

Jasmonic acids and brassinosteroids (BRs) are active in floral development together with stamen development, pollen maturation, and male fertility (Park et al., 2002; Ye et al., 2010). In staminate maize flowers, brassinosteroids promoted pistil abortion. AG controls the maturation and late stages of stamen development in *Arabidopsis* by regulating the biosynthesis of jasmonates (Ito et al., 2007). Reduced JA synthesis in *Jatropha* led to male abortion and downregulation of the genes *DAD1* and *LOX2*. *Arabidopsis*, maize, and tomato mutants with suppressed jasmonate synthesis and brassinosteroid signaling resulted in male sterility (Li et al., 2005; Ye et al., 2010). The *SPL/NZZ*, *Aborted Microspores* (*AMS*) and *Defective in Tapetal Development and Function 1* (*TDF1*) genes are regulated by BRs and are critical for anther and pollen development (Ye et al., 2010). Thus, BRs and JAs promote the development of male organs.

Foliar application of ethylene induced femininity in *Jatropha*. To synthesize ethylene, 1-aminocyclopropane-1-carboxylic acid oxidase 2 (*ACO1*) oxidizes ethylene intermediates. Transgenics plants that overexpressed *ACO2* were male sterile due to suppressed stamens. Little to no activity of *ACO* was observed in *Arabidopsis*, tomato, and tobacco during the development of anthers and pollens (Bartley and Ishida, 2007; Duan et al., 2008; Wang et al., 2010). These experiments have thus shown that ethylene promotes feminism in plants.

Studies have been conducted to see the effect of foliar cytokinin application on the inflorescences. It has been found that 29.99 percent of the total flowers were females in treated inflorescences as compared to 6.96 percent in control. In treated inflorescence, a 4–5-fold increase in the number of seeds was observed but the fruiting rate, seed weight, and oil content decreased (Pan and Xu, 2011; Pan et al., 2014; Chen et al., 2014).

Transcriptomic analysis of *Jatropha* inflorescences treated with cytokinin revealed that genes involved in the initiation of flowers, such as *GI*, *SOC1*, and *LFY,* and the *CYP89A5* gene involved in the development of inflorescences were induced, whereas the *AP1*, *AP2*, *PI*, *AG*, and *SEP1-3* genes were downregulated (Chen et al., 2014; Pan et al., 2014). These developments allowed more time for inflorescence meristems to generate floral primordia. A vital increase in the number of flowers was noted due to *CUC1* upregulation. Application of BA (6-Benzylaminopurine) increased the rate of cell division in inflorescence meristem due to the upregulation of *Cyclin-3-1* (*CycD3;1/2*) and *Cyclin-dependent protein kinase 247* (*CycA3;2*) genes. Li et al. (2010) observed an increase in the number of flowers with an enlarged inflorescence and floral meristem in transgenic *Arabidopsis* overexpressing CK

(cytokinin) biosynthetic gene (*AtIPT4*). Fewer flowers were observed at each inflorescence due to the overexpression of the *CKX* gene (Werner and Schmülling, 2009). Loss-of-function mutation of *LONELY GUY* (*LOG*) (encodes for CK-activating enzyme) gene of rice led to the significant decrease in the number of floral organs (Kurakawa et al., 2007). Chen et al. (2014) reported that BA treatment decreased the expression of *TS2*, which suppresses carpel in maize, leading to increased female to male flower ratios in *Jatropha* (Acosta et al., 2009).

CHALLENGES

Genomic studies on flowering of *Jatropha* and phenotypic changes following the application of PGRs (Plant Growth Regulators) showed an opportunity to increase female flowering, which is one of the aspects for increasing seed yields. There are several challenges to increasing a number of female flowers: (i) manual hormone application to each inflorescence is laborious; (ii) hormone application is not economical; (iii) optimized hormone concentration at one environmental condition may not show the same efficiency under different environmental conditions; (iv) flowering and fruit maturity are not synchronized; and the (v) variation in fruiting rate. Genetic modification of flowering genes or overexpression of genes involved in suppression of male flowers may enable us to overcome these challenges by allowing more female flowers to develop. Other possibilities include enhancing cytokinin synthesis by overexpressing genes associated with cytokinin biosynthesis or suppressing cytokinin breakdown by gene silencing or mutagenesis. Additionally, further research could be carried out on the effect of central carbon flow on the fruiting rate.

CONCLUSION AND PERSPECTIVE

The female to male floral ratio plays a significant role in deciding *Jatropha*'s seed yield. Cytokinin application showed promising results in enhancing the ratio between female and male flowers. Promising approaches to increase the number of female flowers may be to induce the transitioning of male type inflorescences to the middle/intermediate type or to increase male flower abortion rates to allow female flowers to develop. Therefore, genes involved in female flowering or the abortion of male flowers could be targeted for the purpose of increasing female flowers in *Jatropha*.

AUTHOR CONTRIBUTIONS

MG and JS conceived and designed the review manuscript, wrote, read, and approved the manuscript. JS contributed materials or analytical tools and supervised the work.

REFERENCES

Abdelgadir, H. A., Jager, A. K., Johnson, S. D., and Van, S. J. (2010). Influence of plant growth regulators on flowering, fruiting, seed oil content, and oil quality of *Jatropha curcas*. *South Afr. J. Bot.* 76, 440–444.

Achten, W. M. J., Verchot, L., Franken, Y. J., Mathijs, E., Singh, V. P., Aerts, R., et al. (2008). Jatropha bio-diesel production and use. *Biomass Bioenergy* 32, 1063–1084. doi: 10.1016/j.biombioe.2008.03.003

Acosta, I. F., Laparra, H., Romero, S. P., Schmelz, E., Hamberg, M., Mottinger, J. P., et al. (2009). Tasselseed1 is a lipoxygenase affecting jasmonic acid signaling in sex determination of maize. *Science* 323, 262–265. doi: 10.1126/science.1164645

Adam, H., Jouannic, S., Morcillo, F., Verdeil, J. L., Duval, Y., and Tregear, J. W. (2007). Determination of flower structure in elaeisguineensis: do palms use the same homeotic genes as other species? *Ann. Bot.* 100, 1–12. doi: 10.1093/aob/mcm027

Adriano-Anaya, M de. L., Pérez-Castillo, E., Salvador-Figueroa, M., Ruiz-González, S., Vázquez-Ovando, A., Grajales-Conesa, J., et al. (2016). Sex expression and floral diversity in *Jatropha curcas*: a population study in its center of origin. *PeerJ* 4:e2071. doi: 10.7717/peerj.2071

Ainsworth,. C. (1999). *Sex Determination in Plants*. Oxford: BIOS Scientific Publishers

Ainsworth, C. (2000). Boys and girls come out to play: the molecular biology of dioecious plants. *Ann. Bot.* 86, 211–221. doi: 10.1006/anbo.2000.1201

Alvarez, J., and Smyth, D. R. (1999). CRABS CLAW and SPATULA, two *Arabidopsis* genes that control carpel development in parallel with AGAMOUS. *Development* 126, 2377–2386.

Bartley, G. E., and Ishida, B. K. (2007). Ethylene-sensitive and insensitive regulation of transcription factor expression during in vitro tomato sepal ripening. *J. Exp. Bot.* 58, 2043–2051. doi: 10.1093/jxb/erm075

Bentolila, S., Alfonso, A. A., and Hanson, M. R. (2002). A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10887–10892. doi: 10.1073/pnas.102301599

Brownfield, L., Hafidh, S., Durbarry, A., Khatab, H., Sidorova, A., Doerner, P., et al. (2009). *Arabidopsis* DUO POLLEN3 is a key regulator of male germline development and embryogenesis. *Plant Cell* 21, 1940–1956. doi: 10.1105/tpc.109.066373

Brunner, A. M., Rottmann, W. H., Sheppard, L. A., Krutovskii, K., DiFazio, S. P., Leonardi, S., et al. (2000). Structure and expression of duplicate AGAMOUS orthologues in poplar. *Plant Mol. Biol.* 44, 619–634.

Cecchetti, V., Altamura, M. M., Falasca, G., Costantino, P. P., and Cardarelli, M. (2008). Auxin regulates *Arabidopsis* anther dehiscence, pollen maturation, and filament elongation. *Plant Cell* 20, 1760–1774. doi: 10.1105/tpc.107.057570

Chen, M. S., Pan, B. Z., Fu, Q., Tao, Y. B., Martínez-Herrera, J., Niu, L., et al. (2017). Comparative transcriptome analysis between gynoecious and monoecious plants identifies regulatory networks controlling sex determination in *Jatropha curcas*. *Front Plant Sci.* 7:1953.

Chen, M. S., Pan, B. Z., Wang, G. J., Ni, J., Niu, L., and Xu, Z. F. (2014). Analysis of the transcriptional responses in inflorescence buds of *Jatropha curcas* exposed to cytokinin treatment. *BMC Plant Biol.* 14:318.

Chen, Y.-T., Chang, C.-C., Chen, C.-W., Chen, K.-C., and Chu, Y.-W. (2019). MADS-Box gene classification in angiosperms by clustering and machine learning approaches. *Front Genet.* 9:707. doi: 10.3389/fgene.2018.00707

Cheng, Y., Dai, X., and Zhao, Y. (2006). Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in *Arabidopsis*. *Genes Dev.* 20, 1790–1799. doi: 10.1101/gad.1415106

Chu, Z., Yuan, M., Yao, J., Ge, X., Yuan, B., Xu, C., et al. (2006). Promoter mutations of an essential gene for pollen development result in disease resistance in rice. *Genes Dev.* 20, 1250–1255. doi: 10.1101/gad.1416306

Deng, Y., Wang, W., Li, W. Q., Xia, C., Liao, H. Z., Zhang, X. Q., et al. (2010). MALE GAMETOPHYTE DEFECTIVE 2, encoding a sialyltransferase-like protein, is required for normal pollen germination and pollen tube growth in *Arabidopsis*. *J. Integr. Plant Biol.* 52, 829–843. doi: 10.1111/j.1744-7909.2010.00963.x

Duan, Q., Wang, D., Xu, Z., and Bai, S. (2008). Stamen development in *Arabidopsis* is arrested by organ-specific overexpression of a cucumber ethylene synthesis gene CsACO2. *Planta* 228, 537–543. doi: 10.1007/s00425-008-0756-7

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Dittam, G., et al. (2003). MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell* 15, 2603–2611. doi: 10.1105/tpc.015123

Gangwar, M., Shankar, J., and Chauhan, R. S. (2018). *Genomics of Female Flowering and Seed Yield in* Jatropha Curcas L. Ph.D. Thesis, Jaypee University of Information Technology, Solan.

Golovkin, M., and Reddy, A. S. (2003). A calmodulin-binding protein from *Arabidopsis* has an essential role in pollen germination. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10558–10563. doi: 10.1073/pnas.1734110100

Gross, T., Broholm, S., and Becker, A. (2018). CRABS CLAW acts as a bifunctional transcription factor in flower development. *Front. Plant Sci.* 9:835. doi: 10.3389/fpls.2018.00835

Guo, J., Liu, C., Wang, P., Cheng, Q., Sun, L., Yang, W., et al. (2018). The aborted microspores (AMS)-like gene is required for anther and microspore development in pepper (*Capsicum annuum* L.). *Int. J. Mol. Sci.* 19:1341. doi: 10.3390/ijms19051341

Honma, T., and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409, 525–529. doi: 10.1038/35054083

Hu, Y. X., Tao, Y. B., and Xu, Z. F. (2017). Overexpression of Jatropha Gibberellin 2-oxidase 6 (JcGA2ox6) induces dwarfism and smaller leaves, flowers and fruits in *Arabidopsis* and Jatropha. *Front. Plant Sci.* 8:2103.

Huang, J., Li, Z., and Zhao, D. (2016). Deregulation of the OsmiR160 target gene OsARF18 causes growth and developmental defects with an alteration of auxin signaling in rice. *Sci. Rep.* 6:29938.

Hui, W., Yang, Y., Wu, G., Peng, C., Chen, X., and Zayed, M. Z. (2017). Transcriptome profile analysis reveals the regulation mechanism of floral sex differentiation in *Jatropha curcas* L. *Sci. Rep.* 7:16421.

Hui, W. K., Wang, Y., Chen, X. Y., Zayed, M. Z., and Wu, G. J. (2018). Analysis of transcriptional responses of the inflorescence meristems in *Jatropha curcas* following gibberellin treatment. *Int. J. Mol. Sci.* 19:432. doi: 10.3390/ijms19020432

Imaizumi, T. (2010). *Arabidopsis* circadian clock and photoperiodism: time to think about location. *Curr. Opin. Plant Biol.* 13, 83–89. doi: 10.1016/j.pbi.2009.09.007

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

Joshi, G., Shukla, A., and Shukla, A. (2011). Synergistic response of auxin and ethylene on physiology of *Jatropha curcas* L. *Braz. J. Plant Physiol.* 23, 66–77. doi: 10.1590/s1677-04202011000100009

Karlgren, A., Gyllenstrand, N., Clapham, D., and Lagercrantz, U. (2013). FLOWERING LOCUS T/TERMINAL FLOWER1-like genes affect growth rhythm and bud set in Norway spruce. *Plant Physiol.* 163, 792–803. doi: 10.1104/pp.113.224139

Kater, M. M., Franken, J., Carney, K., Colombo, L., and Angenent, G. C. (2001). Sex determination in the monoecious species cucumber is confined to specific floral whorls. *Plant Cell* 13, 481–493. doi: 10.1105/tpc.13.3.481

Kumar, A., and Sharma, S. (2008). An evaluation of multipurpose oil seed crop for industrial uses (*Jatropha curcas* L): a review. *Ind. Crops Prod.* 28, 1–10. doi: 10.1016/j.indcrop.2008.01.001

Kurakawa, T., Ueda, N., Maekawa, M., Kobayashi, K., Kojima, M., Nagato, Y., et al. (2007). Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature* 445:652. doi: 10.1038/nature05504

ACKNOWLEDGMENTS

Leseberg, C. H., Li, A. L., Kang, H., Duvall, M., and Mao, L. (2006). Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*. *Gene* 378, 84–94. doi: 10.1016/j.gene.2006.05.022

Li, C., Fu, Q., Niu, L., Luo, C., Chen, J., and Xu, Z. F. (2017). Three TFL1 homologues regulate floral initiation in the biofuel plant *Jatropha curcas*. *Sci. Rep.* 7:43090.

Li, C., Luo, L., Fu, Q., Niu, L., and Xu, Z. F. (2014). Isolation and functional characterization of JcFT, a flowering LOCUS T (FT) homologous gene from the biofuel plant *Jatropha curcas*. *BMC Plant Biol.* 14:125. doi: 10.1186/1471-2229-14-125

Li, C., Schilmiller, A. L., Liu, G., Lee, G. I., Jayanty, S., Sageman, C., et al. (2005). Role of β-oxidation in jasmonate biosynthesis and systemic wound signaling in tomato. *Plant Cell* 17, 971–986. doi: 10.1105/tpc.104.029108

Li, J., and Li, Q. (2009). The correlatation between flowering sequence and floral gender in the inflorescence of *Jatropha curcas* L. (Euphorbiaceae). *J. Trop. Subtrop. Bot.* 17, 1–4.

Li, X. G., Su, Y. H., Zhao, X. Y., Li, W., Gao, X. Q., and Zhang, X. S. (2010). Cytokinin overproduction caused alteration of flower development is partially mediated by CUC2 and CUC3 in *Arabidopsis*. *Gene* 450, 109–120. doi: 10.1016/j.gene.2009.11.003

Liu, X., Huang, J., Parameswaran, S., Ito, T., Seubert, B., Auer, M., et al. (2009). The SPOROCYTELESS/NOZZLE gene is involved in controlling stamen identity in *Arabidopsis*. *Plant Physiol.* 151, 1401–1411. doi: 10.1104/pp.109.145896

Luo, C., Li, K., Chen, Y., and Yong, S. (2007). Floral display and breeding system of *Jatropha curcas* L. *Forest Stud. China* 9, 114–119. doi: 10.1007/s11632-007-0017-z

Maier, A. T., Stehling-Sun, S., Wollmann, H., Demar, M., Hong, R. L., Haubeiss, S., et al. (2009). Dual roles of the bZIP transcription factor PERIANTHIA in the control of floral architecture and homeotic gene expression. *Development* 136, 1613–1620. doi: 10.1242/dev.033647

Makkena, S., Lee, E., Sack, F. D., and Lamb, R. S. (2012). The R2R3 MYB transcription factors four lips and myb88 regulate female reproductive development. *J. Exp. Bot.* 63, 5545–5558. doi: 10.1093/jxb/ers209

Martínez-Fernández, I., Sanchís, S., Marini, N., Balanzá, V., Ballester, P., Navarrete-Gómez, J., et al. (2014). The effect of NGATHA altered activity on auxin signaling pathways within the *Arabidopsis gynoecium*. *Front. Plant Sci.* 5:210.

Mizoguchi, T., Wright, L., Fujiwara, S., and Cremer, F. (2005). Coupland, distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in *Arabidopsis*. *Plant Cell* 17, 2255–2270. doi: 10.1105/tpc.105.033464

Nakamura, T., Schuster, G., Sugiura, M., and Sugita, M. (2004). Chloroplast RNA binding and pentatricopeptide repeat proteins. *Biochem. Soc. Trans.* 32, 571–574. doi: 10.1042/bst0320571

Pan, B. Z., and Xu, Z. F. (2011). Benzyladenine treatment significantly increases the seed yield of the biofuel plant *Jatropha curcas*. *J. Plant Growth Regul.* 30, 166–174. doi: 10.1007/s00344-010-9179-3

Pan, B. Z., Chen, M. S., Ni, J., and Xu, Z. F. (2014). Transcriptome of the inflorescence meristems of the biofuel plant *Jatropha curcas* treated with cytokinin. *BMC Genomics* 15:974. doi: 10.1186/1471-2164-15-974

Park, J. H., Halitschke, R., Kim, H. B., Baldwin, I. T., Feldmann, K. A., and Feyereisen, R. (2002). A knock-out mutation in allene oxide synthase results in male sterility and defective wound signal transduction in *Arabidopsis* due to a block in jasmonic acid biosynthesis. *Plant J.* 31, 1–12. doi: 10.1046/j.1365-313x.2002.01328.x

Pařenicová, L., de Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., et al. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell* 15, 1538–1551. doi: 10.1105/tpc.011544

Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., and Yanofsky, M. F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* 405, 200–203. doi: 10.1038/35012103

Pinyopich, A., Ditta, G. S., Savidge, B., Liljegren, S. J., Baumann, E., Wisman, E., et al. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* 424, 85–88. doi: 10.1038/nature01741

Prunet, N., Yang, W., Das, P., Meyerowitz, E. M., and Jack, T. P. (2017). SUPERMAN prevents class B gene expression and promotes stem cell termination in the fourth whorl of *Arabidopsis thaliana* flowers. *PNAS* 114, 7166–7171. doi: 10.1073/pnas.1705977114

Ramos, M. J. N., Coito, J. L., Silva, H. G., Cunha, J., Costa, M. M. R., and Rocheta, M. (2014). Flower development and sex specification in wild grapevine. *BMC Genomics* 15:1095. doi: 10.1186/1471-2164-15-1095

Rodríguez-Concepción, M., and Beltrán, J. P. (1995). Repression of the pea lipoxygenase gene loxg is associated with carpel development. *Plant Mol. Biol.* 27, 887–899. doi: 10.1007/bf00037017

Roxrud, I., Lid, S. E., Fletcher, J. C., Schmidt, E. D., and Opsahl-Sorteberg, H. G. (2007). GASA4, one of the 14-member *Arabidopsis* GASA family of small polypeptides, regulates flowering and seed development. *Plant Cell Physiol.* 48, 471–483. doi: 10.1093/pcp/pcm016

Salvador-Figueroa, M., Magaña-Ramos, J., Vázquez-Ovando, J. A., Adriano-Anaya, M. L., and Ovando-Medina, I. (2015). Genetic diversity and structure of *Jatropha curcas* L. in its centre of origin. *Plant Genet Resourc.* 13, 9–17. doi: 10.1017/S1479262114000550

Seesangboon, A., Gruneck, L., Pokawattana, T., Eungwanichayapant, P. D., Tovaranonte, J., and Popluechai, S. (2018). Transcriptome analysis of *Jatropha curcas* L. flower buds responded to the paclobutrazol treatment. *Plant Physiol. Biochem.* 127, 276–286. doi: 10.1016/j.plaphy.2018.03.035

Shafiee, S., and Topal, E. (2009). When will fossil fuel reserves be diminished? *Energy Pol.* 37, 181–189. doi: 10.1016/j.enpol.2008.08.016

Stepanova, A. N., Robertson-Hoyt, J., Yun, J., Benavente, L. M., Xie, D.-Y., Dolezal, K., et al. (2008). TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant development. *Cell* 133, 177–191. doi: 10.1016/j.cell.2008.01.047

Stepanova, A. N., Yun, J., Robles, L. M., Novak, O., He, W., Guo, H., et al. (2011). The *Arabidopsis* YUCCA1 flavin monooxygenase functions in the indole-3-pyruvic acid branch of auxin biosynthesis. *Plant Cell* 23, 3961–3973. doi: 10.1105/tpc.111.088047

Switzenberg, J., Little, H., Hammar, S., and Grumet, R. (2014). Floral primordia-targeted ACS (1-aminocyclopropane-1-carboxylate synthase) expression in transgenic *Cucumis melo* implicates fine tuning of ethylene production mediating unisexual flower development. *Planta* 240, 797–808. doi: 10.1007/s00425-014-2118-y

Tang, M., Tao, Y. B., Fu, Q., Song, Y., Niu, L., and Xu, Z. F. (2016). An ortholog of LEAFY in *Jatropha curcas* regulates flowering time and floral organ development. *Sci. Rep.* 6:37306.

Ueguchi-Tanaka, M., Nakajima, M., Katoh, E., Ohmiya, H., Asano, K., Saji, S., et al. (2007). Molecular interactions of a soluble gibberellin receptor, GID1, with a Rice DELLA protein, SLR1, and gibberellins. *Plant Cell* 19, 2140–2155. doi: 10.1105/tpc.106.043729

Wang, D. H., Li, F., Duan, Q. H., Han, T., Xu, Z. H., and Bai, S. N. (2010). Ethylene perception is involved in female cucumber flower development. *Plant J.* 61, 862–872. doi: 10.1111/j.1365-313x.2009.04114.x

Werner, T., and Schmülling, T. (2009). Cytokinin action in plant development. *Curr. Opin. Plant Biol.* 12, 527–538. doi: 10.1016/j.pbi.2009.07.002

Wu, J., Liu, Y., Tang, L., Zhang, F., and Chen, F. (2011). A study on structural features in early flower development of *Jatropha curcas* L. and the classification of its inflorescence. *Afr. J. Agric. Res.* 6, 275–284.

Xu, G., Huang, J., Yang, Y., and Yao, Y. A. (2016). Transcriptome analysis of flower sex differentiation in *Jatropha curcas* L. using RNA sequencing. *PLoS One* 11:e0145613. doi: 10.1371/journal.pone.0145613

Xu, Y., Prunet, N., Gan, E. S., Wang, Y., Stewart, D., and Wellmer, F. (2018). SUPERMAN regulates floral whorl boundaries through control of auxin biosynthesis. *EMBO J.* 37:e97499.

Yamaki, S., Nagato, Y., Kurata, N., and Nonomura, K. (2011). Ovule is a lateral organ finally differentiated from the terminating floral meristem in rice. *Dev. Biol.* 351, 208–216. doi: 10.1016/j.ydbio.2010.12.006

Yang, J., Yang, M. F., Wen, P. Z., Fan, C., and Shen, S. H. (2011). A putative flowering-time-related Dof transcription factor gene, JcDof3, is controlled by the circadian clock in *Jatropha curcas*. *Plant Sci.* 181, 667–674. doi: 10.1016/j.plantsci.2011.05.003

Ye, J., Geng, Y., Zhang, B., Mao, H., Qu, J., and Chua, N. H. (2014). The jatropha FT ortholog is a systemic signal regulating growth and flowering time. *Biotechnol. Biofuels.* 7:91.

Ye, Q., Zhu, W., Li, L., Zhang, S., Yin, Y., Ma, H., et al. (2010). Brassinosteroids control male fertility by regulating the expression of key genes involved in *Arabidopsis* anther and pollen development. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6100–6105.

Zhu, J., Chen, H., Li, H., Gao, J. F., Jiang, H., Wang, C., et al. (2008). Defective in tapetal development and function 1 is essential for anther development and petal function for microspore maturation in *Arabidopsis*. *Plant J.* 55, 266–277. doi: 10.1111/j.1365-313X.2008.03500.x

Check for
updates

Emerging Technologies to Enable Sustainable Controlled Environment Agriculture in the Extreme Environments of Middle East-North Africa Coastal Regions

Ryan M. Lefers[1,2,3,4], Mark Tester[1,3] and Kyle J. Lauersen[1]*

[1] Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [2] Water Desalination and Reuse Center, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [3] Center for Desert Agriculture, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [4] Texas AgriLife Research and Extension Center at Dallas, Texas A&M University, Dallas, TX, Unites States

Despite global shifts in attitudes toward sustainability and increasing awareness of human impact on the environment, projected population growth and climate change require technological adaptations to ensure food and resource security at a global scale. Although desert areas have long been proposed as ideal sites for solar electricity generation, only recently have efforts shifted toward development of specialized and regionally focused agriculture in these extreme environments. In coastal regions of the Middle East and North Africa (MENA), the most abundant resources are consistent intense sunlight and saline sea water. MENA coastal regions hold incredible untapped potential for agriculture driven by the combination of key emerging technologies in future greenhouse concepts: transparent infrared collecting solar panels and low energy salt water cooling. These technologies can be combined to create greenhouses that drive regionally relevant agriculture in this extreme environment, especially when the target crops are salt-tolerant plants and algal biomass. Future controlled environment agriculture concepts will not compete for municipal fresh water and can be readily integrated into local human/livestock/fisheries food chains. With strategic technological implementation, marginal lands in these environments could participate in production of biomass, sustainable energy generation, and the circular carbon economy. The goal of this perspective is to reframe the idea of these environments as extreme, to having incredible untapped development potential.

Keywords: infrared solar, evaporative desiccant cooling, sustainability, combinatorial farming, algal biotechnology, salt water agriculture

ARTICLE

Global changes in mean surface temperatures are driving increased extreme environmental events which are beyond the tolerance of traditional agricultural practices, creating concerns of food insecurity (Hansen et al., 2012; Rhines and Huybers, 2013; Stone et al., 2013; Sutton et al., 2015; Bathiany et al., 2018; Spinoni et al., 2018). A logical step toward increasing agricultural yield

predictability is to move toward contained agriculture concepts like greenhouses. Glasshouses, or other controlled environment agriculture (CEA) structures, are traditionally applied to extend growth and cultivation periods in cooler climates, reducing frost damage to crops or for specialty/ornamental plant cultivation. CEA enables expansion of horticulture into non-traditional environments and marginal lands, benefits that can contribute to increasing output and food security. The application of CEA in hot or desert environments has been less common than in temperate climates or higher latitudes, owing to the energy required for cooling these structures. The Middle East and North Africa (MENA) regions have some of most extreme desert climates in the world, with high average annual temperatures and very low precipitation (Beck et al., 2018). However, in coastal regions, these environments are rich in two key resources: consistent solar radiation and sea water. Development of technological solutions which work with these resources in this coastal climate to drive sustainable agricultural practices can assist the region to meaningfully contribute to the global bio-economy and local food security.

The consistent solar radiation and strong daily winds of the MENA region can readily be used for sustainable electricity generation by traditional photovoltaics and wind turbines. The global horizontal irradiance (GHI) over the Arabian Peninsula shows significant seasonal variations with an annual mean ranging between 0.75 and 1.06 kW m^{-2}, with higher values over the northwestern region (Dasari et al., 2019). The wind resources over Saudi Arabia exhibit strong spatial variations, with high annual mean wind power density at 80 m height over the northern Red Sea (>0.8 kW m^{-2}), indicating the capacity for these technologies to support local CEA efforts (Langodan et al., 2016). However, for CEA systems to be practical in hot environments, technological solutions are required to minimize operational energy needs, especially for cooling. The combination of low-energy cooling systems with heat-reducing energy-generating technologies will improve overall CEA operational efficiencies and enable their implementation in MENA coastal regions. These systems become even more attractive when mixed salt-tolerant plant species and algal biomass cultivation are combined in high density cultivation concepts to reduce freshwater requirements. Intensive novel crop combinations in future CEA concepts may provide practical regional solutions for food generation while contributing to carbon capture and cycling. This work seeks to shift the perception of coastal MENA regions from extreme and inhospitable environments to locations with ample resources in the form of sunlight and sea water which could be developed into global agricultural powerhouses. These two resources can be combined to drive a regional agricultural revolution when appropriate technologies are implemented with strategic species selection.

Traditional agriculture in MENA regions includes cultivation of date palms (Erskine et al., 2004), and some species of plants which are tolerant to local environment such as *Salvadora* (miswak). Recent developments have seen increased efforts toward aquaculture farming (i.e., NAQUA farms, KSA)[1]. In addition, there are increasing outdoor horticultural efforts in various regions, with Saudi Arabia and Oman being the largest participants (Erskine et al., 2004; Noorka and Heslop-Harrison, 2015). Agriculture in these environments requires large inputs of freshwater, which is dominated by small amounts from desalination and significant extractions from aquifers, the latter being largely unsustainable (Gleeson et al., 2012). Indeed, 80% of freshwater resources in the Gulf region are used for agricultural practices (Erskine et al., 2004; Noorka and Heslop-Harrison, 2015). Hydroponic or CEA systems like those which are practiced in greenhouses use a fraction of the freshwater and waste less fertilizers than field agriculture. Furthermore, there is a sound economic and environmental case for CEA in this region for a large number of vegetables, and even some fresh fruits. We estimate that up to 70% (on a fresh weight basis) of fresh fruits and vegetables could be economically grown locally, primarily facilitated by use of CEA, with an overall lower environmental footprint than from the use of imported food. The contribution of implementation of this practice to local food security cannot be over-stated.

Two key technologies can enable low-energy greenhouses in hot coastal environments: these are efficient organic transparent infrared solar panels and liquid desiccant-based cooling (**Figure 1**). Newly reported advances in efficient infrared organic solar panels have shown key efficiency advances in the capture of latent heat energy to generate electricity (Song et al., 2018). Using a blend of 4% 1-chloronaphthalene as a solvent and the narrow-band-gap non-fullerene acceptor IEICO-4F, a thin material which absorbs maximally at 900 nm was shown to generate 26.8 mA cm^{-2} with photo conversion efficiencies of 12.8% (Song et al., 2018). The narrow bandgap of IEICO-4F allows penetration of photosynthetic wavelengths of light (400–700 nm) unlike commercial silicon technology and is suitable for roll-to-roll production processes. This technology will allow the surfaces of greenhouses and windows to generate electricity, while simultaneously serving as a transparent enclosure to enable plant and algal growth. Capture of infrared energy will also reduce heating effects of sun-light within the CEA structure, thereby reducing the energy required for cooling. Surfaces in MENA regions are also prone to dust buildup and advances in mechanical automated dusting devices can now be implemented to ensure efficient operation of photovoltaics and glasshouse surfaces[2].

Future CEA designs can be even more energy efficient when combined with low-energy air cooling technologies, especially liquid desiccant (Bettahalli et al., 2016; Lefers et al., 2016) or evaporative cooling (Chua et al., 1999; Lefers et al., 2018b; Shahzad et al., 2019). Liquid desiccant cooling relies on highly saline solutions that capture moisture from hot humid air and remove latent heat as moisture is absorbed by the desiccant. Moisture removal from humid air results in a pronounced cooling effect in the air passed through these structures as latent heat is removed (Lefers et al., 2016). Liquid desiccant systems can be further combined with sea water evaporative cooling systems to substitute the latent cooling achieved by the liquid desiccant system pending crop needs. Although sites will not be independent from municipal freshwater use,
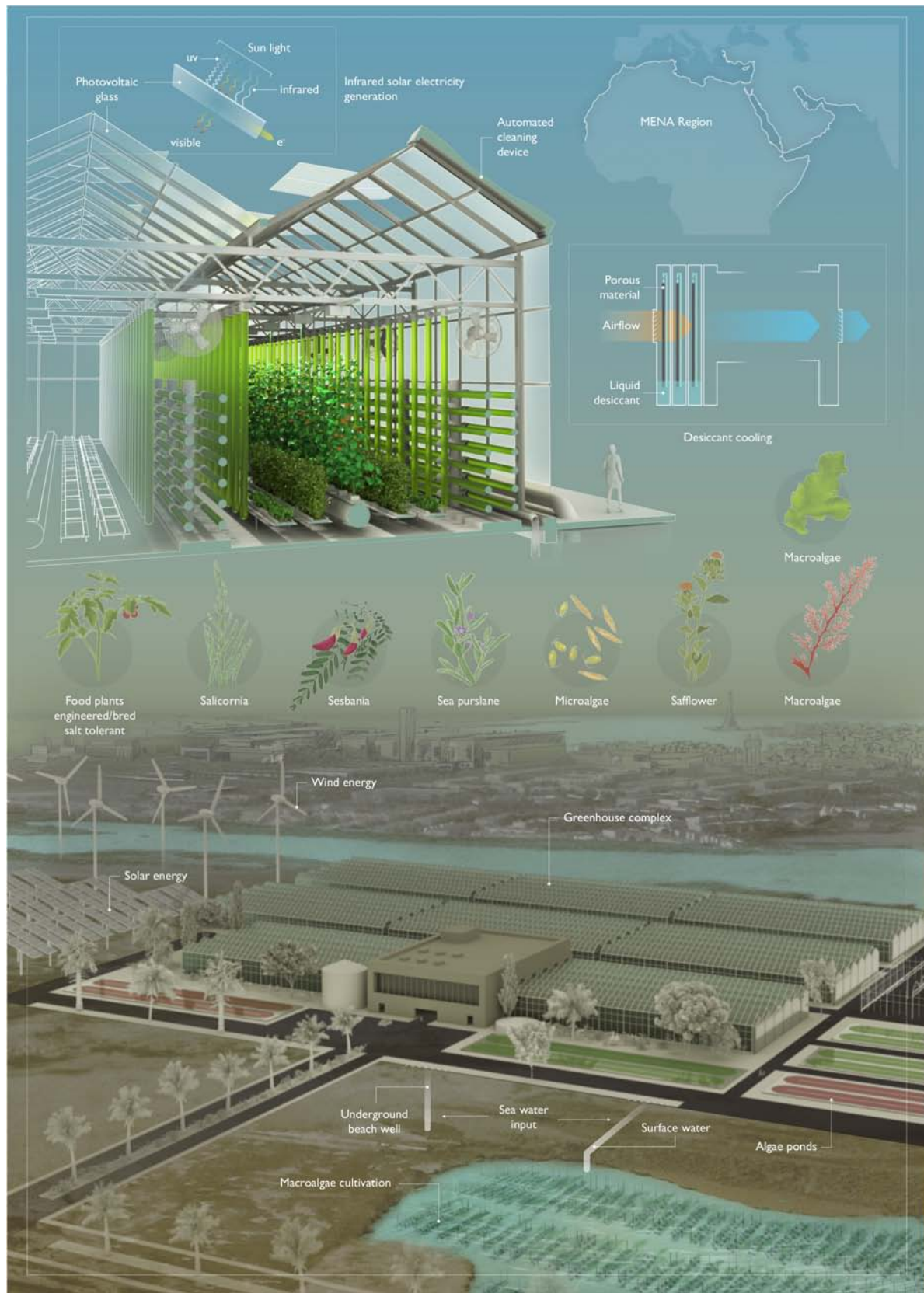
---

[1]http://www.naqua.com.sa

[2]http://www.nomaddesertsolar.com/

FIGURE 1 | Continued

FIGURE 1 | A low-energy glasshouse concept for future agriculture in coastal MENA regions. Future CEA greenhouses will combine infrared solar energy capture and desiccant cooling technologies to create stable contained environments for horticulture in extreme desert coastal environments. Infrared harvesting transparent solar panels allow photosynthetic active radiation (visible spectrum) to penetrate transparent glass surfaces to enable photosynthesis while simultaneously reducing the heating effect. Passive cooling can be achieved by passing hot external humid air through highly saline liquid desiccant solutions in porous matrices which adsorb air moisture, releasing dry, cooler air due to the vapor pressure difference. Coupling these technologies with high density hydroponic cultivation concepts and combined algae photobioreactors (green tubes) will maximize biomass productivity in these systems using seawater as cultivation medium. Macroalgae farming may also be an attractive addition to these concepts and can be coupled in managed pools on land or in the surrounding sea for nutrient removal and intensified biomass production. Plants which are naturally tolerant or those bred/engineered for salinity tolerance (pictured) can be cultivated with locally available sea water resources to minimize fresh-water requirements. Sustainable energy generation by traditional photovoltaics and wind turbines can be combined to support the energy requirements of these facilities.

vacuum regeneration of the liquid desiccant can be applied to yield additional freshwater as a by-product, which can support horticulture within the glasshouse (Lefers et al., 2018a). It is likely that future CEA concepts will be coupled to desalination efforts near local municipalities to supply the freshwater needed for human as well as some plant use. Evaporative cooling technologies utilizing sea water could also provide a lower energy solution and can be coupled to on site desalination plants, such as reverse osmosis and adsorption desalination, to provide more abundant freshwater resources without increasing energy demands and reducing the amount of brine from desalination processes (Ng et al., 2013). The combination of infrared solar cells on greenhouse surfaces with low-energy cooling and desalination technologies provides the preconditions for design of energy efficient and sustainable structures to enable CEA concepts in hot environments. In addition to water-use and temperature control, other challenges to address for CEA in MENA coastal regions include brine management, wastewater reuse, automation, and sustainable sourcing of plant nutrients. Needless to say, for widespread adoption of innovations, demonstration of cost-effectiveness at scale is essential. Our preliminary studies suggest that the modest increases in CapEx for such CEA systems are partially offset by reduced OpEx and reduced transportation. Overall, the extra costs incurred by use of salt water are very modest and do not have a significant effect on the overall business case for such greenhouses.

In addition to structural and technological considerations, selection of appropriate saline tolerant or drought resistant plant species is another key contributing factor in the success of future CEA in MENA coastal regions. New initiatives in the Gulf region are underway to promote desert agriculture using salt tolerant plant species such as *Salicornia*, the group of plants collectively known as sea purslane, *Chenopodium quinoa* (quinoa), miswak, *Sesbania*, *Triticale*, and *Carthamus* (safflower)[3]. The efficiency of water use could also be improved if appropriate saline-tolerant crops are cultivated in full or partial sea-water. Coastal regions are especially valuable in this context as beach wells can provide abundant naturally filtered, thermally consistent, saline water. Combination of freshwater plant species with saline-tolerant crops and algae in combined high-density cultivation concepts has the potential to generate robust biomass production processes in smaller land areas than by traditional field agriculture while using less overall freshwater resources. Plants such as grasses or quinoa are able to be cultivated in marginal lands and arid outdoor environments, and greenhouses are not necessary for their enhanced agricultural production. However, wild, bred or engineered food crops, like recently described saline tolerant tomato varieties (Pailles et al., 2020) and edible greens, will be well suited to high density cultivation in contained greenhouses in MENA coastal regions. The role of genetically modified (GM) crops in this region is a large topic which could be its own article, permissions for various crops have been granted in Egypt, Turkey, Sudan, Iran, and Pakistan[4]. However, the use of GMs is completely prohibited at the time of writing in countries like Saudi Arabia. CEA would allow some amounts of containment and minimize environmental risks/concerns over GM crops, however, improved hardiness would be less important in these controlled environments. GM may be interesting for nutritional improvements of crops grown in CEA, such as increased anthocyanin or phenylpropanoid contents of tomatoes (Butelli et al., 2008; Zhang et al., 2015). Modified traits may be more valuable for microalgal cultivation where increases in lipid content (Ajjawi et al., 2017) or novel traits (Lauersen, 2019; Fabris et al., 2020) are highly desired and add value to the biomass. Ongoing difficulties with consumer acceptance of GM organisms and complex country specific regulatory constraints makes widespread deployment of transgenics unlikely for the foreseeable future, so we do not discuss these issues further here. We also limit the discussion of field agriculture as our work focuses on future CEA. Additional efforts in improving rhizobial interactions and desert-probiotics for crops grown in harsh environments are also steadily developing, with promising results for encouraging heat tolerance and desiccation resistance under outdoor conditions (Bang et al., 2018; Daur et al., 2018; Eida et al., 2018). These efforts can serve as a roadmap for encouraging low-water use in field agriculture for some plants that are grown outdoors in harsh environments (Saad et al., 2020). Neo-domestication of other thermo and saline tolerant plant species as well as selective breeding/engineering could potentially increase productivities of these cultivation concepts and work is accelerating in this field (Lemmon et al., 2018; Li et al., 2018; Dawson et al., 2019; Fernie and Yan, 2019; Shan-e-Ali Zaidi et al., 2019).

In future salt-water driven CEA concepts, it is likely that salt-tolerant and fresh-water cultivars will be alternated to provide balanced supplies of both types of plants for food/feed and minimize fresh water demands. Potential exists here for the introduction of non-traditional agriculture in

---

the form of combinatorial cultivation concepts that integrate algal photobioreactors with higher plant hydroponics. Algae are rapidly growing photosynthetic organisms that can add increased productivity to the CEA system using full sea water as culture medium. Cultivation of algae is practiced both indoors and outdoors, with greenhouses providing improved environmental control as with higher plants (Posten, 2009). Integration of algal photobioreactors into future high-density CEA concepts could provide continual biomass generation for a range of applications and enhance areal carbon turnover rates (Lehr and Posten, 2009; Posten, 2009). Algal biomass can be used in aquaculture, animal feed, bioplastics and cosmetics, or to generate environmentally friendly replacements for plant-based oils (Radmer, 1996; Priyadarshani and Rath, 2012; Gangl et al., 2015; Chen et al., 2019). CEAs that include algal cultivation could act as sustainable sources for the natural inputs to greater bio-based industries, as algal cultivation can provide consistent biomass production at high turnover rates. Some reports even indicate co-cultivation of green algae like *Chlorella* and *Scenedesmus* together with the roots of higher plants in hydroponic systems has a dual benefit for both organisms which share growth promoting factors (Zhang et al., 2017; Barone et al., 2019). No work has yet been performed on combinatorial agriculture with algae and salt-water tolerant plants in controlled settings, which may be a new avenue for bio-fortification in greenhouse concepts. Industrial scale cultivation of marine algal strains such as *Nannochloropsis*, *Dunaliella*, and *Phaeodactylum* are already practiced in many locations globally both in CEA and outdoor cultivation (Laurens, 2017). In the MENA region there is potential to even further develop local strains of interest, for example, a recently described halotolerant *Chloroidium* sp. was isolated in the United Arab Emirates that has similar triacylglycerol profile to palm oil (Nelson et al., 2017). Its intensified cultivation could reduce global impact of palm agriculture as a sustainable alternative. Macroalgae may also be integrated as part of auxiliary value additions to coastal CEA concepts as they can be cultivated near shore or in pond units. Promising productivities in small scale for marine macroalgae *Asparagopsis armata* and *Ulva rigida* have been reported, and pending appropriate water flow rates, can serve as an effective biofilter to capture excess nutrients (nitrogen and phosphorous) in the form of valuable biomass (Mata et al., 2010). Very little work has been done to characterize macroalgae from the MENA region, with some studies of broad population dynamics now emerging (Geraldi et al., 2019; Ortega et al., 2019). It is likely that local bioprospecting will yield further species of interest which are adapted to regional climate conditions and can contribute to enhancing the productivities of coastal CEA hubs.

The circular carbon economy (CCE) is a concept which seeks to capture and capitalize on waste carbon which is otherwise lost to the atmosphere, usually in the form of carbon dioxide (CO_2) and reuse it in a cyclic fashion to minimize the environmental impacts of human activities (Stahel, 2016). This practice is of special importance to the MENA region as global economic and social trends look to the post-oil economies of the future. Plants and algae conduct light driven photosynthesis to generate cellular energy, and through the reactions of the Calvin-Benson-Bassham cycle are able to fix CO_2 into organic sugars for growth.

Photosynthesis-based fixation and cycling of CO_2 to biomass is one part of the greater spectrum of currently developing carbon capture and reuse technologies which are of interest for the development of the CCE. Plant and algal biomass represent a 1.83 weight ratio of fixed CO_2 per unit of biomass (Chisti, 2007). This ratio improves if the biomass is lipid or carbohydrate rich and offers a direct biological route from waste carbon to valuable bio-products. Plant and algal biomass, therefore, are incredible feedstocks for sustainable CCE practices as they represent carbon captured from the atmosphere that can be reused as physical commodities. Intensified CEA concepts which emphasize high biomass productivity in marginal lands will contribute significantly to CCE practices and provide a sustainable source of biological materials for various industries.

CONCLUSION AND OUTLOOK

Although greenhouses are not a new concept, emerging technologies now enable energy efficient and profitable implementation of CEA in extreme desert environments. Low-energy cooling and enhanced energy generation/temperature reduction by transparent infrared harvesting solar cells can be combined to create energy efficient greenhouses primed for future agriculture concepts on marginal coastal lands of the MENA region. The combination of high density hydroponic saline horticulture and algal cultivation can minimize impacts on freshwater water resources and maximize carbon cycling. The increased efficiency of these greenhouses can improve agricultural efforts in the MENA region, while contributing to food security and encouraging development of the CCE. It remains to be seen whether regulatory control and growing demand for locally sourced crops will enable MENA coastal regions to become hubs of future innovative agricultural practices.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Ajjawi, I., Verruto, J., Aqui, M., Soriaga, L. B., Coppersmith, J., Kwok, K., et al. (2017). Lipid production in Nannochloropsis gaditana is doubled by decreasing expression of a single transcriptional regulator. *Nat. Biotechnol.* 35, 647–652. doi: 10.1038/nbt.3865

Bang, C., Dagan, T., Deines, P., Dubilier, N., Duschl, W. J., Fraune, S., et al. (2018). Metaorganisms in extreme environments: do microbes play a role in organismal adaptation? *Zoology* 127, 1–19. doi: 10.1016/j.zool.2018.02.004

Barone, V., Puglisi, I., Fragalà, F., Lo Piero, A. R., Giuffrida, F., and Baglieri, A. (2019). Novel bioprocess for the cultivation of microalgae in hydroponic growing system of tomato plants. *J. Appl. Phycol.* 31, 465–470. doi: 10.1007/s10811-018-1518-y

Bathiany, S., Dakos, V., Scheffer, M., and Lenton, T. M. (2018). Climate models predict increasing temperature variability in poor countries. *Sci. Adv.* 4, 1–11. doi: 10.1126/sciadv.aar5809

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data* 5:180214. doi: 10.1038/sdata.2018.214

Bettahalli, N. M. S., Lefers, R., Fedoroff, N., Leikness, T. O., and Nunes, S. P. (2016). Triple-bore hollow fiber membrane contactor for liquid desiccant based air dehumidification. *J. Memb. Sci.* 514, 135–142. doi: 10.1016/j.memsci.2016.04.059

Butelli, E., Titta, L., Giorgio, M., Mock, H. P., Matros, A., Peterek, S., et al. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* 26, 1301–1308. doi: 10.1038/nbt.1506

Chen, H., Li, T., and Wang, Q. (2019). Ten years of algal biofuel and bioproducts: gains and pains. *Planta* 249, 195–219. doi: 10.1007/s00425-018-3066-8

Chisti, Y. (2007). Biodiesel from microalgae. *Biotechnol. Adv.* 25, 294–306. doi: 10.1016/j.biotechadv.2007.02.001

Chua, H. T., Ng, K. C., Malek, A., Kashiwagi, T., Akisawa, A., and Saha, B. B. (1999). Modeling the performance of two-bed, silica gel-water adsorption chillers. *Int. J. Refrig.* 22, 194–204. doi: 10.1016/S0140-7007(98)00063-2

Dasari, H. P., Desamsetti, S., Langodan, S., Attada, R., Kunchala, R. K., Viswanadhapalli, Y., et al. (2019). High-resolution assessment of solar energy resources over the Arabian Peninsula. *Appl. Energy* 248, 354–371. doi: 10.1016/j.apenergy.2019.04.105

Daur, I., Saad, M. M., Eida, A. A., Ahmad, S., Shah, Z. H., Ihsan, M. Z., et al. (2018). Boosting alfalfa (*Medicago sativa* L.) production with rhizobacteria from various plants in Saudi Arabia. *Front. Microbiol.* 9:477. doi: 10.3389/fmicb.2018.00477

Dawson, I. K., Powell, W., Hendre, P., Banèiè, J., Hickey, J. M., Kindt, R., et al. (2019). The role of genetics in mainstreaming the production of new and orphan crops to diversify food systems and support human nutrition. *New Phytol.* 224, 37–54. doi: 10.1111/nph.15895

Eida, A. A., Ziegler, M., Lafi, F. F., Michell, C. T., Voolstra, C. R., Hirt, H., et al. (2018). Desert plant bacteria reveal host influence and beneficial plant growth properties. *PLoS One* 13:e0208223. doi: 10.1371/journal.pone.0208223

Erskine, W., Moustafa, A. T., Osman, A. E., Lashine, Z., Nejatian, A., Badawi, T., et al. (2004). "Date palm in the GCC countries of the Arabian Peninsula," in *Proceedings of the Regional Workshop on Date Palm Development in the Arabian Peninsula*, Cairo.

Fabris, M., George, J., Kuzhiumparambil, U., Lawson, C. A., Jaramillo-Madrid, A.-C. C., Abbriano, R. M., et al. (2020). Extrachromosomal genetic engineering of the marine diatom phaeodactylum tricornutum enables the heterologous production of monoterpenoids. *ACS Synth. Biol.* 9, 598–612. doi: 10.1021/acssynbio.9b00455

Fernie, A. R., and Yan, J. (2019). De novo domestication: an alternative route toward new crops for the future. *Mol. Plant.* 12, 615–631. doi: 10.1016/j.molp.2019.03.016

Gangl, D., Zedler, J. A. Z., Rajakumar, P. D., Martinez, E. M. R., Riseley, A., Włodarczyk, A., et al. (2015). Biotechnological exploitation of microalgae. *J. Exp. Bot.* 66, 6975–6990. doi: 10.1093/jxb/erv426

Geraldi, N. R., Ortega, A., Serrano, O., Macreadie, P. I., Lovelock, C. E., Krause-Jensen, D., et al. (2019). Fingerprinting blue carbon: rationale and tools to determine the source of organic carbon in marine depositional environments. *Front. Mar. Sci.* 6:263. doi: 10.3389/fmars.2019.00263

Gleeson, T., Wada, Y., Bierkens, M. F. P., and Van Beek, L. P. H. (2012). Water balance of global aquifers revealed by groundwater footprint. *Nature* 488, 197–200. doi: 10.1038/nature11295

Hansen, J., Sato, M., and Ruedy, R. (2012). Public perception of climate change and the new climate dice. *arxiv* [Preprint]. doi: 10.1073/pnas.1205276109

Langodan, S., Viswanadhapalli, Y., Dasari, H. P., Knio, O., and Hoteit, I. (2016). A high-resolution assessment of wind and wave energy potentials in the Red Sea. *Appl. Energy* 181, 244–255. doi: 10.1016/j.apenergy.2016.08.076

Lauersen, K. J. (2019). Eukaryotic microalgae as hosts for light-driven heterologous isoprenoid production. *Planta* 249, 155–180. doi: 10.1007/s00425-018-3048-x

Laurens, L. M. L. (2017). *State of Technology Review - Algae Bioenergy.* Available online at: http://www.ieabioenergy.com/wp-content/uploads/2017/02/IEA-Bioenergy-Algae-report-update-Final-template-20170131.pdf (accessed January 14, 2020).

Lefers, R., Bettahalli, N. M. S., Fedoroff, N., Nunes, S. P., and Leikness, T. O. (2018a). Vacuum membrane distillation of liquid desiccants utilizing hollow fiber membranes. *Sep. Purif. Technol.* 199, 57–63. doi: 10.1016/j.seppur.2018.01.042

Lefers, R., Davies, P. A., Fedoroff, N. V., Almadhoun, N., Tester, M. A., and Leikness, T. (2018b). Proof of concept: pozzolan bricks for saline water evaporative cooling in controlled environment agriculture. *Appl. Eng. Agric.* 34, 929–937. doi: 10.13031/aea.13013

Lefers, R., Bettahalli, N. M. S., Nunes, S. P., Fedoroff, N., Davies, P. A., and Leikness, T. O. (2016). Liquid desiccant dehumidification and regeneration process to meet cooling and freshwater needs of desert greenhouses. *Desalin. Water Treat.* 57, 23430–23442. doi: 10.1080/19443994.2016.1173383

Lehr, F., and Posten, C. (2009). Closed photo-bioreactors as tools for biofuel production. *Curr. Opin. Biotechnol.* 20, 280–285. doi: 10.1016/j.copbio.2009.04.004

Lemmon, Z. H., Reem, N. T., Dalrymple, J., Soyk, S., Swartwood, K. E., Rodriguez-Leal, D., et al. (2018). Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat. Plants* 4, 766–770. doi: 10.1038/s41477-018-0259-x

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., et al. (2018). Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* 36, 1160–1163. doi: 10.1038/nbt.4273

Mata, L., Schuenhoff, A., and Santos, R. (2010). A direct comparison of the performance of the seaweed biofilters, *Asparagopsis armata* and Ulva rigida. *J. Appl. Phycol.* 22, 639–644. doi: 10.1007/s10811-010-9504-z

Nelson, D. R., Khraiwesh, B., Fu, W., Alseekh, S., Jaiswal, A., Chaiboonchoe, A., et al. (2017). The genome and phenome of the green alga chloroidium sp. UTEX 3007 reveal adaptive traits for desert acclimatization. *eLife* 6, 1–24. doi: 10.7554/eLife.25783

Ng, K. C., Thu, K., Kim, Y., Chakraborty, A., and Amy, G. (2013). Adsorption desalination: an emerging low-cost thermal desalination method. *Desalination* 308, 161–179. doi: 10.1016/j.desal.2012.07.030

Noorka, I. R., and Heslop-Harrison, J. S. (2015). "Agriculture and climate change in Southeast Asia and the middle east: breeding, climate change adaptation, agronomy, and water security," in *Handbook of Climate Change Adaptation*, ed. W. Leal Filho (Berlin: Springer), 1511–1519. doi: 10.1007/978-3-642-38670-1_74

Ortega, A., Geraldi, N. R., Alam, I., Kamau, A. A., Acinas, S. G., Logares, R., et al. (2019). Important contribution of macroalgae to oceanic carbon sequestration. *Nat. Geosci.* 12, 748–754. doi: 10.1038/s41561-019-0421-8

Pailles, Y., Awlia, M., Julkowska, M., Passone, L., Zemmouri, K., Negrão, S., et al. (2020). Diverse traits contribute to salinity tolerance of wild tomato seedlings from the Galapagos Islands. *Plant Physiol.* 182, 534–546. doi: 10.1104/pp.19.00700

Posten, C. (2009). Design principles of photo-bioreactors for cultivation of microalgae. *Eng. Life Sci.* 9, 165–177. doi: 10.1002/elsc.200900003

Priyadarshani, I., and Rath, B. (2012). Commercial and industrial applications of micro algae – A review. *J. Algal Biomass Utln* 3, 89–100.

Radmer, R. J. (1996). Algal diversity and commercial algal products. *Bioscience* 46, 263–270. doi: 10.2307/1312833

Rhines, A., and Huybers, P. (2013). Frequent summer temperature extremes reflect changes in the mean not the variance. *Proc. Natl. Acad. Sci. U.S.A.* 110:E546. doi: 10.1073/pnas.1218748110

Saad, M. M., Eida, A. A., and Hirt, H. (2020). Tailoring plant-associated microbial inoculants in agriculture: a roadmap for successful application. *J. Exp. Bot.* 11:eraa111. doi: 10.1093/jxb/eraa111

Shahzad, M. W., Burhan, M., Ybyraiymkul, D., Oh, S. J., and Ng, K. C. (2019). An improved indirect evaporative cooler experimental investigation. *Appl. Energy* 256:113934. doi: 10.1016/j.apenergy.2019.113934

Shan-e-Ali Zaidi, S., Vanderschuren, H., Qaim, M., Mahfouz, M. M., Kohli, A., Mansoor, S., et al. (2019). New plant breeding technologies for food security. *Science* 363, 1390–1391. doi: 10.1126/science.aav6316

Song, X., Gasparini, N., Ye, L., Yao, H., Hou, J., Ade, H., et al. (2018). Controlling blend morphology for ultrahigh current density in nonfullerene acceptor-based organic solar cells. *ACS Energy Lett.* 3, 669–676. doi: 10.1021/acsenergylett.7b01266

Spinoni, J., Vogt, J. V., Naumann, G., Barbosa, P., and Dosio, A. (2018). Will drought events become more frequent and severe in Europe? *Int. J. Climatol.* 38, 1718–1736. doi: 10.1002/joc.5291

Stahel, W. R. (2016). The circular economy. *Nature* 531, 435–438. doi: 10.1038/531435a

Stone, D. A., Paciorek, C. J., Prabhat, Pall, P., and Wehner, M. (2013). Inferring the anthropogenic contribution to local temperature extremes. *Proc. Natl. Acad. Sci. U.S.A.* 110:E1543. doi: 10.1073/pnas.1221461110

Sutton, R., Suckling, E., and Hawkins, E. (2015). What does global mean temperature tell us about local climate?. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 373:20140426. doi: 10.1098/rsta.2014.0426

Zhang, J., Wang, X., and Zhou, Q. (2017). Co-cultivation of *Chlorella* spp and tomato in a hydroponic system. *Biomass Bioenergy* 97, 132–138. doi: 10.1016/j.biombioe.2016.12.024

Zhang, Y., Butelli, E., Alseekh, S., Tohge, T., Rallapalli, G., Luo, J., et al. (2015). Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nat. Commun.* 6, 1–11. doi: 10.1038/ncomms9635

Phosphorylation of ADP-Glucose Pyrophosphorylase During Wheat Seeds Development

Danisa M. L. Ferrero[1], Claudia V. Piattoni[1†], Matías D. Asencion Diez[1], Bruno E. Rojas[1], Matías D. Hartman[1†], Miguel A. Ballicora[2] and Alberto A. Iglesias[1*]

[1] Laboratorio de Enzimología Molecular, Instituto de Agrobiotecnología del Litoral (UNL-CONICET) & FBCB, Santa Fe, Argentina, [2] Department of Chemistry and Biochemistry, Loyola University Chicago, Chicago, IL, United States

Starch is the dominant reserve polysaccharide accumulated in the seed of grasses (like wheat). It is the most common carbohydrate in the human diet and a material applied to the bioplastics and biofuels industry. Hence, the complete understanding of starch metabolism is critical to design rational strategies to improve its allocation in plant reserve tissues. ADP-glucose pyrophosphorylase (ADP-Glc PPase) catalyzes the key (regulated) step in the synthetic starch pathway. The enzyme comprises a small (S) and a large (L) subunit forming an S_2L_2 heterotetramer, which is allosterically regulated by orthophosphate, fructose-6P, and 3P-glycerate. ADP-Glc PPase was found in a phosphorylated state in extracts from wheat seeds. The amount of the phosphorylated protein increased along with the development of the seed and correlated with relative increases of the enzyme activity and starch content. Conversely, this post-translational modification was absent in seeds from *Ricinus communis*. *In vitro*, the recombinant ADP-Glc PPase from wheat endosperm was phosphorylated by wheat seed extracts as well as by recombinant Ca^{2+}-dependent plant protein kinases. Further analysis showed that the preferential phosphorylation takes place on the L subunit. Results suggest that the ADP-Glc PPase is a phosphorylation target in seeds from grasses but not from oleaginous plants. Accompanying seed maturation and starch accumulation, a combined regulation of ADP-Glc PPase by metabolites and phosphorylation may provide an enzyme with stable levels of activity. Such concerted modulation would drive carbon skeletons to the synthesis of starch for its long-term storage, which later support seed germination.

Keywords: crop grasses, glucan accumulation, starch biosynthesis, post-translational modification, enzyme regulation

INTRODUCTION

Starch is a major product of photosynthesis performed by vascular plants and constitutes the foremost storage of carbon and energy in grasses (Jeon et al., 2010; Pandey et al., 2012; Tuncel and Okita, 2013; MacNeill et al., 2017; Goren et al., 2018). Chemically, starch is a mix of two homopolysaccharides, amylose and amylopectin; which are mainly composed of glucose units linked by α-1,4-bonds. Amylose is essentially a linear α-1,4-glucan with scarce branches of

α-1,6-bonds, while amylopectin contains a high number of α-1,6-linked branch points (Pandey et al., 2012; MacNeill et al., 2017; Goren et al., 2018). Cereals (wheat, maize, barley, and rice as more relevant in production) store starch in seed endosperm, and they supply more than half of the caloric demands in the world population (Tuncel and Okita, 2013; Goren et al., 2018). Enhanced biosynthesis of the polysaccharide greatly influences the grain yield of cereals having starch as the principal reserve compound (Jeon et al., 2010; Tuncel and Okita, 2013; MacNeill et al., 2017; Goren et al., 2018; Zi et al., 2018). Improvement (in quantity and quality) in the production of key harvestable grains is a challenge to be solved in the coming decades. Indeed, it is projected a critical demographic expansion together with increasing industrial requirements of feedstock for biofuels, bioplastics, and bioadhesives to cope with climate change by the mid-century (Morell and Myers, 2005; Godfray et al., 2010; Tuncel and Okita, 2013; Iglesias, 2015; Goren et al., 2018). In this scenario, the in-depth understanding of the process of starch biosynthesis is relevant to better design strategies to increase yields in plants of agronomic interest.

Starch accumulates in plastids of both photosynthetic and non-photosynthetic plant cells (Jeon et al., 2010; MacNeill et al., 2017; Goren et al., 2018). The linear α-1,4-glucan is elongated by the starch synthases, which use ADP-glucose (ADP-Glc) as the glycosyl donor molecule. Synthesis of this sugar nucleotide takes place from glucose-1-phosphate (Glc1P) and ATP in a reaction catalyzed by ADP-Glc pyrophosphorylase (EC 2.7.7.27, ADP-Glc PPase) (Ballicora et al., 2003; Ballicora et al., 2004). The metabolic route follows by the action of enzymes implicated in branching, debranching, phosphorylation, and de-phosphorylation of the polymer under formation (Wilkens et al., 2018). Most of these enzymes arise in isoforms exhibiting changes in specificity and ability to interact with different partners forming multi-protein complexes of functional relevance for the production of the polysaccharide (Crofts et al., 2017; Wilkens et al., 2018). The step producing ADP-Glc is rate-limiting in the starch biosynthetic pathway (Kavakli et al., 2001b; Ballicora et al., 2003; Ballicora et al., 2004; Tuncel and Okita, 2013). ADP-Glc PPase is present in bacteria (where it is involved in glycogen synthesis) and green plants. The enzyme from different sources is allosterically regulated by metabolites that are critical intermediates of the central carbon energy metabolism operating in the respective organism (Ballicora et al., 2003; Ballicora et al., 2004). ADP-Glc PPase from cyanobacteria (Iglesias et al., 1991), green algae (Iglesias et al., 1994), and higher plants (Ballicora et al., 2004) have 3P-glycerate (3PGA) and inorganic orthophosphate (Pi) as the principal allosteric activator and inhibitor, respectively; with some enzyme promiscuity toward being activated by hexose-Ps (Kuhn et al., 2013).

The ADP-Glc PPase presents in green algae and higher plants is a heterotetramer (S_2L_2), composed of small (S, 50-53 kDa) and large (L, 54-60 kDa) subunits (Ballicora et al., 2003; Ballicora et al., 2004). Subunits S and L are homologous proteins, where the L polypeptide has emerged in different species *via* gene duplication followed by subfunctionalization (Ballicora et al.,

2005; Georgelis et al., 2008; Georgelis et al., 2009; Kuhn et al., 2009; Ferrero et al., 2018; Figueroa et al., 2018). In potato tuber, the S subunit retained the catalytic function, whereas the L subunit specialized in modulating the regulation of the former (Ballicora et al., 2005). Even so, different alternatives exist where the L subunit is also catalytic, or it presents isoforms within the same organism (Crevillén et al., 2003; Ventriglia et al., 2008; Kuhn et al., 2009). The interaction between both subunits is determinant for the enzyme activity and regulatory responses (Kavakli et al., 2001a; Baris et al., 2009). As demonstrated in previous studies (Crevillén et al., 2003; Ferrero et al., 2018), the L subunit confers the sensitivity to allosteric regulators exhibited by heterotetrameric ADP-Glc PPases from Arabidopsis and wheat. Excluding the enzyme from monocot (as wheat) endosperm, the ADP-Glc PPase S subunit from leaves and other plant tissues has an N-terminal cysteine residue that is critical for redox regulation. This latter involves the formation of a disulfide bridge between the S subunits in the heterotetramer, mediated by the thioredoxin system (Ballicora et al., 1999; Ballicora et al., 2000; Tiessen et al., 2002). In addition, as it has been reviewed previously (Ballicora et al., 2004), ADP-Glc PPases found in cereal endosperm have distinct regulatory properties. In the barley and maize forms, 3PGA modifies the relative affinity for substrates (Plaxton and Preiss, 1987; Kleczkowski et al., 1993). Also, in wheat (*Triticum aestivum*) endosperm enzyme, neither 3PGA nor fructose-6P (Fru6P) modify the V_{max}, but they increase the relative affinity for Glc1P by 2-fold (Gómez-Casati and Iglesias, 2002; Ferrero et al., 2018). On the other hand, the 3PGA and Fru6P revert the inhibition by Pi (Gómez-Casati and Iglesias, 2002).

A substantial body of experimental evidence is giving support to the modulation of starch biosynthesis by the combined action of post-translational mechanisms, including redox modification, protein complex formation, and protein phosphorylation (Lohrig et al., 2009; Kötting et al., 2010; Geigenberger, 2011; Ma et al., 2014; Tetlow et al., 2015; Goren et al., 2018). In this context, proteomic information obtained in the last decade draws attention to ADP-Glc PPase as a putative target of protein kinases (Lohrig et al., 2009; Kötting et al., 2010; Ma et al., 2014). Specifically, a proteomic analysis of maize endosperm identified the phosphorylation of the small subunit of the enzyme (Yu et al., 2019). However, all these predictive results lack functional and developmental evidence of the tangible presence of the enzyme at a phosphorylated state *in planta*. The latter is critical for the complete understanding of factors affecting plant productivity, thus limiting the design of better strategies for its improvement. Herein, we report the *in vivo* phosphorylation of ADP-Glc PPase associated with the development of wheat seeds. This modification was further analyzed by *in vitro* studies working with recombinant enzymes, specifically wheat endosperm ADP-Glc PPase, as well as Ca^{2+}-dependent SOS2 and CDPK plant protein kinases. Results suggest that phosphorylation of the enzyme involved in the limiting step of starch built-up would be functionally relevant for the yield of grains in grass crops.

MATERIALS AND METHODS

Chemicals

ATP, Fru6P, 3PGA, and Glc1P were from Sigma Aldrich (St. Louis, MO, USA). All other reagents were of the highest quality available.

Seeds Harvest

Seeds harvesting was as described previously (Piattoni et al., 2017). Briefly, *Triticum aestivum* L. cv. Baguette 11 samples collected at 3, 6, 10, 14, 17, and 27 days post-anthesis (DPA), and spikes were frozen immediately in liquid nitrogen. This sampling was based on reports indicating that the complete growth of wheat seeds is reached at 45 DPA, establishing the following respective phases of development: cell proliferation 0–10 DPA, accumulation or reserves 11–30 DPA, and seed maturation and desiccation 30–45 DPA (http://bio-gromit.bio.bris.ac.uk/cerealgenomics/WheatBP/Documents/DOC_WheatBP.php). Wheat samples were grains taken from the central part of the frozen spike (between the fifth and tenth spike) stored at -80°C until analysis. Castor (*Ricinus communis*) seeds were collected at 5, 10, 20, 25, 32, and 40 days post-pollination (DPP). The sampling criterion considered that castor oil seed development completes in ~60 days (Canvin, 1963), after which six groups classified according to the morphology exhaustively described in (Greenwood and Bewley, 1982). Sampled seeds were dissected from the capsule, frozen immediately in liquid nitrogen, and store at -80°C until analysis.

Soluble Protein Extraction From Seeds

Seeds whole protein extraction was made by triplicate (using independent biological replicates) as reported elsewhere (Piattoni et al., 2017). Frozen seeds of wheat or castor oil were ground to a fine powder in liquid nitrogen using a mortar and pestle. It followed the addition of 1 µl cold fresh prepared extraction buffer per 1 mg of frozen powdered tissue. Composition of the extraction buffer was 50 mM MOPS pH 8.0, 1 mM EDTA, 1 mM EGTA, 25 mM NaF, 0,1% (v/v) Triton X-100, 20% (v/v) glycerol, 10 mM MgCl$_2$, 2 mM DTT, 4% (p/v) PEG-8000, 2 mM PMSF, 5 mM malic acid, and 1% (p/v) polyvinylpyrrolidone (Turner et al., 2005). This buffer was supplemented with 2 mM aminocaproic acid, 1 mM benzamidine, 10 mM NaF, 1 mM Na$_2$MoO$_4$, 1 mM Na$_2$VO$_4$, and SETIII (1X) protease inhibitor cocktail EDTA-Free (Calbiochem). The mixture was incubated 20 min on ice with constant homogenization. Extracts were centrifuged 30 min at 4°C and 15,000 × g. The supernatant was recovered and used immediately for the different assays.

Starch and Lipid Quantification

Quantification of starch from seeds was as indicated in (Piattoni et al., 2017), following a protocol that combines two previously described procedures (Reibach and Benedict, 1982; Baud and Graham, 2006). Plant tissue (100 mg) ground in a mortar under liquid nitrogen was soaked with 500 ml of ethanol 95% (v/v) at 4°C and then centrifuged 10 min at 15,000 × g and 4°C.

The supernatant was discarded and the extraction repeated thrice to eliminate soluble sugars. The pellets were dried at 60°C, weighed, and then dissolved in 10 ml of distilled H$_2$O per mg of extracted material. The resulting tubes (hermetically closed) boiled for 1 h to solubilize the starch, and then centrifuged during 10 min at 15,000 ×g and 4°C. The soluble fraction (20 µl) was added to 200 µl of 100 mM sodium acetate pH 4.5 plus 70 U of amyloglucosidase (1,4-α-D-glucan glucohydrolase) and incubated 16 h at 55–60°C for starch digestion. After centrifuging for 10 min at 15,000 × g and 4°C, the resulting soluble sugars were quantified by an enzymatic colorimetric assay where the H$_2$O$_2$ produced by glucose oxidase is measured by a peroxidase coupled to a colorimetric compound. The reaction mixture (100 µl) consisted of 70 µl of the commercial reactive (10 kU/l glucose oxidase, 1 kU/l peroxidase, 0.5 mM 4-aminophenazone, 100 mM phosphate buffer pH 7.0, and 12 mM 4-hydroxybenzoates) and 30 µl of sample conveniently diluted. The reaction was developed for 10 min at 37°C and the product quantified at 492 nm. To correlate the quantity of starch and the concentration of soluble sugars, we constructed a calibration curve [glucose (mg/ml) versus starch (mg)] with a standard starch solution treated identically to the sample.

To determine the contents of TAGs in seeds, we utilized protocols already described (Folch et al., 1957; Piattoni et al., 2017). Plant tissue (200 mg) was ground to a fine powder in liquid nitrogen and lipids extracted with 0.2 ml of MilliQ H$_2$O and 3.8 ml of chloroform/methanol: 2/1 (v/v) solution. The extraction was performed in hermetically closed tubes incubated 2 h at room temperature with gentle mixing every 30 min. Samples were filtered on oil-free filter paper previously washed with the extraction solution, placed in previously weighed tubes, and vigorously mixed with 0.7 ml of 0.02% CaCl$_2$ solution in chloroform/methanol/H$_2$O: 3/48/47 (v/v). After centrifugation at 3,000 × g for 5 min and discard the upper phase, 0.7 ml of chloroform/methanol/H$_2$O: 3/48/47 (v/v) were added, following vigorous mixing. Samples were centrifuged as before until phase separation, and the upper phase was discarded by suction, then evaporating chloroform at 45–50°C to obtain the lipids. After weighing, the lipids dissolved in isopropanol served to quantify TAGs by an enzymatic colorimetric assay. This method was based on the treatment of TAGs with lipase, then converting the produced glycerol into glycerol-3P by a specific kinase coupled to the reaction of glycerol-3P oxidase that produces H$_2$O$_2$. The latter was quantified by employing a peroxidase generating a compound that absorbs at 492 nm. The reaction mixture contained 50 mM PIPES pH 7.5, 5 mM 4-chlorophenol, 15 kU/l lipase, 1 kU/l glycerol-3P kinase, 2.5 U/ml glycerol-3P oxidase, 0.44 kU/l peroxidase, 0.7 mM 4-aminophenasone, 0.18 mM ATP, and the lipid sample.

All values of starch and lipids contents are the mean of at least three independent determinations, and reproducible within ±10%.

Enzyme Activity Assay

ADP-Glc PPase activity was determined at 37°C in ADP-Glc synthesis direction by following Pi formation (after hydrolysis of

PP$_i$ by inorganic pyrophosphatase) using the highly sensitive colorimetric method previously described (Fusari et al., 2006). The reaction mixture contained 100 mM MOPS pH 8.0, 7 mM MgCl$_2$, 1.5 mM ATP, 1.0 mM Glc1P, 0.2 mg/ml BSA, 0.5 U/ml yeast inorganic pyrophosphatase and a proper sample enzyme dilution. Assays initiated by the addition of Glc1P at a final concentration of 1.5 mM in a total volume of 50 µl. The reaction mixture was incubated for 10 min at 37°C and terminated by adding the Malachite Green reactive. The complex formed with the released P$_i$ was measured at 630 nm. One unit of activity (U) is the amount of enzyme catalyzing the formation of 1 µmol of ADP-Glc per minute under the described conditions. All determinations are the mean of at least three independent sets of data that were reproducible within ±10%.

Protein Methods

Total protein quantification was determined by the method of Bradford (Bradford, 1976) using BSA as a standard. Protein electrophoresis in polyacrylamide gels under denatured conditions (SDS-PAGE) was performed as previously described by Laemmli (Laemmli, 1970). Western blotting was performed by transferring the proteins resolved by SDS-PAGE to nitrocellulose membranes using a Mini-PROTEAN II (Bio-Rad) apparatus. The membrane was blocked 2 h at room temperature and subsequently incubated with primary antibody during 16 h at 15°C with agitation. The primary antibodies were raised in rabbits against the ADP-Glc PPase purified from spinach leaves (Gómez-Casati and Iglesias, 2002) or against each subunit of the wheat endosperm enzyme (*Tae*S and *Tae*L) produced in our laboratory. After intensive washing, membranes were incubated with HRP −conjugated anti−rabbit secondary antibody (Sigma) for 1 h. Bands were visualized using the ECL method and detection reagents (Thermo Scientific).

For the phosphorylated protein mobility delay assay, Phos-tag™ (Wako Chemicals) was added (100 µM) to the 10% (w/v) SDS-PAGE acrylamide gel. The acrylamide-pendant Phos-tag ligand provides a molecule with a functional affinity to interact with phosphate groups, after which produce electrophoretic mobility-shifts of phosphorylated proteins (Kinoshita et al., 2015). The phosphorylation reaction was carried out as described below except for the use of non-radioactive ATP and modifying the reaction time to 2 h with the addition of the recombinant kinase (and the corresponding amount of buffer according to the final volume change) every 30 min. Then, samples were denatured with SDS-PAGE buffer [1% SDS (w/v), 100 mM β-mercaptoethanol, in 50 mM TRIS-HCl pH 6.8] at 100°C for 5 min. After the electrophoretic run (30 mA per gel), gels were washed twice in methanol-free Tris-Gly transfer buffer and 1 mM EDTA, for 10 min each time with gentle agitation, followed other 10 min in methanol-free transfer buffer without EDTA. It followed electrotransference of the proteins and immunodetection with specific antibodies.

Phosphorylated Proteins Purification and De-Phosphorylation

Purification of phosphorylated proteins was performed by Fe^{3+}-Immobilized metal affinity chromatography (IMAC-Fe^{3+}), as

previously described (Muszyńska et al., 1986). Total proteins (1.2 mg) in 50 mM MES-NaOH pH 6.0 were loaded onto 100 ml of iminodiacetic acid-Fe^{3+} previously equilibrated with the same buffer. After 1 h incubation at room temperature and constant homogenization, non-adsorbed proteins were washed out twice with 2 ml of 50 mM MES-NaOH, pH 6.0. The increase of pH in three steps was then employed to elute the adsorbed. First, 5 volumes of 50 mM PIPES-HCl pH 7.2 were applied, then the adsorbed proteins were washed out with three volumes of 50 mM Tris-HCl pH 8.0, and finally phosphorylated proteins (tightly bound to the matrix) eluted with two volumes of 50 mM of Tris-HCl pH 9.0. SDS-PAGE served to resolve the samples eluted at the different pH conditions and for the analysis of the phosphorylated proteins.

The protein purification was performed from three independent biological replicates to enable an assessment of significance.

Protein de-phosphorylation experiments followed a protocol described elsewhere (Bustos and Iglesias, 2002). Proteins (600 µg) extracted from seeds were diluted in 50 mM Tris–HCl (pH 8.5), 1 mM EDTA, 10 mM MgCl$_2$, 1.2 mM CaCl$_2$, 20 mM 2-mercaptoethanol, 1mM PMSF, and 20 U of alkaline phosphatase (Promega). Samples were incubated at 37°C for 5 h, and the reaction stopped by the addition of 4X SDS-PAGE sample buffer. Then, samples were analyzed by electrophoresis followed by western blotting and immunodetection.

ADP-Glc PPase and Protein Kinases Cloning, Expression, and Purification

The genes coding for the small (*Tae*S) and large (*Tae*L) subunits of wheat endosperm ADP-Glc PPase were synthesized *de novo* from DNA sequence reported previously (Ainsworth et al., 1993 and Ainsworth et al., 1995) and subcloned into pET28c as reported elsewhere (Ferrero et al., 2018). This recombinant enzyme displayed similar properties to the ADP-Glc PPase purified from the wheat endosperm (Gómez-Casati and Iglesias, 2002). For individual expression, each subunit was subcloned into the pET19TEV vector using *Nde*I and *Xho*I. For coexpression, *Tae*S was subcloned from the pET28c plasmid into the pCDFDuet-1 vector using *Nde*I and *Xho*I sites. By combining the pCDFDuet-1/ *Tae*S and pET19TEV/*Tae*L constructs, we obtained the heterotetrameric wheat endosperm ADP-Glc PPase (*Tae*SL) with a His-tag only in the L subunit. All the sequences were confirmed by the University of Chicago DNA Sequencing Facility (Chicago, IL, United States).

The *Ath*SnRK1α1 gene amplified from *A. thaliana* cDNA served to generate the S198D mutation [required for its activity, as reported by (Shen et al., 2009)], by employing the QuickChange Site-Directed Mutagenesis kit (Agilent) (Kunkel et al., 1991). The mutated gene was subcloned in the pETDuet-1 vector to produce a protein with an N-term His-tag. The *Mdo*SOS2 gene was synthesized *de novo* with a sequence and codon usage optimized for its expression in *E. coli*. The T168D mutation [required for activity (Guo, 2001)] was introduced by quick-change mutagenesis and then subcloned into a pET28b vector to produce a protein with an N-term His-tag. Molecular

cloning of the gene coding for *Stu*CDPK1 was as already reported (Santin et al., 2017; Rojas et al., 2018).

Recombinant *Tae*S, *Tae*L and *Tae*SL were obtained from *E. coli* BL21(DE3) RIL cells transformed with [pET19TEV/*Tae*S], [pET19TEV/*Tae*L] and co-transformed with [pCDFDuet-1/ *Tae*S + pET19/*Tae*L], respectively. Recombinant *Ath*SnRK1α1 was obtained from *E. coli* BL21 Shuffle cells transformed with [pETDuet-1/*Ath*SnRK1α1]; *Mdo*SOS2 was obtained from *E. coli* BL21 Codon Plus cells transformed with [pET28b/*Mdo*SOS2] and *Stu*CDPK1 was obtained from *E. coli* BL21 Shuffle cells transformed with [pET22$^{(+)}$/*Stu*CDPK1]. Transformed cells were grown in 1 L of LB medium supplemented with the appropriate antibiotic (100 μg/ml ampicillin for pET19TEV, pET22$^{(+)}$, and pETDuet-1; 50 μg/ml kanamycin for pET28b, and 100 μg/ml for pCDFDuet-1) at 37 °C and 200 rpm until OD600 reached ~1.2. Cells were induced with 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) at 25°C and 180 rpm for 16 h and then harvested by centrifuging 10 min at 4°C and 5000 x*g*. The cells were resuspended in Buffer H [50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5% (v/v) glycerol] and disrupted by sonication on ice (4 s pulse on with intervals of 3 s pulse off for a total time of 5 min). After centrifuging twice at 30,000 x*g* for 10 min, the supernatant (crude extract) was loaded onto a 1 ml HisTrap column (GE Healthcare) previously equilibrated with Buffer H. The recombinant protein was eluted with a linear gradient from 10 to 300 mM imidazole in Buffer H, and fractions containing the highest activity (or more pure proteins) were pooled with 10% (v/v) glycerol and stored at -80°C until use. The pools of *Tae*SL, *Tae*L and *Tae*S, were dialyzed against Buffer X [50 mM MOPS pH 8.0, 0.1 mM EDTA, 20% (w/v) sucrose, 5 mM MgCl$_2$] and concentrated using an Amicon Ultra-4 30 K unit (Millipore, Billerica, MA, USA). Under these conditions the enzyme was fast frozen and stored at -80°C until use, being fully actives for at least 1 year (in the case of *Tae*SL).

In Vitro Phosphorylation Assays

For *in vitro* phosphorylation of recombinant forms of the wheat endosperm ADP-Glc PPase [the heterotetrameric (*Tae*SL), small subunit (*Tae*S), or large subunit (*Tae*L); see (Ferrero et al., 2018), the respective purified enzyme (2 μg) was incubated under conditions established for the activity of different protein kinases from plants (Piattoni et al., 2011; Rojas et al., 2018)]. The three (I–III) specific phosphorylation conditions used for assay crude extracts from seeds or purified recombinant protein kinase were as follows. I. For SnRK or Ca^{2+}-independent kinases: 100 mM HEPES pH 7.3, 5 mM DTT, 10 mM MgCl$_2$, 0.05 mM ATP, 0.5 mM EGTA. II. For CDPK or Ca^{2+}-dependent kinases: 20 mM Tris-HCl pH 7.5, 10 mM MgCl$_2$, 5 mM DTT, 1 mM CaCl$_2$, 0.05 mM ATP. III. For SOS2: 20 mM Tris-HCl pH 7.2, 5 mM MgCl$_2$, 5 mM DTT, 0.5 mM CaCl$_2$, 0.01 mM ATP, 2.5 mM MgCl$_2$. Worthy of mention is that the critical difference between these protein kinases is related to requirement (or not) of calcium. In contrast, all of them need reducing condition (5 mM DTT) for optimal activity (Raíces et al., 2001; Crozet et al., 2010; Gong et al., 2019). Each reaction (final volume 20 μl) was incubated at 30°C for 30 min after the addition of 1 μCi of [^{32}P]-γ-ATP (Migliore-Lacaustra) and initiated by adding

wheat endosperm extract as kinase resource, or an appropriate aliquot of the recombinant protein kinase [0.4 mU (determined for AMARA peptide)]. After the reaction, the protein mixtures were denatured with SDS-PAGE buffer [1% SDS (w/v), 100 mM β-mercaptoethanol, in 50 mM Tris-HCl pH 6.8]. It followed the resolution of the proteins by electrophoresis [SDS-PAGE with 10% (w/v) acrylamide and run at 30 mA/gel]. Then, gels were stained with Coomassie Brilliant Blue R-250, dried, and radioactivity incorporation detected by Storing Phospho-screen (GE Healthcare) exposure and scanning with the Typhoon^{TM} system (GE Healthcare). All phosphorylation assays were performed from three independent technical replicates to enable an assessment of significance.

RESULTS AND DISCUSSION

ADP-Glc PPase Is Phosphorylated Along With Development of the Seed in Wheat but Not in Castor Bean
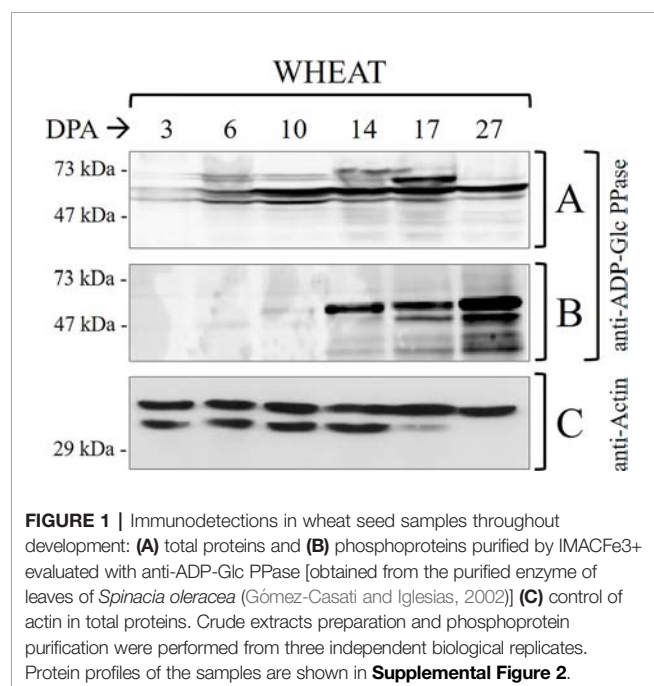
Seeds of grasses are relevant components within the world crops providing food and energy, with wheat being one of the principal edible grains. From this, it is clear the relevance of starch as a major staple. Another primary product of agriculture is TAGs accumulated in seeds of oleaginous plants. To better understand starch biosynthesis (and its regulation) in seeds, we explored about contents of the polysaccharide and TAGs, as well as ADP-Glc PPase activity and phosphorylation profiles along with the development of *Triticum aestivum* (wheat) seeds (**Table 1** and **Figure 1**). For comparison, we performed similar studies with seeds of the oleaginous *Ricinus communis* (castor bean) (**Supplemental Table 1** and **Supplemental Figure 1**). As shown in **Table 1**, wheat seeds stored significantly more starch than TAGs, with levels of the polysaccharide progressively increasing along the stages of development, reaching levels up to 30% of the seed weight during accumulation of reserves (between 11–30 DPA). These profiles are in agreement with values reported in previous studies on developing wheat endosperm (Briarty et al., 1979; Dai, 2010; Piattoni et al., 2017). Instead, castor bean seeds exhibited a progressive increase in the amount of starch during cell proliferation (0–20 DPP), but an abrupt decrease during accumulation of reserves (20–40 DPP) and maturation plus desiccation (30–50 DPP). As reported for the oleaginous plant (Canvin, 1963; Greenwood and Bewley, 1982), TAGs remained at a low level during cell proliferation switching to an increase during accumulation to reach up to 30% of the seed weight at 50 DPP (**Supplemental Table 1**).

It is worth correlating levels of starch and ADP-Glc PPase activity in **Table 1** with profiles of the enzyme immunodetected in extracts (either whole protein or phosphoprotein enriched after IMAC-Fe^{3+}) from wheat seeds at the different development stages detailed in **Figure 1**. As expected from the critical role played by ADP-Glc PPase in starch biosynthesis, the specific activity determined for the enzyme followed a moderate increase between the cell proliferation, transition, and maturation steps of development of wheat grains, accompanying levels of the

TABLE 1 | Determinations of weight, starch, and TAG contents, ADP-Glc PPase activity and total soluble proteins in wheat seeds at different development stages. DPA, days post-anthesis.

Determinations	WHEAT (DPA)					
	3	6	10	14	17	27
Seed weight (mg/seed)	1.39	1.07	2.67	6.94	11.23	22.61
Starch (mg/seed)	0.02	0.08	0.22	0.75	1.98	7.01
TAG (mg/seed)	1.10^{-4}	1.10^{-4}	0.001	0.008	0.004	0.107
ADP-Glc PPase activity (U/mg)	0.131	0.303	0.232	0.408	0.616	0.597
Total soluble proteins (mg/seed)	0.023	0.016	0.032	0.090	0.127	0.314
Events		*Cell proliferation*		*Transition*		*Storage/ Maturation*

The determinations are the mean of at least three independent set of data that were reproducible within ±10%.

FIGURE 1 | Immunodetections in wheat seed samples throughout development: **(A)** total proteins and **(B)** phosphoproteins purified by IMACFe3+ evaluated with anti-ADP-Glc PPase [obtained from the purified enzyme of leaves of *Spinacia oleracea* (Gómez-Casati and Iglesias, 2002)] **(C)** control of actin in total proteins. Crude extracts preparation and phosphoprotein purification were performed from three independent biological replicates. Protein profiles of the samples are shown in **Supplemental Figure 2**.

polysaccharide accumulation (**Table 1**). Thus, the enzyme activity increased ~2-fold between 3 and 10 DPA and other ~2.5-fold between 10 and 27 DPA. The profile of the enzyme immunodetection showed a continued increase in protein levels between 3 and 10 DPA, then remaining almost similar between 10 and 17 DPA with a slight decrease at 27 DPA (**Figure 1A**). It is also evident that since 14 DPA (with small increases at 17 and 27 DPA), ADP-Glc PPase was found as a phosphoprotein in wheat seed extracts (**Figure 1B**). Concerning castor bean seeds, values of the specific activity (**Supplemental Table 1**) and protein levels (**Supplemental Figure 1A**) for ADP-Glc PPase accompanied amounts of starch accumulated, being higher at the first stage and continuously decreasing along with development. A remarkable difference compared to wheat grains is that in castor bean, the ADP-Glc PPase exhibited no phosphorylation at any DPP stage of seeds (**Supplemental Figure 1B**).

It is worth noting that our results reveal the proteolytic degradation of control actin at 17–27 DPA in wheat (**Figure 1C**) and 35-50 DPP in castor bean (**Supplemental Figure 1C**). This is in agreement with reports indicating that at the advanced stages of development the seed may contain high levels of proteolytic enzymes, which with the accumulation of storage products could constitute a detriment for the stability of proteins (Ahn and Chen, 2007; Nadaud et al., 2010; Nogueira et al., 2012; Piattoni et al., 2017). The important issue concerning ADP-Glc PPase is that: (i) in the case of castor bean the absence of phosphorylation is apparent even when no proteolytic activity is observed (5–35 DPP) (**Supplemental Figure 1**); and (ii) still, its presence (phosphorylated) in wheat was observed even when there were signs of degradation (17–27 DPA) (**Figure 1**). These results support that the phosphorylation of ADP-Glc PPase would be of relevance for starch synthesis in tissues where the polysaccharide accumulation is the long-term metabolic goal for carbon partitioning. Besides, the metabolic picture is markedly different in oleaginous seeds, this fact linked with the storage product in these plants.

It was relevant to corroborate that ADP-Glc PPase recovery after IMAC-Fe^{3+} chromatography of the total protein extract from wheat grains was a consequence of enzyme phosphorylation, rather than an unspecific interaction of its non-phosphorylated form. For this, we performed a de-phosphorylation treatment of the whole protein extract obtained from wheat seeds at 17 DPA with alkaline phosphatase (**Figure 2**). **Figure 2A** shows that the ADP-Glc PPase from the untreated protein extract loaded for IMAC-Fe^{3+} interacted with the matrix, because of its immunodetection after elution. Conversely, for extracts incubated with alkaline phosphatase, no immunoreactive protein bands were observed after elution from the affinity chromatography (**Figure 2B**). These results confirmed that the phosphoprotein enrichment is specifically related to the presence of a post-translationally modified ADP-Glc PPase in the wheat seeds.

Ca^{2+}-Dependent Protein Kinases Are Involved in Phosphorylation of ADP-Glc PPase in Wheat Seeds

To advance in the characterization of the modification by phosphorylation of ADP-Glc PPase from wheat endosperm by protein kinases at the molecular level, we performed *in vitro* studies using enzymes produced recombinantly with high purity degree. The experimental approach considered that plant protein kinases
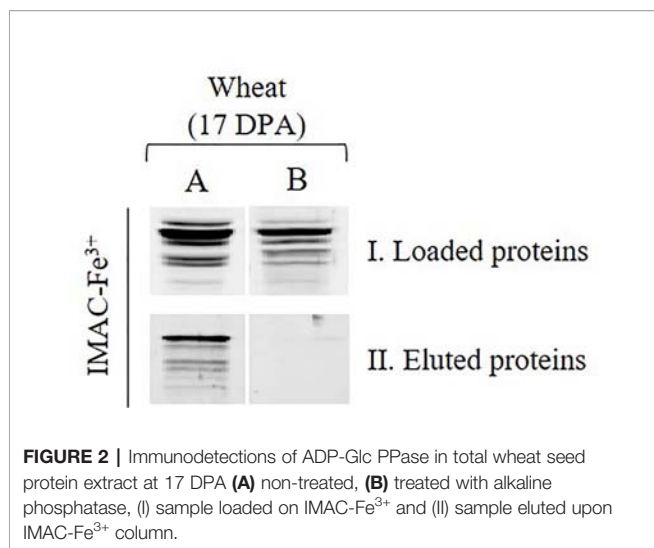
FIGURE 2 | Immunodetections of ADP-Glc PPase in total wheat seed protein extract at 17 DPA **(A)** non-treated, **(B)** treated with alkaline phosphatase, (I) sample loaded on IMAC-Fe^{3+} and (II) sample eluted upon IMAC-Fe^{3+} column.
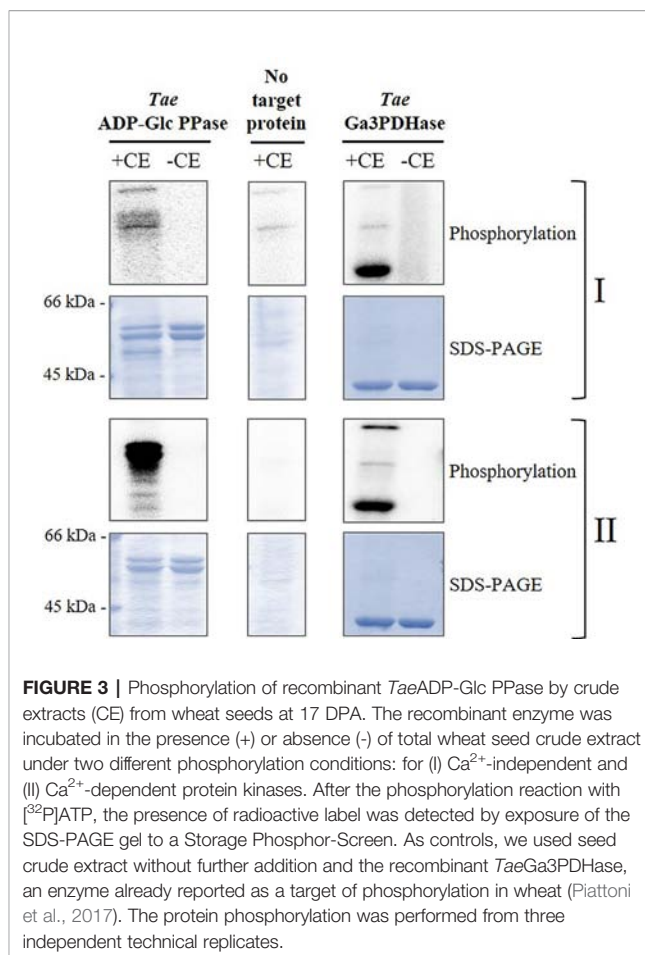

FIGURE 3 | Phosphorylation of recombinant *Tae*ADP-Glc PPase by crude extracts (CE) from wheat seeds at 17 DPA. The recombinant enzyme was incubated in the presence (+) or absence (-) of total wheat seed crude extract under two different phosphorylation conditions: for (I) Ca^{2+}-independent and (II) Ca^{2+}-dependent protein kinases. After the phosphorylation reaction with [^{32}P]ATP, the presence of radioactive label was detected by exposure of the SDS-PAGE gel to a Storage Phosphor-Screen. As controls, we used seed crude extract without further addition and the recombinant *Tae*Ga3PDHase, an enzyme already reported as a target of phosphorylation in wheat (Piattoni et al., 2017). The protein phosphorylation was performed from three independent technical replicates.

can be classified based on structure-to-function relationships. Purposely, the classification establishes families relating protein sequence and specificity for substrates, cofactors (e.g., Ca^{2+}-dependence), and other reaction conditions (Hardie, 1999). Within this categorization, it is common that the properties are similar among different plant species. In this context, the plant-specific superfamily emerges by grouping protein kinases of the type SNF1-related and Ca^{2+}-dependent (SnRK-CDPK) (Vlad et al., 2008).

Based on the above detailed, we first incubated the recombinant heterotetrameric wheat endosperm ADP-Glc PPase (*Tae*ADP-Glc PPase) with crude extracts from wheat seeds at the 17 DPA stage of development in media containing [^{32}P]ATP and optimal conditions for the activity of different plant protein kinases. These conditions allowed to assay either, Ca^{2+}-independent (SnRK1) or Ca^{2+}-dependent (CDPK and SOS2) protein kinases (Piattoni et al., 2011; Piattoni et al., 2017; Rojas et al., 2018). As shown in **Figure 3**, *Tae*SL was poorly phosphorylated under conditions promoting SnRK1 activity (left panels), even when this kinase was able to modify the wheat NAD$^+$-dependent glyceraldehyde-3P dehydrogenase (EC 1.2.1.12, *Tae*Ga3PDHase) used as control of positive phosphorylation (right panels) (Piattoni et al., 2011). Conversely, marked phosphorylation of *Tae*ADP-Glc PPase was evident in media favoring Ca^{2+}-dependent protein kinases (**Figure 3**).

The above described higher capacity of Ca^{2+}-dependent protein kinases was confirmed performing *in vitro* experiments using highly purified, recombinant forms of SnRK1α1 (Ca^{2+}-independent), SOS2, and CDPK1 (Ca^{2+}-dependent) kinases from different plants to phosphorylate *Tae*ADP-Glc PPase. Results illustrated by **Figure 4** indicate that the incorporation of radioactive phospho-moieties from [^{32}P]-ATP into the ADP-Glc PPase was significantly higher when the enzyme was incubated with SOS2 and CDPK1 in comparison with that performed with the SnRK1α1 plant protein kinase. These data support a scenario where the phosphorylation of ADP-Glc PPase by Ca^{2+}-dependent protein kinases would contribute to direct metabolism toward the active synthesis of

starch in wheat seeds. This fact would be relevant in a plant tissue that, at complete development, mainly drives assimilated carbon to produce the polysaccharide for a long-term reserve component.

The L Subunit of ADP-Glc PPase Is Majorly Phosphorylated in Wheat Seeds

Since the functional ADP-Glc PPase found in the wheat endosperm is a heterotetrameric enzyme, we sought to determine if its phosphorylation involves the indistinct modification of both subunits. For such a purpose, we performed the incubation of *Tae*ADP-Glc PPase with SnRK1α1, SOS2, or CDPK1 under conditions of phosphorylation using non-radioactive ATP. Afterward, the samples were analyzed by SDS-PAGE (either alone or containing Phos-tag, to explore the lower mobility of phosphorylated proteins) followed by electro-transference and detection using antibodies specific for each ADP-Glc PPase S or L subunits. **Figure 5** shows that after treatment assuring modification by the Ca^{2+}-dependent protein kinases, the delay in migration was only evident for the L polypeptide, suggesting that this is the subunit that is the target of phosphorylation.

Further experiments gave support to the specificity of the protein kinases to modify the L subunit of *Tae*ADP-Glc PPase. Indeed, the incubation of the purified recombinant *Tae*S or *Tae*L forms with crude extracts from wheat seeds and [^{32}P]ATP rendered SDS-PAGE
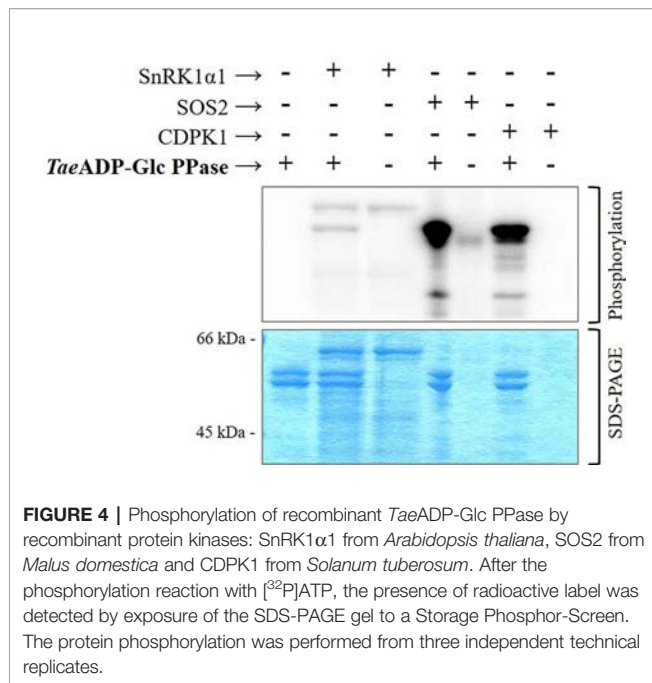
FIGURE 4 | Phosphorylation of recombinant *Tae*ADP-Glc PPase by recombinant protein kinases: SnRK1α1 from *Arabidopsis thaliana*, SOS2 from *Malus domestica* and CDPK1 from *Solanum tuberosum*. After the phosphorylation reaction with [^{32}P]ATP, the presence of radioactive label was detected by exposure of the SDS-PAGE gel to a Storage Phosphor-Screen. The protein phosphorylation was performed from three independent technical replicates.

FIGURE 5 | Immunodetection of S and L subunits of phosphorylated recombinant *Tae*ADP-Glc PPase. The heterotetrameric enzyme was phosphorylated with the respective recombinant protein kinase and then resolved by SDS-PAGE with or without Phos-tag, with subsequent electrotransfer and immunodetection using specific antibodies anti-*Tae*L or anti-*Tae*S. Lanes without *Tae*ADP-Glc PPase are shown as control. The white arrows indicate the phosphorylated delayed peptides. The protein phosphorylation was performed from three independent technical replicates.

profiles revealing phosphorylation of both subunits but at a different degree. As illustrated by **Supplemental Figure 3**, significantly higher incorporation of radioactivity took place on the *Tae*L protein. Also, in the analysis of the samples from the incubation of *Tae*S or *Tae*L subunits with the recombinant plant protein kinases using the approach of western-blots from SDS-PAGE with Phos-tag, the lower mobility was only observed for L polypeptide (**Supplemental Figure 4**). These results strongly agree with the key (tissue-specific) regulatory role assigned to the L subunit of heterotetrameric ADP-Glc PPases (Crevillén et al., 2003; Ballicora et al., 2004; Ventriglia et al., 2008; Ferrero et al., 2018; Figueroa et al., 2018), suggesting that its post-translational modification by phosphorylation would be relevant for starch synthesis in wheat seeds.

CONCLUSIONS

In this work, we report that wheat seed ADP-Glc PPase undergoes progressive phosphorylation along with grain development. The modification of the enzyme increased, as well as the specific activity, during the transition and the beginning of the maturation steps of development. These changes followed the pattern of active synthesis of starch, which is the principal carbon reserve in this grain. The plant protein kinases involved in the phosphorylation of *Tae*ADP-Glc PPase are of the type Ca^{2+}-dependent, and studies with recombinant enzymes support the specific action of SOS2 and CDPK1 kinases. Notably, the post-translational modification of ADP-Glc PPase was absent in seeds of castor bean, which accumulate lipids instead of carbohydrates as a reserve. Results reported herein suggest that phosphorylation of ADP-Glc PPase in seeds of grasses would be relevant for the synthesis of starch. Future studies will shed light on the actual effects of the phosphorylation on
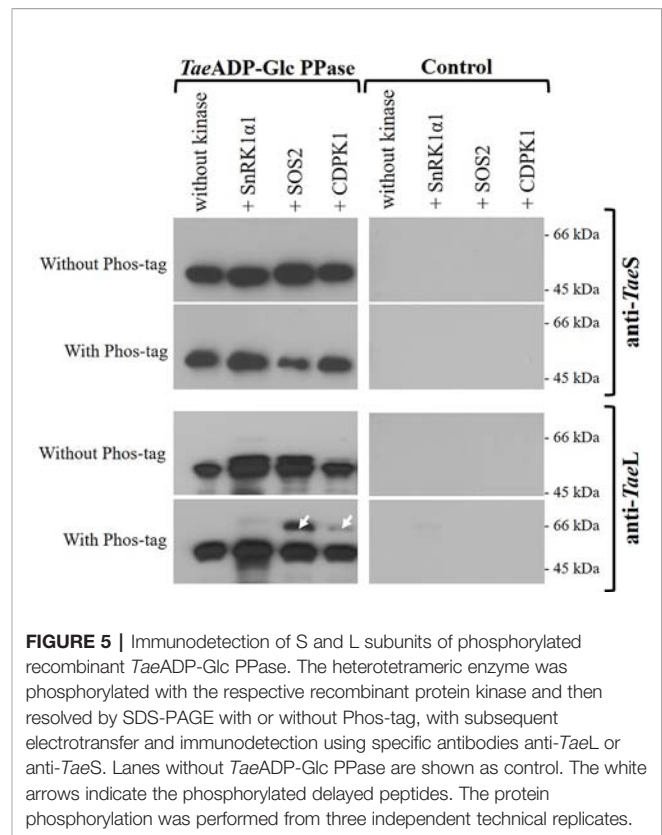
the stability or allosteric regulation (or both) of the wheat endosperm ADP-Glc PPase. A critical role of phosphorylation on the modulation of this enzyme is in good agreement with the limiting role it plays for the production of the main reserve polysaccharide in bacteria and plants (Ballicora et al., 2003; Ballicora et al., 2004). The characterization of the post-translational modification at a molecular level indicated that in the heterotetrameric *Tae*ADP-Glc PPase, the L subunit is at least the major target of phosphorylation. In other plant species, the L subunit also has a primary function in tissue-specific regulation of the catalytic activity exerted by the S subunit (Crevillén et al., 2003; Ballicora et al., 2004; Ballicora et al., 2005; Ventriglia et al., 2008; Ferrero et al., 2018; Figueroa et al., 2018).

This work provides evidence on the post-translational phosphorylation of ADP-Glc PPase in wheat seeds, thus complementing previous predictive phosphoproteomic studies (Lohrig et al., 2009; Kötting et al., 2010; Ma et al., 2014). This modification of the enzyme limiting the pathway of starch biosynthesis would be part of a proposed general mechanism in which protein kinases critically phosphorylate different enzymes involved in the metabolism of the polysaccharide in plants (Kötting et al., 2010; Geigenberger, 2011; Goren et al., 2018). In this framework, phosphorylation would chiefly orchestrate the functioning of the starch metabolism in combination with redox modification and the formation of complexes between different proteins (Lohrig et al., 2009; Kötting et al., 2010; Geigenberger, 2011; Ma et al., 2014; Tetlow et al., 2015; Goren et al., 2018). These combined processes would operate as an effective mechanism to

optimize the accumulation of the polysaccharide in seeds of wheat and other grass crops. Concerning post-translational modification, the wheat endosperm ADP-Glc PPase was distinctively characterized as insensitive to the redox regulation demonstrated for the enzyme from other plant sources (Ballicora et al., 2004; Linebarger et al., 2005; Tuncel et al., 2014; Ferrero et al., 2018). Thus, results reported herein provide new information about the mechanism involved in starch synthesis in grasses.

The gained information reported at present opens many research approaches in the way to reach a better understanding (at the molecular level) of starch synthesis in wheat (and other cereals producing the polysaccharide as a prime component). A critical issue to investigate is related to protein chemistry studies (including mass spectrometry) to reach the identification of the specific residues in the enzyme that are the target of protein kinases. This latter characterization could be followed by the production of phosphomimetic forms of the *Tae*ADP-Glc PPase, which would allow analyzing how phosphorylation modifies its kinetic, regulatory, and stability properties. In this scenario, the possibility that phosphorylated *Tae*ADP-Glc PPase would better interact with other enzymes and proteins is an issue needing attention in future research. Another topic to be explored refers to gaining detail of the specific Ca^{2+}-dependent protein kinase from wheat involved in the post-translational modification. This subject is of complex understanding, considering that *T. aestivum* is a hexaploid organism, a product of broad hybrid mix in the breeding process (Appels et al., 2018). At present, genomic information on *Triticum aestivum* L. identified fifteen and twenty genes coding for SnRK1 (Perochon et al., 2019) and CDPK (Li et al., 2008), respectively.

Future studies will be central for the design of strategies and biotechnological tools to improve yields in agriculture. Starch, as the primary feedstock of many crops, is currently the leading world supplier for caloric demands of animals (including humans). Besides, starch is a natural product of high relevance for the development of biofuels and bioplastics, which are critical products to cope with the world-wide challenges tied to demographic expansion and climate change.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

DF, CP, MA, BR, MH, MB, and AI conceived and designed the experiments. AI wrote the paper. DF, BR, and MH performed the experiments. DF, CP, MA, MB, and AI analyzed the data. AI, MA and MB contributed reagents, materials, and analysis tools.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.01058/full#supplementary-material

SUPPLEMENTAL TABLE 1 | Determinations of weight, starch, and TAG contents, ADP-Glc PPase activity and total soluble proteins in castor bean at different development stages. DPP, days post-pollination. The determinations are the mean of at least three independent set of data that were reproducible within ±10%.

SUPPLEMENTAL FIGURE 1 | Immunodetections in castor bean samples throughout development: **(A)** total proteins and **(B)** phosphoproteins purified by IMAC-Fe^{3+} evaluated with anti-ADP-Glc PPase obtained from the purified enzyme of leaves of *Spinacia oleracea* (Gómez-Casati and Iglesias, 2002) and **(C)** control of actin in total proteins. Crude extracts preparation and phosphoprotein purification were performed from three independent biological replicates. Protein profiles of the samples are shown in Supplemental Figure 3.

SUPPLEMENTAL FIGURE 2 | SDS-PAGE of wheat and castor bean seed samples throughout development: (A) total proteins (same amount in all cases) and (B) phosphoproteins purified by IMAC-Fe^{3+} (same volume of eluted protein of IMAC-Fe^{3+} in all cases).

SUPPLEMENTAL FIGURE 3 | Phosphorylation of *Tae*S and *Tae*L recombinant versions by crude extracts (CE) from wheat seeds at 17 DPA. The recombinant enzyme was incubated in the presence (+) or absence (-) of total wheat seed crude extract under two different phosphorylation conditions: for (I) Ca^{2+}-independent and (II) Ca^{2+}-dependent protein kinases. After the phosphorylation reaction with $[^{32}P]$ ATP, the presence of radioactive label was detected by exposure of the SDS-PAGE gel to a Storage Phosphor-Screen. The protein phosphorylation was performed from three independent technical replicates

SUPPLEMENTAL FIGURE 4 | Immunodetection of phosphorylated *Tae*L and *Tae*S subunits. The subunits were phosphorylated with the respective recombinant protein kinase and then resolved by SDS-PAGE with or without Phos-tag, with subsequent electrotransfer and immunodetection using specific antibodies anti-*Tae*L or anti-*Tae*S. Lanes without recombinant subunit are shown as control. The white arrows indicate the phosphorylated delayed peptides. The protein phosphorylation was performed from three independent technical replicates.

REFERENCES

Ahn, Y. J., and Chen, G. Q. (2007). Temporal and spatial expression of 2S albumin in castor (*Ricinus communis* L.). *J. Agric. Food Chem.* 55, 10043–10049. doi: 10.1021/jf071272p

Ainsworth, C., Tarvis, M., and Clark, J. (1993). Isolation and analysis of a cDNA clone encoding the small subunit of ADP-glucose pyrophosphorylase from wheat. *Plant Mol. Biol.* 23, 23–33. doi: 10.1007/BF00021416

Ainsworth, C., Hosein, F., Tarvis, M., Weir, F., Burrell, M., Devos, K. M., et al. (1995). Adenosine diphosphate glucose pyrophosphorylase genes in wheat: differential expression and gene mapping. *Planta.* 197, 1–10. doi: 10.1007/BF00239933

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science.* 361 (6403), 1–13. doi: 10.1126/science.aar7191

Ballicora, M. A., Fu, Y., Frueauf, J. B., and Preiss, J. (1999). Heat stability of the potato tuber ADP-glucose pyrophosphorylase: role of Cys residue 12 in the small subunit. *Biochem. Biophys. Res. Commun.* 257, 782–786. doi: 10.1006/bbrc.1999.0469

Ballicora, M. A., Frueauf, J. B., Fu, Y., Schürmann, P., Preiss, J., Ballicora, M. A., et al. (2000). Activation of the Potato Tuber ADP-glucose Pyrophosphorylase by Thioredoxin. *J. Biol. Chem.* 275, 1–7. doi: 10.1074/jbc.275.2.1315

Ballicora, M. A., Iglesias, A. A., and Preiss, J. (2003). ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis. *Microbiol. Mol. Biol. Rev.* 67, 213–225. doi: 10.1128/MMBR.67.2.213-225.2003

Ballicora, M. A., Iglesias, A. A., and Preiss, J. (2004). ADP-glucose pyrophosphorylase: A regulatory enzyme for plant starch synthesis. *Photosynth. Res.* 79, 1–24. doi: 10.1023/B:PRES.0000011916.67519.58

Ballicora, M. A., Dubay, J. R., Devillers, C. H., and Preiss, J. (2005). Resurrecting the ancestral enzymatic role of a modulatory subunit. *J. Biol. Chem.* 280, 10189–10195. doi: 10.1074/jbc.M413540200

Baris, I., Tuncel, A., Ozber, N., Keskin, O., and Kavakli, I. H. (2009). Investigation of the interaction between the large and small subunits of potato ADP-glucose pyrophosphorylase. *PloS Comput. Biol.* 5, 1–14. doi: 10.1371/journal.pcbi.1000546

Baud, S., and Graham, I. A. (2006). A spatiotemporal analysis of enzymatic activities associated with carbon metabolism in wild-type and mutant embryos of Arabidopsis using in situ histochemistry. *Plant J.* 46, 155–169. doi: 10.1111/j.1365-313X.2006.02682.x

Bradford, M. M. (1976). A Rapid and Sensitive Method for the Quantitation Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* 254, 248–254. doi: 10.1016/0003-2697(76)90527-3

Briarty, L. G., Hughes, C. E., and Evers, A. D. (1979). The Developing Endosperm of Wheat - A Stereological Analysis. *Ann. Bot.* 44, 641–658. doi: 10.1093/oxfordjournals.aob.a085779

Bustos, D. M., and Iglesias, A. A. (2002). Non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase is post-translationally phosphorylated in heterotrophic cells of wheat (Triticum aestivum). *FEBS Lett.* 530, 169–173. doi: 10.1016/S0014-5793(02)03455-5

Canvin, D. T. (1963). Formation of oil in the seed of *Ricinus communis* L. *Can. J. Biochem. Physiol.* 41, 1879–1885. doi: 10.1139/y63-214

Crevillén, P., Ballicora, M. A., Mérida, Á, Preiss, J., and Romero, J. M. (2003). The different large subunit isoforms of Arabidopsis thaliana ADP-glucose pyrophosphorylase confer distinct kinetic and regulatory properties to the heterotetrameric enzyme. *J. Biol. Chem.* 278, 28508–28515. doi: 10.1074/jbc.M304280200

Crofts, N., Nakamura, Y., and Fujita, N. (2017). Critical and speculative review of the roles of multi-protein complexes in starch biosynthesis in cereals. *Plant Sci.* 262, 1–8. doi: 10.1016/j.plantsci.2017.05.007

Crozet, P., Jammes, F., Valot, B., Ambard-Bretteville, F., Nessler, S., Hodges, M., et al. (2010). Cross-phosphorylation between *Arabidopsis thaliana* sucrose nonfermenting 1-related protein kinase 1 (AtSnRK1) and its activating kinase (AtSnAK) determines their catalytic activities. *J. Biol. Chem.* 285, 12071–12077. doi: 10.1074/jbc.M109.079194

Dai, Z. (2010). Activities of enzymes involved in starch synthesis in wheat grains differing in starch content. *Russ. J. Plant Physiol.* 57, 74–78. doi: 10.1134/S1021443710010103

Ferrero, D. M. L., Asencion Diez, M. D., Kuhn, M. L., Falaschetti, C. A., Piattoni, C. V., Iglesias, A. A., et al. (2018). On the roles of wheat endosperm ADP-glucose pyrophosphorylase subunits. *Front. Plant Sci.* 9, 1498. doi: 10.3389/fpls.2018.01498

Figueroa, C. M., Kuhn, M. L., Hill, B. L., Iglesias, A. A., and Ballicora, M. A. (2018). Resurrecting the Regulatory Properties of the Ostreococcus tauri ADP-Glucose Pyrophosphorylase Large Subunit. *Front. Plant Sci.* 9, 1564. doi: 10.3389/fpls.2018.01564

Folch, J., Lees, M., and Sloane Stanley, G. H. (1957). A simple method for the isolation and purification of total lipids from animal tissues. *J. Biol. Chem.* 226, 497–509.

Fusari, C., Demonte, A. M., Figueroa, C. M., Aleanzi, M., and Iglesias, A. A. (2006). A colorimetric method for the assay of ADP-glucose pyrophosphorylase. *Anal. Biochem.* 352, 145–147. doi: 10.1016/j.ab.2006.01.024

Gómez-Casati, D. F., and Iglesias, A. A. (2002). ADP-glucose pyrophosphorylase from wheat endosperm. Purification and characterization of an enzyme with novel regulatory properties. *Planta* 214, 428–434. doi: 10.1007/s004250100634

Geigenberger, P. (2011). Regulation of starch biosynthesis in response to a fluctuating environment. *Plant Physiol.* 155, 1566–1577. doi: 10.1104/pp.110.170399

Georgelis, N., Braun, E. L., and Hannah, L. C. (2008). Duplications and functional divergence of ADP-glucose pyrophosphorylase genes in plants. *BMC Evol. Biol.* 8, 232. doi: 10.1186/1471-2148-8-232

Georgelis, N., Shaw, J. R., and Hannah, L. C. (2009). Phylogenetic analysis of ADP-glucose pyrophosphorylase subunits reveals a role of subunit interfaces in the allosteric properties of the enzyme. *Plant Physiol.* 151, 67–77. doi: 10.1104/pp.109.138933

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food security: The challenge of feeding 9 billion people. *Science* 327, 812–818. doi: 10.1126/science.1185383

Gong, D., Guo, Y., Jagendorf, A. T., and Zhu, J. (2019). Biochemical Characterization of the Arabidopsis Protein Kinase SOS2 That Functions in Salt Tolerance 1. *Plant Physiol.* 130, 256–264. doi: 10.1104/pp.004507.256

Goren, A., Ashlock, D., and Tetlow, I. J. (2018). Starch formation inside plastids of higher plants. *Protoplasma* 255, 1855–1876. doi: 10.1007/s00709-018-1259-4

Greenwood, J. S., and Bewley, J. D. (1982). Seed development in *Ricinus communis* (castor bean) I.Descriptive morphology. *Can. J. Bot.* 60, 1751–1760. doi: 10.1139/b82-222

Guo, Y. (2001). Molecular Characterization of Functional Domains in the Protein Kinase SOS2 That Is Required for Plant Salt Tolerance. *Plant Cell* 13, 1383–1400. doi: 10.1105/tpc.13.6.1383

Hardie, D. G. (1999). Plant Protein Serine/Threonine Kinases: Classification and Functions. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 97–131. doi: 10.1146/annurev.arplant.50.1.97

Iglesias, A., Kakefuda, G., and Preiss, J. (1991). Regulatory and Structural Properties of the Cyanobacterial ADP-glucose Pyrophosphorylases. *Plant Physiol.* 97, 1187–1195. doi: 10.1104/pp.97.3.1187

Iglesias, A. A., Charng, Y. Y., Ball, S., and Preiss, J. (1994). Characterization of the Kinetic, Regulatory, and Structural-Properties of ADP-glucose Pyrophosphorylase from *Chlamydomonas reinhardtii*. *Plant Physiol.* 104, 1287–1294. doi: 10.1104/pp.104.4.1287

Iglesias, A. A. (2015). Re-Paving the Road Built by Chemistry: A Challenge to Biochemistry and Biotechnology. *Am. J. Biochem. Biotechnol.* 11, 3–4. doi: 10.3844/ajbbsp.2015.3.4

Jeon, J. S., Ryoo, N., Hahn, T. R., Walia, H., and Nakamura, Y. (2010). Starch biosynthesis in cereal endosperm. *Plant Physiol. Biochem.* 48, 383–392. doi: 10.1016/j.plaphy.2010.03.006

Kötting, O., Kossmann, J., Zeeman, S. C., and Lloyd, J. R. (2010). Regulation of starch metabolism: The age of enlightenment? *Curr. Opin. Plant Biol.* 13, 321–329. doi: 10.1016/j.pbi.2010.01.003

Kavakli, I. H., Greene, T. W., Salamone, P. R., Choi, S. B., and Okita, T. W. (2001a). Investigation of subunit function in ADP-glucose pyrophosphorylase. *Biochem. Biophys. Res. Commun.* 281, 783–787. doi: 10.1006/bbrc.2001.4416

Kavakli, I. H., Park, J. S., Slattery, C. J., Salamone, P. R., Frohlick, J., and Okita, T. W. (2001b). Analysis of Allosteric Effector Binding Sites of Potato ADP-glucose Pyrophosphorylase through Reverse Genetics. *J. Biol. Chem.* 276, 40834–40840. doi: 10.1074/jbc.M106310200

Kinoshita, E., Kinoshita-Kikuta, E., and Koike, T. (2015). Advances in Phos-tag-based methodologies for separation and detection of the phosphoproteome. *Biochim. Biophys. Acta - Proteins Proteomics* 1854, 601–608. doi: 10.1016/j.bbapap.2014.10.004

Kleczkowski, L., Villand, P., Lüthi, E., Olsen, O., and Preiss, J. (1993). Insensitivity of barley endosperm ADP-glucose pyrophosphorylase to 3-phosphoglycerate and orthophosphate regulation. *Plant Physiol.* 101, 179–186. doi: 10.1104/pp.101.1.179

Kuhn, M. L., Falaschetti, C. A., and Ballicora, M. A. (2009). *Ostreococcus tauri* ADP-glucose pyrophosphorylase reveals alternative paths for the evolution of subunit roles. *J. Biol. Chem.* 284, 34092–34102. doi: 10.1074/jbc.M109.037614

Kuhn, M. L., Figueroa, C. M., Iglesias, A. A., and Ballicora, M. A. (2013). The ancestral activation promiscuity of ADP-glucose pyrophosphorylases from oxygenic photosynthetic organisms. *BMC Evol. Biol.* 13, 51. doi: 10.1186/1471-2148-13-51

Kunkel, T. A., Bebenek, K., and McClary, J. (1991). Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.* 204, 125–139. doi: 10.1016/0076-6879(91)04008-C

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685. doi: 10.1038/227680a0

Li, A. L., Zhu, Y. F., Tan, X. M., Wang, X., Wei, B., Guo, H. Z., et al. (2008). Evolutionary and functional study of the CDPK gene family in wheat (*Triticum aestivum* L.). *Plant Mol. Biol.* 66, 429–443. doi: 10.1007/s11103-007-9281-5

Linebarger, C. R., Boehlein, S. K., Sewell, A. K., Shaw, J., and Hannah, L. C. (2005). Heat stability of maize endosperm ADP-glucose pyrophosphorylase is enhanced by insertion of a cysteine in the N terminus of the small subunit. *Plant Physiol.* 139, 1625–1634. doi: 10.1104/pp.105.067637

Lohrig, K., Müller, B., Davydova, J., Leister, D., and Wolters, D. A. (2009). Phosphorylation site mapping of soluble proteins: Bioinformatical filtering reveals potential plastidic phosphoproteins in *Arabidopsis thaliana.* *Planta* 229, 1123–1134. doi: 10.1007/s00425-009-0901-y

Ma, C., Zhou, J., Chen, G., Bian, Y., Lv, D., Li, X., et al. (2014). iTRAQ-based quantitative proteome and phosphoprotein characterization reveals the central metabolism changes involved in wheat grain development. *BMC Genomics* 15, 1–20. doi: 10.1186/1471-2164-15-1029

MacNeill, G. J., Mehrpouyan, S., Minow, M. A. A., Patterson, J. A., Tetlow, I. J., and Emes, M. J. (2017). Starch as a source, starch as a sink: The bifunctional role of starch in carbon allocation. *J. Exp. Bot.* 68, 4433–4453. doi: 10.1093/jxb/erx291

Morell, M. K., and Myers, A. M. (2005). Towards the rational design of cereal starches. *Curr. Opin. Plant Biol.* 8, 204–210. doi: 10.1016/j.pbi.2005.01.009

Muszyńska, G., Andersson, L., and Porath, J. (1986). Selective adsorption of phosphoproteins on gel-immobilized ferric chelate. *Biochemistry* 25, 6850–6853. doi: 10.1021/bi00370a018

Nadaud, I., Girousse, C., Debiton, C., Chambon, C., Bouzidi, M. F., Martre, P., et al. (2010). Proteomic and morphological analysis of early stages of wheat grain development. *Proteomics* 10, 2901–2910. doi: 10.1002/pmic.200900792

Nogueira, F. C. S., Palmisano, G., Soares, E. L., Shah, M., Soares, A. A., Roepstorff, P., et al. (2012). Proteomic profile of the nucellus of castor bean (*Ricinus communis* L.) seeds during development. *J. Proteomics* 75, 1933–1939. doi: 10.1016/j.jprot.2012.01.002

Pandey, M. K., Rani, N. S., Madhav, M. S., Sundaram, R. M., Varaprasad, G. S., Sivaranjani, A. K. P., et al. (2012). Different isoforms of starch-synthesizing enzymes controlling amylose and amylopectin content in rice (*Oryza sativa* L.). *Biotechnol. Adv.* 30, 1697–1706. doi: 10.1016/j.biotechadv.2012.08.011

Perochon, A., Váry, Z., Malla, K. B., Halford, N. G., Paul, M. J., and Doohan, F. M. (2019). The wheat SnRK1α family and its contribution to Fusarium toxin tolerance. *Plant Sci.* 288, 110217. doi: 10.1016/j.plantsci.2019.110217

Piattoni, C., Bustos, D., Guerrero, S., and Iglesias, A. (2011). Nonphosphorylating Glyceraldehyde-3-Phosphate Dehydrogenase Is Phosphorylated in Wheat Endosperm at Serine-404 by an SNF1-Related Protein Kinase Allosterically Inhibited by Ribose-5-Phosphate. *Plant Physiol.* 156, 1337–1350. doi: 10.1104/pp.111.177261

Piattoni, C. V., Ferrero, D. M. L., Dellaferrera, I., Vegetti, A., and Iglesias, A. A. (2017). Cytosolic Glyceraldehyde-3-Phosphate Dehydrogenase Is Phosphorylated during Seed Development. *Front. Plant Sci.* 8, 1–16. doi: 10.3389/fpls.2017.00522

Plaxton, W., and Preiss, J. (1987). Purification and properties of nonproteolytic degraded ADP-glucose pyrophosphorylase from maize endosperm. *Plant Physiol.* 83, 105–112. doi: 10.1104/pp.83.1.105

Raíces, M., Chico, J. M., Tellez-Iñón, M. T., and Ulloa, R. M. (2001). Molecular characterization of StCDPK1, a calcium-dependent protein kinase from

Solanum tuberosum that is induced at the onset of tuber development. *Plant Mol. Biol.* 46, 591–601. doi: 10.1023/A:1010661304980

Reibach, P. H., and Benedict, C. R. (1982). Biosynthesis of Starch in Proplastids of Germinating *Ricinus communis* Endosperm Tissue. *Plant Physiol.* 70, 252–256. doi: 10.1104/pp.70.1.252

Rojas, B. E., Santin, F., Ulloa, R. M., Iglesias, A. A., and Figueroa, C. M. (2018). A fluorometric method for the assay of protein kinase activity. *Anal. Biochem.* 557, 120–122. doi: 10.1016/j.ab.2018.07.018

Santin, F., Bhogale, S., Fantino, E., Grandellis, C., Banerjee, A. K., and Ulloa, R. M. (2017). *Solanum tuberosum* StCDPK1 is regulated by miR390 at the posttranscriptional level and phosphorylates the auxin efflux carrier StPIN4 in vitro, a potential downstream target in potato development. *Physiol. Plant* 159, 244–261. doi: 10.1111/ppl.12517

Shen, W., Reyes, M.II, and Hanley-Bowdoin, L. (2009). Arabidopsis Protein Kinases GRIK1 and GRIK2 Specifically Activate SnRK1 by Phosphorylating Its a Activation Loop. *Plant Physiol.* 150, 996–1005. doi: 10.1104/pp.108.132787

Tetlow, I. J., Liu, F., and Emes, M. J. (2015). "Protein-Protein Interactions During Starch Biosynthesis," in *Starch: Metabolism and Structure*, Ed. Y. Nakamura, (Springer) 1–451. doi: 10.1007/978-4-431-55495-0

Tiessen, A., Hendriks, J. H. M., Stitt, M., Branscheid, A., Gibon, Y., Farré, E. M., et al. (2002). Starch synthesis in potato tubers is regulated by post-translational redox modification of ADP-glucose pyrophosphorylase. *Plant Cell* 14, 2191–2213. doi: 10.1105/tpc.003640.2192

Tuncel, A., and Okita, T. W. (2013). Improving starch yield in cereals by over-expression of ADPglucose pyrophosphorylase: Expectations and unanticipated outcomes. *Plant Sci.* 211, 52–60. doi: 10.1016/j.plantsci.2013.06.009

Tuncel, A., Kawaguchi, J., Ihara, Y., Matsusaka, H., Nishi, A., Nakamura, T., et al. (2014). The rice endosperm ADP-glucose pyrophosphorylase large subunit is essential for optimal catalysis and allosteric regulation of the heterotetrameric enzyme. *Plant Cell Physiol.* 55, 1169–1183. doi: 10.1093/pcp/pcu057

Turner, W. L., Knowles, V. L., and Plaxton, W. C. (2005). Cytosolic pyruvate kinase: Subunit composition, activity, and amount in developing castor and soybean seeds, and biochemical characterization of the purified castor seed enzyme. *Planta* 222, 1051–1062. doi: 10.1007/s00425-005-0044-8

Ventriglia, T., Kuhn, M. L., Ruiz, M. T., Ribeiro-Pedro, M., Valverde, F., Ballicora, M. A., et al. (2008). Two Arabidopsis ADP-glucose pyrophosphorylase large subunits (APL1 and APL2) are catalytic. *Plant Physiol.* 148, 65–76. doi: 10.1104/pp.108.122846

Vlad, F., Turk, B. E., Peynot, P., Leung, J., and Merlot, S. (2008). A versatile strategy to define the phosphorylation preferences of plant protein kinases and screen for putative substrates. *Plant J.* 55, 104–117. doi: 10.1111/j.1365-313X.2008.03488.x

Wilkens, C., Svensson, B., and Møller, M. S. (2018). Functional roles of starch binding domains and surface binding sites in enzymes involved in starch biosynthesis. *Front. Plant Sci.* 9, 1652. doi: 10.3389/fpls.2018.01652

Yu, G., Lv, Y., Shen, L., Wang, Y., Qing, Y., Wu, N., et al. (2019). The proteomic analysis of maize endosperm protein enriched by Phos-tag™ reveals the phosphorylation of Brittle-2 subunit of ADP-Glc pyrophosphorylase in starch biosynthesis process. *Int. J. Mol. Sci.* 20, 1–17. doi: 10.3390/ijms20040986

Zi, Y., Ding, J., Song, J., Humphreys, G., Peng, Y., and Li, C. (2018). Grain Yield, Starch Content and Activities of Key Enzymes of Waxy and Non-waxy Wheat (*Triticum aestivum* L.). *Sci. Rep.* 8, 1–12. doi: 10.1038/s41598-018-22587-0

Global Role of Crop Genomics in the Face of Climate Change

Mohammad Pourkheirandish[†], Agnieszka A. Golicz[†], Prem L. Bhalla and Mohan B. Singh*

Plant Molecular Biology and Biotechnology Laboratory, Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, VIC, Australia

The development of climate change resilient crops is necessary if we are to meet the challenge of feeding the growing world's population. We must be able to increase food production despite the projected decrease in arable land and unpredictable environmental conditions. This review summarizes the technological and conceptual advances that have the potential to transform plant breeding, help overcome the challenges of climate change, and initiate the next plant breeding revolution. Recent developments in genomics in combination with high-throughput and precision phenotyping facilitate the identification of genes controlling critical agronomic traits. The discovery of these genes can now be paired with genome editing techniques to rapidly develop climate change resilient crops, including plants with better biotic and abiotic stress tolerance and enhanced nutritional value. Utilizing the genetic potential of crop wild relatives (CWRs) enables the domestication of new species and the generation of synthetic polyploids. The high-quality crop plant genome assemblies and annotations provide new, exciting research targets, including long non-coding RNAs (lncRNAs) and cis-regulatory regions. Metagenomic studies give insights into plant-microbiome interactions and guide selection of optimal soils for plant cultivation. Together, all these advances will allow breeders to produce improved, resilient crops in relatively short timeframes meeting the demands of the growing population and changing climate.

Keywords: domestication, genomics, climate change, crops, transcriptomics, abiotic stress

INTRODUCTION

The world will require a dramatic increase in food production in the next 30 years. Global food security is one of the key challenges of this century with the current human population of 7.7 billion expected to reach 8.6 billion in 2030 and 10 billion by 2050 (Tomlinson, 2013). The increase in population has led to an increase in urbanization, which is directly and indirectly, reducing our access to suitable land for agriculture (Satterthwaite et al., 2010). Simultaneously, the effects of climate change, including but not limited to increased temperature, changing patterns of rainfall, and increased levels of CO_2 and ozone, impose further pressure on agriculture *via* drought and salinity that limit agricultural land and water use (Godfray et al., 2010).

Population growth is not the only reason we will need to increase food production. Significant income growth in rapidly developing economies gave rise to an emerging middle class, accelerating the dietary transition toward higher consumption of meat, eggs, and dairy products and boosting the need to grow more grain to feed more cattle, pigs, and poultry (Tilman and Clark, 2014). Agriculture in 2050 will need to produce almost 60–100% more food and feed than it is doing now (Tilman et al., 2011). This goal must be achieved despite the increase in global

temperatures associated with climate change and growing scarcity of water and land, which are predicted to have significant impacts on the yield of all major crops.

In the last few centuries, plant breeders successfully used crossing and selection to improve the agronomic character of cultivated crops, such as wheat, maize, rice, barley, and others, resulting in dramatic increases in food production. However, agriculture has shifted to monoculture, resulting in the significant reduction of genetic diversity with today's global agricultural food depending on a few key plant species (Khoury et al., 2014).

The genetic gains achieved by conventional crop breeding and advanced agronomic practices have led to more than a double increase in crop yields between 1960 and 2015. The development of dwarf varieties of rice and wheat coupled with greater use of synthetic fertilizers and irrigation led to the first green revolution. However, the yield increases due to the green revolution are declining and/or beginning to plateau for the major food crops (Grassini et al., 2013). After years of improvement, we are getting close to the final capacity of these few crops on yield and their tolerance to biotic and abiotic stresses. The current trend of annual yield increases for major crops of between 0.9 and 1.6% is insufficient to meet requirements in the near future (Ray et al., 2013). It has been estimated that about 2.4% annual yield gain is required to meet the global food demand (Ray et al., 2013). Thus, development of high-yielding climate change resilient crops with enhanced tolerance to water deficit, temperature, and biotic stresses is critical for increasing productivity to keep pace with the increasing human population.

The challenge of feeding the increasing human population under climate change conditions is unlikely to be met by conventional breeding technologies alone. Plant breeding must adopt new, multidisciplinary approaches to enhance the rate of genetic gain (Varshney et al., 2018). Fortunately, the science underpinning plant breeding is being revolutionized by the recent conceptual and technological innovations including the development of rapid, cheap sequencing technologies and the rise of genomics allowing for the detailed analysis of plant genomes and dissection of the genetic basis of agronomic traits. Genomics is now at the core of crop improvement, including the identification of genetic variation underlying differences in phenotypes, identification of additional sources of variation and novel traits, and characterization of molecular pathways involved in biotic and abiotic stress tolerance.

Recently the development of genome editing technologies, especially CRISPR/Cas9, opened new routes of fast and precise genome modification promising rapid translation of knowledge from the lab to the field. Genome editing allows introduction of insertions/deletions or an entirely new sequence at a desired location in the target genome (Scheben et al., 2017). Known genes controlling important traits can be selectively modified using genome editing, allowing for manipulation of phenotypes. In recent years, several genome edited crop plants entered final stages of commercialization in the United States of America including drought and salt tolerant soybean, *Camelina* with increased oil content, and waxy corn (Waltz, 2018).

Considering the urgent need for crop plant improvement and the new, exciting technological and conceptual developments, this review outlines the potential of genomic approaches (**Table 1**) for the development of climate change resilient crops.
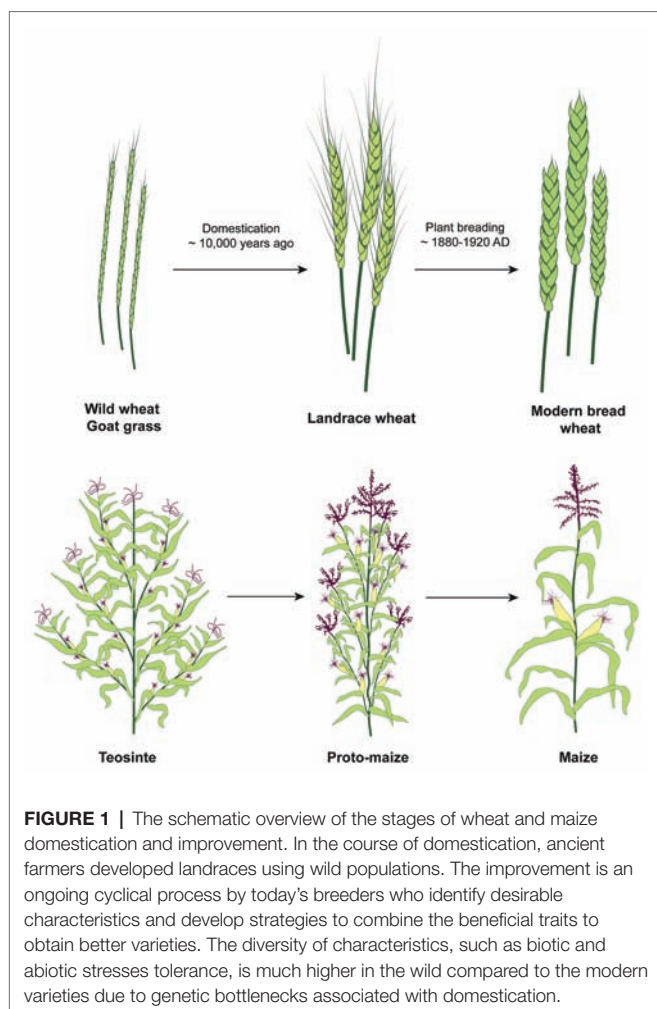
USING GENOMICS TO IMPROVE CROP PLANT DIVERSITY AND RESILIENCE

Accessing Genetic Diversity of CWRs

Wild plants have survived under a changing climate for millions of years, during which they have been subjected to selective pressure by biotic and abiotic factors. This natural selection has led to the accumulation of genes allowing plants to resist, tolerate, or avoid extreme temperatures, draught, or flooding, as well as pests and diseases. However, during subsequent domestication, many of those, now important, traits and associated genetic material were lost, transforming some of the plants into our current remarkably productive crops with limited genetic diversity (**Figure 1**). The remainder of the genetic resources was left behind and mostly treated as a weed. Insights gained from the genome sequencing projects of different crops

TABLE 1 | Summary of different approaches, which can be used to improve crop diversity and resilience.

Approach	Desired outcome
Using genomics to improve crop plant diversity and resilience	
Accessing genetic diversity of crop wild relatives (CWRs)	Diversification of the existing breeding resources
De novo crop domestication	Domestication of completely new crops using wild species
Engineering polyploidy	Controlled genome duplication or bridging the genomes of two related species
Harnessing plant-microbe interactions	Optimal choice of suitable crops for the specific soil type and geographic location
The challenge of climate change and plant diseases	Prediction of pathogen evolution and prevalence and deployment of suitable protective measures ahead of time
Genome editing for nutritionally enhanced crops	Editing of target genes to improve crop nutritional value
Accessing new breeding targets using genomic technologies	
Third-generation sequencing	Use of long sequencing reads for higher quality reference genome construction
Accurate gene prediction and functional annotation	Precise candidate gene identification
Analysis of the non-coding part of genome	Identification of new functional genomic sequences and breeding targets
Pangenome as a reference sequence	Inclusion of species-wide genomic variation in the analysis
Pairing genomics with other emerging technologies	
Machine learning and crop plant genomics	Use of artificial intelligence for crop genotype and phenotype prediction
Speed Breeding	Shortening the breeding cycle
High-throughput phenotyping	Increased resolution, accuracy and speed of plant phenotyping

FIGURE 1 | The schematic overview of the stages of wheat and maize domestication and improvement. In the course of domestication, ancient farmers developed landraces using wild populations. The improvement is an ongoing cyclical process by today's breeders who identify desirable characteristics and develop strategies to combine the beneficial traits to obtain better varieties. The diversity of characteristics, such as biotic and abiotic stresses tolerance, is much higher in the wild compared to the modern varieties due to genetic bottlenecks associated with domestication.

FIGURE 2 | *De novo* domestication compared to conventional breeding. Comparison of the traditional breeding approach using backcrossing of the elite line with its wild relative with the molecular *de novo* domestication approach using the genome editing technique. Examples of potential target genes to be edited in the process of *de novo* domestication of wild relatives are indicated below the arrow.

(Zhou et al., 2015b; Mascher et al., 2017; Appels et al., 2018; Springer et al., 2018; Zhao et al., 2018b) demonstrated the narrow germplasm of our modern crops and emphasized their vulnerabilities to climate change. However, all modern crop plants were domesticated from crop wild relatives (CWRs), which are still found in the wild, and provide a rich pool of genetic material, which is often excluded from the existing breeding programs (Brozynska et al., 2016).

Elite cultivated crops, such as wheat, maize, rice, and barley, are often dependant on farmer supplied resources, including water *via* irrigation, nutrition *via* fertilizers, and resistance to biotic stresses through the use of pesticides. This has led to the elite varieties becoming less resilient compared to their wild counterparts. In addition, strong artificial selection for a handful of crucial traits resulted in reduced diversity and restriction of the gene pool available within breeding programs. CWRs constitute an additional source of genetic diversity, which can be utilized during crop improvement programs, with as much as 30% of the increases in crop yields during the late 20th century being attributed to the use of CWRs in plant breeding programmes (Pimentel et al., 1997; Brozynska et al., 2016).

Over 1,500 CWRs of food crops have been identified as a potential source of genetic diversity for 173 globally important crops (Vincent et al., 2013). Advances in sequencing technologies facilitated construction of CRW reference genomes, which in turn can be used in comparative genomics analyses, allowing for the identification of novel genes controlling key traits (Brozynska et al., 2016).
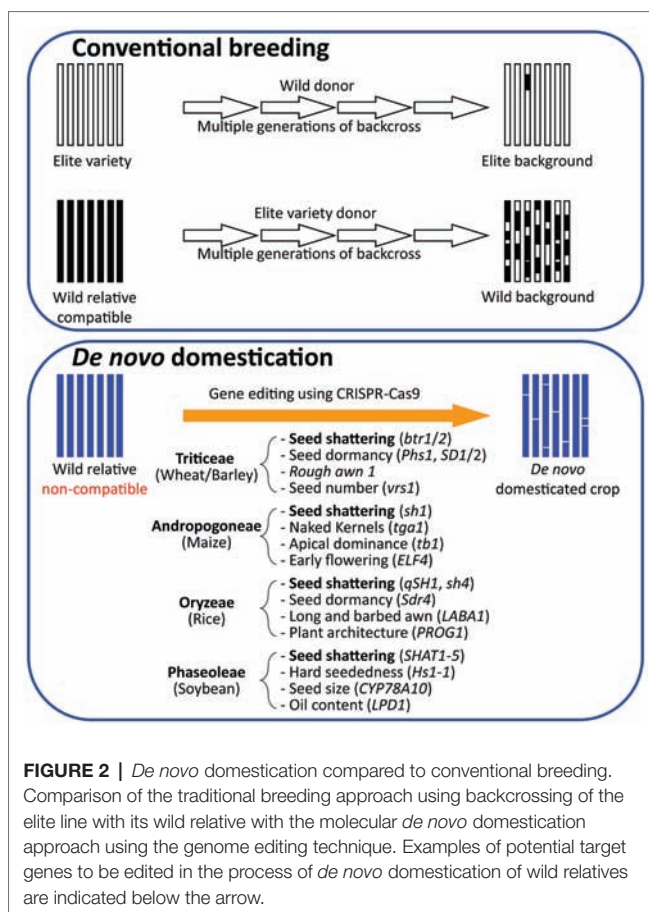
Traditionally, the new genetic material was transferred from CWRs to crop plants by introgression of new genes into elite cultivar background (**Figure 2**; Dempewolf et al., 2017). Genomic resources have been widely used to speed up the process *via* marker assisted selection, including transfer of disease resistance genes in grape vine, apple, and banana (Migicovsky and Myles, 2017). Despite its obvious success, especially in transfer of major genes, the method is time consuming and restricted to sexually compatible species. For example, *Hordeum vulgare* (cultivated barley) has been extensively crossed with cross-compatible *Hordeum spontaneum* (wild progenitor), and there has been limited success crossing cultivated barley with *Hordeum bulbosum*, where chromosome segment from *H. bulbosum* can be transferred to the chromosomes of cultivated barley (Westerbergh et al., 2018). There are however 32 species in the genus *Hordeum*, including diploid, tetraploid, and hexaploid varieties (Bothmer et al., 1995), and the vast majority of *Hordeum* species cannot be used due to crossing barriers. However, once the candidate genes have been identified, transgenics and genome editing technologies can be used to transfer the desirable genetic material between species regardless of natural crossing barriers. To aid improvement,

rich CWR genomic resources for many key crop species have been developed including soybean, rice, and maize.

Glycine soja – a wild relative of cultivated soybean (*Glycine max*) – has been shown to have a much more diverse gene pool compared to *G. max*, due to artificial selection during domestication and further loss as a result of modern breeding practices (Hyten et al., 2006; Kofsky et al., 2018). Wild and cultivated soybean differ in a number of agriculturally important traits, including pod shattering (Dong et al., 2014), determinate growth habit (Tian et al., 2010), and seed size (Zhou et al., 2015a; Kofsky et al., 2018). Additionally, it was shown that half of the annotated resistance-related sequences in *G. soja* were absent in both the landraces and cultivars (Zhou et al., 2015b). Despite the phenotypic differences, *G. soja* and *G. max* are cross-compatible, facilitating the transfer of desirable traits.

In maize (*Zea mays*), lowland teosinte (*Z. mays* ssp. *parviglumis*), highland teosinte (*Z. mays* ssp. *mexicana*), and the genus *Tripsacum* comprising nine species of warm-season, perennial grasses have been characterized as donors of important traits, which could be used for improvement (Mammadov et al., 2018). Genome-wide studies demonstrated that over 10% of the maize genome shows evidence of introgression from the *mexicana* genome, suggesting its contribution to adaptation and improvement (Hufford et al., 2013; Yang et al., 2017).

Rice (*Oryza sativa* L.) belongs to the genus *Oryza*, encompassing over 20 species, two of which are cultivated (*O. sativa* L. and *Oryza glaberrima* S.). The species are subdivided into several groups, and not all are cross-compatible. In recent analyses, *Oryza rufipogon*, a wild species believed to be the immediate progenitor of *O. sativa*, showed higher sequence diversity and harbored sequence and genes completely missing from the population of cultivated rice (Huang et al., 2012; Xu et al., 2012; Zhao et al., 2018b), highlighting the potential of its use for modern rice improvement.

In *Brassica*, a comparison of a CWR *Brassica macrocarpa*, with nine cultivated lines of *Brassica oleracea* showed that the former harbored unique disease resistance genes most likely lost during the domestication and improvement of elite *B. oleracea* germplasm (Golicz et al., 2016b).

The increasing abundance of genomic resources for CWRs will significantly aid future breeding efforts, helping identify the optimal crosses and genome editing targets.

De novo Crop Domestication

Another strategy for utilization of the wild plant resources is new crop (*de novo*) domestication. The domestication syndrome refers to a unique collection of phenotypic traits associated with the genetic change of an organism from a wild progenitor to a domesticated one. Most of the changes linked to the domestication syndrome, such as grain dispersal in wheat, barley, and rice; apical dominance in maize; fruit size in tomato; and grain quality in wheat, result from modification of a single or few genes (Frary et al., 2000; Clark et al., 2004; Konishi et al., 2006; Uauy et al., 2006; Dubcovsky and Dvorak, 2007; Pourkheirandish et al., 2015, 2018). Also, most of them are due to a loss of function mutation in the causal gene (Komatsuda et al., 2007; Ramsay et al., 2011; Ishimaru et al., 2013; Pourkheirandish et al., 2015).

For example, wheat, barley, rice, maize, and sorghum were selected for inflorescence that retained the grains, which made it easy to harvest. This characteristic results from the loss of function mutations in the genes controlling shattering (Konishi et al., 2006; Lin et al., 2012; Pourkheirandish et al., 2015). Similarly, a domestication associated NAC gene controlling pod shattering resistance has been identified in soybean (Dong et al., 2014). Advances in genomics provided the necessary platform to facilitate gene discovery and identification such as detection of genes associated with non-brittle rachis in pasta wheat and seed filling in maize using whole-genome sequencing (Sosso et al., 2015; Avni et al., 2017); smooth awn in barley using genotyping by sequencing (Milner et al., 2019); seed quality in soybean; and cutin responsible for water retention in barley using RNA sequencing (Li et al., 2012, 2017a; Gao et al., 2018).

The syntenic and orthologous gene relationships among plant genomes are well demonstrated (Devos, 2005; Tang et al., 2008). Synteny allows identification of homologous genes and has been used to identify genes with similar functions in related species (Pourkheirandish et al., 2007; Chen et al., 2009; Sakuma et al., 2010; Ning et al., 2013). For example, grain retention in both wheat and barley results from a mutation in homologous genes *brittle rachis* 1 (Pourkheirandish et al., 2018). The *brittle rachis* 1 homologues appear to have a similar role in grain dispersal in wild progenitors of wheat and barley. A loss of function mutation in this gene results in spike stiffness in the domesticated lines. As the same gene controls brittleness in both wheat and barley, *brittle rachis* 1 most likely evolved before the divergence of *Triticum* (wheat genus) and *Hordeum* (barley genus) over 5 Mya (Middleton et al., 2014). This suggests that the other non-domesticated species within *Hordeum* and *Triticum* that are not cross fertile with cultivated wheat and barley probably carry the *brittle rachis* 1, which controls their mode of grain dispersal. Recently a study involving crop plant species from multiple families used genome-wide association study (GWAS) to identify a domestication-related gene controlling seed dormancy in soybean and then showed that orthologs of this gene in rice and tomato also display evidence of selection during domestication. Analysis of transgenic plants confirmed the conservation of function in soybean, rice, and *Arabidopsis*, highlighting the power of comparative genomics in new domestication target gene identification (Wang et al., 2018a).

A pre-existing knowledge of target gene makes further crop domestication speedy and feasible. Domestication of a new crop species allows access to a novel gene pool with the potential for generating new crops, which are productive, resilient, and nutritious. Recent successes in wild tomato domestication by editing loci important for yield and productivity provide a proof of concept (Li et al., 2018; Zsögön et al., 2018). For example, targeting of *brittle rachis* 1 gene in any wild species of *Hordeum* or *Triticum* using gene editing would disrupt its function and result in a significant step toward domestication of a new species. It is important to note that the ease of genome editing and therefore its use for crop *de novo* domestication and other applications is related to plant ploidy. Gene knockout efficiency is lower in polyploids compared to diploids, as multiple alleles must be edited simultaneously to achieve a similar effect (Zhang et al., 2019b).

Engineering Polyploidy

Polyploid plants possess three or more sets of homologous chromosomes stemming either from the duplication of a single genome (autopolyploidy) or hybridization followed by doubling of two diverged genomes (allopolyploidy; Comai, 2005). Many of the agriculturally important crop plants and staple food species are natural polyploids, including: bread wheat (allo-hexaploid; $6\times = 42$), pasta wheat (allo-tetraploid; $4\times = 28$), strawberry (allo-octaploid; $8\times = 56$), potato (auto-tetraploid; $4\times = 48$), and banana (auto-triploid; $3\times = 33$). Recent modeling work linked the occurrence of polyploidy to domestication (Salman-Minkov et al., 2016). Higher genome copy number masks deleterious mutations, increases the adaptive potential, and provides the opportunity for genes to gain new function. Thus, polyploidy is considered a major driver of evolution (Sattler et al., 2016). Induced polyploidy has also been used by breeders to develop new crops and flowers, such as triploid watermelon (seedless), hexaploid Triticale (a hybrid of wheat and rye), triploid tulips, roses, and many more ornamental flowers (Sattler et al., 2016). Polyploid plants tend to display hybrid vigor and improved abiotic stress tolerance (Chen, 2010; Tamayo-Ordóñez et al., 2016), with different manifestations of traits observed depending on the level of ploidy. For example, a study in *Arabidopsis*, which performed a rigorous comparison of plants with different somatic ploidy levels ($2\times$, $4\times$, $6\times$, and $8\times$) observed significant differences in phenotypes (Corneillie et al., 2019).

The engineering of polyploid plants has been proposed as one of the routes for the generation of improved crop varieties (Katche et al., 2019). However, a better understanding of the causes and effects of polyploidy is a necessary prerequisite. Two major routes of polyploid plant formation are *via* unreduced gametes or somatic doubling (Ramsey and Schemske, 1998; Tamayo-Ordóñez et al., 2016). In laboratory conditions, polyploidy can be induced by application of antimicrotubule drugs such as colchicine. The viability of polyploid plants depends on stabilization of mitotic and meiotic divisions (Comai, 2005). Understanding of the molecular mechanisms behind cell cycle control, homologous chromosome pairing, and meiotic crossover formation is therefore paramount. Molecular mechanisms controlling cell cycle progression are deeply conserved and rely on cyclins (CYCs) and cyclin dependent kinases (CKDs). Previous studies in *Arabidopsis thaliana* identified seven classes of CDKs, named CDKA through CDKF, but CDKA and CDKB were identified as major drivers of cell cycle in plants (Menges et al., 2005; Tank and Thaker, 2011; Tamayo-Ordóñez et al., 2016). An extensive literature search compiled a list of over a 100 meiosis-related genes in *Arabidopsis* (Gaebelein et al., 2019). Comparative genomics approaches can be used to find orthologs of those genes in other species and perform further characterization. For example, a recent study of synthetic allohexaploid *Brassica* hybrids ($2n = 6\times = $ AABBCC) identified genomic regions associated with fertility, which harbored orthologs of *A. thaliana* genes involved in meiosis (Gaebelein et al., 2019).

In addition, plant genomes are known to undergo extensive structural rearrangements and methylation changes upon polyploidization. A study of resynthesized *Brassica napus* lines demonstrated extensive restructuring of the merged genomes in the early generations following hybridization (Szadkowski et al., 2010). Many hybrids and recent allopolyploids display genome dominance, resulting in sub-genome biases in gene content and expression (Bird et al., 2018). Genomics can be used to track post-hybridization structural re-arrangements and the establishment of sub-genome dominance to better understand plant genome evolution post-hybridization (Edger et al., 2017). It can also help predict the optimal combination of different wild species to construct new synthetic crops that can diversify our agriculture and bring resilience to climate change.

As an example, bread wheat (*Triticum aestivum*), a major crop accounting for 20% of world daily food consumption, is an allohexaploid plant originated *via* multiple hybridizations. The most accepted hypothesis of its origin is demonstrated in **Figure 3** (Haider, 2013). Because the bread wheat carries genes of three different genomes (A, B, and D), it is robust and has been able to adapt to different climatic zones. Today, bread wheat (AABBDD), which originated from fertile crescent (30–35°N), can grow from Sweden (65°N) to Argentina or New Zealand (45°S), a cultivation zone much broader than that pasta wheat (AABB; Feuillet et al., 2008). Another example of a widely known polyploid plant is the octoploid strawberry (Edger et al., 2019). The modern strawberry arose from a series of hybridization events between diploid, tetraploid, and hexaploid species spanning Eurasia and North America (**Figure 4**).
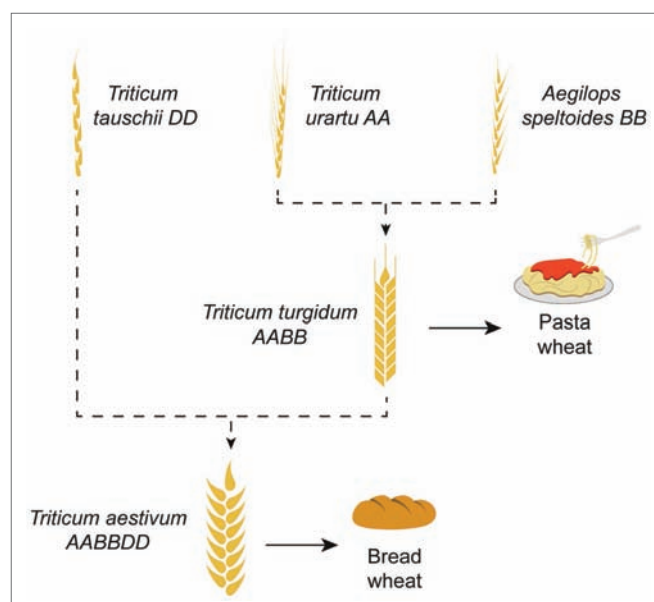
FIGURE 3 | The origin of bread wheat. Bread wheat (*Triticum aestivum*), is an allo-hexaploid originated *via* multiple hybridizations. The most accepted hypothesis of its origin is based on the hybridization of *Triticum urartu* (AA; $2\times = 14$) and *Aegilops speltoides* (BB; $2\times = 14$), resulting in tetraploid pasta wheat (AABB; $4\times = 28$). At the next step, hybridization of the tetraploid wheat with *Aegilops tauschii* (DD; $2\times = 14$) resulted in the emergence of the hexaploid bread wheat (AABBDD; $6\times = 42$).

FIGURE 4 | The origin of strawberry. Strawberry is an allo-octoploid originated from multiple hybridization events in Eurasia and America. Modern variety *Fragaria × ananassa* resulted from cross between two octoploid genotypes *Fragaria virginiana* and *Fragaria chiloensis*. Figure adapted from Bertioli (2019).
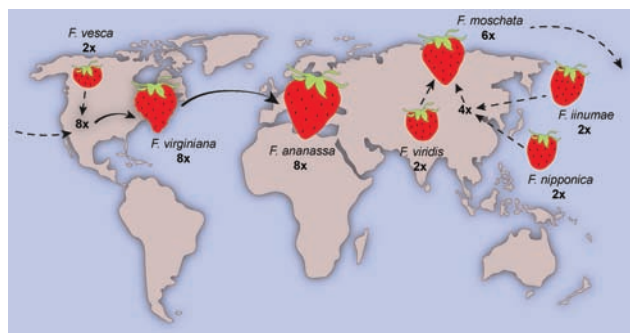
Harnessing Plant-Microbe Interactions to Boost Agricultural Output

Microbes which live within (endosphere) and surrounding plant roots in the soil (rhizosphere) have a significant impact on the host, including health and fitness, productivity, and responses to climate change (Wei et al., 2019). Plants and microbes interact *via* signaling molecules originating from both organisms (Leach et al., 2017; Chagas et al., 2018). Some bacterial communities have been shown to manipulate the plant potential to use soil resources, promote plant biotic and abiotic stress tolerance, and stimulate growth and nutrient uptake (Mendes et al., 2011; Berendsen et al., 2012; Fellbaum et al., 2012; Fitzpatrick et al., 2018; Paredes et al., 2018). At the other end of the spectrum, pathogenic microbes also exist, which negatively affect plant health (**Figure 5**; Pusztahelyi et al., 2015; Chagas et al., 2018).

Soil microorganisms are in constant competition to accumulate around the root and access the plant secreted carbohydrates (Venturi and Keel, 2016). Plants and microbes evolved together, resulting in beneficial microbes being attracted to a specific root exudate profile and forming a community of microbes in the rhizosphere (Garbeva et al., 2008). Plant species can therefore shape the composition of their rhizosphere's microbiome. Crops grown in the soils with microbial profiles similar to their native environment are expected to have a better chance of forming beneficial plant-microbiome interactions (Pérez-Jaramillo et al., 2018), promoting tolerance to biotic and abiotic stress. The advances in genomic technologies resulted in the sequencing of numerous soil microorganisms and improvement of our understanding of the soil microbial communities (Jansson and Hofmockel, 2018). For example, the availability of genomic sequences for nitrogen-fixing and phosphate-solubilizing bacteria expanded significantly (Zekic et al., 2017; Basenko et al., 2018; Jeong et al., 2018; Ormeño-Orrillo et al., 2018). Combined analysis of microbiome genomic and metabolomic data provides an accurate tool necessary to understand plant-microbe interactions and predict the most favorable crop plant–soil microbiome combinations, allowing for mapping of suitable crops to specific locations.
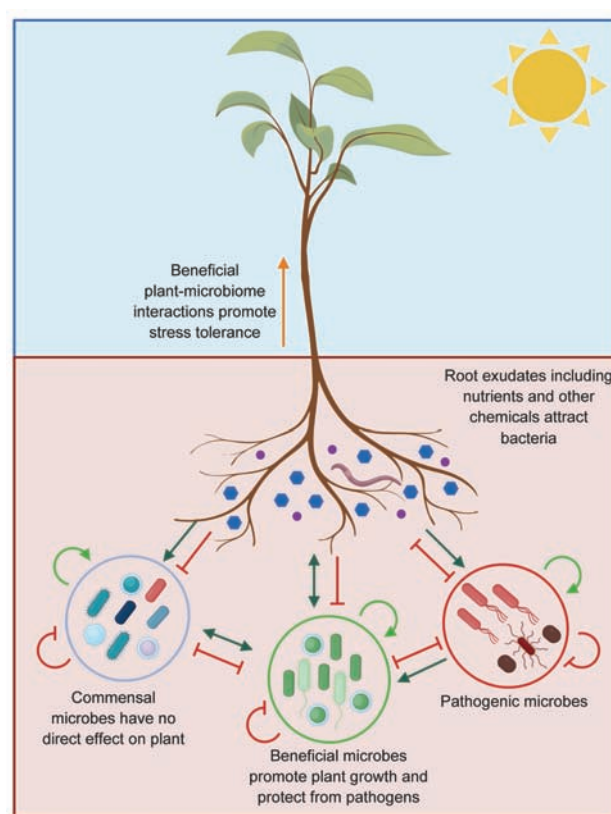

FIGURE 5 | Plant-microbe interactions in the rhizosphere. Plants can influence the composition of microbiome surrounding plant roots through exudation of compounds that stimulate (green arrows) or inhibit (red blocked arrows) microbes. A wide range of pathogens living in the soil can also affect plant health. Being able to attract the beneficial microbes will limit the success of the pathogenic microbes due to resource competition or by enhancing the plant immune system. The commensal microbes do not affect the plant or the pathogen directly. Microbiome-plant interactions are presented as described by Berendsen et al. (2012).

The Challenge of Climate Change and Plant Diseases

Crop plant pathogens are considered a major threat to modern agriculture (Oerke and Dehne, 2004; Savary et al., 2012; Nelson et al., 2018). The ongoing battle between plants and pathogens resulted in their co-evolution and shaped the genetic diversity of both (Jones and Dangl, 2006; Tiffin and Moeller, 2006; Dodds and Rathjen, 2010; Hartmann et al., 2017). Diseases generally result from a specific interaction between host and pathogen (Veresoglou and Rillig, 2014; Põlme et al., 2018). For example, the wheat leaf rust pathogen *Puccinia triticina*, one of the most common diseases of wheat globally, does not affect rice, maize, or any other crop. World-wide over-cultivation of a few crops (wheat, maize, rice, soybean, and barley) with low genetic diversity has led to the increased pathogen inoculum and accelerated pathogen evolution, promoting its spread globally (Savary et al., 2019).

Climate change affects the epidemiology of pathogens at specific locations and the geographic distribution of plant

diseases (Barford, 2013; Chakraborty, 2013). Increasing the crop plant diversity by cultivation of orphan crops and the domestication of new crops will result in reduced selective pressure on pathogen populations; thus, the life of genetic resistance is expected to be longer (Cook, 2006; Hajjar et al., 2008; Storkey et al., 2019). The life extension of genetic resistance could be an effective and ecologically sustainable way to control diseases. Climate change affects not only the crop but also the pathogen survival and reproduction. One of the expected impacts of climate change on plant disease is the migration of pathogens to latitudes beyond their historical range, examples of which have already been documented (Barford, 2013; Chakraborty, 2013). An increase in temperature would result in pathogen movement and spread of disease further from north in the northern hemisphere and south in the southern hemisphere to geographical locations in which they previously have not been able to reproduce effectively nor infect the plant. Recent genomic advances have resulted in the prediction and isolation of several resistance genes from crops and the identification of the corresponding genes from the pathogen (Fu et al., 2009; Mago et al., 2015; Moore et al., 2015; Sperschneider et al., 2015; Krattinger et al., 2016; Wan et al., 2019). These advances have provided a snapshot of resistance mechanisms that crops have developed during the long co-evolutionary history. The discovery of the genes underlying resistance has led to an improved understanding of their molecular function and established an entry point for studies of the defense pathways.

In addition, genome sequencing provides a rapid method of pathogen identification (Boykin et al., 2019), outbreak progression, and tracking of its spread to new locations. In fact, the development of third-generation sequencing technologies, especially Oxford Nanopore, resulted in the introduction of small, affordable, mobile sequencing instruments perfectly suited for in-field diagnostic system. Oxford Nanopore MinION technology has already been used for real-time diagnostics of human pathogens including Ebola (Quick et al., 2016) and Zika viruses (Faria et al., 2016) with protocols for identification of plant pathogens and pests under active development. For example, a recent proof-of-concept study has shown that using portable sequencing technology diagnostic test, it is possible to deliver test results within 48 h and, thus, greatly reduce the risk of community crop failure (Boykin et al., 2019).

Genome Editing for Nutritionally Enhanced Crop Production

Augmentation of crop nutritional value plays a central role in ensuring global food security. Breeding of crops for enhanced nutrient content has been a long standing goal of plant research (DellaPenna, 1999; Welch and Graham, 2002). Plants are a key source of macro- and micro-nutrients, but many of the staple foods, including cassava, wheat, rice, and maize are poor source of some macro-nutrients and many essential micro-nutrients (DellaPenna, 1999). However, nutrient profile can be altered by manipulation of biochemical pathways involved in macro- and micro-nutrient biosynthesis. Advances in genome sequencing and annotation provided the necessary resource

to identify the candidate genes involved in plant metabolism. As a result, genome editing technologies could be used to modify nutritional profiles of crops, for example producing soybeans with high oleic acid and low linoleic acid content (Haun et al., 2014; Demorest et al., 2016) and reducing anti-nutritional phytic acid content in maize (Liang et al., 2014). Nutritional enhancement of crops can also be achieved using transgenic technologies (Hefferon, 2015). In addition, genome editing facilitated *de novo* domestication of new nutrient rich crops could lead to a more diversified and healthier diet.

ACCESSING NEW BREEDING TARGETS USING GENOMIC TECHNOLOGIES

Third-Generation Sequencing for Improved Reference Genomes

The beginning of the twenty-first century saw rapid development of new sequencing methods. Second-generation sequencing technologies, including Illumina, allowed assembly of over 200 plant genomes (Chen et al., 2018) with much more ambitious plans of generating 10,000 draft genome assemblies by 2025 (Cheng et al., 2018). The main challenge posed by second-generation sequencing technologies was short-read length, making them unable to bridge over long stretches of repetitive sequences, resulting in fragmented assemblies. However, the introduction of third-generation sequencing and long reads produced by PacBio and Oxford Nanopore now allows for chromosomal level assemblies of plant genomes (Belser et al., 2018). The long-read sequencing technologies are often combined with optical mapping and conformation capture, achieving draft genomes of unprecedented contiguity (Belser et al., 2018; Shi et al., 2019). Importantly, the sequencing strategy used and the resulting contiguity and completeness of the assembly have been shown to impact downstream evolutionary and functional analyses. For example, comparative analysis of two *Brassica rapa* assemblies, one built using Illumina sequencing data and the other one using a PacBio, optical mapping (BioNano) and conformation capture (Hi-C) revealed that the latter harbored ~3,000 assembly specific genes as well as over 500 previously unidentified transposable element (TE) families (Zhang et al., 2018). The availability of high-quality, chromosome scale genome assemblies substantially improves the accuracy of the downstream genomic analysis, including gene and regulatory region annotation, GWAS, gene expression quantification, and homologue detection.

Accurate Gene Prediction and Functional Annotation for Precise Candidate Gene Identification

The explosion of plant genome sequencing was accompanied by extensive annotation efforts aiming to generate a comprehensive catalog of gene models for a given species. Gene model is defined as a region of the genome, which is believed to be transcribed into protein-coding messenger RNA (mRNA) or one of the classes of non-coding RNAs (ncRNA; Schnable, 2019). Gene models are often built using a combination of *ab*

initio gene prediction and homology-based methods that take advantage of sequence similarity to known transcripts or proteins (Campbell et al., 2014; Klasberg et al., 2016). Early on gene expression evidence was mostly derived from expressed sequence tags (ESTs) and extended by full-length sequencing *via* cloning followed by Sanger sequencing. Later, the information was supplemented by RNASeq data from diverse tissues, and it was shown that gene models and isoforms with highly tissue-specific expression were underrepresented in exiting annotations (Cheng et al., 2017; Golicz et al., 2018b; Van Bel et al., 2019). Currently, addition of long reads generated by PacBio or Oxford Nanopore sequencing technologies allows for recovery of full-length transcripts, providing new insights into the extent of alternative splicing and transcriptome diversity (Cook et al., 2019). Annotation of loci harboring non-coding transcripts is also becoming routine, further improving our understanding of the complexity of plant transcriptomes (Van Bel et al., 2019).

Despite the availability of genome annotations, functional characterization of annotated genes, which allows for the direct connection between genome and phenome, poses a key challenge in molecular breeding pipelines (Scheben and Edwards, 2018). In the key experimental model plant species, *A. thaliana*, >90% of genes have been annotated with putative functions and ~50% of genes have annotation supported by experimental evidence (Van Bel et al., 2019). However, for most of the crop plants, gene functional annotations rely on homology-based inference and are performed by transfer of annotation from most similar genes in model plants like *Arabidopsis* and rice, with very little direct experimental support. Annotation transfer is further complicated by plant evolutionary history, where successive rounds of polyploidy and subsequent diploidization lead to gene redundancy, differential loss, and neo- and sub-functionalization (Jiao and Paterson, 2014; Salman-Minkov et al., 2016). However, rapid progress in application of CRISPR/Cas9 genome editing will soon allow construction of genome-wide mutant libraries for key crops, significantly contributing to the functional annotation efforts. In fact, such libraries are already available for rice (Lu et al., 2017; Meng et al., 2017). Integrative genomics approaches have also been used to facilitate discovery of top candidates. For example, specialized databases integrating genotypic, phenotypic, and association data have been developed for rice (SNP-Seek), soybean (SoyBase), and wheat (T3; Grant et al., 2010; Blake et al., 2016; Mansueto et al., 2017). Beyond specialized database, tools like KnetMiner and MCRiceRepGP were developed aiming to rank candidate genes involved in biological processes of interest using multicriteria decision analysis (Hassani-Pak and Rawlings, 2017; Golicz et al., 2018b).

Non-coding Part of Genome as a Reservoir of New Breeding Targets

Only several percent of most large crop plant genomes encode protein-coding genes and the remainder is made up of non-coding sequences. For a long time, the non-coding stretches of DNA were considered to have little function; however, recent technological and conceptual developments revealed that plant genomes encode thousands of potentially functional ncRNAs as well as prevalence of distant regulatory elements including

enhancers (Weber et al., 2016). The ncRNAs encompass several classes of transcripts, including not only the relatively well characterized ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nucleolar RNAs (snoRNAs), and micro RNA (miRNAs) but also much more poorly understood long non-coding RNAs (lncRNAs). LncRNAs are transcripts over 200 base pairs in length without discernible protein coding protential, identified from RNASeq data, and have been shown to be involved in a range of biological processes, including flowering time regulation, stress tolerance, and gamete formation (Golicz et al., 2018a). At least some of the lncRNAs are likely to be functional, as evidenced by mutant phenotypes of knock-outs of newly discovered lncRNAs (Huang et al., 2018). Interestingly, lncRNAs have a strong bias toward transcription in reproductive tissues (Zhang et al., 2014; Golicz et al., 2018b; Johnson et al., 2018), suggesting involvement in plant sexual reproduction, a critical process affecting flowering, fruit, and grain formation. Newly characterized lncRNA, which affect important traits, can become genome editing targets. For example, a rice lncRNA LDMAR was shown to be involved in control of photoperiod-sensitive male sterility (PSMS), a key trait which contributed to the development of hybrid rice (Ding et al., 2012).

Another promising category of non-coding DNA sequences are cis-regulatory elements (CREs, promoters and enhancers/silencers), capable of recruiting transcription factors and promoting gene expression. Changes in CREs are considered one of the key evolutionary mechanisms underlying, for example, emergence of novel morphological forms (Stern and Orgogozo, 2008; Weber et al., 2016) and the divergence of cis-regulatory regions associated with domestication underscore their important roles in control of traits targeted by artificial selection (Lemmon et al., 2014; Wang et al., 2017). Several enhancers have been identified that modulate the expression of genes involved in the control of important traits, like anthocyanin content in maize and flowering time in *Arabidopsis* (Chua et al., 2003; Louwers et al., 2009; Adrian et al., 2010). SNPs corresponding to the different rapeseed ecotype groups were also identified in the promoter regions of FLOWERING LOCUS T and FLOWERING LOCUS C orthologs (two key genes controlling flowering time; Wu et al., 2019). An SNP corresponding to spatial expression of a homeobox transcription factor was selected for during the selection of non-shattering rice (Konishi et al., 2006). An insertion of TE in the CRE of *teosinte branched* 1 gene was discovered as the reason for apical dominance in maize also selected in the course of plant domestication (Studer et al., 2011). In the last few years, a significant progress has been made in identification of plant CREs with studies of the model plant species *Arabidopsis* as well as rice, maize, and cotton (Zhang et al., 2012; Pajoro et al., 2014; Zhu et al., 2015; Rodgers-Melnick et al., 2016; Oka et al., 2017; Wang et al., 2017; Bajic et al., 2018; Tannenbaum et al., 2018; Zhao et al., 2018a; Yan et al., 2019). The rapid developments are due to adoption of DNase-Seq and ATAC-Seq techniques in plant research, which measure DNA "openness" as a proxy for the accessibility of DNA to transcription factors, RNA polymerase, and other protein complexes involved in gene expression (Pajoro et al., 2014; Wang et al., 2016). Improved understanding of

the function of the non-coding elements of the genome will provide a new, yet untapped pool of breeding targets.

Beyond Single Reference Genomics – The Pan-Genome Approach

Generation of the reference genomes and subsequent large-scale re-sequencing of hundreds to thousands of individuals per species revealed extensive genomic diversity, including large-scale presence/absence variation (Golicz et al., 2016a; Varshney et al., 2017, 2019; Fuentes et al., 2019; Wu et al., 2019). As our knowledge of genomic variation increased, it become apparent that a single reference sequence is insufficient to represent the extent of genomic variation found within species, resulting in the introduction and adoption of the pangenome concept (**Figure 6**; Golicz et al., 2020). Pangenome represents the entirety of the genomic sequence and gene content found within a species rather than a single individual. First introduced in bacteria (Tettelin et al., 2005), it is highly relevant to plant research with more than 50% of genes in some species being variable (accessory), found in some individuals but not others (Golicz et al., 2020). Pangenomes have been constructed for key crop species, such as rice, soybean, bread wheat, and oilseed rape (Li et al., 2014; Golicz et al., 2016b; Contreras-Moreira et al., 2017; Gordon et al., 2017; Montenegro et al., 2017; Zhou et al., 2017; Hurgobin et al., 2018; Ou et al., 2018; Zhao et al., 2018b; Gao et al., 2019; Zhang et al., 2019a). Plant accessory genes have been shown to be over-represented in functions related to signaling and disease resistance as well as abiotic stress response (Golicz et al., 2016b; Montenegro et al., 2017; Hurgobin et al., 2018; Wang et al., 2018b), perhaps contributing to environmental adaptation and phenotypic plasticity and providing promising targets for crop improvement. Especially, since some of the accessory genes may be completely missing from the elite germplasm.

In addition, the pangenome offers a natural replacement for the current paradigm of using a single reference genome, as the choice of the reference affects downstream genomic analyses, including GWAS and gene expression quantification (Gage et al., 2019). Using pangenome as a reference improves read mapping and variant calling accuracy (Eggertsson et al., 2017; Garrison et al., 2018; Kim et al., 2019; Tian et al., 2019). The adoption

FIGURE 6 | Schematic representation of the plant pangenome. Pangenome represents the entire genomic sequence found in the species.

of the pangenome reference will also allow the inclusion of variants beyond SNPs in GWAS. Several studies in both plants and humans showed that the inclusion of structural variants in association studies could help identify causal variants (Chiang et al., 2017; Fuentes et al., 2019). For example, the use of sequence presence/absence variation allowed the identification of missing quantitative trait locus (QTLs) associated with disease resistance in oilseed rape (Gabur et al., 2018).

PAIRING GENOMICS WITH OTHER EMERGING TECHNOLOGIES TO MAXIMIZE THEIR POTENTIAL

Machine Learning and Crop Plant Genomics

Almost all aspects of genomic analyses can now be supported by the development and implementation of machine learning algorithms. Machine learning algorithms find new patterns and "learn" the necessary predictive features from the data, rather than rely on pre-existing criteria. This property makes them suitable for analysis of complex, multilayer datasets, where expert knowledge is incomplete or inaccurate, and when the amount of data is too large to be handled manually (Yip et al., 2013). Several promising applications of machine learning to plant genomics exits. As discussed above, the functional non-coding portions of plant genomes remain largely poorly understood. In animal research, machine learning and deep learning based methods have been particularity successful in genomic feature annotation, including regulatory regions like promoters, enhancers, and transcription factor binding sites (Xu and Jackson, 2019). The use of machine learning improved the quality of feature annotation, helped uncover the underlying sequence characteristics of the regulatory regions, and even allowed prediction of variant impact (Zhou and Troyanskaya, 2015; Kelley et al., 2016). However, the limited availability of large-scale epigenetic modification and chromatin accessibility datasets may delay similar studies in crop plants. While the lack of suitable datasets may be hampering regulatory region annotation, hundreds of sequenced and assembled genomes are readily available for comparative analyses. Identification of conserved and unique elements is one of the primary aims of comparative genomics. To date, sequence comparisons are mostly based on local or whole-genome alignments and limited by sensitivity of alignment tools. However, machine learning algorithms are being developed, which are capable of computing probability of sequence conservation for any query of interest (Joly-Lopez et al., 2016; Li et al., 2017b), providing new, exciting avenues for plant comparative genomics. Finally, plant phenotyping has legged significantly behind genotyping, requiring considerable resources and specialized equipment (Scheben and Edwards, 2018). One proposed application of machine learning is the prediction of phenotype from genotype and complementation of the more traditional genomic prediction models (Ma et al., 2018). Taken together, machine learning methods have the potential to add significant value to the existing genomic resources and methodologies.
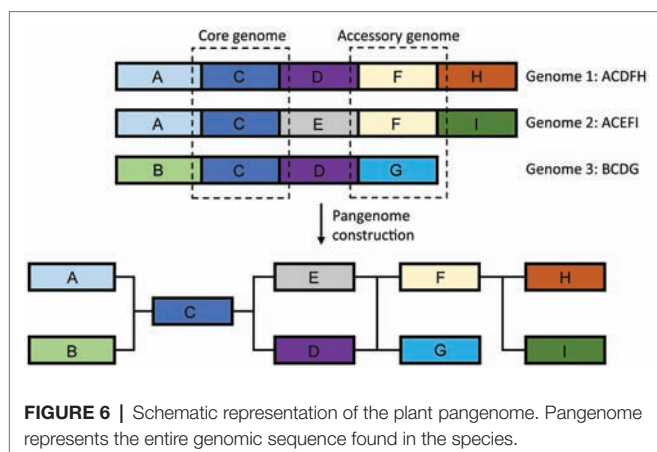
Speed Breeding to Accelerate the Development of New Crops

Advances in molecular and genomic technologies resulted in isolation and characterization of many agronomically important genes, for example ones controlling seed shattering, dormancy, increasing seed number, and size (Doebley et al., 2006). Improved understanding of the molecular function of these genes makes the new crop domestication and improvement of orphan crops feasible. However, the generation of new crops or improved crop varieties using traditional breeding techniques requires a lengthy process of recurrent selection, which can take many years (Gorjanc et al., 2018). One of the limiting factors in the process is the plant generation time, from seed germination to the harvest. The plant generation cycle takes up to 4 months in wheat and barley and even longer in others (Watson et al., 2018). Domestication of new crops would require numerous generations to stack the edited genes before crop release. Speed breeding is a procedure, which accelerates crop generation time by changing growth conditions, such as day length and temperature (Hickey et al., 2017). Growing long-day species under extended photoperiod (22 h light/2 h dark) and controlled temperature stimulates rapid flowering and maturation. The technology successfully shortened the plant generation time of some of the world's major agri-food crops, such as bread wheat, pasta wheat, barley, and canola (Watson et al., 2018). Production of up to six generations for wheat and barley is documented using speed breeding, which is much more efficient compared to two generations per year in traditional methods. Speed breeding protocols have also been successfully applied to orphan crops, such as chickpea, peanut, grass pea, lentil, and quinoa (O'Connor et al., 2013; Chiurugwi et al., 2019). The successful application of speed breeding to orphan crops indicates its flexibility and possible application for new crop domestication. A combination of speed breeding with our current knowledge about the target genes and genomic tools such as precision genome editing by CRISPR would make the new crop domestication feasible in a short time. Speed breeding can also be paired with genomic selection (GS), allowing further reductions in plant breeding cycles. GS is a modern breeding technology, which uses genome-wide markers to estimate the breeding values (EBV) and allows simultaneous selection for multiple traits. A recent study combined multivariate GS and speed breeding for yield prediction in spring wheat (Watson et al., 2019). Even though the current speed breeding protocols are limited to the long-day species, new protocols are expected for the short-day crops in the near future. Coupling speed breeding with genomics will make the GS for breeding and *de novo* domestication feasible.

High-Throughput Phenotyping

Plant phenotyping refers to the measurement of any morphological or physiological characteristics of plants. The phenotype can result from the action of individual genes, gene-by-gene, or gene-by-environment interactions. Many agronomically essential traits, such as yield and its components and drought/salt tolerance, are controlled by multiple genes with small effects and their interactions with the environment (Mickelbart et al., 2015).
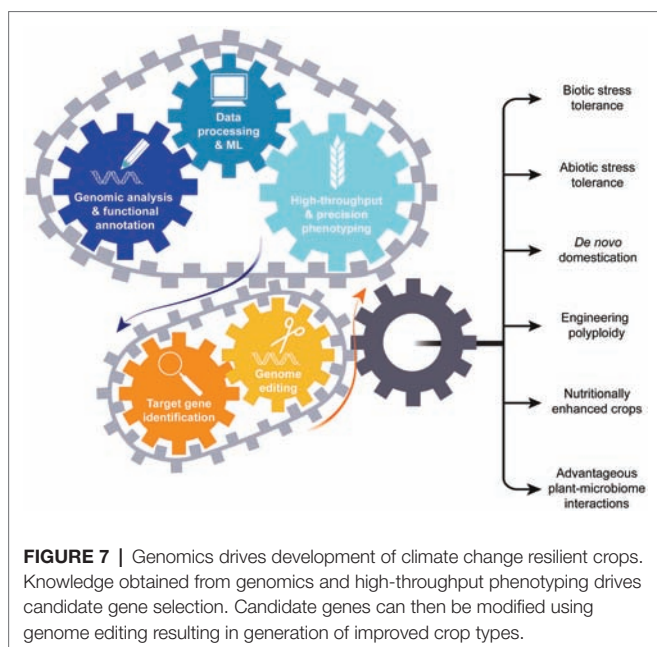
For practical reasons, many research groups focus on a controlled environment to grow plants and study their response to biotic and abiotic stresses (Velásquez et al., 2018). This includes stress induced by temperature, humidity, light, and other environmental factors. However, in farming, the environment and microclimate change dynamically during the day and affect the plant unevenly, for example due to shading. Moreover, controlled light conditions are hardly equivalent to the irradiance levels and spectral quality typical of natural sun conditions. There is a great need to study plant stresses in dynamic environmental conditions to thoroughly understand the complete picture of plant-stress responses. As the genotypic information is now available for hundreds or thousands of breeding lines in different species, collection and analysis of the corresponding high-throughput phenotyping data is one of the significant tasks ahead (Araus et al., 2018).

High-throughput phenotyping platforms, which employ robotics and spectral-based imaging technologies, are rapid and reliable (Galvez et al., 2019). The main limitation is the controlled environment, which is different from the natural growth conditions in the field. The introduction of hyperspectral imaging technology combined with drones and manned aircrafts provides an opportunity for high-throughput in-field phenotyping of traits, such as canopy temperature, chlorophyll fluorescence, as well as other biochemical plant characteristics (Camino et al., 2019). This technology increases the resolution and accuracy of measurements and is becoming cost-effective. The main challenge of using airborne platforms would be the analysis of large quantity of data in a short time frame (Singh et al., 2016; Taghavi Namin et al., 2018). However, machine learning based methods have shown promise in high-throughput phenotyping data processing. In-field high-throughput phenotyping is perfectly suited for evaluation of the complex physiological traits such as abiotic stresses tolerance.

CONCLUSION

Recent advances in genome sequencing, assembly, and annotation allowed unprecedented access to crop plant genomic information. High-throughput phenotyping techniques have been significantly advanced through the introduction of hyperspectral cameras and specialized processing software. Integration of genomic and phenomic data provides an opportunity to identify new agronomically relevant genes and characterize their functions. This knowledge has direct practical implications and can be translated to crop plant improvement using genome editing. While genome editing is currently applied in major crops and model plants, the technique has the potential to accelerate *de novo* domestication and allow rapid improvement of orphan crop plants, targeting the current and future climate challenges. The success of genomics in crop improvement is also influenced by the type of trait under investigation. For example, traits strongly affected by the environment and the interaction between genotype and the environment are more challenging to study and modify.

Disease resistance and dwarfing genes were introduced into crops such as wheat and rice during the green revolution (Khush, 2001).

FIGURE 7 | Genomics drives development of climate change resilient crops. Knowledge obtained from genomics and high-throughput phenotyping drives candidate gene selection. Candidate genes can then be modified using genome editing resulting in generation of improved crop types.

Breeders developed the high yielding varieties using the extra supply of nitrogen fertilizers in the presence of sufficient water under the climate conditions of the 1950–1960's. The equation is different today as climate change causes water shortages and temperature increases. However, the information gained from genomics and phenomics will drive candidate gene identification and enable genome editing (**Figure 7**), initiating the new crop plant breeding revolution.

AUTHOR CONTRIBUTIONS

MP and AG contributed equally. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Adrian, J., Farrona, S., Reimer, J. J., Albani, M. C., Coupland, G., and Turck, F. (2010). Cis–regulatory elements and chromatin state coordinately control temporal and spatial expression of FLOWERING LOCUS T in *Arabidopsis*. *Plant Cell* 22, 1425–1440. doi: 10.1105/tpc.110.074682

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., and Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23, 451–466. doi: 10.1016/j.tplants.2018.02.001

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S. O., Gundlach, H., et al. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* 357, 93–97. doi: 10.1126/science.aan0032

Bajic, M., Maher, K. A., and Deal, R. B. (2018). "Identification of open chromatin regions in plant genomes using ATAC-Seq" in *Plant chromatin dynamics: Methods and protocols*. eds. M. Bemer and C. Baroux (New York, NY: Springer New York), 183–201.

Barford, E. (2013). Crop pests advancing with global warming. *Nature* doi: 10.1038/nature.2013.13644

Basenko, E. Y., Pulman, J. A., Shanmugasundram, A., Harb, O. S., Crouch, K., Starns, D., et al. (2018). FungiDB: an integrated bioinformatic resource for fungi and oomycetes. *J. fungi* 4:39. doi: 10.3390/jof4010039

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F. -C., Falentin, C., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* 4, 879–887. doi: 10.1038/s41477-018-0289-4

Berendsen, R. L., Pieterse, C. M. J., and Bakker, P. A. H. M. (2012). The rhizosphere microbiome and plant health. *Trends Plant Sci.* 17, 478–486. doi: 10.1016/j.tplants.2012.04.001

Bertioli, D. J. (2019). The origin and evolution of a favorite fruit. *Nat. Genet.* 51, 372–373. doi: 10.1038/s41588-019-0365-3

Bird, K. A., Vanburen, R., Puzey, J. R., and Edger, P. P. (2018). The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* 220, 87–93. doi: 10.1111/nph.15256

Blake, V. C., Birkett, C., Matthews, D. E., Hane, D. L., Bradbury, P., and Jannink, J. -L. (2016). The *Triticeae* toolbox: combining phenotype and genotype data to advance small-grains breeding. *Plant Genome* 9. doi: 10.3835/plantgenome2014.12.0099

Bothmer, R., Jacobsen, N., Baden, C., Jørgensen, R., and Linde-Laursen, I. (1995). *An ecogeographical study of the genus Hordeum*. Rome: International Plant Genetic Resources Institute.

Boykin, L. M., Sseruwagi, P., Alicai, T., Ateka, E., Mohammed, I. U., Stanton, J. L., et al. (2019). Tree lab: portable genomics for early detection of plant viruses and pests in sub-Saharan Africa. *Genes* 10:632. doi: 10.3390/genes10090632

Brozynska, M., Furtado, A., and Henry, R. J. (2016). Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol. J.* 14, 1070–1085. doi: 10.1111/pbi.12454

Camino, C., Gonzalez-Dugo, V., Hernandez, P., and Zarco-Tejada, P. J. (2019). Radiative transfer Vcmax estimation from hyperspectral imagery and SIF retrievals to assess photosynthetic performance in rainfed and irrigated plant phenotyping trials. *Remote Sens. Environ.* 231:111186. doi: 10.1016/j.rse.2019.05.005

Campbell, M. S., Law, M., Holt, C., Stein, J. C., Moghe, G. D., Hufnagel, D. E., et al. (2014). MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.* 164, 513–524. doi: 10.1104/pp.113.230144

Chagas, F. O., Pessotti, R. C., Caraballo-Rodriguez, A. M., and Pupo, M. T. (2018). Chemical signaling involved in plant-microbe interactions. *Chem. Soc. Rev.* 47, 1652–1704. doi: 10.1039/C7CS00343A

Chakraborty, S. (2013). Migrate or evolve: options for plant pathogens under climate change. *Glob. Chang. Biol.* 19, 1985–2000. doi: 10.1111/gcb.12205

Chen, Z. J. (2010). Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci.* 15, 57–71. doi: 10.1016/j.tplants.2009.12.003

Chen, F., Dong, W., Zhang, J., Guo, X., Chen, J., Wang, Z., et al. (2018). The sequenced angiosperm genomes and genome databases. *Front. Plant Sci.* 9:418. doi: 10.3389/fpls.2018.00418

Chen, G. X., Komatsudu, T., Pourkheirandish, M., Sameri, M., Sato, K., Krugman, T., et al. (2009). Mapping of the eibi1 gene responsible for the drought hypersensitive cuticle in wild barley (*Hordeum spontaneum*). *Breed. Sci.* 59, 21–26. doi: 10.1270/jsbbs.59.21

Cheng, C. -Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., and Town, C. D. (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 89, 789–804. doi: 10.1111/tpj.13415

Cheng, S., Melkonian, M., Smith, S. A., Brockington, S., Archibald, J. M., Delaux, P. -M., et al. (2018). 10KP: a phylodiverse genome sequencing plan. *GigaScience* 7, 1–9. doi: 10.1093/gigascience/giy013

Chiang, C., Scott, A. J., Davis, J. R., Tsang, E. K., Li, X., Kim, Y., et al. (2017). The impact of structural variation on human gene expression. *Nat. Genet.* 49, 692–699. doi: 10.1038/ng.3834

Chirurgwi, T., Kemp, S., Powell, W., and Hickey, L. T. (2019). Speed breeding orphan crops. *Theor. Appl. Genet.* 132, 607–616. doi: 10.1007/s00122-018-3202-7

Chua, Y. L., Watson, L. A., and Gray, J. C. (2003). The transcriptional enhancer of the pea plastocyanin gene associates with the nuclear matrix and regulates gene expression through histone acetylation. *Plant Cell* 15, 1468–1479. doi: 10.1105/tpc.011825

Clark, R. M., Linton, E., Messing, J., and Doebley, J. F. (2004). Pattern of diversity in the genomic region near the maize domestication gene tb1. *Proc. Natl. Acad. Sci. U. S. A.* 101, 700–707. doi: 10.1073/pnas.2237049100

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711

Contreras-Moreira, B., Cantalapiedra, C. P., García-Pereira, M. J., Gordon, S. P., Vogel, J. P., Igartua, E., et al. (2017). Analysis of plant pan-genomes and transcriptomes with GET_HOMOLOGUES-EST, a clustering solution for sequences of the same species. *Front. Plant Sci.* 8:184. doi: 10.3389/fpls.2017.00184

Cook, R. J. (2006). Toward cropping systems that enhance productivity and sustainability. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18389–18394. doi: 10.1073/pnas.0605946103

Cook, D. E., Valle-Inclan, J. E., Pajoro, A., Rovenich, H., Thomma, B. P. H. J., and Faino, L. (2019). Long-read annotation: automated eukaryotic genome annotation based on long-read cDNA sequencing. *Plant Physiol.* 179, 38–54. doi: 10.1104/pp.18.00848

Corneillie, S., De Storme, N., Van Acker, R., Fangel, J. U., De Bruyne, M., De Rycke, R., et al. (2019). Polyploidy affects plant growth and alters cell wall composition. *Plant Physiol.* 179, 74–87. doi: 10.1104/pp.18.00967

Dellapenna, D. (1999). Nutritional genomics: manipulating plant micronutrients to improve human health. *Science* 285, 375–379. doi: 10.1126/science.285.5426.375

Demorest, Z. L., Coffman, A., Baltes, N. J., Stoddard, T. J., Clasen, B. M., Luo, S., et al. (2016). Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. *BMC Plant Biol.* 16:225. doi: 10.1186/s12870-016-0906-1

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Sci.* 57, 1070–1082. doi: 10.2135/cropsci2016.10.0885

Devos, K. M. (2005). Updating the 'Crop circle'. *Curr. Opin. Plant Biol.* 8, 155–162. doi: 10.1016/j.pbi.2005.01.005

Ding, J., Lu, Q., Ouyang, Y., Mao, H., Zhang, P., Yao, J., et al. (2012). A long non-coding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc. Natl. Acad. Sci. U. S. A.* 109, 2654–2659. doi: 10.1073/pnas.1121374109

Dodds, P. N., and Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat. Rev. Genet.* 11, 539–548. doi: 10.1038/nrg2812

Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321. doi: 10.1016/j.cell.2006.12.006

Dong, Y., Yang, X., Liu, J., Wang, B. -H., Liu, B. -L., and Wang, Y. -Z. (2014). Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. *Nat. Commun.* 5:3352. doi: 10.1038/ncomms4352

Dubcovsky, J., and Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316, 1862–1866. doi: 10.1126/science.1143986

Edger, P. P., Poorten, T. J., Vanburen, R., Hardigan, M. A., Colle, M., Mckain, M. R., et al. (2019). Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* 51, 541–547. doi: 10.1038/s41588-019-0356-4

Edger, P. P., Smith, R., Mckain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y., et al. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* 29, 2150–2167. doi: 10.1105/tpc.17.00010

Eggertsson, H. P., Jonsson, H., Kristmundsdottir, S., Hjartarson, E., Kehr, B., Masson, G., et al. (2017). Graphtyper enables population-scale genotyping using pangenome graphs. *Nat. Genet.* 49, 1654–1660. doi: 10.1038/ng.3964

Faria, N. R., Sabino, E. C., Nunes, M. R. T., Alcantara, L. C. J., Loman, N. J., and Pybus, O. G. (2016). Mobile real-time surveillance of Zika virus in Brazil. *Genome Med.* 8:97. doi: 10.1186/s13073-016-0356-2

Fellbaum, C. R., Gachomo, E. W., Beesetty, Y., Choudhari, S., Strahan, G. D., Pfeffer, P. E., et al. (2012). Carbon availability triggers fungal nitrogen uptake and transport in arbuscular mycorrhizal symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* 109, 2666–2671. doi: 10.1073/pnas.1118650109

Feuillet, C., Langridge, P., and Waugh, R. (2008). Cereal breeding takes a walk on the wild side. *Trends Genet.* 24, 24–32. doi: 10.1016/j.tig.2007.11.001

Fitzpatrick, C. R., Copeland, J., Wang, P. W., Guttman, D. S., Kotanen, P. M., and Johnson, M. T. J. (2018). Assembly and ecological function of the root microbiome across angiosperm plant species. *Proc. Natl. Acad. Sci. U. S. A.* 115, E1157–E1165. doi: 10.1073/pnas.1717617115

Frary, A., Nesbitt, T. C., Grandillo, S., Knaap, E., Cong, B., Liu, J., et al. (2000). fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289, 85–88. doi: 10.1126/science.289.5476.85

Fu, D., Uauy, C., Distelfeld, A., Blechl, A., Epstein, L., Chen, X., et al. (2009). A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. *Science* 323, 1357–1360. doi: 10.1126/science.1166289

Fuentes, R. R., Chebotarov, D., Duitama, J., Smith, S., De La Hoz, J. F., Mohiyuddin, M., et al. (2019). Structural variants in 3000 rice genomes. *Genome Res.* 29, 870–880. doi: 10.1101/gr.241240.118

Gabur, I., Chawla, H. S., Liu, X., Kumar, V., Faure, S., Von Tiedemann, A., et al. (2018). Finding invisible quantitative trait loci with missing data. *Plant Biotechnol. J.* 16, 2102–2112. doi: 10.1111/pbi.12942

Gaebelein, R., Schiessl, S. V., Samans, B., Batley, J., and Mason, A. S. (2019). Inherited allelic variants and novel karyotype changes influence fertility and genome stability in *Brassica* allohexaploids. *New Phytol.* 223, 965–978. doi: 10.1111/nph.15804

Gage, J. L., Vaillancourt, B., Hamilton, J. P., Manrique-Carpintero, N. C., Gustafson, T. J., Barry, K., et al. (2019). Multiple maize reference genomes impact the identification of variants by genome-wide association study in a diverse inbred panel. *Plant Genome* 12. doi: 10.3835/plantgenome2018.09.006

Galvez, S., Merida-Garcia, R., Camino, C., Borrill, P., Abrouk, M., Ramirez-Gonzalez, R. H., et al. (2019). Hotspots in the genomic architecture of field drought responses in wheat as breeding targets. *Funct. Integr. Genomic.* 19, 295–309. doi: 10.1007/s10142-018-0639-3

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051. doi: 10.1038/s41588-019-0410-2

Gao, H., Wang, Y., Li, W., Gu, Y., Lai, Y., Bi, Y., et al. (2018). Transcriptomic comparison reveals genetic variation potentially underlying seed developmental evolution of soybeans. *J. Exp. Bot.* 69, 5089–5104. doi: 10.1093/jxb/ery291

Garbeva, P., Van Elsas, J. D., and Van Veen, J. A. (2008). Rhizosphere microbial community and its response to plant species and soil history. *Plant Soil* 302, 19–32. doi: 10.1007/s11104-007-9432-0

Garrison, E., Siren, J., Novak, A. M., Hickey, G., Eizenga, J. M., Dawson, E. T., et al. (2018). Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat. Biotechnol.* 36, 875–879. doi: 10.1038/nbt.4227

Godfray, H. C. J., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Nisbett, N., et al. (2010). The future of the global food system. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 365, 2769–2777. doi: 10.1098/rstb.2010.0180

Golicz, A. A., Batley, J., and Edwards, D. (2016a). Towards plant pangenomics. *Plant Biotechnol. J.* 14, 1099–1105. doi: 10.1111/pbi.12499

Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H., Martinez, P. A., et al. (2016b). The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat. Commun.* 7:13390. doi: 10.1038/ncomms13390

Golicz, A. A., Bayer, P. E., Bhalla, P. L., Batley, J., and Edwards, D. (2020). Pangenomics comes of age: from bacteria to plant and animal applications. *Trends Genet.* 36, 132–145. doi: 10.1016/j.tig.2019.11.006

Golicz, A. A., Bhalla, P. L., and Singh, M. B. (2018a). lncRNAs in plant and animal sexual reproduction. *Trends Plant Sci.* 23, 195–205. doi: 10.1016/j.tplants.2017.12.009

Golicz, A. A., Bhalla, P. L., and Singh, M. B. (2018b). MCRiceRepGP: a framework for the identification of genes associated with sexual reproduction in rice. *Plant J.* 96, 188–202. doi: 10.1111/tpj.14019

Gordon, S. P., Contreras-Moreira, B., Woods, D. P., Des Marais, D. L., Burgess, D., Shu, S., et al. (2017). Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nat. Commun.* 8:2184. doi: 10.1038/s41467-017-02292-8

Gorjanc, G., Gaynor, R. C., and Hickey, J. M. (2018). Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theor. Appl. Genet.* 131, 1953–1966. doi: 10.1007/s00122-018-3125-3

Grant, D., Nelson, R. T., Cannon, S. B., and Shoemaker, R. C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* 38, D843–D846. doi: 10.1093/nar/gkp798

Grassini, P., Eskridge, K. M., and Cassman, K. G. (2013). Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nat. Commun.* 4:2918. doi: 10.1038/ncomms3918

Haider, N. (2013). The origin of the B-genome of bread wheat (*Triticum aestivum* L.). *Russ. J. Genet.* 49, 263–274. doi: 10.1134/S1022795413030071

Hajjar, R., Jarvis, D. I., and Gemmill-Herren, B. (2008). The utility of crop genetic diversity in maintaining ecosystem services. *Agric. Ecosyst. Environ.* 123, 261–270. doi: 10.1016/j.agee.2007.08.003

Hartmann, F. E., Sánchez-Vallet, A., Mcdonald, B. A., and Croll, D. (2017). A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME J.* 11, 1189–1204. doi: 10.1038/ismej.2016.196

Hassani-Pak, K., and Rawlings, C. (2017). Knowledge discovery in biological databases for revealing candidate genes linked to complex phenotypes. *J. Integr. Bioinform.* 14:20160002. doi: 10.1515/jib-2016-0002

Haun, W., Coffman, A., Clasen, B. M., Demorest, Z. L., Lowy, A., Ray, E., et al. (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnol. J.* 12, 934–940. doi: 10.1111/pbi.12201

Hefferon, K. L. (2015). Nutritionally enhanced food crops; progress and perspectives. *Int. J. Mol. Sci.* 16, 3895–3914. doi: 10.3390/ijms16023895

Hickey, L. T., German, S. E., Pereyra, S. A., Diaz, J. E., Ziems, L. A., Fowler, R. A., et al. (2017). Speed breeding for multiple disease resistance in barley. *Euphytica* 213:64. doi: 10.1007/s10681-016-1803-2

Huang, L., Dong, H., Zhou, D., Li, M., Liu, Y., Zhang, F., et al. (2018). Systematic identification of long non-coding RNAs during pollen development and fertilization in *Brassica rapa*. *Plant J.* 96, 203–222. doi: 10.1111/tpj.14016

Huang, X., Kurata, N., Wei, X., Wang, Z. -X., Wang, A., Zhao, Q., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497–501. doi: 10.1038/nature11532

Hufford, M. B., Lubinksy, P., Pyhäjärvi, T., Devengenzo, M. T., Ellstrand, N. C., and Ross-Ibarra, J (2013). The genomic signature of crop-wild introgression in maize. *PLoS Genet.* 9:e1003477. doi: 10.1371/annotation/2eef7b5b-29b2-412f-8472-8fd7f9bd65ab

Hurgobin, B., Golicz, A. A., Bayer, P. E., Chan, C. K. K., Tirnaz, S., Dolatabadian, A., et al. (2018). Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnol. J.* 16, 1265–1274. doi: 10.1111/pbi.12867

Hyten, D. L., Song, Q., Zhu, Y., Choi, I. -Y., Nelson, R. L., Costa, J. M., et al. (2006). Impacts of genetic bottlenecks on soybean genome diversity. *Proc. Natl. Acad. Sci. U. S. A.* 103, 16666–16671. doi: 10.1073/pnas.0604379103

Ishimaru, K., Hirotsu, N., Madoka, Y., Murakami, N., Hara, N., Onodera, H., et al. (2013). Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. *Nat. Genet.* 45, 707–711. doi: 10.1038/ng.2612

Jansson, J. K., and Hofmockel, K. S. (2018). The soil microbiome—from metagenomics to metaphenomics. *Curr. Opin. Microbiol.* 43, 162–168. doi: 10.1016/j.mib.2018.01.013

Jeong, J. -J., Sang, M. K., Pathiraja, D., Park, B., Choi, I. -G., and Kim, K. D. (2018). Draft genome sequence of phosphate-solubilizing *Chryseobacterium* sp. strain ISE14, a biocontrol and plant growth-promoting rhizobacterium isolated from cucumber. *Genome Announc.* 6, e00612–e00618. doi: 10.1128/genomeA.00612-18

Jiao, Y., and Paterson, A. H. (2014). Polyploidy-associated genome modifications during land plant evolution. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 369:20130355. doi: 10.1098/rstb.2013.0355

Johnson, C., Conrad, L. J., Patel, R., Anderson, S., Li, C., Pereira, A., et al. (2018). Reproductive long intergenic noncoding RNAs exhibit male gamete specificity and polycomb repressive complex 2-mediated repression. *Plant Physiol.* 177, 1198–1217. doi: 10.1104/pp.17.01269

Joly-Lopez, Z., Flowers, J. M., and Purugganan, M. D. (2016). Developing maps of fitness consequences for plant genomes. *Curr. Opin. Plant Biol.* 30, 101–107. doi: 10.1016/j.pbi.2016.02.008

Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286

Katche, E., Quezada-Martinez, D., Katche, E. I., Vasquez-Teuber, P., and Mason, A. S. (2019). Interspecific hybridization for *Brassica* crop improvement. *Crop Breed. Genet. Genom.* 1:e190007. doi: 10.20900/cbgg20190007

Kelley, D. R., Snoek, J., and Rinn, J. L. (2016). Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res.* 26, 990–999. doi: 10.1101/gr.200535.115

Khoury, C. K., Bjorkman, A. D., Dempewolf, H., Ramirez-Villegas, J., Guarino, L., Jarvis, A., et al. (2014). Increasing homogeneity in global food supplies and the implications for food security. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4001–4006. doi: 10.1073/pnas.1313490111

Khush, G. S. (2001). Green revolution: the way forward. *Nat. Rev. Genet.* 2, 815–822. doi: 10.1038/35093585

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Klasberg, S., Bitard-Feildel, T., and Mallet, L. (2016). Computational identification of novel genes: current and future perspectives. *Bioinform. Biol. Insights* 10, 121–131. doi: 10.4137/BBI.S39950

Kofsky, J., Zhang, H., and Song, B. -H. (2018). The untapped genetic reservoir: the past, current, and future applications of the wild soybean (*Glycine soja*). *Front. Plant Sci.* 9:949. doi: 10.3389/fpls.2018.00949

Komatsuda, T., Pourkheirandish, M., He, C., Azhaguvel, P., Kanamori, H., Perovic, D., et al. (2007). Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc. Natl. Acad. Sci. U. S. A.* 104, 1424–1429. doi: 10.1073/pnas.0608580104

Konishi, S., Izawa, T., Lin, S. Y., Ebana, K., Fukuta, Y., Sasaki, T., et al. (2006). An SNP caused loss of seed shattering during rice domestication. *Science* 312, 1392–1396. doi: 10.1126/science.1126410

Krattinger, S. G., Sucher, J., Selter, L. L., Chauhan, H., Zhou, B., Tang, M., et al. (2016). The wheat durable, multipathogen resistance gene Lr34 confers partial blast resistance in rice. *Plant Biotechnol. J.* 14, 1261–1268. doi: 10.1111/pbi.12491

Leach, J. E., Triplett, L. R., Argueso, C. T., and Trivedi, P. (2017). Communication in the phytobiome. *Cell* 169, 587–596. doi: 10.1016/j.cell.2017.04.025

Lemmon, Z. H., Bukowski, R., Sun, Q., and Doebley, J. F. (2014). The role of cis regulatory evolution in maize domestication. *PLoS Genet.* 10:e1004745. doi: 10.1371/journal.pgen.1004745

Li, C., Chen, G., Mishina, K., Yamaji, N., Ma, J. F., Yukuhiro, F., et al. (2017a). A GDSL-motif esterase/acyltransferase/lipase is responsible for leaf water retention in barley. *Plant Direct* 1:e00025. doi: 10.1002/pld3.25

Li, Y., Quang, D., and Xie, X. (2017b). "Understanding sequence conservation with deep learning" in *Proceedings of the 8th ACM international conference on bioinformatics, computational biology, and health informatics.* Boston, Massachusetts, USA: ACM.

Li, C., Wang, A., Ma, X., Pourkheirandish, M., Sakuma, S., Wang, N., et al. (2012). An eceriferum locus, cer-zv, is associated with a defect in cutin responsible for water retention in barley (*Hordeum vulgare*) leaves. *Theor. Appl. Genet.* 126, 637–646. doi: 10.1007/s00122-012-2007-3

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., et al. (2018). Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* 36, 1160–1163. doi: 10.1038/nbt.4273

Li, Y. H., Zhou, G., Ma, J., Jiang, W., Jin, L. G., Zhang, Z., et al. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* 32, 1045–1052. doi: 10.1038/nbt.2979

Liang, Z., Zhang, K., Chen, K., and Gao, C. (2014). Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. *J. Genet. Genomics* 41, 63–68. doi: 10.1016/j.jgg.2013.12.001

Lin, Z., Li, X., Shannon, L. M., Yeh, C. T., Wang, M. L., Bai, G., et al. (2012). Parallel domestication of the shattering1 genes in cereals. *Nat. Genet.* 44, 720–724. doi: 10.1038/ng.2281

Louwers, M., Bader, R., Haring, M., Van Driel, R., De Laat, W., and Stam, M. (2009). Tissue- and expression level–specific chromatin looping at maize epialleles. *Plant Cell* 21, 832–842. doi: 10.1105/tpc.108.064329

Lu, Y., Ye, X., Guo, R., Huang, J., Wang, W., Tang, J., et al. (2017). Genome-wide targeted mutagenesis in rice using the CRISPR/Cas9 system. *Mol. Plant* 10, 1242–1245. doi: 10.1016/j.molp.2017.06.007

Ma, W., Qiu, Z., Song, J., Li, J., Cheng, Q., Zhai, J., et al. (2018). A deep convolutional neural network approach for predicting phenotypes from genotypes. *Planta* 248, 1307–1318. doi: 10.1007/s00425-018-2976-9

Mago, R., Zhang, P., Vautrin, S., Simkova, H., Bansal, U., Luo, M. C., et al. (2015). The wheat Sr50 gene reveals rich diversity at a cereal disease resistance locus. *Nat. Plants* 1:15186. doi: 10.1038/nplants.2015.186

Mammadov, J., Buyyarapu, R., Guttikonda, S. K., Parliament, K., Abdurakhmonov, I. Y., and Kumpatla, S. P. (2018). Wild relatives of maize, rice, cotton, and soybean: treasure troves for tolerance to biotic and abiotic stresses. *Front. Plant Sci.* 9:886. doi: 10.3389/fpls.2018.00886

Mansueto, L., Fuentes, R. R., Borja, F. N., Detras, J., Abriol-Santos, J. M., Chebotarov, D., et al. (2017). Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic Acids Res.* 45, D1075–D1081. doi: 10.1093/nar/gkw1135

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544, 427–433. doi: 10.1038/nature22043

Mendes, R., Kruijt, M., De Bruijn, I., Dekkers, E., Van Der Voort, M., Schneider, J. H. M., et al. (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* 332, 1097–1100. doi: 10.1126/science.1203980

Meng, X., Yu, H., Zhang, Y., Zhuang, F., Song, X., Gao, S., et al. (2017). Construction of a genome-wide mutant library in rice using CRISPR/Cas9. *Mol. Plant* 10, 1238–1241. doi: 10.1016/j.molp.2017.06.006

Menges, M., De Jager, S. M., Gruissem, W., and Murray, J. A. H. (2005). Global analysis of the core cell cycle regulators of *Arabidopsis* identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control. *Plant J.* 41, 546–566. doi: 10.1111/j.1365-313X.2004.02319.x

Mickelbart, M. V., Hasegawa, P. M., and Bailey-Serres, J. (2015). Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. *Nat. Rev. Genet.* 16, 237–251. doi: 10.1038/nrg3901

Middleton, C. P., Senerchia, N., Stein, N., Akhunov, E. D., Keller, B., Wicker, T., et al. (2014). Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the *Triticeae* tribe. *PLoS One* 9:e85761. doi: 10.1371/journal.pone.0085761

Migicovsky, Z., and Myles, S. (2017). Exploiting wild relatives for genomics-assisted breeding of perennial crops. *Front. Plant Sci.* 8:460. doi: 10.3389/fpls.2017.00460

Milner, S. G., Jost, M., Taketa, S., Mazon, E. R., Himmelbach, A., Oppermann, M., et al. (2019). Genebank genomics highlights the diversity of a global barley collection. *Nat. Genet.* 51, 319–326. doi: 10.1038/s41588-018-0266-x

Montenegro, J. D., Golicz, A. A., Bayer, P. E., Hurgobin, B., Lee, H., Chan, C. K. K., et al. (2017). The pangenome of hexaploid bread wheat. *Plant J.* 90, 1007–1013. doi: 10.1111/tpj.13515

Moore, J. W., Herrera-Foessel, S., Lan, C., Schnippenkoetter, W., Ayliffe, M., Huerta-Espino, J., et al. (2015). A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. *Nat. Genet.* 47, 1494–1498. doi: 10.1038/ng.3439

Nelson, R., Wiesner-Hanks, T., Wisser, R., and Balint-Kurti, P. (2018). Navigating complexity to breed disease-resistant crops. *Nat. Rev. Genet.* 19, 21–33. doi: 10.1038/nrg.2017.82

Ning, S., Wang, N., Sakuma, S., Pourkheirandish, M., Wu, J., Matsumoto, T., et al. (2013). Structure, transcription and post-transcriptional regulation of the bread wheat orthologs of the barley cleistogamy gene Cly1. *Theor. Appl. Genet.* 126, 1273–1283. doi: 10.1007/s00122-013-2052-6

O'connor, D. J., Wright, G. C., Dieters, M. J., George, D. L., Hunter, M. N., Tatnell, J. R., et al. (2013). Development and Application of Speed Breeding technologies in a commercial peanut breeding program. *Peanut Sci.* 40, 107–114. doi: 10.3146/PS12-12.1

Oerke, E. C., and Dehne, H. W. (2004). Safeguarding production—losses in major crops and the role of crop protection. *Crop Prot.* 23, 275–285. doi: 10.1016/j.cropro.2003.10.001

Oka, R., Zicola, J., Weber, B., Anderson, S. N., Hodgman, C., Gent, J. I., et al. (2017). Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol.* 18:137. doi: 10.1186/s13059-017-1273-4

Ormeño-Orrillo, E., Aguilar-Cuba, Y., and Zúñiga-Dávila, D. (2018). Draft genome sequence of *Rhizobium sophoriradicis* H4, a nitrogen-fixing bacterium associated with the leguminous plant *Phaseolus vulgaris* on the coast of Peru. *Genome Announc.* 6, e00241–e00218. doi: 10.1128/genomeA.00241-18

Ou, L., Li, D., Lv, J., Chen, W., Zhang, Z., Li, X., et al. (2018). Pan-genome of cultivated pepper (Capsicum) and its use in gene presence–absence variation analyses. *New Phytol.* 220, 360–363. doi: 10.1111/nph.15413

Pajoro, A., Madrigal, P., Muiño, J. M., Matus, J. T., Jin, J., Mecchia, M. A., et al. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol.* 15:R41. doi: 10.1186/gb-2014-15-3-r41

Paredes, S. H., Gao, T. X., Law, T. F., Finkel, O. M., Mucyn, T., Teixeira, P. J. P. L., et al. (2018). Design of synthetic bacterial communities for predictable plant phenotypes. *PLoS Biol.* 16:e2003962. doi: 10.1371/journal.pbio.2003962

Pérez-Jaramillo, J. E., Carrión, V. J., De Hollander, M., and Raaijmakers, J. M. (2018). The wild side of plant microbiomes. *Microbiome* 6:143. doi: 10.1186/s40168-018-0519-z

Pimentel, D., Wilson, C., Mccullum, C., Huang, R., Dwen, P., Flack, J., et al. (1997). Economic and environmental benefits of biodiversity. *Bioscience* 47, 747–757. doi: 10.2307/1313097

Põlme, S., Bahram, M., Jacquemyn, H., Kennedy, P., Kohout, P., Moora, M., et al. (2018). Host preference and network properties in biotrophic plant–fungal associations. *New Phytol.* 217, 1230–1239. doi: 10.1111/nph.14895

Pourkheirandish, M., Dai, F., Sakuma, S., Kanamori, H., Distelfeld, A., Willcox, G., et al. (2018). On the origin of the non-brittle rachis trait of domesticated Einkorn wheat. *Front. Plant Sci.* 8:2031. doi: 10.3389/fpls.2017.02031

Pourkheirandish, M., Hensel, G., Kilian, B., Senthil, N., Chen, G., Sameri, M., et al. (2015). Evolution of the grain dispersal system in barley. *Cell* 162, 527–539. doi: 10.1016/j.cell.2015.07.002

Pourkheirandish, M., Wicker, T., Stein, N., Fujimura, T., and Komatsuda, T. (2007). Analysis of the barley chromosome 2 region containing the six-rowed spike gene vrs1 reveals a breakdown of rice-barley micro collinearity by a transposition. *Theor. Appl. Genet.* 114, 1357–1365. doi: 10.1007/s00122-007-0522-4

Pusztahelyi, T., Holb, I. J., and Pocsi, I. (2015). Secondary metabolites in fungus-plant interactions. *Front. Plant Sci.* 6:573. doi: 10.3389/fpls.2015.00573

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., et al. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530, 228–232. doi: 10.1038/nature16996

Ramsay, L., Comadran, J., Druka, A., Marshall, D. F., Thomas, W. T., Macaulay, M., et al. (2011). INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene *teosinte branched* 1. *Nat. Genet.* 43, 169–172. doi: 10.1038/ng.745

Ramsey, J., and Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29, 467–501. doi: 10.1146/annurev.ecolsys.29.1.467

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8:e66428. doi: 10.1371/journal.pone.0066428

Rodgers-Melnick, E., Vera, D. L., Bass, H. W., and Buckler, E. S. (2016). Open chromatin reveals the functional maize genome. *Proc. Natl. Acad. Sci. U. S. A.* 113, E3177–E3184. doi: 10.1073/pnas.1525244113

Sakuma, S., Pourkheirandish, M., Matsumoto, T., Koba, T., and Komatsuda, T. (2010). Duplication of a well-conserved homeodomain-leucine zipper transcription factor gene in barley generates a copy with more specific functions. *Funct. Integr. Genomic.* 10, 123–133. doi: 10.1007/s10142-009-0134-y

Salman-Minkov, A., Sabath, N., and Mayrose, I. (2016). Whole-genome duplication as a key factor in crop domestication. *Nat. Plants* 2:16115. doi: 10.1038/nplants.2016.115

Satterthwaite, D., Mcgranahan, G., and Tacoli, C. (2010). Urbanization and its implications for food and farming. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 365, 2809–2820. doi: 10.1098/rstb.2010.0136

Sattler, M. C., Carvalho, C. R., and Clarindo, W. R. (2016). The polyploidy and its key role in plant breeding. *Planta* 243, 281–296. doi: 10.1007/s00425-015-2450-x

Savary, S., Ficke, A., Aubertot, J. N., and Hollier, C. (2012). Crop losses due to diseases and their implications for global food production losses and food security. *Food Sec.* 4, 519–537. doi: 10.1007/s12571-012-0200-5

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., Mcroberts, N., and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* 3, 430–439. doi: 10.1038/s41559-018-0793-y

Scheben, A., and Edwards, D. (2018). Bottlenecks for genome-edited crops on the road from lab to farm. *Genome Biol.* 19:178. doi: 10.1186/s13059-018-1555-5

Scheben, A., Wolter, F., Batley, J., Puchta, H., and Edwards, D. (2017). Towards CRISPR/Cas crops—bringing together genomics and genome editing. *New Phytol.* 216, 682–698. doi: 10.1111/nph.14702

Schnable, J. C. (2019). Genes and gene models, an important distinction. *New Phytol.* doi: 10.1111/nph.16011

Shi, J., Ma, X., Zhang, J., Zhou, Y., Liu, M., Huang, L., et al. (2019). Chromosome conformation capture resolved near complete genome assembly of broomcorn millet. *Nat. Commun.* 10:464. doi: 10.1038/s41467-018-07876-6

Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* 21, 110–124. doi: 10.1016/j.tplants.2015.10.015

Sosso, D., Luo, D., Li, Q. -B., Sasse, J., Yang, J., Gendrot, G., et al. (2015). Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. *Nat. Genet.* 47, 1489–1493. doi: 10.1038/ng.3422

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Manners, J. M., Singh, K. B., and Taylor, J. M. (2015). Advances and challenges in computational prediction of effectors from plant pathogenic fungi. *PLoS Pathog.* 11:e1004806. doi: 10.1371/journal.ppat.1004806

Springer, N. M., Anderson, S. N., Andorf, C. M., Ahern, K. R., Bai, F., Barad, O., et al. (2018). The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat. Genet.* 50, 1282–1288. doi: 10.1038/s41588-018-0158-0

Stern, D. L., and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution* 62, 2155–2177. doi: 10.1111/j.1558-5646.2008.00450.x

Storkey, J., Bruce, T. J. A., Mcmillan, V. E., and Neve, P. (2019). "Chapter 12—the future of sustainable crop protection relies on increased diversity of cropping systems and landscapes" in *Agroecosystem diversity*. eds. G. Lemaire, P. C. D. F. Carvalho, S. Kronberg and S. Recous (Cambridge, MA, USA: Academic Press), 199–209.

Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat. Genet.* 43, 1160–1163. doi: 10.1038/ng.942

Szadkowski, E., Eber, F., Huteau, V., Lodé, M., Huneau, C., Belcram, H., et al. (2010). The first meiosis of resynthesized *Brassica napus*, a genome blender. *New Phytol.* 186, 102–112. doi: 10.1111/j.1469-8137.2010.03182.x

Taghavi Namin, S., Esmaeilzadeh, M., Najafi, M., Brown, T. B., and Borevitz, J. O. (2018). Deep phenotyping: deep learning for temporal phenotype/genotype classification. *Plant Methods* 14:66. doi: 10.1186/s13007-018-0333-4

Tamayo-Ordóñez, M. C., Espinosa-Barrera, L. A., Tamayo-Ordóñez, Y. J., Ayil-Gutiérrez, B., and Sánchez-Teyer, L. F. (2016). Advances and perspectives in the generation of polyploid plant species. *Euphytica* 209, 1–22. doi: 10.1007/s10681-016-1646-x

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320, 486–488. doi: 10.1126/science.1153917

Tank, J. G., and Thaker, V. S. (2011). Cyclin dependent kinases and their role in regulation of plant cell cycle. *Biol. Plant.* 55, 201–212. doi: 10.1007/s10535-011-0031-9

Tannenbaum, M., Sarusi-Portuguez, A., Krispil, R., Schwartz, M., Loza, O., Benichou, J. I. C., et al. (2018). Regulatory chromatin landscape in *Arabidopsis thaliana* roots uncovered by coupling INTACT and ATAC-seq. *Plant Methods* 14:113. doi: 10.1186/s13007-018-0381-9

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc. Natl. Acad. Sci. U. S. A.* 102, 13950–13955. doi: 10.1073/pnas.0506758102

Tian, X., Li, R., Fu, W., Li, Y., Wang, X., Li, M., et al. (2019). Building a sequence map of the pig pan-genome from multiple de novo assemblies and Hi-C data. *Sci. China Life Sci.* 63, 750–763. doi: 10.1007/s11427-019-9551-7

Tian, Z., Wang, X., Lee, R., Li, Y., Specht, J. E., Nelson, R. L., et al. (2010). Artificial selection for determinate growth habit in soybean. *Proc. Natl. Acad. Sci. U. S. A.* 107, 8563–8568. doi: 10.1073/pnas.1000088107

Tiffin, P., and Moeller, D. A. (2006). Molecular evolution of plant immune system genes. *Trends Genet.* 22, 662–670. doi: 10.1016/j.tig.2006.09.011

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20260–20264. doi: 10.1073/pnas.1116437108

Tilman, D., and Clark, M. (2014). Global diets link environmental sustainability and human health. *Nature* 515, 518–522. doi: 10.1038/nature13959

Tomlinson, I. (2013). Doubling food production to feed the 9 billion: a critical perspective on a key discourse of food security in the UK. *J. Rural. Stud.* 29, 81–90. doi: 10.1016/j.jrurstud.2011.09.001

Uauy, C., Distelfeld, A., Fahima, T., Blechl, A., and Dubcovsky, J. (2006). A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science* 314, 1298–1301. doi: 10.1126/science.1133649

Van Bel, M., Bucchini, F., and Vandepoele, K. (2019). Gene space completeness in complex plant genomes. *Curr. Opin. Plant Biol.* 48, 9–17. doi: 10.1016/j.pbi.2019.01.001

Varshney, R. K., Saxena, R. K., Upadhyaya, H. D., Khan, A. W., Yu, Y., Kim, C., et al. (2017). Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nat. Genet.* 49, 1082–1088. doi: 10.1038/ng.3872

Varshney, R. K., Singh, V. K., Kumar, A., Powell, W., and Sorrells, M. E. (2018). Can genomics deliver climate-change ready crops? *Curr. Opin. Plant Biol.* 45, 205–211. doi: 10.1016/j.pbi.2018.03.007

Varshney, R. K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H. D., Yang, W., et al. (2019). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat. Genet.* 51, 857–864. doi: 10.1038/s41588-019-0401-3

Velásquez, A. C., Castroverde, C. D. M., and He, S. Y. (2018). Plant–pathogen warfare under changing climate conditions. *Curr. Biol.* 28, R619–R634. doi: 10.1016/j.cub.2018.03.054

Venturi, V., and Keel, C. (2016). Signaling in the rhizosphere. *Trends Plant Sci.* 21, 187–198. doi: 10.1016/j.tplants.2016.01.005

Veresoglou, S. D., and Rillig, M. C. (2014). Do closely related plants host similar arbuscular mycorrhizal fungal communities? A meta-analysis. *Plant Soil* 377, 395–406. doi: 10.1007/s11104-013-2008-2

Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N. P., et al. (2013). A prioritized crop wild relative inventory to help underpin global food security. *Biol. Conserv.* 167, 265–275. doi: 10.1016/j.biocon.ssss2013.08.011

Waltz, E. (2018). With a free pass, CRISPR-edited plants reach market in record time. *Nat. Biotechnol.* 36, 6–7. doi: 10.1038/nbt0118-6b

Wan, L., Koeck, M., Williams, S. J., Ashton, A. R., Lawrence, G. J., Sakakibara, H., et al. (2019). Structural and functional insights into the modulation of the activity of a flax cytokinin oxidase by flax rust effector AvrL567-A. *Mol. Plant Pathol.* 20, 211–222. doi: 10.1111/mpp.12749

Wang, Y., Jiang, R., and Wong, W. H. (2016). Modeling the causal regulatory network by integrating chromatin accessibility and transcriptome data. *Natl. Sci. Rev.* 3, 240–251. doi: 10.1093/nsr/nww025

Wang, M., Li, W., Fang, C., Xu, F., Liu, Y., Wang, Z., et al. (2018a). Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nat. Genet.* 50, 1435–1441. doi: 10.1038/s41588-018-0229-2

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., et al. (2018b). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49. doi: 10.1038/s41586-018-0063-9

Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., et al. (2017). Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat. Genet.* 49, 579–587. doi: 10.1038/ng.3807

Watson, A., Ghosh, S., Williams, M. J., Cuddy, W. S., Simmonds, J., Rey, M. D., et al. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat. Plants* 4, 23–29. doi: 10.1038/s41477-017-0083-8

Watson, A., Hickey, L. T., Christopher, J., Rutkoski, J., Poland, J., and Hayes, B. J. (2019). Multivariate genomic selection and potential of rapid indirect selection with speed breeding in spring wheat. *Crop Sci.* 59, 1945–1959. doi: 10.2135/cropsci2018.12.0757

Weber, B., Zicola, J., Oka, R., and Stam, M. (2016). Plant enhancers: a call for discovery. *Trends Plant Sci.* 21, 974–987. doi: 10.1016/j.tplants.2016.07.013

Wei, Z., Gu, Y., Friman, V. -P., Kowalchuk, G. A., Xu, Y., Shen, Q., et al. (2019). Initial soil microbiome composition and functioning predetermine future plant health. *Sci. Adv.* 5:eaaw0759. doi: 10.1126/sciadv.aaw0759

Welch, R. M., and Graham, R. D. (2002). Breeding crops for enhanced micronutrient content. *Plant Soil* 245, 205–214. doi: 10.1023/A:1020668100330

Westerbergh, A., Lerceteau-Köhler, E., Sameri, M., Bedada, G., and Lundquist, P. -O. (2018). Towards the development of perennial barley for cold temperate climates—evaluation of wild barley relatives as genetic resources. *Sustainability* 10:1969. doi: 10.3390/su10061969

Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., et al. (2019). Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. *Mol. Plant* 12, 30–43. doi: 10.1016/j.molp.2018.11.007

Xu, C., and Jackson, S. A. (2019). Machine learning and complex biological data. *Genome Biol.* 20:76. doi: 10.1186/s13059-019-1689-0

Xu, X., Liu, X., Ge, S., Jensen, J. D., Hu, F., Li, X., et al. (2012). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.* 30, 105–111. doi: 10.1038/nbt.2050

Yan, W., Chen, D., Schumacher, J., Durantini, D., Engelhorn, J., Chen, M., et al. (2019). Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. *Nat. Commun.* 10:1705. doi: 10.1038/s41467-019-09513-2

Yang, N., Xu, X. -W., Wang, R. -R., Peng, W. -L., Cai, L., Song, J. -M., et al. (2017). Contributions of *Zea mays* subspecies mexicana haplotypes to modern maize. *Nat. Commun.* 8:1874. doi: 10.1038/s41467-017-02063-5

Yip, K. Y., Cheng, C., and Gerstein, M. (2013). Machine learning and genome annotation: a match meant to be? *Genome Biol.* 14:205. doi: 10.1186/gb-2013-14-5-205

Zekic, F., Weselowski, B., and Yuan, Z. -C. (2017). Complete genome sequence of *Burkholderia cenocepacia* CR318, a phosphate-solubilizing bacterium isolated from corn root. *Genome Announc.* 5, e00490–e00417. doi: 10.1128/genomeA.00490-17

Zhang, L., Cai, X., Wu, J., Liu, M., Grob, S., Cheng, F., et al. (2018). Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation capture technologies. *Hortic. Res.* 5:50. doi: 10.1038/s41438-018-0071-9

Zhang, Y. -C., Liao, J. -Y., Li, Z. -Y., Yu, Y., Zhang, J. -P., Li, Q. -F., et al. (2014). Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* 15:512. doi: 10.1186/s13059-014-0512-1

Zhang, Y., Malzahn, A. A., Sretenovic, S., and Qi, Y. (2019b). The emerging and uncultivated potential of CRISPR technology in plant science. *Nat. Plants* 5, 778–794. doi: 10.1038/s41477-019-0461-5

Zhang, W., Zhang, T., Wu, Y., and Jiang, J. (2012). Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. *Plant Cell* 24, 2719–2731. doi: 10.1105/tpc.112.098061

Zhang, B., Zhu, W., Diao, S., Wu, X., Lu, J., Ding, C., et al. (2019a). The poplar pangenome provides insights into the evolutionary history of the genus. *Commun. Biol.* 2:215. doi: 10.1038/s42003-019-0474-7

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., et al. (2018b). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* 50, 278–284. doi: 10.1038/s41588-018-0041-z

Zhao, H., Zhang, W., Chen, L., Wang, L., Marand, A. P., Wu, Y., et al. (2018a). Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. *Plant Physiol.* 176, 2789–2803. doi: 10.1104/pp.17.01467

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., et al. (2015b). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33, 408–414. doi: 10.1038/nbt.3096

Zhou, P., Silverstein, K. A. T., Ramaraj, T., Guhlin, J., Denny, R., Liu, J., et al. (2017). Exploring structural variation and gene family architecture with De Novo assemblies of 15 Medicago genomes. *BMC Genomics* 18:261. doi: 10.1186/s12864-017-3654-1

Zhou, J., and Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning–based sequence model. *Nat. Methods* 12, 931–934. doi: 10.1038/nmeth.3547

Zhou, L., Wang, S. -B., Jian, J., Geng, Q. -C., Wen, J., Song, Q., et al. (2015a). Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. *Sci. Rep.* 5:9350. doi: 10.1038/srep09350

Zhu, B., Zhang, W., Zhang, T., Liu, B., and Jiang, J. (2015). Genome-wide prediction and validation of intergenic enhancers in *Arabidopsis* using open chromatin signatures. *Plant Cell* 27, 2415–2426. doi: 10.1105/tpc.15.00537

Zsögön, A., Čermák, T., Naves, E. R., Notini, M. M., Edel, K. H., Weinl, S., et al. (2018). *De novo* domestication of wild tomato using genome editing. *Nat. Biotechnol.* 36, 1211–1216. doi: 10.1038/nbt.4272

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

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