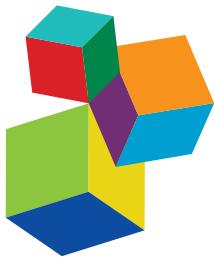


# PHENOMICS

EDITED BY: John Doonan and Marcos Egea-Cortines  
PUBLISHED IN: Frontiers in Plant Science



#### Frontiers Copyright Statement

© Copyright 2007-2018 Frontiers Media SA. All rights reserved.

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

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

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

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

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

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

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

ISSN 1664-8714  
ISBN 978-2-88945-607-9  
DOI 10.3389/978-2-88945-607-9

#### About Frontiers

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

#### Frontiers Journal Series

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

#### Dedication to Quality

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

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

#### What are Frontiers Research Topics?

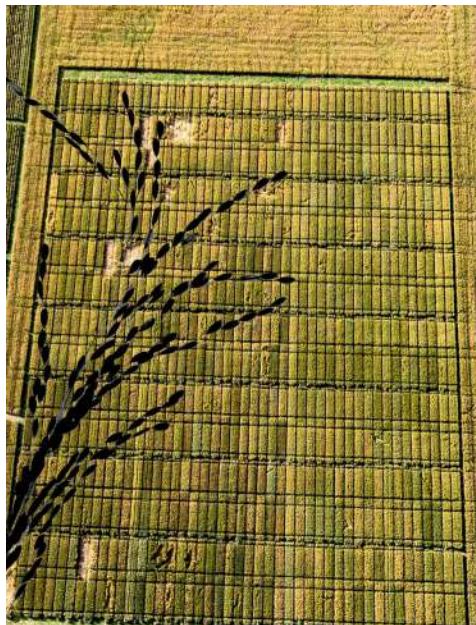
Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most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

# PHENOMICS

Topic Editors:

**John Doonan**, Aberystwyth University, United Kingdom

**Marcos Egea-Cortines**, Universidad Politécnica de Cartagena, Spain



Phenotyping across scales: genetically diverse trial plots imaged from UAV with super-imposed silhouette.

Image by Jason Brook and Wanneng Yang.

"Phenomics" is an emerging area of research whose aspiration is the systematic measurement of the physical, physiological and biochemical traits (the phenotype) belonging to a given individual or collection of individuals. Non-destructive or minimally invasive techniques allow repeated measurements across time to follow phenotypes as a function of developmental time. These longitudinal traits promise new insights into the ways in which crops respond to their environment including how they are managed.

To maximize the benefit, these approaches should ideally be scalable so that large populations in multiple environments can be sampled repeatedly at reasonable cost. Thus, the development and validation of non-contact sensing technologies remains an area of intensive activity that ranges from Remote Sensing of crops within the landscape to high resolution at the subcellular level. Integration of this potentially highly dimensional data and linking it with variation at the genetic level is an ongoing challenge that promises to release the potential of both established and under-exploited crops.

**Citation:** Doonan, J., Egea-Cortines, M., eds. (2018). Phenomics. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-607-9

# Table of Contents

- 05 Editorial: Phenomics**  
Marcos Egea-Cortines and John H. Doonan
- 07 Deciphering Phosphate Deficiency-Mediated Temporal Effects on Different Root Traits in Rice Grown in a Modified Hydroponic System**  
Manisha Negi, Raghavendarao Sanagala, Vandna Rai and Ajay Jain
- 21 SeedUSoon: A New Software Program to Improve Seed Stock Management and Plant Line Exchanges Between Research Laboratories**  
Céline Charavay, Stéphane Segard, Nathalie Pochon, Laurent Nussaume and Hélène Javot
- 37 Estimation of Wheat Plant Density at Early Stages Using High Resolution Imagery**  
Shouyang Liu, Fred Baret, Bruno Andrieu, Philippe Burger and Matthieu Hemmerlé
- 47 Evaluation of the SeedCounter, A Mobile Application for Grain Phenotyping**  
Evgenii Komyshhev, Mikhail Genaev and Dmitry Afonnikov
- 56 Non-destructive Phenotypic Analysis of Early Stage Tree Seedling Growth Using an Automated Stereovision Imaging Method**  
Antonio Montagnoli, Mattia Terzaghi, Nicoletta Fulgaro, Borys Stoew, Jan Wipenmyr, Dag Ilver, Cristina Rusu, Gabriella S. Scippa and Donato Chiatante
- 68 Exploring Relationships Between Canopy Architecture, Light Distribution, and Photosynthesis in Contrasting Rice Genotypes Using 3D Canopy Reconstruction**  
Alexandra J. Burgess, Renata Retkute, Tiara Herman and Erik H. Murchie
- 83 UAV-Based Thermal Imaging for High-Throughput Field Phenotyping of Black Poplar Response to Drought**  
Riccardo Ludovisi, Flavia Tauro, Riccardo Salvati, Sacha Khoury, Giuseppe Scarascia Mugnozza and Antoine Harfouche
- 101 Non-destructive Phenotyping of Lettuce Plants in Early Stages of Development With Optical Sensors**  
Ivan Simko, Ryan J. Hayes and Robert T. Furbank
- 120 Accurate Digitization of the Chlorophyll Distribution of Individual Rice Leaves Using Hyperspectral Imaging and an Integrated Image Analysis Pipeline**  
Hui Feng, Guoxing Chen, Lizhong Xiong, Qian Liu and Wanneng Yang
- 134 High-Throughput and Computational Study of Leaf Senescence Through a Phenomic Approach**  
Jae IL Lyu, Seung Hee Baek, Sukjoon Jung, Hyosub Chu, Hong Gil Nam, Jeongsik Kim and Pyung Ok Lim
- 142 Arabidopsis Seed Content QTL Mapping Using High-Throughput Phenotyping: The Assets of Near Infrared Spectroscopy**  
Sophie Jasinski, Alain Lécureuil, Monique Durandet, Patrick Bernard-Moulin and Philippe Guerche

- 153 Spectral Knowledge (SK-UTALCA): Software for Exploratory Analysis of High-Resolution Spectral Reflectance Data on Plant Breeding**  
Gustavo A. Lobos and Carlos Poblete-Echeverría
- 168 A Comprehensive Phenotypic Investigation of the "Pod-Shattering Syndrome" in Common Bean**  
Maria L. Murgia, Giovanna Attene, Monica Rodriguez, Elena Bitocchi, Elisa Bellucci, Davide Fois, Laura Nanni, Tania Gioia, Diego M. Albani, Roberto Papa and Domenico Rau
- 183 Genomic and Phenomic Screens for Flower Related RING Type Ubiquitin E3 Ligases in Arabidopsis**  
Mirko Pavicic, Katriina Mouhu, Feng Wang, Marcelina Bilicka, Erik Chovanček and Kristiina Himanen
- 200 Automated Method to Determine Two Critical Growth Stages of Wheat: Heading and Flowering**  
Pouria Sadeghi-Tehran, Kasra Sabermanesh, Nicolas Virlet and Malcolm J. Hawkesford
- 214 Plant Recycling for Molecular Biofarming to Produce Recombinant Anti-Cancer mAb**  
Deuk-Su Kim, Ilchan Song, Jinhee Kim, Do-Sun Kim and Kisung Ko



# Editorial: Phenomics

Marcos Egea-Cortines<sup>1\*</sup> and John H. Doonan<sup>2</sup>

<sup>1</sup> Genética Molecular, Instituto de Biotecnología Vegetal, Universidad Politécnica de Cartagena, Cartagena, Spain, <sup>2</sup> National Plant Phenomics Centre, Aberystwyth University, Aberystwyth, United Kingdom

**Keywords:** phenomics, artificial vision, RGB data, RGB image analysis, multispectral imaging

## Editorial on the Research Topic

### Phenomics

Developments in high throughput molecular technologies, from DNA sequencing to metabolite analysis and proteomics, have opened up new and previously undreamt of vistas in biology. Previously, one often had to make a difficult choice between longitudinal and cross-sectional studies but, with these highly scalable technologies, the number of individuals that can be screened has increased so dramatically that temporal studies are possible on whole populations. The technologies tend not to be specific to a given species, allowing us to sample the entire tree of life. One consequence of this technological explosion is that measurement of phenotypic traits for large populations over developmental time and in response to environmental variable has become highly desirable, if not a necessity (Houle et al., 2010). In this context, the development of new technologies to obtain reliable phenotypic data is a pre-requisite to approaching the overall challenge. As compared to the genotype (or even the proteome), the phenotype is highly dimensional to the extent that measuring all possible phenotypic traits is not feasible. However, the concept of "phenomics" has been proposed to cover sets of technologies devised to obtain phenotypic data in an analogous way to 'omics associated with the various molecular technologies. Phenomics therefore includes a vast array of approaches that, in most cases, include some sort of automatic sampling or non-invasive methods to obtain repeated sampling from an individual or population.

### OPEN ACCESS

#### Edited by:

Fabio Marroni,  
University of Udine, Italy

#### Reviewed by:

Ulrich Schurr,  
Forschungszentrum Jülich, Germany

#### \*Correspondence:

Marcos Egea-Cortines  
marcos.egea@upct.es

#### Specialty section:

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

Received: 23 March 2018

Accepted: 03 May 2018

Published: 23 May 2018

#### Citation:

Egea-Cortines M and Doonan JH  
(2018) Editorial: Phenomics.  
*Front. Plant Sci.* 9:678.  
doi: 10.3389/fpls.2018.00678

The general requirement for reproducibility is an additional driver for phenomics. While commercial phenotyping platforms can be very powerful (Virlet et al., 2017), the economic aspects of purchasing and maintenance and the lack of flexibility (in what are emerging technologies) has fostered in-house developments (Navarro et al., 2012; Lou et al., 2014). In plant biology, growth conditions (the environment) play a key role in the final phenotype of a plant and having well-defined growth parameters is not yet the rule (despite what the material and methods section of a typical peer-reviewed research paper might imply). In this special topic on Phenomics, Negi et al. addressed reproducibility of growth conditions, developing a modified hydroponic system to test for phosphate deficiency on rice root traits. The digital nature of the data is a major advantage as it allows sharing and re-use, both key to the success of the other 'omics technologies. An open-source software tool (Seedusoon) allows management of germplasm gathering together phenotypic and genetic data for a given accession (Charavay et al.).

Many of the non-destructive phenomic approaches rely on image analysis systems to acquire and process images. While the approach may seem straightforward, quantitative extraction of interesting features, such as intensity of image pixels, geometry of pixels or textures, remains challenging, and trade-offs between the ideal and the affordable are commonplace. For example, a key decision involves the type of camera used for data capture as that can limit the band width used to measure a given trait. This will have knock-on consequences, affecting the procedures used for image analysis (Navarro et al., 2016; Perez-Sanz et al., 2017). In the current edition, several publications address issues associated with the analysis of a variety of plants using different image acquisition devices. Standard cameras including those found in smartphones perform

image acquisition with an RED-GREEN-BLUE or RGB sensor. One study utilizes RGB images to determine wheat density at early stages of development (Liu et al.). There is an increasing number of publicly available libraries that facilitate image analysis (see Perez-Sanz et al., 2017 for a review). OpenCV, a widely used image processing library, underpinned development of SeedCounter (Komyshev et al.). This free Android App for mobile phone and pads, provides seed and grain morphometry under lab and field conditions, with much of the functionality of much more expensive equipment.

Stereo-vision is a long-established technique that uses two carefully positioned RGB cameras to capture 3-D information. Growth has been monitored in four species of tree seedling using the green channel and a stereo-vision approach (Montagnoli et al.). A regression model between the level of “greenness” and the real biomass obtained by destructive measures gave R values ranging between 0.67 for *Fagus sylvatica* and 0.95 for *Quercus ilex*, again showing actual differences between plants for a given setup. The interaction between canopy structure and photosynthesis has been studied by coupling 3-D reconstruction with gas exchange analysis showing that even complex traits such as 3-D structures can be related to photosynthesis efficiency (Burgess et al.).

The non-visible wavelengths can provide additional information on physiology and function. Thermal infrared imaging devices mounted on unmanned aerial vehicles (UAV) enables high throughput analysis of *Populus nigra* populations for dynamic responses to drought stress (Ludovisi et al.). Combined hyperspectral and thermal imaging of lettuce reveals how these plants adapt to multiple stresses (Simko et al.). Hyperspectral imaging has high information content and can measure several parameters simultaneously when calibrated. Thus, parallel analysis of chlorophyll a, chlorophyll b, total chlorophyll, and carotenoid in rice showed high correlation with hand measurements is 0.827–0.928 at the tillering stage, illustrating great potential to screen large populations (Feng et al.).

Using a combination of five non-invasive camera-based imaging units equipped with fluorescent, RGB Visible Near

Infrared (VNIR), Short Wave Infrared and three dimensional imaging, Lyu et al. determined a total of 200 quantitative traits during leaf senescence. This illustrates the enormous potential of phenomic approaches to have a comprehensive understanding of biological variation.

High-throughput screening of combinations of traits is the immediate promise of phenomics and is further exemplified by the use of near-infrared reflectance spectroscopic (NIRS) to undertake a coordinated analysis of oil, protein, carbon, and nitrogen content in *Arabidopsis* seeds. As a result, a set of QTLs controlling these traits, and the variance component of genotype, culture, Genetic by Environment interaction, and residual effect have been determined (Jasinski et al.).

Image-based approaches can be compromised by the quality of the signal obtained. This is an ongoing problem common to many ‘omics technologies where assessment of quality plays a key role in downstream data analysis. Directly addressing this problem (Lobos and Poblete-Echeverría) developed software to assess the quality of spectral reflectance data. As spectral reflectance data are widely used to obtain crop performance indices such as NDVI, this type of exploratory data analysis is essential for evaluating data quality.

## AUTHOR CONTRIBUTIONS

ME-C wrote draft. JD corrected it and both agreed on final version.

## ACKNOWLEDGMENTS

Work in the lab of ME-C is funded by MICINN BFU2017-88300-C2-1-R and Fundación Séneca 19398/PI/14. Work in the lab of JHD is funded by BBSRC grants BB/J004464/1; BB/M018407/1; BB/L009889/1; BB/CAP1730/1 and the European Union (EPPN Grant Agreement No. 284443 and EPPN2020 grant agreement No. 731013.

## REFERENCES

- Houle, D., Govindaraju, D., and Omholt, S. (2010). Phenomics: the next challenge. *Nat. Rev. Genet.* 11, 855–866. doi: 10.1038/nrg2897
- Lou, L., Liu, Y., Han, J., and Doonan, J. H. (2014). Accurate multi-view stereo 3D reconstruction for cost-effective plant phenotyping. *Lect. Notes Comput. Sci.* 8815, 349–356. doi: 10.1007/978-3-319-11755-3\_39
- Navarro, P. J. P. J., Fernández, C., Weiss, J., and Egea-Cortines, M. (2012). Development of a configurable growth chamber with a vision system to study circadian rhythm in plants. *Sensors* 12, 15356–15375. doi: 10.3390/s121115356
- Navarro, P. J., Pérez, F., Weiss, J., and Egea-Cortines, M. (2016). Machine learning and computer vision system for phenotype data acquisition and analysis in plants. *Sensors* 16:E641. doi: 10.3390/s16050641
- Perez-Sanz, F., Navarro, P. J., and Egea-Cortines, M. (2017). Plant phenomics: an overview of image acquisition technologies and image data analysis algorithms. *Gigascience* 6, 1–18. doi: 10.1093/gigascience/gix092

Virlet, N., Sabermanesh, K., Sadeghi-Tehran, P., and Hawkesford, M. J. (2017). Field Scanalyzer: an automated robotic field phenotyping platform for detailed crop monitoring. *Funct. Plant Biol.* 44:143. doi: 10.1071/FP16163

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

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



# Deciphering Phosphate Deficiency-Mediated Temporal Effects on Different Root Traits in Rice Grown in a Modified Hydroponic System

Manisha Negi, Raghavendarao Sanagala, Vandna Rai and Ajay Jain\*

National Research Centre on Plant Biotechnology, Lal Bahadur Shastri Building, New Delhi, India

## OPEN ACCESS

**Edited by:**

John Doonan,  
Aberystwyth University, UK

**Reviewed by:**

Philippe Nacry,  
French National Institute  
for Agricultural Research, France  
Ankush Prasad,  
Biomedical Engineering Research  
Center – Tohoku Institute  
of Technology, Japan

**\*Correspondence:**

Ajay Jain  
ajay1762jain@gmail.com

**Specialty section:**

This article was submitted to  
Technical Advances in Plant Science,  
a section of the journal  
*Frontiers in Plant Science*

**Received:** 24 October 2015

**Accepted:** 11 April 2016

**Published:** 04 May 2016

**Citation:**

Negi M, Sanagala R, Rai V and Jain A  
(2016) Deciphering Phosphate Deficiency-Mediated Temporal Effects on Different Root Traits in Rice Grown in a Modified Hydroponic System.  
*Front. Plant Sci.* 7:550.  
doi: 10.3389/fpls.2016.00550

Phosphate (Pi), an essential macronutrient for growth and development of plant, is often limiting in soils. Plants have evolved an array of adaptive strategies including modulation of root system architecture (RSA) for optimal acquisition of Pi. In rice, a major staple food, RSA is complex and comprises embryonically developed primary and seminal roots and post-embryonically developed adventitious and lateral roots. Earlier studies have used variant hydroponic systems for documenting the effects of Pi deficiency largely on primary root growth. Here, we report the temporal effects of Pi deficiency in rice genotype MI48 on 15 ontogenetically distinct root traits by using easy-to-assemble and economically viable modified hydroponic system. Effects of Pi deprivation became evident after 4 days- and 7 days-treatments on two and eight different root traits, respectively. The effects of Pi deprivation for 7 days were also evident on different root traits of rice genotype Nagina 22 (N22). There were genotypic differences in the responses of primary root growth along with lateral roots on it and the number and length of seminal and adventitious roots. Notably though, there were attenuating effects of Pi deficiency on the lateral roots on seminal and adventitious roots and total root length in both these genotypes. The study thus revealed both differential and comparable effects of Pi deficiency on different root traits in these genotypes. Pi deficiency also triggered reduction in Pi content and induction of several Pi starvation-responsive (PSR) genes in roots of MI48. Together, the analyses validated the fidelity of this modified hydroponic system for documenting Pi deficiency-mediated effects not only on different traits of RSA but also on physiological and molecular responses.

**Keywords:** *Oryza sativa*, phosphate deficiency, aerated hydroponic system, root system architecture, Pi content, Pi starvation-responsive genes

## INTRODUCTION

Rice, a major staple food in Asia, is grown largely under rain-fed ecosystem on soils that are naturally low in phosphorus (P) (Gamuyao et al., 2012). P is an essential macronutrient required for growth and development of plant (López-Arredondo et al., 2014). Root system plays a key role in acquisition of inorganic phosphate (Pi); a readily bioavailable source of P in rhizosphere

(Marschner, 1995). However, Pi is often limiting due to its slow diffusion rate and interactions with different soil constituents (Raghothama, 1999). Therefore, soil exploration by roots is critically important for optimal acquisition of Pi (Lynch, 2015).

*Arabidopsis thaliana*, a favored model plant species, has been extensively used for elucidation of Pi deficiency-mediated responses of root system architecture (RSA; Sánchez-Calderón et al., 2005; Gruber et al., 2013; Kellermeier et al., 2014). Conventionally, effects of Pi deprivation on RSA are documented by either growing on agar plate (Williamson et al., 2001; López-Bucio et al., 2002; Kellermeier et al., 2014) or hydroponically (Jain et al., 2009; Alatorre-Cobos et al., 2014) under aseptic condition. Pi deficiency induces inhibitory effects on the developmental responses of both embryonically and post-embryonically developed primary and lateral roots, respectively (Sánchez-Calderón et al., 2005; Jain et al., 2007).

Root system of *Oryza sativa* (rice) is relatively complex, comprising embryonically developed primary and seminal roots with post-embryonic adventitious roots making up bulk of the root system (Hochholdinger and Zimmermann, 2008). While primary and seminal roots play important roles during seedling stage, adventitious roots dominate the functional root system in mature plant (Hochholdinger et al., 2004). Different types of hydroponic system have been used for determining the effects of Pi deficiency on root development (Yi et al., 2005; Zhou et al., 2008; Dai et al., 2012, 2016; Wang et al., 2015). The effects of Pi deprivation have largely been focused on only a few root traits, i.e., total root length (Yi et al., 2005), primary root length (Shimizu et al., 2004; Zhou et al., 2008; Zheng et al., 2009; Hu et al., 2011; Dai et al., 2012, 2016; Wang S. et al., 2014; Wang et al., 2015), lateral root number on primary root (Wang S. et al., 2014), lateral root length (Yang et al., 2014), seminal root length (Ogawa et al., 2014); and number and/or length of adventitious roots (Zhou et al., 2008; Hu et al., 2011; Dai et al., 2012, 2016; Wang et al., 2015). None of these studies provided a holistic overview of the effects of Pi deficiency on different root traits. Different concentrations of Pi considered as P+ and P- media, variation in the duration of Pi deficiency treatment and use of different rice genotypes in these studies further makes it difficult to draw any explicit conclusion on the global effects of Pi deprivation on the developmental responses of ontogenetically distinct root traits.

In this study, we used modified hydroponic system for deciphering the effects of Pi deficiency on the developmental responses of primary, seminal and adventitious roots and also of lateral roots on each of them in rice cv. MI48 and N22. The modified hydroponic system was equally efficient for generating tissues for elucidation of Pi deficiency-mediated physiological and molecular responses.

## MATERIALS AND METHODS

### Plant Material and Seed Germination

Seeds of rice (*O. sativa* L. ssp. *indica*) genotype MI48 and Nagina 22 (N22) were used for this study. In Petri plate (110 mm × 25 mm), lined with filter paper and wetted with sterile water, 10 seeds were placed equi-distant and wrapped

in aluminum foil for maintaining dark condition. For each experiment, about 150–200 seeds were used. Petri plates were then placed in an incubator set at 28°C for 4 days. After germination, seedlings were transferred to Petri plate containing 1% (w/v) agar and scanned at 600 dots per inch (dpi) by using a desktop scanner. Scanned images were then used for documenting the radicle length by using ImageJ; a Java image-processing program (Collins, 2007). Seedlings often show significant variation in their radicle length. Therefore, for minimizing the effects of intrinsic variability on subsequent treatment under different Pi regime, only those seedlings with radicle length in the range of 2–3 cm were selected.

### Modified Hydroponic System

Autoclavable hydroponic system was assembled by easily available components, i.e., a standard polycarbonate transparent Magenta (GA-7) box (width × length × height = 75 mm × 74 mm × 138 mm), support made of polycarbonate sheet (0.030" thick), a polypropylene mesh (250 µm mesh size, width × length = 24" × 12" by Small Parts and available at amazon.com), aquarium air pump (power 5 W and pressure 2 MPa × 0.02 MPa), flexible air line tubing (3 mm in diameter) and tee connector. Polycarbonate sheets were cut into 80 mm × 40 mm rectangular pieces and notched at midpoint up to 20 mm so that the two pieces could fit together into an X-shaped support. Polypropylene mesh sheet was cut into a square piece (50 mm × 50 mm) and four holes (4 mm in diameter) were punched toward the perimeter for facilitating the penetration of radicle through the mesh into the nutrient medium. For experiments where rice seedlings are to be grown for a longer duration, the height of the X-shaped support could be easily increased up to 120 mm to ensure that root tip does not come in contact with the bottom of the Magenta box. Also, the number of seedlings in each magenta box could be reduced from four to a lesser number by punching the required number of holes in the mesh. Wedge support was placed into the Magenta box, filled with enough deionized water so that the level remained above the X-shaped wedge support and autoclaved. After autoclaving, water was removed from the Magenta box. To avoid warping, cut mesh pieces (10–15) were stacked and wrapped in aluminum foil and autoclaved separately. On each mesh, four germinated seedlings were placed close to the hole to facilitate penetration of radicle through the mesh and lowered gently on the wedge support placed in the Magenta box. Nutrient medium was then added (about 200 ml) to the hydroponic system to ensure that its level remained 2–3 mm above the X-shaped wedge support. Nutrient media (P+ and P-) were prepared as described (Jia et al., 2011) and buffered to pH 5.7 with 0.5 mM 2-(*N*-morpholino) ethanesulfonic acid (MES). P+ and P- represented 0.3 mM NaH<sub>2</sub>PO<sub>4</sub> and 0 mM NaH<sub>2</sub>PO<sub>4</sub>, respectively. Hydroponic system was placed under controlled growth condition in the greenhouse (16-h day/8-h night cycle, 28 ± 2°C and relative humidity was maintained at ~60–70%).

<sup>1</sup><http://rsb.info.nih.gov/ij>

## Quantification of Total Shoot Area and RSA Parameters

Seedlings grown under P+ and P- conditions in the hydroponic system were removed along with the mesh sequentially 2, 4, and 7 days after treatment. Seedlings were transferred in an inverted position to a Petri plate containing pool of water. Under stereomicroscope, root and shoot were separated at shoot:hypocotyl junction. Carefully, primary, seminal and adventitious roots along with their lateral roots were separated from each other. Dissected roots were then transferred immediately to 70% (v/v) ethanol to avoid any desiccation and for subsequent documentation. Shoots were gently spread and pasted with glue stick on white sheet of paper. For revealing RSA, dissected roots were transferred from 70% (v/v) ethanol to a Petri plate containing 1% (w/v) agar. Lateral roots on primary, seminal and adventitious roots were spread gently with a camel hair brush under stereomicroscope ensuring no overlap. Glued shoots on paper and spread out roots on agar plates were scanned at 600 dpi using a desktop scanner. Scanned images were used for documenting 15 different RSA parameters (**Figure 1**) and total shoot area by using ImageJ program.

## Soluble Pi Content

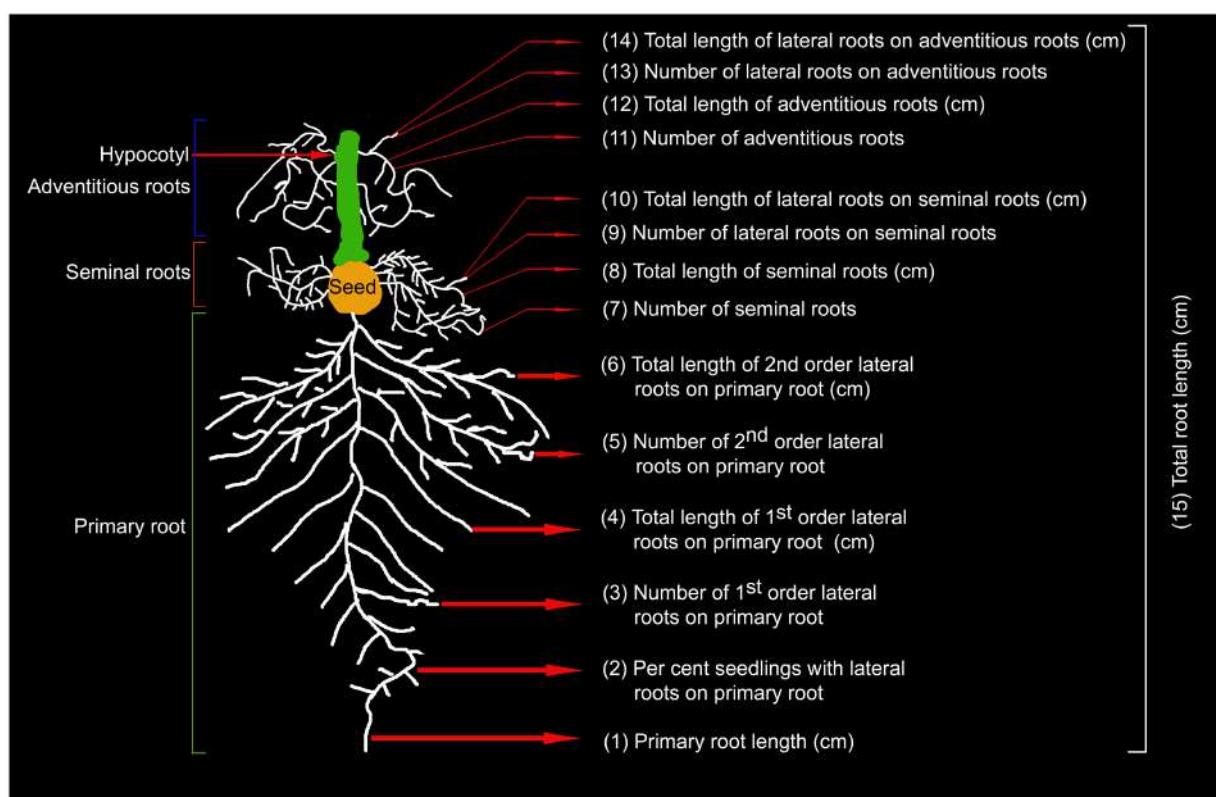
Harvested roots were rinsed 5–6 times in deionized water, blot-dried gently, frozen in liquid nitrogen and ground to a

fine powder and stored at  $-80^{\circ}\text{C}$  till further use. Ground tissue (25–50 mg) was homogenized with 250  $\mu\text{l}$  of 1% (v/v) glacial acetic acid, vortexed and centrifuged at 10,000 rpm for 5 min. Supernatant was collected for assaying Pi content by phosphomolybdate colorimetric assay as described (Ames, 1966). A standard curve generated with  $\text{KH}_2\text{PO}_4$  was used for determining the concentration of soluble Pi.

## Real-Time PCR

The root samples collected from two independent biological experiments were pooled for isolating total RNA by using Spectrum<sup>TM</sup> Plant Total RNA kit as described (Sigma, USA). DNase treatment was given for removing trace amount of DNA. RNA was quantified by NanoDrop 1000 Spectrophotometer (Thermo Scientific, USA) and its quality was assessed on 1.2% (w/v) denatured agarose gel. First-strand cDNA was synthesized from the total RNA (1  $\mu\text{g}$ ) using SuperScript<sup>®</sup> III first-strand synthesis system (Invitrogen, USA). Real-time PCR was performed on Stratagene MX 3005P (Agilent Technologies, USA) using SYBR GreenER<sup>™</sup> qPCR Universal SuperMix (Invitrogen, USA). Gene-specific primers were designed using PrimerQuest software<sup>2</sup>. *OsRubQ1* was used as an internal control. Amplicons were subjected to meltcurve analysis for checking the specificity

<sup>2</sup><https://www.idtdna.com>



**FIGURE 1 | Schematic overview of rice RSA.** Temporal effects of Pi deficiency was quantified on developmental responses of 15 roots traits comprising primary, seminal and adventitious roots and lateral roots on each of them.

of amplified products. Relative expression levels of the genes were computed by  $2^{-\Delta\Delta C_T}$  method of relative quantification (Livak and Schmittgen, 2001). List of primers used for real-time PCR is given in Supplementary Table S1.

## Statistics

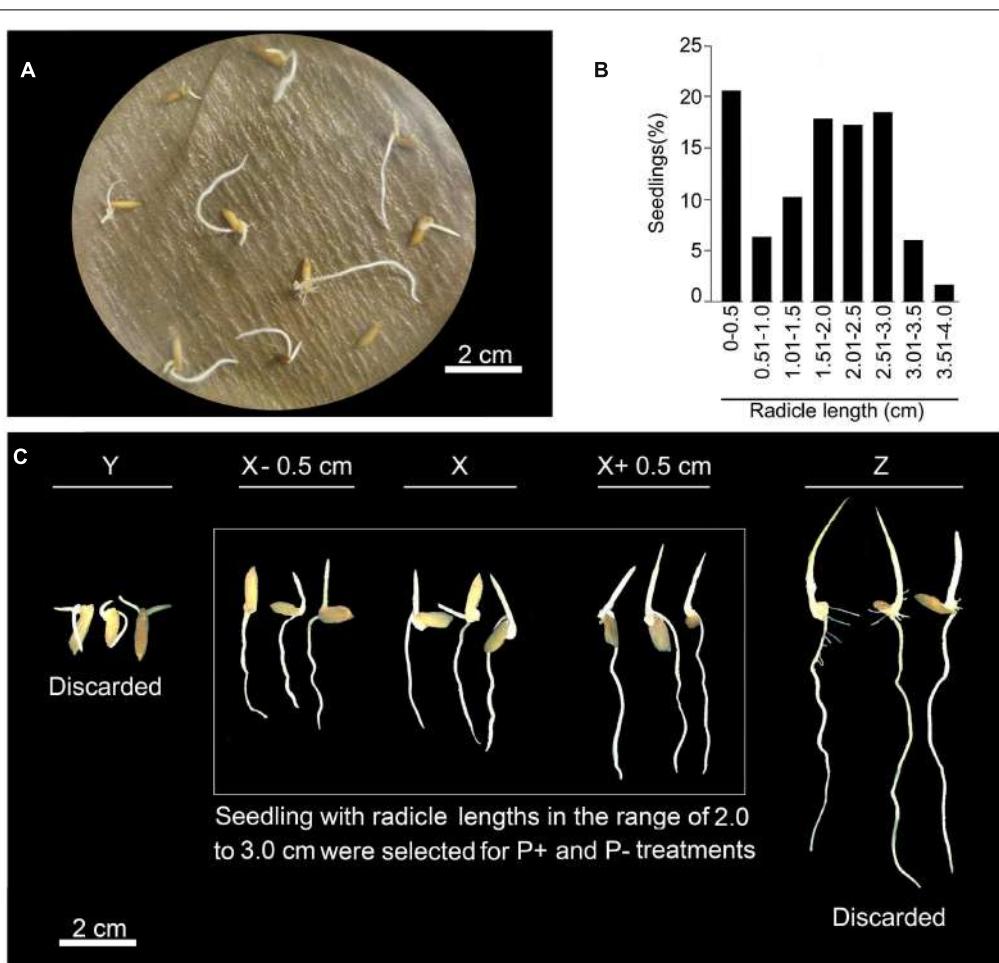
For each experiment, data were collected from 2 to 3 independent biological experiments. Statistical significance of differences between mean values was determined using Student's *t*-test. Different letters on histograms were used for indicating means that were statistically different at  $P < 0.05$ .

## RESULTS AND DISCUSSION

### Selection of Seedlings Prior to Treatment under Different Pi Regime

Radicle length of germinated rice seedlings varies significantly across different genotypes. For Pi deficiency treatment, uniformly

grown seedlings are normally selected based on eyeballing, which could often lead to an erroneous selection. Therefore, to minimize the effects of intrinsic variability on radicle length during subsequent Pi deficiency treatment, a more pragmatic approach was adopted. Around 200 seeds were distributed uniformly in Petri plates (10 seeds/Petri plate) lined with wet filter paper and kept for germination at 28°C for 4 days (Figure 2A). Germinated seedling was transferred to 1% (w/v) agar plate and scanned. Scanned image was used for measuring radicle length using ImageJ program. Based on radicle length, seedlings were grouped into different size ranges of 0.5 cm each and computed per cent seedlings falling in each of these groups (Figure 2B). Radicles of several seedlings (~20%) exhibited stunted growth with their lengths falling in the range of 0–0.5 cm. Per cent seedlings with radicle length in other size categories varied from ~2 to 18%. It was interesting to note that ~5% seedlings revealed an exaggerated radicle growth (~3–4 cm). It was evident from this analysis that extensive variation in radicle length of rice seedlings could exert significant erroneous influence



**FIGURE 2 | Elimination of intrinsic variability in radicle length. (A)** Seeds of rice genotype MI48 were germinated in a Petri plate lined with wet filter paper at 28°C for 4 days in dark. **(B)** Radicle lengths of germinated seedlings were measured by ImageJ program and categorized into different size ranges of 0.5 cm each. Histogram represents per cent seedlings in different size ranges. **(C)** About 30–40% seedlings falling in the size range of 2.0–3.0 cm were selected and transferred to hydroponic set up for temporal treatment under P+ and P- conditions.

on Pi deficiency-mediated effects on different root traits. To circumvent this problem, only those seedlings were selected whose radicle length was in the range of 2–3 cm (**Figure 2C**). Rest of the seedlings outside this size range was discarded. Although, several studies have reported the effects of Pi deprivation on different root traits (Shimizu et al., 2004; Zhou et al., 2008; Dai et al., 2012; Ogawa et al., 2014; Yang et al., 2014), it is not evident from any of these studies as how the likely erroneous influence of intrinsic variability in radicle length of rice seedlings prior to P+ and P− treatments was addressed.

## Modified Hydroponic System

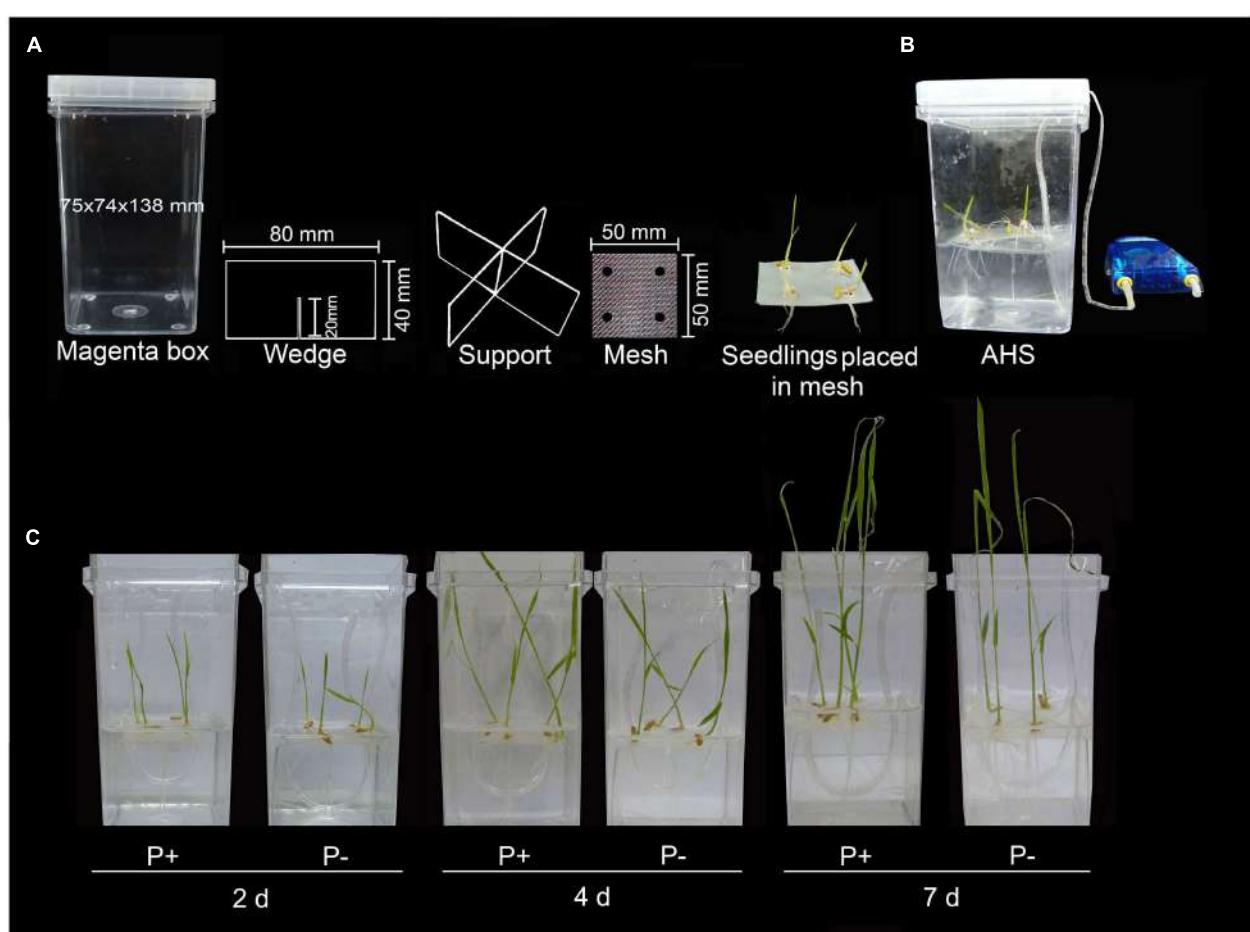
Conventionally rice is grown in a hydroponic system maintained under green house condition for deciphering Pi deficiency-mediated effects on the developmental responses of different root traits (Zhou et al., 2008; Dai et al., 2012; Yang et al., 2014). However, nutrient-rich medium of hydroponic system is often susceptible to elemental contamination, which often results in erroneous interpretations on the effect of Pi deficiency on various morphophysiological and molecular traits (Jain et al., 2009).

In addition, growth of algal bloom, fungi, and bacteria in the medium aggravates the problem.

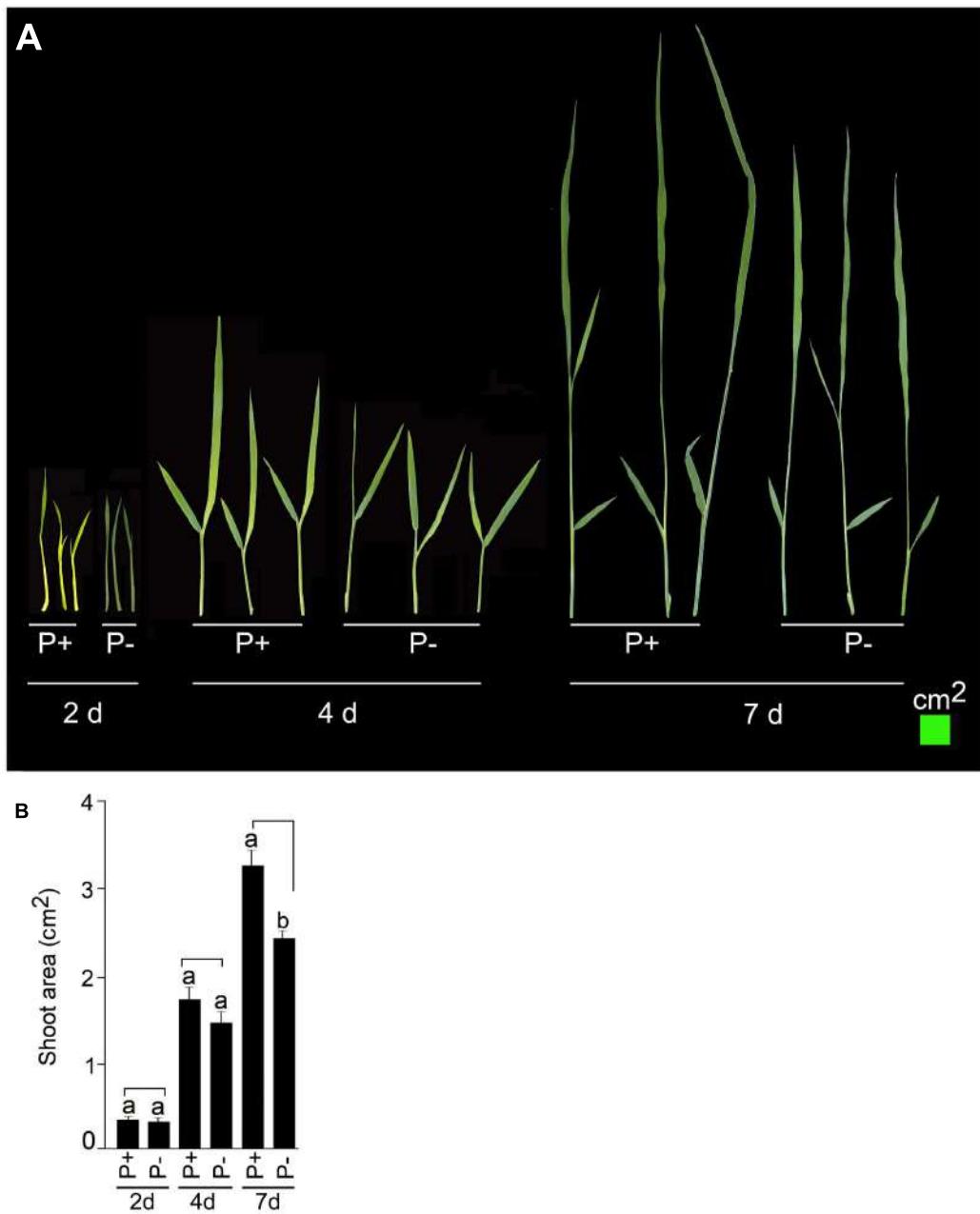
To circumvent these multitude of problems, hydroponic system was modified for growing rice under P+ and P− conditions by assembling easily available autoclavable components (Magenta box, polycarbonate X-shaped wedge support and polypropylene mesh; **Figure 3A**). Further, hydroponic set-up was aerated using aquarium air pump for proper oxygenation and nutrient circulation (**Figure 3B**). Non-aerated hydroponic system could limit oxygen availability to plant roots, which could trigger ethylene production and may exert adverse affects on root growth (Barrett-Lennard and Dracup, 1988). Modified hydroponic system was used for studying the temporal (2, 4, and 7 days) effects of Pi deficiency on morphophysiological and molecular responses (**Figure 3C**).

## Pi Deficiency-Mediated Affects on Phenotypic Traits

Temporal effects of Pi deficiency was determined on shoot phenotype and its total area (**Figure 4**). Pi deprivation for 2



**FIGURE 3 | Modified hydroponic system.** **(A)** Modified hydroponic system made of autoclavable Magenta box, polycarbonate wedge support, polypropylene mesh and germinated rice seedlings placed on the mesh with radicle traversing through the hole punched around its perimeter. **(B)** Complete aerated hydroponic system (AHS). **(C)** Seedlings were grown in AHS under P+ (0.3 mM NaH<sub>2</sub>PO<sub>4</sub>) and P− (0 mM NaH<sub>2</sub>PO<sub>4</sub>) conditions for 2, 4, and 7 days.

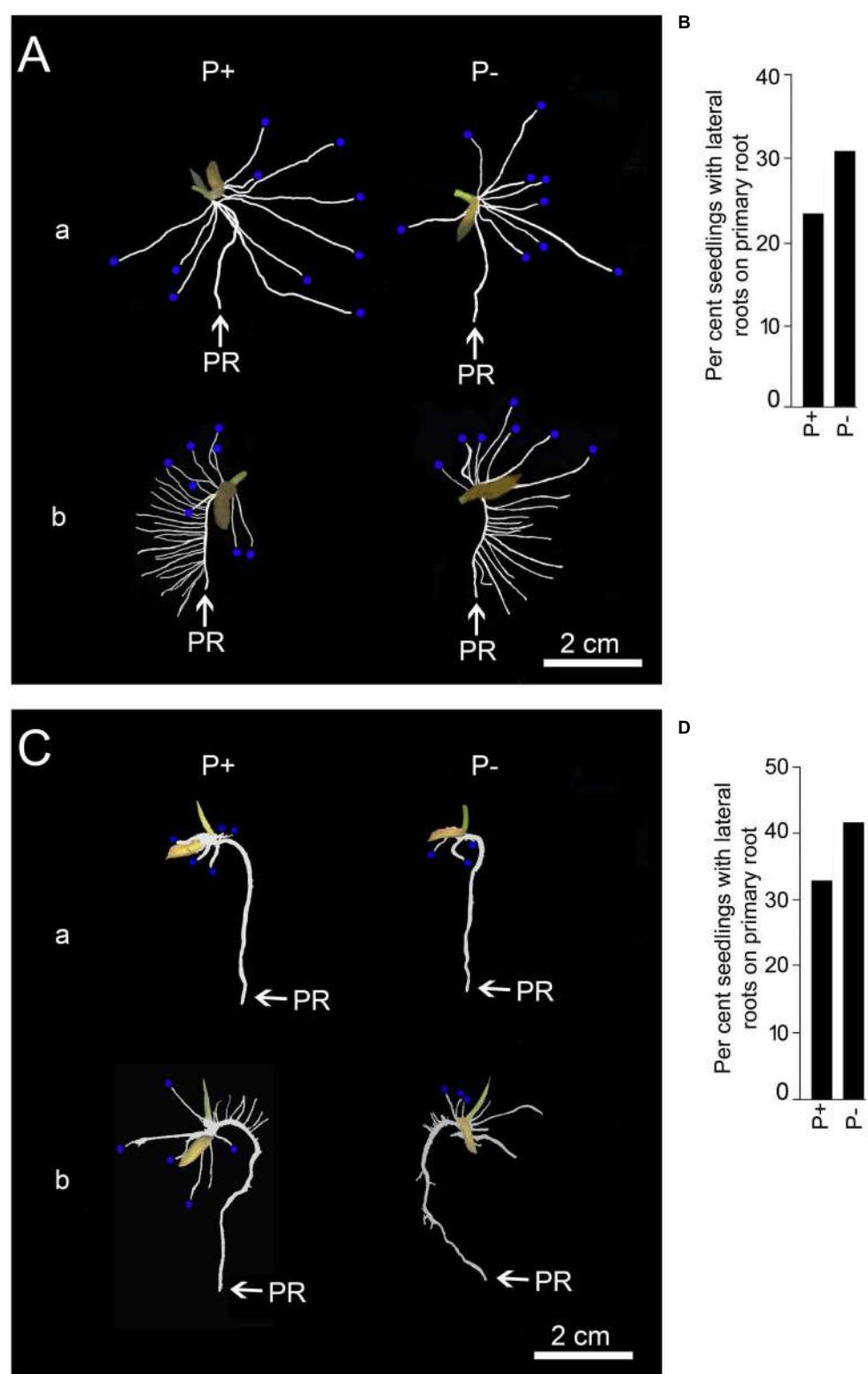


**FIGURE 4 | Temporal effect of Pi deficiency on shoot area.** MI48 seedlings (4-days-old) were grown under P+ and P- conditions for 2, 4, and 7 days. **(A)** Harvested shoots were gently spread and scanned to reveal the details of leaves. **(B)** Data presented for total shoot area. Values are mean  $\pm$  SE and different letters indicate that the means differ significantly ( $P < 0.05$ ).

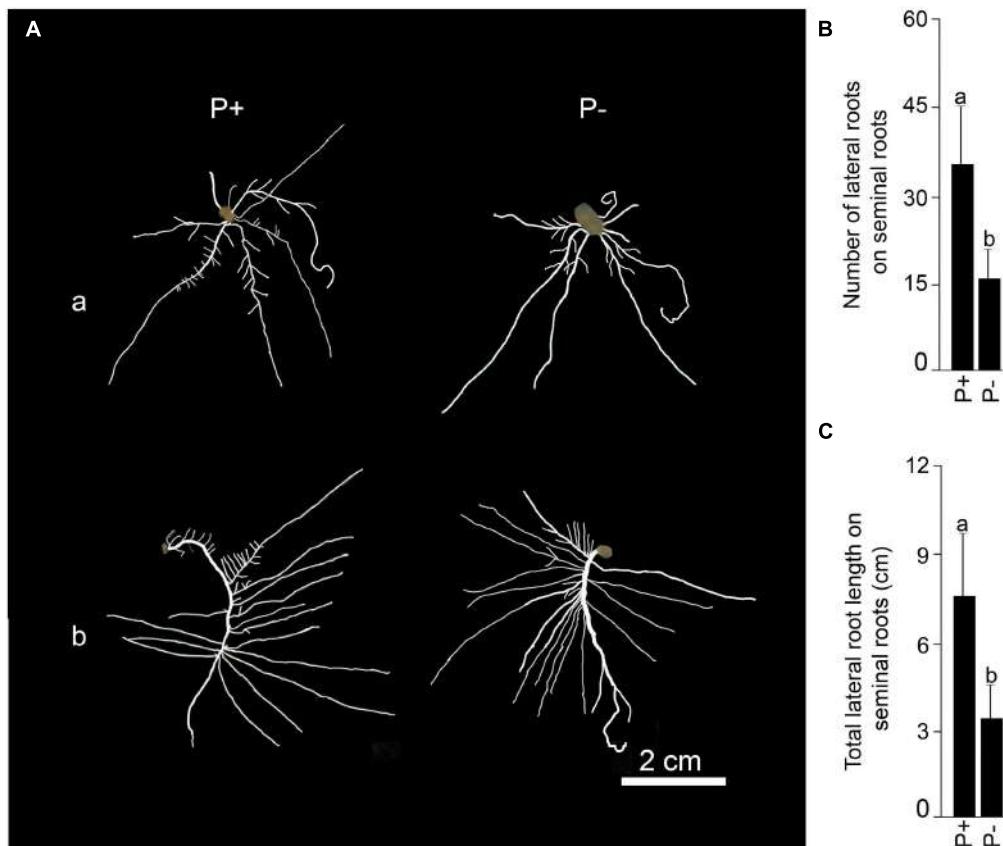
and 4 days did not exert any significant ( $P < 0.05$ ) influence on shoot phenotype (Figure 4A) and its total area (Figure 4B). The effect of Pi deficiency became evident on shoot growth only after 7 days treatment. Growth of P+ shoot was more vigorous compared with P- shoot (Figure 4A) and also area of P- shoot was  $\sim 25\%$  lower compared with P+ (Figure 4B). The result was consistent with an earlier study, which also showed attenuating effect of Pi deprivation on shoot length in rice (Yang et al., 2014). This suggested the suitability of modified hydroponic

system for generating shoot tissues for Pi deficiency-mediated responses.

Temporal effects of Pi deprivation was investigated on the responses of embryonically (primary and seminal) and post-embryonically (adventitious and lateral) developed roots. Two distinct root phenotypes were observed for both P+ and P- seedlings grown for 2 days (Figure 5). Although, majority of the primary roots of P+ and P- seedlings did not show lateral root growth (Figure 5Aa), 25–30% of them revealed



**FIGURE 5 | Effects of Pi deficiency for 2 days on RSA under different growth conditions.** MI48 seedlings (4-days-old) were grown under P+ and P- conditions for 2 days in a modified hydroponic system as shown in **Figure 3C** and in Petri plates lined with blotting paper. **(A,C)** Harvested roots were spread to reveal the details of RSA showing (a) lack and (b) exuberant growth of lateral roots on primary root of P+ and P- seedlings during growth on modified hydroponic system **(A)** and in Petri plates lined with blotting paper **(C)**. Seminal roots **(A,C)** are indicated with blue dots at their tips. **(B,D)** Data presented for per cent seedlings showing lateral root development on primary root during growth on modified hydroponic system **(B)** and in Petri plates lined with blotting paper **(D)**.

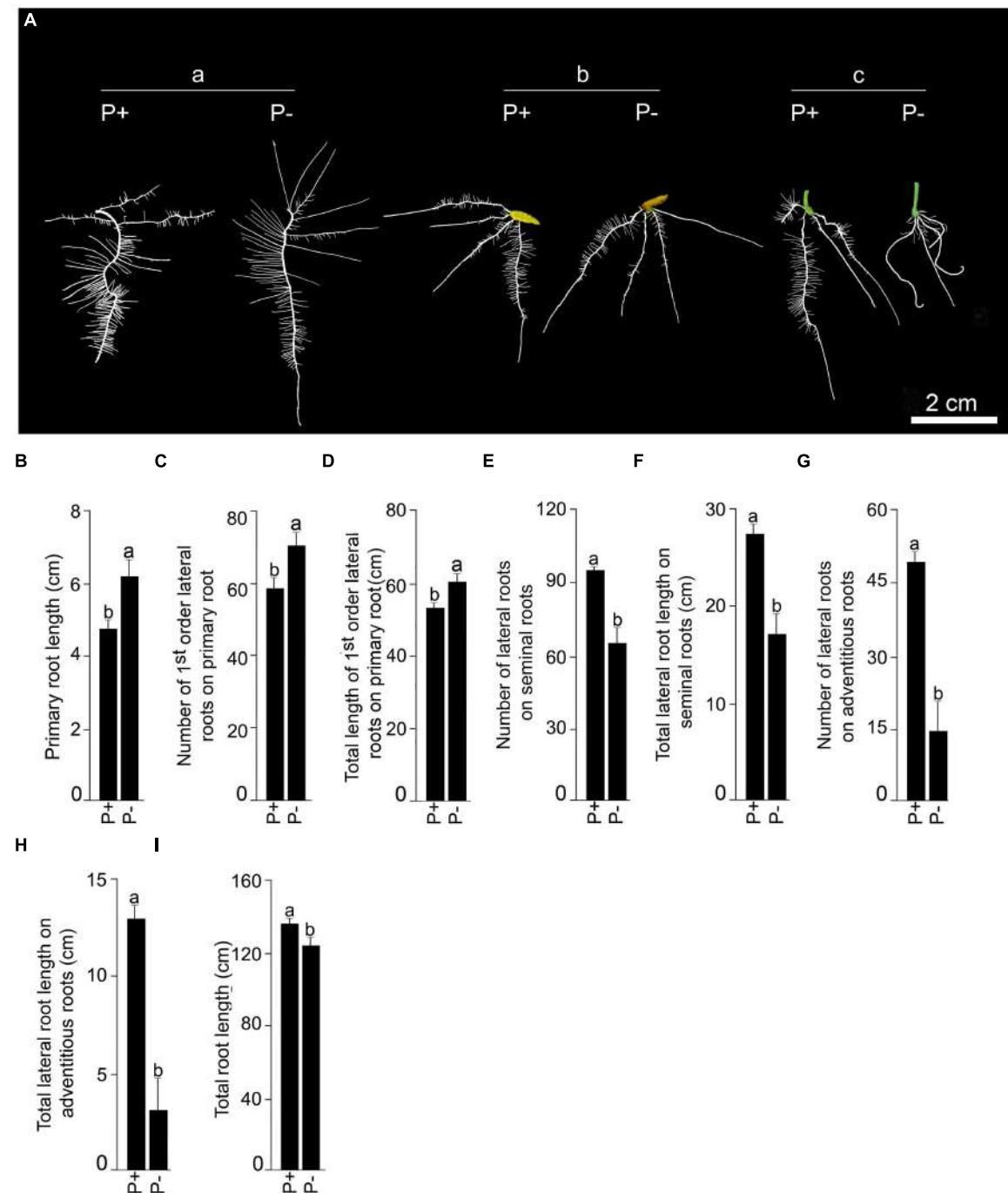


**FIGURE 6 | Effects of Pi deficiency for 4 days on RSA.** MI48 seedlings (4-days-old) were grown under P+ and P- conditions for 4 days. **(A)** Harvested roots were spread to reveal the details of RSA showing (a) seminal and (b) primary roots with well developed lateral roots of both P+ and P- seedlings. Data are presented for **(B)** number and **(C)** total length of lateral roots on seminal roots. Values **(B,C)** are mean  $\pm$  SE and  $n = 12$  replicates. Different letters on the histogram indicate that the means differ significantly ( $P < 0.05$ ).

their development (Figure 5Ab). However, there were significant variations in both the number (P+,  $2.58 \pm 1.71$  [SE]; P-,  $4.01 \pm 2.09$  [SE]) and total length (P+,  $2.98 \text{ cm} \pm 1.85 \text{ cm}$  [SE]; P-,  $4.09 \text{ cm} \pm 2.07 \text{ cm}$  [SE]) of these lateral roots. This highlighted the prevalence of extensive variability in the developmental responses of lateral roots on primary root irrespective of Pi regime. There was no significant ( $P < 0.05$ ) difference in primary root length of P+ and P- seedlings and was comparable to the radicle length before the treatment (Figure 2C), which indicated no significant ( $P < 0.05$ ) increment in this root trait during 2 days treatment. There was thus an apparent lack of correlation between growth responses of primary root and occasional post-embryonically developed lateral roots on them in P+ and P- seedlings. Number (P+,  $7.01 \pm 0.81$  [SE]; P-,  $6.75 \pm 0.48$  [SE]) and length (P+,  $10.57 \text{ cm} \pm 1.85 \text{ cm}$  [SE]; P-,  $10.95 \text{ cm} \pm 1.14 \text{ cm}$  [SE]) of seminal roots varied but the effect of Pi deficiency was not apparent. Neither lateral roots on seminal roots nor adventitious roots in P+ and P- seedlings could be detected after 2 days treatment.

To ensure that the responses of the root system under P+ and P- conditions in the modified hydroponic system was not

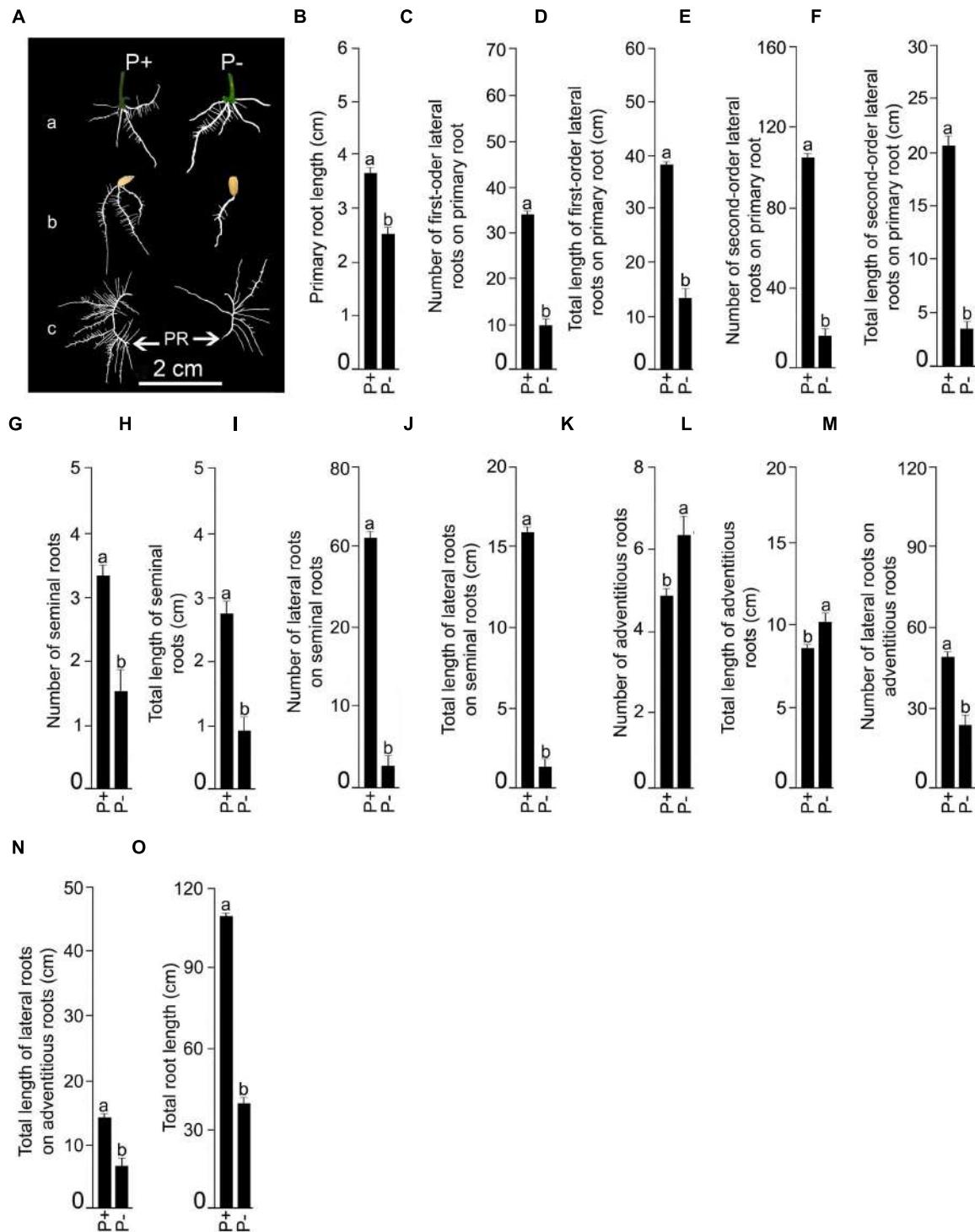
an artifact, MI48 seedlings with radicle length in the range of 2–3 cm (Figure 2C) were also grown on square Petri plate (115 mm  $\times$  115 mm) lined with blotting paper kept moist with these nutrient solutions for 2 days. Figure 5C presents the RSA of these seedlings. Differences in primary root length of P+ ( $2.14 \text{ cm} \pm 0.06 \text{ cm}$  [SE]) and P- ( $2.21 \text{ cm} \pm 0.06 \text{ cm}$  [SE]) seedlings were insignificant ( $P < 0.05$ ) and also the corresponding values were comparable with those grown in the modified hydroponic system. About 33% of both P+ and P- seedlings developed lateral roots (Figure 5D). Sporadic development of lateral roots on P+ and P- primary roots on blotting paper was similar to that observed in P+ and P- seedlings grown in the modified hydroponic system (Figure 5B). These lateral roots exhibited significant variations in both the number (P+,  $3.08 \pm 1.43$  [SE]; P-,  $4.75 \pm 1.81$  [SE]) and total length (P+,  $0.48 \text{ cm} \pm 0.27 \text{ cm}$  [SE]; P-,  $0.56 \text{ cm} \pm 0.28 \text{ cm}$  [SE]); a feature also observed with the seedlings grown in the modified hydroponic system. Overall, these root traits (primary root length, number and length of lateral roots) of MI48 showed comparable responses irrespective of Pi regime or growth conditions (hydroponics and blotting paper). This suggested that



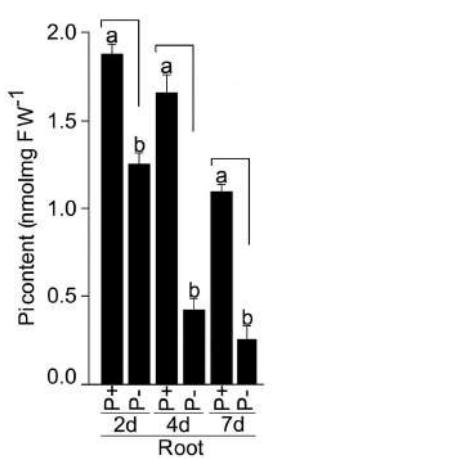
**FIGURE 7 | Effects of Pi deficiency for 7 days on MI48 RSA.** MI48 seedlings (4-days-old) were grown under P+ and P- conditions for 7 days. **(A)** Harvested roots were spread to reveal the details of RSA of P+ and P- seedlings showing (a) primary, (b) seminal and (c) adventitious roots along with their lateral roots. Data are presented for primary root length **(B)**, number **(C)**, and total length **(D)** of 1st order lateral roots on primary root, number **(E)**, and total length **(F)** of lateral roots on seminal roots, number **(G)** and total length **(H)** of lateral roots on adventitious roots and total root length **(I)**. Values **(B–I)** are mean  $\pm$  SE and  $n = 12$  replicates. Different letters on the histogram indicate that the means differ significantly ( $P < 0.05$ ).

developmental responses of these root traits under P+ and P- conditions are not artifacts of the modified hydroponic system. Pi deficiency also did not exert any significant ( $P < 0.05$ ) influence on the number (P+,  $4.25 \pm 0.65$  [SE]; P-,  $4.17 \pm 0.49$

[SE]) and total length (P+,  $1.33 \text{ cm} \pm 0.29 \text{ cm}$  [SE]; P-,  $1.19 \text{ cm} \pm 0.19 \text{ cm}$  [SE]) of seminal roots during growth on the blotting paper. Although, the developmental responses of seminal roots were not influenced by the Pi regime during



**FIGURE 8 | Effects of Pi deficiency for 7 days on N22 RSA.** N22 seedlings (4-days-old) were grown under P+ and P- conditions for 7 days. **(A)** Harvested roots were spread to reveal the details of RSA of P+ and P- seedlings showing (a) adventitious roots, (b) seminal and (c) primary roots along with their lateral roots. Data are presented for primary root length **(B)**, number **(C)**, and total length **(D)** of 1st order lateral roots on primary root, number **(E)** and total length **(F)** of 2nd order lateral roots on primary root, number **(G)** and total length **(H)** of seminal roots, number **(I)** and total length **(J)** of lateral roots on seminal roots, number **(K)** and total length **(L)** of adventitious roots, number **(M)** and total length **(N)** of lateral roots on adventitious roots, and total root length **(O)**. Values **(B–O)** are mean  $\pm$  SE and  $n = 12$  replicates. Different letters on the histogram indicate that the means differ significantly ( $P < 0.05$ ).



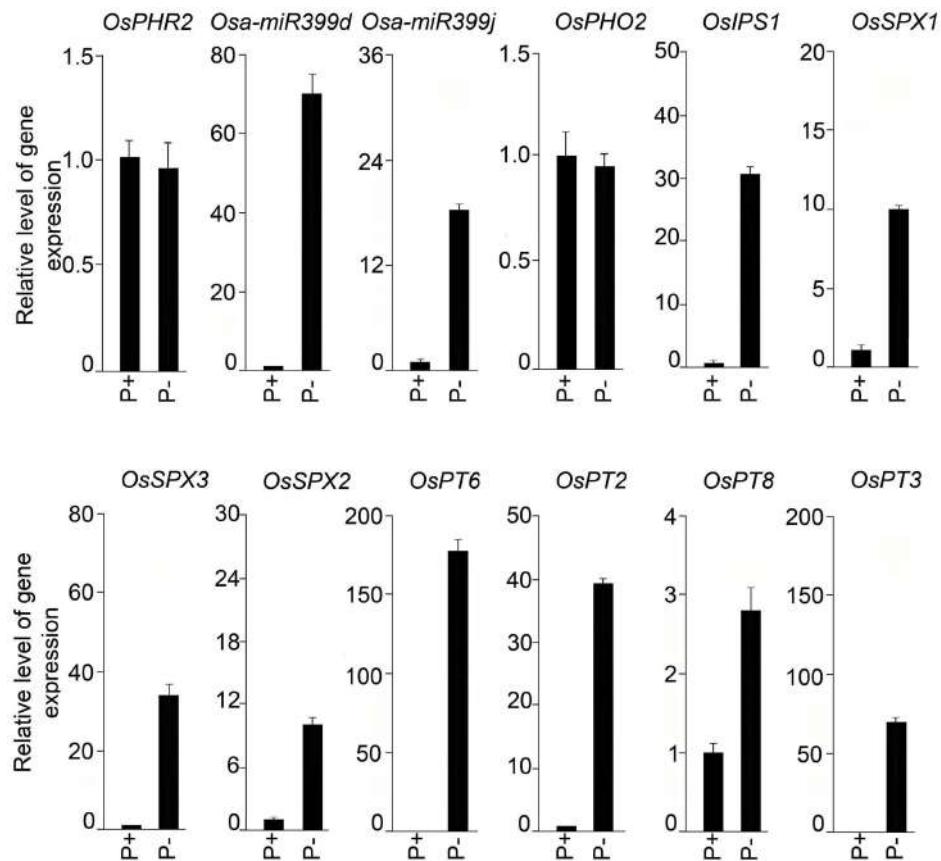
**FIGURE 9 | Effects of Pi deficiency on soluble Pi content.** Mi48 seedlings (4-days-old) were grown under P+ and P- conditions for 2, 4, and 7 days. Roots were harvested for determining soluble Pi content. Values are mean  $\pm$  SE and  $n = 12$  replicates. Different letters on the histogram indicate that the means differ significantly ( $P < 0.05$ ).

growth in the modified hydroponic system and on the blotting paper, the corresponding values (number and total length) were significantly ( $P < 0.05$ ) higher in the former. This suggested better growth and development of root system of rice under P+ and P- conditions in the modified hydroponic system compared with the blotting paper. A lower total root length of P+ ( $3.94 \text{ cm} \pm 0.52 \text{ cm} [\text{SE}]$ ) and P- ( $3.84 \text{ cm} \pm 0.41 \text{ cm} [\text{SE}]$ ) seedlings during growth on the blotting paper compared with modified hydroponic system further provided evidence toward the efficacy of the latter. Therefore, the modified hydroponic system was employed for subsequent analysis of the effects of Pi deprivation for 4 and 7 days on different root traits.

Seedlings were grown under P+ and P- conditions for 4 days and their RSA are presented in Figure 6A. Although, there was no significant ( $P < 0.05$ ) difference in primary root length of P+ ( $5.22 \text{ cm} \pm 0.84 \text{ cm} [\text{SE}]$ ) and P- ( $4.57 \text{ cm} \pm 1.39 \text{ cm} [\text{SE}]$ ) seedlings, it was significantly ( $P < 0.05$ ) higher compared with corresponding 2 days seedlings. This revealed a progressive increment in primary root length of both P+ and P- seedlings over a period of time. There was a significant ( $P < 0.05$ ) increase in the percentage of P+ and P- primary roots with well developed lateral roots from  $\sim 25$  (2 days) to  $\sim 75$  (4 days) suggesting a temporal delay in the development of lateral roots irrespective of Pi regime. However, there were substantial variations in both the number (P+,  $28.92 \pm 11.35$  [SE]; P-,  $19.58 \pm 6.17$  [SE]) and length (P+,  $23.23 \text{ cm} \pm 8.91 \text{ cm} [\text{SE}]$ ; P-,  $23.06 \text{ cm} \pm 7.47 \text{ cm} [\text{SE}]$ ) of 1st-order lateral roots of P+ and P- seedlings. A similar trend was also observed for the number (P+,  $8.42 \pm 4.16$  [SE]; P-,  $3.67 \pm 1.89$  [SE]) and length (P+,  $1.37 \text{ cm} \pm 0.65 \text{ cm} [\text{SE}]$ ; P-,  $0.85 \text{ cm} \pm 0.44 \text{ cm} [\text{SE}]$ ) of 2nd-order lateral roots of these seedlings. Therefore, it was not surprising to see a lack of significant ( $P < 0.05$ ) differences in any of these lateral root traits under different Pi regime. The effect of Pi deprivation

was also not evident on the number (P+,  $8 \pm 0.57$  [SE]; P-,  $8 \pm 0.61$  [SE]) and total length (P+,  $20.29 \text{ cm} \pm 2.99 \text{ cm} [\text{SE}]$ ; P-,  $18.33 \text{ cm} \pm 1.88 \text{ cm} [\text{SE}]$ ) of seminal roots. Although, both primary and seminal roots are embryonic in origin, the latter contributed significantly toward the total root length under both P+ and P- conditions. There is a positive correlation between seminal root length and the ability of rice genotypes to produce deep roots and high yield (Rahman and Musa, 2009). Lateral root development on the seminal roots became apparent only 4 days after treatment and both the number and total length were significantly ( $P < 0.05$ ) higher in P+ seedlings compared with P- seedlings (Figures 6B,C). Adventitious roots were not detected in these seedlings. Despite some effects of Pi deprivation on the developmental responses of lateral roots on seminal roots, differences in the total root length of P+ ( $58.01 \text{ cm} \pm 6.55 \text{ cm} [\text{SE}]$ ) and P- ( $48.81 \text{ cm} \pm 6.21 \text{ cm} [\text{SE}]$ ) seedlings were insignificant ( $P < 0.05$ ).

Finally, the effects of Pi deficiency for 7 days on different RSA traits were determined. Details of P+ and P- RSA are presented in Figure 7A. Pi deficiency triggered significant ( $P < 0.05$ ) increase ( $\sim 20\%$ ) in primary root length compared with P+ seedling (Figure 7B). The result was consistent with earlier studies reporting Pi deficiency-mediated accentuated growth response of primary root in *O. sativa* ssp. *indica* genotype Kasalath (Shimizu et al., 2004) and *O. sativa* ssp. *japonica* genotypes Zhonghua10 (Dai et al., 2012, 2016) and Nipponbare (Zhou et al., 2008; Torabi et al., 2009; Zheng et al., 2009; Hu et al., 2011). Interestingly, a similar trend was also observed in NIL6-4 derived from Pi deficiency-intolerant Nipponbare  $\times$  Pi deficiency-tolerant Kasalath (Torabi et al., 2009). Primary root of P+ and P- seedlings exhibited exuberant growth of lateral roots. There were significant ( $P < 0.05$ ) increases in both the number and total length of 1st-order lateral roots of P- seedlings compared with P+ seedlings (Figures 7C,D). The result was consistent with earlier studies on *japonica* genotypes Dongjin and Nipponbare exhibiting augmented number and/or length of lateral roots on primary root of Pi-deprived seedling (Wang S. et al., 2014; Yang et al., 2014). Some of the older 1st-order lateral roots of P+ and P- seedlings developed 2nd-order lateral roots but differences in their number (P+,  $52.5 \pm 5.99$  [SE]; P-,  $44.58 \pm 4.18$  [SE]) and total length (P+,  $11.95 \text{ cm} \pm 1.48 \text{ cm} [\text{SE}]$ ; P-,  $11.43 \text{ cm} \pm 3.57 \text{ cm} [\text{SE}]$ ) were insignificant ( $P < 0.05$ ). Differences in the number (P+,  $3.75 \pm 0.22$  [SE]; P-,  $4.17 \pm 0.52$  [SE]) and total length (P+,  $13.98 \text{ cm} \pm 0.48 \text{ cm} [\text{SE}]$ ; P-,  $16.39 \text{ cm} \pm 1.92 \text{ cm} [\text{SE}]$ ) of seminal roots were insignificant ( $P < 0.05$ ) under P+ and P-conditions. However, attenuating effects of Pi deprivation were evident on both the number and total length of lateral roots on seminal roots (Figures 7E,F). An earlier study has also shown the attenuating effects of Pi deficiency on seminal root length of *O. rufipogon* (wild rice species) and Curinga (*tropical japonica*; Ogawa et al., 2014). There was also development of adventitious roots from hypocotyls of P+ and P- seedlings. Their number (P+,  $4.25 \pm 0.22$  [SE]; P-,  $4.08 \pm 0.49$  [SE]) and length (P+,  $11.48 \text{ cm} \pm 1.31 \text{ cm} [\text{SE}]$ ; P-,  $8.73 \text{ cm} \pm 0.87 \text{ cm} [\text{SE}]$ ) were not significantly ( $P < 0.05$ ) influenced by Pi status of the nutrient medium. Wang S. et al. (2014) also



**FIGURE 10 | Effects of Pi deficiency-mediated responses of genes involved in Pi homeostasis.** Real-time PCR analysis of relative expression levels in roots of MI48 seedlings grown under P+ and P- conditions for 7 days. *OsRubQ1* was used as an internal control. Data presented are means of six technical replicates  $\pm$  SE.

did not observe any significant effect of Pi deficiency on the number of adventitious root of genotypes Dongjin and Nipponbare. On the contrary, in other studies Pi deficiency was found to exert either inhibitory (genotype Zhonghua10, Dai et al., 2016) or stimulatory (genotype Zhonghua10, Dai et al., 2012; genotype Dongjin, Wang et al., 2015) effects on the developmental responses of adventitious root. This clearly suggested the influence of the genotype on Pi deficiency-mediated effects on the number and/or length of adventitious root. Inhibitory effects of Pi deprivation were evident on both the number and total length of lateral roots on adventitious roots compared with P+ seedlings (Figures 7G,H). Overall, total root length of P- seedling was significantly ( $P < 0.05$ ) lower compared with P+ seedling (Figure 7I). Further, effects of Pi deficiency for 7 days on different root traits were evaluated in Nagina22 (N22) (Figure 8). Details of P+ and P- RSA are presented in Figure 8A. Pi deficiency exerted attenuating effects on the primary root length (Figure 8B), which was consistent with an earlier study (Panigrahy et al., 2014). Interestingly though, the response was contrary to the stimulatory effect on this root trait in MI48. Shimizu et al. (2004) also reported stimulation and no effect on the primary

root growth during Pi deprivation in Kasalath (indica) and Gimbozu (japonica), respectively. Pi deficiency also exerted inhibitory effects on both the number (Figure 8C) and total length (Figure 8D) of 1st-order lateral roots on primary root of N22, which was contrary to MI48. A similar inhibitory influence of Pi deficiency was also evident on the number (Figure 8E) and total length (Figure 8F) of 2nd-order lateral roots on primary root, and number (Figure 8G) and total length (Figure 8H) of seminal roots of N22. Comparatively, none of these traits were significantly ( $P < 0.05$ ) affected by Pi deficiency in MI48. Although, Pi deficiency did not have significant influence on both the number and total length of adventitious roots in MI48, values for both these traits were significantly ( $P < 0.05$ ) higher in P- seedlings compared with P+ seedlings of N22 (Figures 8K,L). Comparative analysis of the effects of Pi deficiency on different root traits of MI48 and N22 clearly revealed the genotypic differences. In addition, there were attenuating effects of Pi deprivation on the number (Figures 8I,M) and length (Figures 8J,N) of lateral roots on seminal and adventitious roots and total root length (Figure 8O) of N22. A similar effect of Pi deficiency was also observed for these root traits in MI48. The study thus indicated the efficacy of

the modified hydroponic system in delineating both differential and comparable effects of Pi deficiency on different root traits in these genotypes.

## Pi Deficiency-Mediated Molecular Responses

Roots of seedlings grown under P+ and P- conditions for 2, 4, and 7 days were analyzed for soluble Pi content (**Figure 9**). Compared with P+ roots, Pi content in P- roots was significantly ( $P < 0.05$ ) reduced by 33, 73, and 77% after Pi deprivation for 2, 4, and 7 days, respectively. Several genes have been identified in rice that play pivotal roles in the maintenance of Pi homeostasis in rice (Wu et al., 2013). The effects of Pi deficiency on the relative expression levels of these genes in the roots of seedlings grown under P+ and P- conditions for 7 days were assayed by real-time PCR (**Figure 10**). There was no induction of transcription factor OsPHR2 in response to Pi deficiency and was consistent with an earlier study (Zhou et al., 2008). Whereas, there were 69- and 18-fold induction in the relative expression levels of OsmiR399d and OsmiR399j, respectively. An earlier study had also reported Pi deficiency-mediated induction of OsmiR399s; a pivotal component of Pi sensing and signaling pathway downstream of OsPHR2 (Zhou et al., 2008). miRNA399 targets E2 ubiquitin-conjugase OsPHO2, which is expressed constitutively irrespective of the Pi regime (Hu et al., 2011). Consistent with this report, the relative expression levels of OsPHO2 were comparable in P+ and P- roots. Further, Pi deficiency triggered 31-fold induction in the relative expression level of OsIPSI. Hou et al. (2005) also reported rapid induction of OsIPSI in Pi-deprived roots of rice and has been implicated in potentially mimicking OsmiR399 target thereby attenuating its suppressive effect (Wu et al., 2013). Proteins harboring the SPX domain (OsSPX1–6) play key roles in the maintenance of Pi homeostasis (Secco et al., 2012). In P- roots, there were significant increases in the relative expression levels of OsSPX1 (~10-fold), OsSPX2 (~10-fold), and OsSPX3 (~34-fold) compared with P+ roots suggesting their roles in the maintenance of Pi homeostasis and were coherent with an earlier study demonstrating their significant induction during Pi deficiency (Wang et al., 2009). SPX1 and SPX2 act as Pi-dependent inhibitors of OsPHR2 activity (Wang Z. et al., 2014). Pi deficiency also triggered significant increases in the relative expression levels of Pi transporters OsPT2 (~40-fold), OsPT3 (~70-fold), OsPT6 (~170-fold), and OsPT8 (~3-fold) in P- roots compared with P+ roots. OsPT2, a low-affinity Pi transporter, plays a role in mobilizing stored Pi in plants and high-affinity Pi transporter OsPT6 has been implicated in uptake and translocation of Pi throughout the plant (Ai et al., 2009). OsPT8, another high-affinity Pi transporter, is essential for the maintenance of Pi homeostasis and proper growth and

development of plant (Jia et al., 2011). OsPT3 has not yet been functionally characterized.

## CONCLUSION

The modified hydroponic system was amenable for detailed analysis of the temporal effects of Pi deprivation on the developmental responses of primary, seminal and adventitious roots and also of the lateral roots on each of them of rice genotypes MI48 and N22. The data generated on Pi deficiency-mediated effects on different root traits could be employed for mathematical simulation and modeling. The modified hydroponic system also facilitated generation of tissues for physiological and molecular analyses. It is equally conducive for studying the effects of other nutrient deficiencies or cross talk between different nutrients on morphophysiological and molecular responses of rice genotypes. This modified hydroponic system would also facilitate in rapid identification of Pi deficiency-responsive root traits in a large number of genotypes for genome-wide association study (GWAS).

## AUTHOR CONTRIBUTIONS

AJ conceived the project and designed the experiments. MN, RS, and VR contributed toward execution of all the experiments. AJ wrote the manuscript. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

This work was supported by the Ministry of Science and Technology, Department of Biotechnology (DBT) Government of India (Ramalingaswamy Fellowship to AJ [BT/HRD/35/02/26/2009] and grant awarded to AJ [BT/PR8254/AGII/103/876/2013]). We are also thankful to DBT for the financial support to MN as Research Associate in a project sanctioned to AJ. We thank Dr. Viswanathan Satheesh for his useful suggestions throughout the course of this study. We also extend our appreciation to Amit Kumar and Rajesh Kumar for their technical help during course of the experiments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00550>

## REFERENCES

- Ai, P., Sun, S., Zhao, J., Fan, X., Xin, W., Guo, Q., et al. (2009). Two rice phosphate transporters, OsPht1;2 and OsPht1;6, have different functions and kinetic properties in uptake and translocation. *Plant J.* 57, 798–809. doi: 10.1111/j.1365-313X.2008.03726.x

- Alatorre-Cobos, F., Calderón-Vázquez, C., Ibarra-Laclette, E., Yong-Villalobos, L., Pérez-Torres, C.-A., Oropeza-Aburto, A., et al. (2014). An improved, low-cost, hydroponic system for growing Arabidopsis and other plant species under aseptic conditions. *BMC Plant Biol.* 14:69. doi: 10.1186/1471-2229-14-69
- Ames, B. N. (1966). Assay of inorganic phosphate, total phosphate and phosphatases. *Methods Enzymol.* 8, 115–118. doi: 10.1016/0076-6879(66)08014-5

- Barrett-Lennard, E. G., and Dracup, M. (1988). A porous agar medium for improving the growth of plants under sterile conditions. *Plant Soil* 108, 294–298. doi: 10.1007/BF02375663
- Collins, T. J. (2007). ImageJ for microscopy. *Biotechniques* 43, S25–S30. doi: 10.2144/000112517
- Dai, X., Wang, Y., Yang, A., and Zhang, W.-H. (2012). OsMYB2P-1, an R2R3-MYB transcription factor, is involved in the regulation of phosphate-starvation responses and root architecture in rice. *Plant Physiol.* 159, 169–183. doi: 10.1104/pp.112.194217
- Dai, X., Wang, Y., and Zhang, W. H. (2016). OsWRKY74, WRKY transcription factor, modulates tolerance to phosphate starvation in rice. *J. Exp. Bot.* 67, 947–960. doi: 10.1093/jxb/erv515
- Gamuyao, R., Chin, J. H., Pariasca-Tanaka, J., Pesaresi, P., Catausan, S., Dalid, C., et al. (2012). The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. *Nature* 488, 535–539. doi: 10.1038/nature11346
- Gruber, B. D., Giehl, R. F. H., Friedel, S., and von Wirén, N. (2013). Plasticity of the *Arabidopsis* root system under nutrient deficiencies. *Plant Physiol.* 163, 161–179. doi: 10.1104/pp.113.218453
- Hochholdinger, F., Park, W. J., Sauer, M., and Woll, K. (2004). From weeds to crops: genetic analysis of root development in cereals. *Trends Plant Sci.* 9, 42–48. doi: 10.1016/j.tplants.2003.11.003
- Hochholdinger, F., and Zimmermann, R. (2008). Conserved and diverse mechanisms in root development. *Curr. Opin. Plant Biol.* 11, 70–74. doi: 10.1016/j.pbi.2007.10.002
- Hou, X. L., Wu, P., Jiao, F. C., Jia, Q. J., Chen, H. M., Yu, J., et al. (2005). Regulation of the expression of *OsIPS1* and *OsIPS2* in rice via systemic and local Pi signalling and hormones. *Plant Cell Environ.* 28, 353–364. doi: 10.1111/j.1365-3040.2005.01272.x
- Hu, B., Zhu, C., Li, F., Tang, J., Wang, Y., Lin, A., et al. (2011). *LEAF TIP NECROSIS1* plays a pivotal role in the regulation of multiple phosphate starvation responses in rice. *Plant Physiol.* 156, 1101–1115. doi: 10.1104/pp.110.170209
- Jain, A., Poling, M. D., Karthikeyan, A. S., Blakeslee, J. J., Peer, W. A., Titapiwatankun, B., et al. (2007). Differential effects of sucrose and auxin on localized phosphate deficiency-induced modulation of different traits of root system architecture in *Arabidopsis*. *Plant Physiol.* 144, 232–247. doi: 10.1104/pp.106.092130
- Jain, A., Poling, M. D., Smith, A. P., Nagarajan, V. K., Lahner, B., Meagher, R. B., et al. (2009). Variations in the composition of gelling agents affect morphophysiological and molecular responses to deficiencies of phosphate and other nutrients. *Plant Physiol.* 150, 1033–1049. doi: 10.1104/pp.109.136184
- Jia, H., Ren, H., Gu, M., Zhao, J., Sun, S., Zhang, X., et al. (2011). The phosphate transporter gene *OsPht1;8* is involved in phosphate homeostasis in rice. *Plant Physiol.* 156, 1164–1175. doi: 10.1104/pp.111.175240
- Kellermeier, F., Armengaud, P., Seditas, T. J., Danku, J., Salt, D. E., and Amtmann, A. (2014). Analysis of the root system architecture of *Arabidopsis* provides a quantitative readout of crosstalk between nutritional signals. *Plant Cell* 26, 1480–1496. doi: 10.1105/tpc.113.122101
- 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
- López-Arredondo, D. L., Leyva-González, M. A., González-Morales, S. I., López-Bucio, J., and Herrera-Estrella, L. (2014). Phosphate nutrition: improving low-phosphate tolerance in crops. *Annu. Rev. Plant Biol.* 65, 95–123. doi: 10.1146/annurev-arplant-050213-035949
- López-Bucio, J., Hernández-Abreu, E., Sánchez-Calderón, L., Nieto-Jacobo, M. F., Simpson, J., and Herrera-Estrella, L. (2002). Phosphate availability alters architecture and causes changes in hormone sensitivity in the *Arabidopsis* root system. *Plant Physiol.* 129, 244–256. doi: 10.1104/pp.010934
- Lynch, J. P. (2015). Root phenes that reduce the metabolic costs of soil exploration: opportunities for 21st century agriculture. *Plant Cell Environ.* 38, 1775–1784. doi: 10.1111/pce.12451
- Marschner, H. (1995). *Mineral Nutrition of Higher Plants*. London: Academic Press.
- Ogawa, S., Selvaraj, M. G., Fernando, A. J., Lorieux, M., Ishitani, M., McCouch, S., et al. (2014). N- and P-mediated seminal root elongation response in rice seedlings. *Plant Soil* 375, 303–315. doi: 10.1007/s11104-013-1955-y
- Panigrahy, M., Rao, D. N., Yugandhar, P., Raju, N. S., Krishnamurthy, P., Voleti, S. R., et al. (2014). Hydroponic experiment for identification of tolerance traits developed by rice Nagina 22 mutants to low-phosphorus in field condition. *Arch. Agron. Soil Sci.* 60, 565–576. doi: 10.1080/03650340.2013.821197
- Raghothama, K. G. (1999). Phosphate acquisition. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 665–693. doi: 10.1146/annurev.aplant.50.1.665
- Rahman, Z. A., and Musa, M. H. (2009). Upland rice root characteristics and their relationship to nitrogen uptake. *Pertanika J. Trop. Agric. Sci.* 32, 261–266.
- Sánchez-Calderón, L., López-Bucio, J., Chacón-López, A., Cruz-Ramírez, A., Nieto-Jacobo, F., Dubrovsky, J. G., et al. (2005). Phosphate starvation induces a determinate developmental program in the roots of *Arabidopsis thaliana*. *Plant Cell Physiol.* 46, 174–184. doi: 10.1093/pcp/pci011
- Secco, D., Wang, C., Arpat, B. A., Wang, Z., Poirier, Y., Tyerman, S. D., et al. (2012). The emerging importance of the SPX domain-containing proteins in phosphate homeostasis. *New Phytol.* 193, 842–851. doi: 10.1111/j.1469-8137.2011.04002.x
- Shimizu, A., Yanagihara, S., Kawasaki, S., and Ikehashi, H. (2004). Phosphorus deficiency-induced root elongation and its QTL in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 109, 1361–1368. doi: 10.1007/s00122-004-1751-4
- Torabi, S., Wissuwa, M., Heidari, M., Naghavi, M. R., Gilany, K., Hajirezaei, M. R., et al. (2009). A comparative proteome approach to decipher the mechanism of rice adaptation to phosphorous deficiency. *Proteomics* 9, 159–170. doi: 10.1002/pmic.200800350
- Wang, H., Sun, R., Cao, Y., Pei, W., Sun, Y., Zhou, H., et al. (2015). OsSIZ1, a SUMO E3 ligase gene, is involved in the regulation of the responses to phosphate and nitrogen in rice. *Plant Cell Physiol.* 56, 2381–2395. doi: 10.1093/pcp/pcv162
- Wang, S., Zhang, S., Sun, C., Xu, Y., Chen, Y., Yu, C., et al. (2014). Auxin response factor (OsARF12), a novel regulator for phosphate homeostasis in rice (*Oryza sativa*). *New Phytol.* 201, 91–103. doi: 10.1111/nph.12499
- Wang, Z., Hu, H., Huang, H., Duan, K., Wu, Z., and Wu, P. (2009). Regulation of OsSPX1 and OsSPX3 on expression of OsSPX domain genes and Pi-starvation signaling in rice. *J. Integr. Plant Biol.* 51, 663–674. doi: 10.1111/j.1744-7909.2009.00834.x
- Wang, Z., Ruan, W., Shi, J., Zhang, L., Xiang, D., Yang, C., et al. (2014). Rice SPX1 and SPX2 inhibit phosphate starvation responses through interacting with PHR2 in a phosphate-dependent manner. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14953–14958. doi: 10.1073/pnas.1404680111
- Williamson, L. C., Ribrioux, S. P. C. P., Fitter, A. H., and Leyser, H. M. O. (2001). Phosphate availability regulates root system architecture in *Arabidopsis*. *Plant Physiol.* 126, 875–882. doi: 10.1104/pp.126.2.875
- Wu, P., Shou, H., Xu, G., and Lian, X. (2013). Improvement of phosphorus efficiency in rice on the basis of understanding phosphate signaling and homeostasis. *Curr. Opin. Plant Biol.* 16, 205–212. doi: 10.1016/j.pbi.2013.03.002
- Yang, W. T., Baek, D., Yun, D.-J., Hwang, W. H., Park, D. S., Nam, M. H., et al. (2014). Overexpression of OsMYB4P, an R2R3-type MYB transcriptional activator, increases phosphate acquisition in rice. *Plant Physiol. Biochem.* 80, 259–267. doi: 10.1016/j.plaphy.2014.02.024
- Yi, K., Wu, Z., Zhou, J., Du, L., Guo, L., Wu, Y., et al. (2005). OsPTF1, a novel transcription factor involved in tolerance to phosphate starvation in rice. *Plant Physiol.* 138, 2087–2096. doi: 10.1104/pp.105.063115
- Zheng, L., Huang, F., Narsai, R., Wu, J., Giraud, E., He, F., et al. (2009). Physiological and transcriptome analysis of iron and phosphorus interaction in rice seedlings. *Plant Physiol.* 151, 262–274. doi: 10.1104/pp.109.141051
- Zhou, J., Jiao, F., Wu, Z., Li, Y., Wang, X., He, X., et al. (2008). OsPHR2 is involved in phosphate-starvation signaling and excessive phosphate accumulation in shoots of plants. *Plant Physiol.* 146, 1673–1686. doi: 10.1104/pp.107.111443

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

Copyright © 2016 Negi, Sanagala, Rai and Jain. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# SeedUSoon: A New Software Program to Improve Seed Stock Management and Plant Line Exchanges between Research Laboratories

Céline Charavay<sup>1,2,3</sup>, Stéphane Segard<sup>1,2,3</sup>, Nathalie Pochon<sup>4,5,6</sup>, Laurent Nussaume<sup>4,5,6</sup> and Hélène Javot<sup>4,5,6\*</sup>

<sup>1</sup> Institut de Biosciences et Biotechnologies de Grenoble-Laboratoire Biologie à Grande Échelle, Université Grenoble Alpes, Grenoble, France, <sup>2</sup> Institut de Biosciences et Biotechnologies de Grenoble-Laboratoire Biologie à Grande Échelle-Groupe Informatique pour les Scientifiques du Sud Est, Commissariat à l'Energie Atomique et aux Énergies Alternatives (CEA), Grenoble, France, <sup>3</sup> Laboratoire Biologie à Grande Échelle, Institut National de la Santé et de la Recherche Médicale (INSERM), Grenoble, France, <sup>4</sup> Laboratoire Biologie Develop Plantes, Institut de Biosciences et Biotechnologies, Commissariat à l'Energie Atomique et aux Énergies Alternatives (CEA), Saint-Paul-lez-Durance, France, <sup>5</sup> Centre National de la Recherche Scientifique (CNRS) , UMR 7265 Biologie Végétale et Microbiologie Environnementales, Saint-Paul-lez-Durance, France, <sup>6</sup> Aix Marseille Université, BVME UMR 7265, Marseille, France

## OPEN ACCESS

### Edited by:

John Doonan,  
Aberystwyth University, UK

### Reviewed by:

Gavin Mager George,  
Stellenbosch University, South Africa  
Yuhui Chen,  
Samuel Roberts Noble Foundation,  
USA

### \*Correspondence:

Hélène Javot  
SeedUSoon@cea.fr;  
helene.javot@cea.fr

### Specialty section:

This article was submitted to  
Technical Advances in Plant Science,  
a section of the journal  
*Frontiers in Plant Science*

Received: 31 August 2016

Accepted: 04 January 2017

Published: 20 January 2017

### Citation:

Charavay C, Segard S, Pochon N, Nussaume L and Javot H (2017) SeedUSoon: A New Software Program to Improve Seed Stock Management and Plant Line Exchanges between Research Laboratories. *Front. Plant Sci.* 8:13. doi: 10.3389/fpls.2017.00013

Plant research is supported by an ever-growing collection of mutant or transgenic lines. In the past, a typical basic research laboratory would focus on only a few plant lines that were carefully isolated from collections of lines containing random mutations. The subsequent technological breakthrough in high-throughput sequencing, combined with novel and highly efficient mutagenesis techniques (including site-directed mutagenesis), has led to a recent exponential growth in plant line collections used by individual researchers. Tracking the generation and genetic properties of these genetic resources is thus becoming increasingly challenging for researchers. Another difficulty for researchers is controlling the use of seeds protected by a Material Transfer Agreement, as often only the original recipient of the seeds is aware of the existence of such documents. This situation can thus lead to difficult legal situations. Simultaneously, various institutions and the general public now demand more information about the use of genetically modified organisms (GMOs). In response, researchers are seeking new database solutions to address the triple challenge of research competition, legal constraints, and institutional/public demands. To help plant biology laboratories organize, describe, store, trace, and distribute their seeds, we have developed the new program SeedUSoon, with simplicity in mind. This software contains data management functions that allow the separate tracking of distinct mutations, even in successive crossings or mutagenesis. SeedUSoon reflects the biotechnological diversity of mutations and transgenes contained in any specific line, and the history of their inheritance. It can facilitate GMO certification procedures by distinguishing mutations on the basis of the presence/absence of a transgene, and by recording the technology used for their generation. Its interface can be customized to match the context and rules of any laboratory. In addition, SeedUSoon includes functions to help the laboratory protect

intellectual property, export data, and facilitate seed exchange between laboratories. The SeedUSoon program, which is customizable to match individual practices and preferences, provides a powerful toolkit to plant laboratories searching for innovative approaches in laboratory management.

**Keywords:** software, database, plant, seed, genetics, genealogy, MTA

## INTRODUCTION

Basic research in plant biology frequently relies on plants whose genomes have been engineered for distinct purposes. For example, biotechnological applications derived from the machinery of the plant pathogen *Agrobacterium* allows the now routine insertion of T-DNA from specific vectors into the plant genome (Hellens et al., 2000). Inserted sequences can allow the expression of a vast array of constructs of interest (such as RNAi or antisense transcripts, GFP-protein fusions, over-expressed genes, biosensors, reporters, and antibiotic or herbicide resistance cassettes). In addition, the insertion of T-DNA into the genome is used to generate libraries of knock-out (KO) mutants. These insertions occur randomly in the plant genome, even though collections of T-DNA insertional mutants will often be enriched for insertions occurring within transcriptionally active parts of the genome (Ortega et al., 2002; Kim and Gelvin, 2007). Similarly, libraries of insertional mutants have been built on the ability of transposons to replicate and insert randomly into the plant genome (Greco et al., 2001; Wegmuller et al., 2008). In addition to collections of T-DNA or transposon insertional mutants, researchers have access to libraries of plant lines containing point mutations or deletions that randomly affect endogenous gene sequences through EMS treatments or irradiations (Li et al., 2002; Kim et al., 2006; Svistoonoff et al., 2007).

Recently, the panel of available mutations was further expanded by engineering site-specific nucleases derived from CRISPR/Cas9, TALEN, ZFN, and meganucleases (Fauser et al., 2014; Baltes and Voytas, 2015). These techniques now make it possible to generate random or precise mutations within specific gene loci in plants, by performing site-directed mutagenesis. Each mutagenesis often results in the generation of a whole set of mutant alleles for a single targeted sequence. Due to their simplicity, some of these applications are becoming routine methods for synthetic biology applications and basic research purposes. These genetic modifications are not only restricted to DNA, as pentatricopeptide repeat proteins also make it possible to alter RNA (Yagi et al., 2014).

The wide availability of efficient and affordable cloning, mutagenesis and transformation techniques has accelerated the generation of transgenic plants. In parallel, the availability of mutant libraries combined with the development of high-throughput sequencing methods has impressively facilitated the precise genotyping of KO mutants. As a consequence, the size of plant and seed collections has dramatically expanded in the past 10 years for many typical research laboratories. These collections contain lines derived from a large diversity of mutagenesis methods, reflecting the ever-growing power

of genetics. Plant lines can also combine several mutations, and today it is common to analyze triple, quadruple or quintuple mutants for different loci that were obtained through combinations of different mutagenesis technologies. A clear understanding of the genetic diversity and biotechnological origin of these seed collections is becoming more and more crucial, as each technique presents different risks of artifacts. For instance, CRISPR/Cas9 mutagenesis presents risks of off-target mutations (Baltes and Voytas, 2015), and EMS-mutagenized collections often contain numerous point mutations in a single plant (Henikoff et al., 2004). This can complicate phenotype studies, giving undesired effects that are unrelated to the targeted gene. In addition, some particular sequences (such as the 35S promoter) are known to trigger progressive T-DNA silencing after each successive generation (Mlotshwa et al., 2010). This underscores the importance of maintaining a clear overview of the progeny of a seed (including amplifications), as well as a record of the history of T-DNA inheritance through crosses with other plant lines, or through secondary mutagenesis of lines already containing a T-DNA insert.

Another critical factor is the maintaining of an accurate record of all stored plant lines to comply with procedures linked to genetically modified organism (GMO) certifications. Although the definition of GMOs itself is a matter of debate, the current European regulation distinguished GMOs based on the techniques used for the biotechnological engineering of plants (Hartung and Schiemann, 2014). It will distinguish between plants that contain recombinant DNA from other organisms (classified as GMOs) and plants that contain only point mutations of their native DNA (considered non-GMOs). Transposons are a specific case, depending on whether or not the sequence of the native transposons has been engineered (Greco et al., 2001; Wegmuller et al., 2008). The ability to distinguish mutations on a biotechnology basis (i.e., the presence or not of T-DNA or transposon transgenes versus point mutations) would be a first step toward the improved tracking of plant lines for GMO certifications.

Seed collections are often the result of combined efforts from several people and different laboratories. Resources can be obtained through seed stock centers or by directly contacting the laboratories that generated them. Although the use of most plant lines is often unrestricted, some lines are protected under a Material Transfer Agreement (MTA) signed between research institutions, defining a strict set of acceptable uses of the seeds. It is important to track the original plants protected by an MTA or under the control of a GMO certification as well as their progeny, through all the series of successive seed amplifications and crosses with other lines. The possible use for all of these

related plants is equally constrained within the limits of signed MTAs or GMO certification documents. New tools that facilitate tracking of GMOs and MTAs for scientists would greatly improve the compliance within administrative and legal contexts.

Several affordable or free software programs are presently available to improve plant line management. However, to our knowledge none of them are capable of reflecting the inheritance patterns of individual mutations through the successive rounds of seed amplifications, line crossing or mutagenesis encountered in a typical research laboratory. Indeed, most programs have been designed for managers of plant transformation or greenhouse facilities that use standardized procedures (Scott et al., 2003; Henry et al., 2008; Kohl and Gremmels, 2010; Hanke et al., 2014), or for plant breeding laboratories facing large sets of phenotyping and genotyping data derived from accession sequencing or QTL mapping (Lee et al., 2005; Jung et al., 2011; Milc et al., 2011; Love et al., 2012). Nevertheless, several software programs have been developed to track plant lines in basic research laboratories, such as PlantDB and Phytotrapper (Exner et al., 2008; Nieuwland et al., 2012). Although they include such functions as plant, seed and plasmid management modules along with genotyping indications, these programs are not capable of independently tracking individual mutations through successive crossings and seed amplifications. In fact, they can only follow the general relationship between seed batches, tracking all seed batches derived from an individual plant.

Despite the presence of these programs, we could not identify any software designed to specifically track the inheritance of mutations or transgenes within the complex history of seed collections, which would also be capable of reflecting the ever-growing diversity of biotechnological applications for plant mutagenesis and transgenesis. We therefore decided to create a new seed and plant database solution that utilizes strong genetic concepts, including mutation inheritance and independent genotyping of each mutation. At the same time, we wanted to provide a simple and intuitive interface that would respect the habits of individual laboratories and their members.

To answer this need, we have developed the “SeedUSoon” software. Its intuitive and flexible user interface permits the tracking of plant lines along with plants and seed batches, and it includes a graphical representation of the genetic link between related plant lines arising from crosses or secondary mutagenesis. Mutations inherited from parental lines can easily be identified using our software, and transgenic (GMO)/non-transgenic (non-GMO) types of mutations can be color-coded for fast visual identification.

The program can be easily customized to the needs of each laboratory through an administrative module, for use with different plant models or mutagenesis techniques, for instance. Users can also decide whether they enter each seed generation, or only important seed batches. Other functions include the uploading of genotyping protocols (for instance PCR primers and programs used to identify a mutation), articles, genotyping, and the phenotyping results of individual samples, microscopy images, etc. We have also developed export/import functions to facilitate seed exchange between laboratories, and MTA tracking functions for improved intellectual property

management practices. Altogether, the SeedUSoon software is an attractive and free solution for plant laboratories facing the challenge of keeping accurate seed collection records.

## MATERIALS AND METHODS

### Implementation

We developed the SeedUSoon user interface (version 1.1.0) using the platform-independent Java programming language<sup>1</sup>. This choice allows the software to operate on any system running Java 1.8 or higher. It has been tested with the Windows XP, Windows 7, Windows 8, and Mac OS X (up to Mavericks) operating systems.

The SeedUSoon software needs to connect to a database (client/server architecture) that can be present on the same computer, or preferably on a server for multiple user access. The computer hosting the database must run MySQL (version 5.5.35 or higher). Extended computer knowledge is only necessary for the database installation.

### Software and Start-Up Database Availability

SeedUSoon is distributed under a proprietary license, and is free exclusively for academic purposes (i.e., non-profit institutions). Non-academics interested in the program cannot access the software and must contact us directly.

Academics can sign the proprietary license agreement through the project website<sup>2</sup>. Once completed, this provides access to the download page for the SeedUSoon software, a start-up database, and the installation procedures.

An example of exported data, a template form to load customized laboratory information, demonstration movies, FAQs and access to updated versions of the software will be posted on the project website. Specific questions can be directed to the project leaders by using the dedicated email address SeedUSoon@cea.fr.

## RESULTS

### SeedUSoon Concepts

#### “Line” Concept

SeedUSoon is designed around the core concept of “plant lines,” whose definition is very similar to the one used by many plant science researchers when referring to the series of successive generations derived from particular plants. A “Plant line” is defined by a set of unique traits (mutations or transgenes, named “Genetic features” in SeedUSoon; **Figures 1A,B**) in a biological context (species and ecotype). All plants and seeds arising from selfcrosses or backcrosses are still considered part of the same “Plant line,” so that under a single “Plant line” entry, the user can record as many seeds or plants as desired. The precise genotyping (such as the heterozygous/homozygous state of each mutation, or

<sup>1</sup><https://www.java.com>

<sup>2</sup><http://biam.ceaf.fr/drif/biam/Pages/laboratoires/lbdp/SeedUSoon.aspx>

**A**

General information	Genetic features	Tree
Plants table	Seed batches table	

**B**

<b>Transgenesis</b>	<b>Endogenous gene mutagenesis</b>
<b>TRANSPOSONS</b> <i>Native</i> <i>Engineered</i> <b>T-DNA</b> <i>Random insertions</i> <i>Transcript overexpression</i> <i>CRISPR /Talen/ZNF machinery</i> <i>GFP-fusions ...</i>	<b>DIRECTED MUTAGENESIS</b> <i>CRISPR, Talen, ZNF targets...</i> <b>RANDOM MUTAGENESIS</b> <i>EMS</i> <i>Gamma irradiation</i> <i>Natural variants ...</i>

**C**

The screenshot shows the SeedUSoon software interface. The main window has several panels:

- General Information (h):** Contains fields for Line name, Secondary names, Source, MTA details, Record date, Period, Species, Ecotype, Article reference, and Article reference file.
- Plants (b):** Shows a table with columns: Actions, Identifier, Generation, Phenotyping, Genotyping. It lists two entries: T1 (pale green) and T2 (dark green).
- Seed batches (c):** Shows a table with columns: Actions, ID, Identifier, Generation, Doseant type, Parent plant, Harvest, Storage date, Storage place, Quantity, Germassate. It lists three entries: A-4 (Self-cross, A, 2015-07-22, Lab stock, box 3, about 100 seeds), A-5 (Self-cross, 2016-07-14, Lab stock Box 16), and A-6 (Self-cross, 2016-07-14, Lab stock Box 16).
- Search plant lines by criteria (g):** Includes fields for Gene/Line/Genetic feature and Line.
- Genetic features (a):** Shows a table with columns: Line name, Secondary names, species, Nb of genetic features. It lists one entry: 35S:GUS [from Pro35S:GUS].
- Tree (i):** A hierarchical tree diagram.
- Add a transgenesis (k):** Button for adding new transgenesis.
- Add an endogenous gene mutagenesis (l):** Button for adding new endogenous gene mutagenesis.
- Export buttons (m, n):** Buttons for exporting data.
- Access buttons (o, p):** Buttons for access to different modes.
- User mode interface (e, f, j):** Buttons for user mode, administrative mode, and genealogy tree.

**FIGURE 1 | User mode interface.** **(A)** General organization of a plant line datasheet. **(B)** The two categories of genetic features (transgenesis/endogenous gene mutagenesis) with examples of corresponding applications. **(C)** Detailed user mode organization; **a:** genetic features table, **b:** plant table, **c:** seed batches table, **d:** access to the customized laboratory guidelines, **e:** access to the user mode, **f:** access to the administrative mode, **g:** search engine, **h:** plant line datasheet, **i:** general information, **j:** genealogy tree, **k:** access to plant line generation wizards (new, crossing, mutagenesis, and import), **l:** addition of new genetic features, **m,n:** lock buttons, **o,p:** export buttons.

the segregation pattern of seed batches) can be recorded for each individual plant or seed batch entry.

However, if crosses are performed with plants carrying other types or mutations or with plants from another ecotype, the genetic properties of the resulting line will significantly differ from the original plant. For this reason, the progeny of these types of crossings must be entered as a new “Plant line” entry. Similarly, if a previously mutagenized plant line is subjected to a secondary mutagenesis or transgenesis, the resulting progeny will constitute a new “Plant line.”

A “Plant line” datasheet is organized in five different parts (**Figure 1A**). The line name and general information associated with this line are displayed in the upper left area, including the plant species and ecotype, its origin and the existence of any MTA protecting the material (“General information” fields, **Figures 1A,C**, Supplementary Figure 1). In the center of the screen, a table presents blue or green boxes containing the names of all individual mutations or transgenes (i.e., “Genetic features”; **Figures 1A–C**) along with their origin (in which “from...” indicates the original plant line containing this mutation). The content of each box can be expanded in order to access the individual properties of each “Genetic feature” (“Genetic features fields,” Supplementary Figure 1). If no green or blue box appears inside this table, the genotype of the corresponding plant line is considered to be WT. At the upper right corner, the parental lineage and the mutagenesis/genetic history of the opened plant line datasheet is represented by a tree.

Finally, two tables for plants (**Figures 1A,C**, left) and seed batches (**Figures 1A,C**, right) are located in the lower part of the datasheet. Generations of plants and seeds can be recorded at any time, including the skipping of generations, which allows users to avoid a strict “generation workflow.” The user enters the generation stage and can record for each plant the precise genotyping of each “Genetic feature” listed in the upper table (or the segregation profile for each seed batch). Consequently, by using a “Plant line” as the entry point, the user can access all recorded generations of seeds or plants that share the same overall genetic background.

By organizing the datasheet into five parts (general information, genetic features, tree, the plants table, and the seed batches table), users can focus on the core properties of each line (genetic context and history, ecotype, etc.) before searching through all available seeds or plants corresponding to these criteria. The software also allows users to decide whether they want to track all successive generations of a plant line, or to only record a subset of particularly valuable seed batches.

During the development of our software, the ability to compare the behavior and properties of successive seed generations in a single table was a recurring request from researchers that we spoke with. One reason for this is that epigenetic phenomena can affect the behavior of descendants of seemingly identical plants, in particular through DNA methylation (Mlotshwa et al., 2010; Diez et al., 2014). A table comparing the properties of distinct generations can thus be instrumental in identifying the appearance or loss of specific phenotypes, or the progressive silencing of T-DNA expression. Similarly, the lack of a link between phenotype and desired

mutations might suggest the presence of an unknown off-target mutation (frequent with CRISPR/Cas9 or EMS mutagenesis).

When performing T-DNA transformation for specific purposes (such as RNAi silencing or expression of GFP-protein fusions), it is convenient to visualize all available independent transformants at once (along with their descendants). In this case, a software user often prefers to record all independent transformants within a single “Plant line” rather than as separate “plant lines.” Properly speaking, the independent transformants do not share the exact same genetic properties. The T-DNA insertion sites are variable in these lines, and have a putative impact on the properties of the resulting plants (Kohli et al., 2006). Nevertheless, comparing all independent transformants within a single plant line can be advantageous for tracking the outcome of a T-DNA transformation; this approach can also quickly identify undesirable effects, such as construct silencing, patchy expression, etc. With SeedUSoon, laboratories can decide if they want to record independent plant line datasheets for each insertion or use a single datasheet for all independent transgene insertion events, since T-DNA insertion sites can be defined in two locations within the datasheet: either in the transgene sequence section in the “Genetic feature” table (**Figure 1A**), or for each of the individual plants recorded in the lower table (i.e., different insertion sites can be recorded in the plant table specifically for each plant entry; **Figures 1A,C-a,b**). Finally, seed batches can be linked to these individual plants.

## Two Categories of “Genetic Features”

Mutations or transgenes present in the genome of a “Plant line” are recorded as unique “Genetic Features,” and listed in the corresponding table of the plant line datasheet (**Figure 1A**). There are two distinct feature categories, “Transgenesis” and “Endogenous gene mutagenesis,” with the latter one corresponding to point mutations, nucleic deletions and insertions affecting an endogenous genomic sequence that does not involve the insertion of a transgene (T-DNA or transposon).

The “Transgenesis” feature must be selected for T-DNA or transposon mutagenesis (Clough and Bent, 1998; Greco et al., 2001; Gomez et al., 2009), and corresponding mutations will appear in green in the features table (**Figure 1B**). “Endogenous gene mutagenesis” corresponds to mutations in endogenous genes with no transgene insertion, such as EMS, gamma irradiation, or natural variants (Kim et al., 2006; Svistoonoff et al., 2007; Fauser et al., 2014). These will appear in blue in the features table (**Figure 1B**). In the case of CRISPR, TALEN, or ZFN mutagenesis, the plant line will contain both a “Transgenesis” box in green (i.e., the T-DNA containing the mutagenic machinery) and an “Endogenous gene mutagenesis” box in blue [i.e., the targeted endogenous gene loci (Fauser et al., 2014)].

The blue/green color code allows the user to quickly recognize the “Transgenesis” from “Endogenous gene mutagenesis” features. Plants potentially containing transgenes (i.e., GMOs) can thus be immediately distinguished from all other mutation categories (**Figure 1A**).

Each feature category will call for a specific set of information fields that are ready to be completed by the user (Supplementary Figure 1). In particular, a single sequence can be recorded

for “Endogenous mutagenesis” (the mutated genomic locus), whereas a “Transgenesis” feature can record the transgene sequence (in the “Genetic features” properties) as well as several independent genomic insertion sites (in the plant table entries; **Figure 1A**).

## Easy Customization of SeedUSoon: Adaptation to the Laboratory Context

### Personalized “lab” guidelines

SeedUSoon contains a customizable document that will provide laboratory members with specific guidelines and rules decided within their own laboratory. An “Our lab” icon is always visible on the SeedUSoon main page and provides access to this document (**Figure 1C-d**). This document specifies how to name lines and successive generations, and describes which file formats are acceptable for upload into the database. In addition, it provides details on the organization of the laboratory’s common seed stock, how to store seeds, and protocols for seed selection, transformation, etc.

A window asking the user to upload the manual will appear following the first activation of the “Our lab” icon. After this initial upload, the document will automatically open whenever any user clicks on “Our lab.” Newer versions of the manual can then be uploaded by following the path: Tools tab/Options/Labo/User manual.

A document containing an example of laboratory guidelines is provided for use as a template (see Supplementary Data Sheet 1 and the project website).

### Customization of the user module

Parameters and methods susceptible to change between laboratories are presented in scroll-down menus when in the user mode. These menu options are customizable, but can only be edited by the database administrator in the administrative mode (see the corresponding section for details). This allows the software options to closely match the habits and protocols of each laboratory, while preserving a certain consistency.

Through these scroll-down menus, the user will have access to a specific selection of laboratory member’s names, plant species, ecotypes, strains, plant resistances, and mutation methods in use in the laboratory. New entries or modifications to the scroll-down menu options can be made at any time during the database lifetime, and corresponding plant lines will be updated accordingly.

## User Mode

After starting the software, the user interface can be accessed from the home page by clicking on “User” (**Figure 1C-e**).

### Built-In Pop-Ups

Scrolling the mouse pointer over most fields or icons activates pop-ups that provide more information to the user about the purpose of these functions (**Figure 2**). In some cases, pop-ups will recommend reading the “Our lab” document mentioned in the previous section, to ensure that users will follow the specific rules that have been decided for their laboratory.

**FIGURE 2 | Search engine interface.** Key words can be entered in a generic field (“Gene/Line/Genetic feature” field) or in specific sub-categories to narrow the search. A “%” symbol can be included to replace any number of characters in the query. Including a “\_” indicates the presence of a single missing unknown character in the query. Adding a “\$” before “\_” or “%” in the query field will allow the user to search for items containing the characters “\_” or “%,” bypassing the use of these symbols as replacements of any characters in the query.

## Searching for Available Plant Lines or Seed Batches

A search engine is located at the bottom of the user interface (**Figures 1C-g and 2**), which can provide access to all “Plant lines” present in the database (by clicking on “Show all lines”), or only a subset of lines when “Search lines by criteria” is selected (**Figure 2**). The first field (“Gene, Line, Genetic feature”) can be used to search a keyword throughout all plant line names, gene names and genetic features recorded in the database. Alternatively, users can select more restrictive query criteria by completing the fields specifically associated with the 4 individual sub-parts of a plant line datasheet. These fields can be among the general properties of the line, genetic feature properties, and plant or seed batch properties (including seed batch name or ID, person involved, etc.; **Figures 1A and 2**; Supplementary Figure 1).

When there is some uncertainty regarding the exact spelling of a query, a “%” symbol can be included at the beginning or end of the word (**Figure 2**). This will identify any lines containing the searched criteria, including any number of characters appearing before or after the searched word (i.e., “%” = any number of characters).

Clicking on the name of a plant line in the search engine result table will open the corresponding datasheet (**Figure 1C-h**).

## Creating a Plant Line

There are four possible ways to create a new line in SeedUSoon: through the generation of a “New record,” crossing, secondary mutagenesis or by import. A button corresponding to each mode is located at the upper left corner of the user interface (**Figure 1C-k**).

### New record in the database

The user can create a “Plant line” record *de novo*, by entering any available information in the empty fields of the new database entry (**Figure 3**). Most fields are optional and can be completed later (mandatory and facultative fields are listed in Supplementary Figure 1) to avoid any wrong assumptions when recording data (arising from erroneous “guess work”). Data can easily be saved, completed or modified at any time by clicking on the lock buttons (**Figures 1C-m,n and 3**).

When starting the “new” plant line wizard, the user will only be required to complete three mandatory fields: the line name, the person associated with this datasheet, and the plant species. No other additional information is needed in the case of a WT plant.

If the plant line contains one or several mutations (or transgenesis), the user can click on “Add a transgenesis” or “Add an endogenous gene mutagenesis.” This will select the correct category of “Genetic Features” to appear in the “Genetic features” table (**Figures 1B,C-a**) with new empty fields related to these specific mutations (such as gene, mutagenesis method, transgene or mutation sequence, attached sequence files, selectable marker in plants, etc. see Supplementary Figure 1). Only the “designation” (i.e., a mutation name) is required for each genetic feature, which allows their quick recognition in the “Genetic features” table (for example: Pro35S:GUS). In the “Genotyping protocol” field, the user can type or upload a standard genotyping method for this particular feature (including PCR primers, PCR programs, a picture of a typical gel, etc.).

The user should add as many “Genetic features” as the line contains mutations or transgenesis. The different mutations can all be recorded during the “New” plant line wizard, or can be completed after a plant line is created (by clicking on the buttons located over the genetic features table; **Figure 3**).

### Crossing two previously recorded plant lines

SeedUSoon is capable of predicting the genetic configuration of plants resulting from the crossing of two plant lines that are already present in the database. The software will import all important properties from the parental lines and create a new plant line that combines this data (mix of ecotypes, set of combined genetic features, etc.). Only intra-species crosses are permitted by the software.

After starting the “Cross” wizard (**Figure 4**), the user will only need to specify which parents were used as male/female; optionally, the seed batches used for the crossing can be included. This will generate a new datasheet, to which the user can allocate corresponding seeds or plants (the user can also specify whether the mutations are homozygous or heterozygous). To avoid mistakes, the user cannot modify the inherited properties,

as these come from the parental lines. Any modifications should therefore be entered in the parental line, and all descendants will be updated accordingly. All non-inherited fields can be edited.

### Secondary mutagenesis of a recorded plant line

The “MUTAG” wizard can be used if a new mutagenesis has been applied to a plant line previously recorded in the database (**Figure 5**). As with the “Cross” wizard, “MUTAG” will import all the genetic properties from the mother line into the new line (species, ecotype, and genetic features). The user will only need to complete the fields corresponding to the new “Genetic feature” selected for the secondary mutagenesis (**Figure 5**).

Similarly to the “Cross” wizard, the data inherited from the parental line cannot be edited in the datasheet of the resulting line.

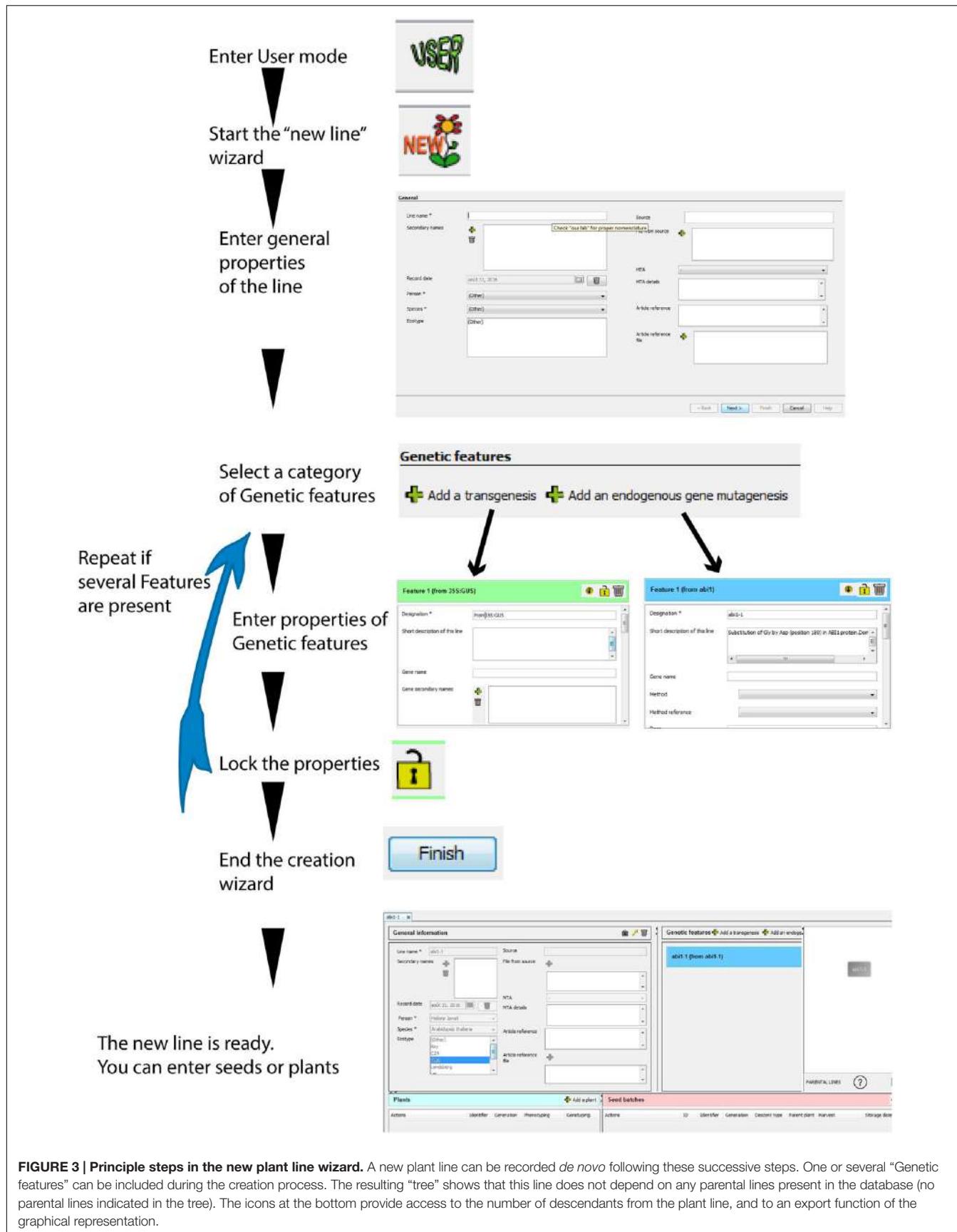
### Importing/exporting a line

It is possible to export a database entry from a “Plant line” (with or without the corresponding seed batches) into a single file that can easily be sent to collaborators (**Figure 6**). This file is generated using an exchange format (.json). Although this format can only be partially read in a text processor such as WordPad, it permits a very complete data exchange between two SeedUSoon databases, including all attached files (plasmid sequence, phenotyping results, etc.).

The export can be performed either from the Research result table (**Figure 6A**) or directly from the opened plant line datasheets (**Figure 6B**). The simultaneous export of several lines is only possible through the Research result table, however, no seed batches can be exported along with the line information in this case. When exporting directly from the plant line datasheet, it is possible to assemble the information from one or several seed batches (with the exception of sensitive/personal information; see Supplementary Figure 1). Any plant (from the plant table) that is linked to a seed batch (identified as its mother plant) will be exported as well.

Reciprocally, a plant line can be created by importing data from other databases. To import data, the user must click on the “Import arrow” at the upper left corner (**Figures 1C-k and 6C**), specify a name associated with this new database entry, and select the .json file. If the import contains options that are not available in the database scroll-down menus, a pop-up window will warn the user that the administrator must create the corresponding choices via the administration mode. Alternatively, the missing entry can be edited directly by opening the .json file in a text processor software; this can also serve as a temporary solution if the administrator is not present. For instance, a missing ecotype can be temporarily changed to “(Other)”; this option is included by default in the scroll-down menu.

When lines are imported, all links to parental lines are severed. However, the imported line will contain the proper list of genetic features. The graphical representation of the genealogical tree (refer to the section below) will not be lost: it will be exported as an image, and will be uploaded as a “File from source” in the “General information” panel (**Figure 1C-i**).

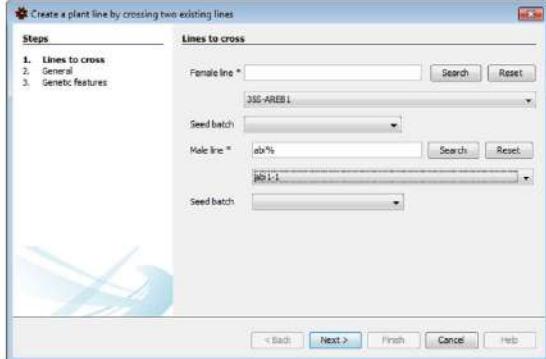


**FIGURE 3 | Principle steps in the new plant line wizard.** A new plant line can be recorded de novo following these successive steps. One or several "Genetic features" can be included during the creation process. The resulting "tree" shows that this line does not depend on any parental lines present in the database (no parental lines indicated in the tree). The icons at the bottom provide access to the number of descendants from the plant line, and to an export function of the graphical representation.

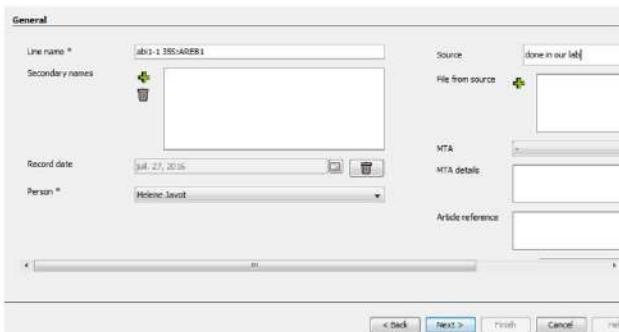
Enter User mode and start the "crossing" wizard




Select the parental lines



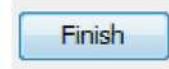
Enter the general properties of the resulting line



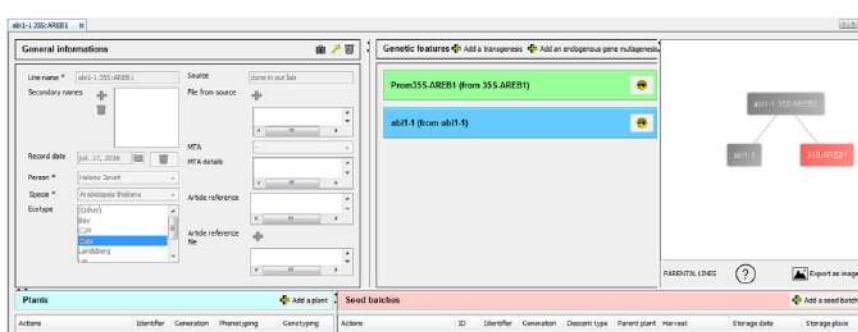
*Optional for pollen mutagenesis: include an extra Genetic feature*



End the creation wizard



The new line is ready. You can enter seeds or plants in the tables

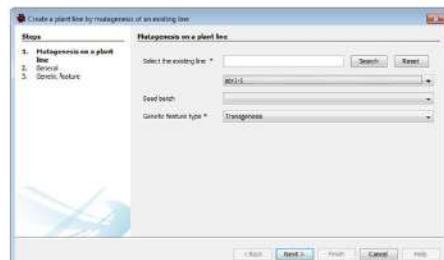


**FIGURE 4 | Simulated crossing of two lines using the "cross" wizard.** A new plant line can be created using the simulated crossing between two parental lines present in the database. This line will inherit all genetic features from the parents. It is possible to add an extra genetic feature during the crossing process for the specific case of pollen mutagenesis, where crossings are combined with additional mutagenesis. The resulting "genealogical tree" reflects the relationship between the parental and resulting plant lines. A plant line protected by an Material Transfer Agreement (MTA) is indicated in red. The icons at the bottom provide access to the number of descendants from the plant line, and to an export function of the graphical representation.

Enter User mode and start the “mutagenesis” wizard



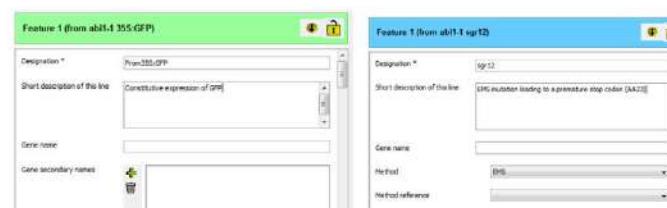
Select the parental line and the category of the new Genetic feature



Enter the general properties of the resulting line



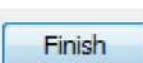
Enter the properties of the new Genetic feature



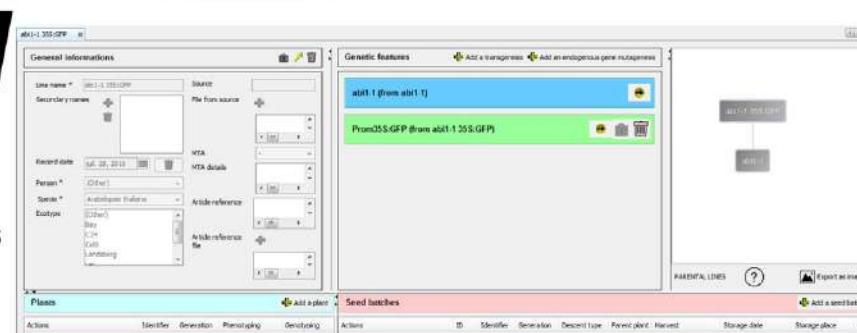
Lock the properties



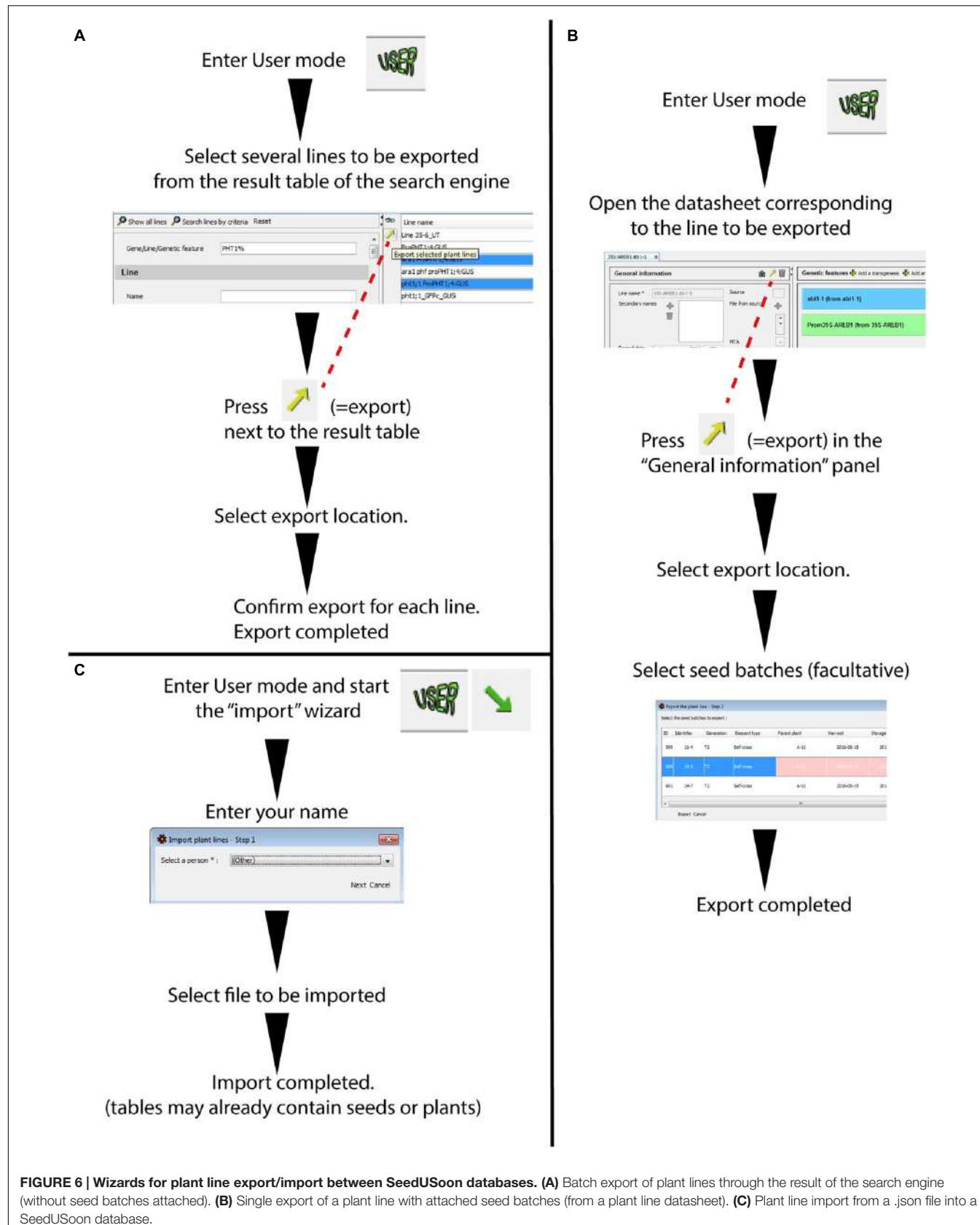
End the creation wizard



The new line is ready. You can enter seeds or plants in the tables



**FIGURE 5 | Creation of a line using the “MUTAG” wizard in order to apply a secondary mutagenesis to an existing plant line.** A new plant line can be created by secondary mutagenesis (addition of a genetic feature) to a parental line already present in the database. The new line will inherit all genetic features from this parental line, and combine it to the new genetic feature. The resulting “genealogical tree” reflects the relationship between the parental and resulting plant lines. The icons at the bottom provide access to the number of descendants from the plant line, and to an export function of the graphical representation.



**FIGURE 6 | Wizards for plant line export/import between SeedUSoon databases. (A)** Batch export of plant lines through the result of the search engine (without seed batches attached). **(B)** Single export of a plant line with attached seed batches (from a plant line datasheet). **(C)** Plant line import from a .json file into a SeedUSoon database.

**FIGURE 7 | Two tables for recording biological material corresponding to a plant line.** Samples are separated into two tables: plants (**A**) and seed batches (**B**). In the example shown here, the researcher decided to record different T1 plants derived from a T-DNA transformation within a single plant line datasheet. The number of insertion sites and location/sequences of independent insertions can be stored separately for each plant. Only the major fields are shown in the columns; detailed information and attached files can be accessed by clicking on the eye icon. It is possible to link the mother plants to their seed batches. The “Copy” button permits fast duplication of plant or seed batch data.

Most of the entries related to general information, genetic features and seed batches will be imported (**Figures 1A,C-i** and Supplementary Figure 1). However, any sensitive/personal information (such as any personal name, notebook information, storage place, or MTA details) will not be included in the export format file.

#### Graphical representation of the genealogy of a line

On the right part of any “Plant line” datasheet is displayed a graphical representation of the history (“genealogy” or “tree”) of this plant line (**Figure 1A**) in relation to other plant lines. The purpose of this is not to track successive generations of a single plant line, but rather to represent the links toward the parental “Plant lines” and visualize when new “Genetic features” were brought into the genome of the plants.

A plant line created *de novo* (using the “New” plant line wizard) will be represented by a simple rectangle containing its name (**Figure 3**). If a line is crossed to another one (through the “Cross” wizard), or if it was recorded for a secondary mutagenesis (using the “MUTAG” wizard), both the parental lines and the resulting line will appear in the graphics (**Figures 4 and 5**). The graphical representation will reflect the complete origin of a plant line, even for plant lines resulting from several rounds of successive crossings and mutagenesis.

Although this tree does not directly track the individual mutations, it is easy to infer from the adjacent “Genetic features” table whether any mutations or transgenes are inherited from a parental line. First, inherited genetic features cannot be edited (no “lock” or “trash” icons appear in their corresponding boxes). Second, the boxes contain the name of the plant line from which the corresponding genetic feature (indicated by the name of the feature followed by “from...”) originated. The color code of the boxes (green/blue) permits the fast identification of the category of the genetic feature (transgenic vs. non-transgenic).

The parental line can be directly accessed by clicking on its name in the graphics. In addition, scrolling the mouse pointer

over the line connecting two lines displays the seed batches used for their generation in a pop-up (if previously recorded).

The graphics only represent the parental lines of the line of interest. However, placing the mouse pointer over the question mark located beneath the graphic will display a pop-up indicating the number of descendants derived from this specific line (**Figure 1C**).

The parental lines are located at the bottom of the “genealogy” tree in this software version. The tree orientation can be inverted or modified by dragging its individual components inside the graphics window with the mouse.

Graphics can be exported as an image (in .png format) by clicking on “Export as image” underneath the tree (**Figure 1C**), for inclusion in notebooks or PowerPoint presentations. Furthermore, when exporting a line using the SeedUSoon export/import format, the graphical representation of the genealogical tree will be included as an image in the “File from source” field, in the “General information” panel (**Figure 1A**).

#### Recording Plants and Seeds

Following the creation of a “Plant line,” it is possible to record the corresponding seed batches or individual plants in two dedicated tables (**Figures 1A,C-b,c** and **7A,B**). These tables can contain any generation of seed batches or plants sharing the same ecotype and set of mutations (i.e., “Genetic features”), including descendants of self-crossed or back-crossed plants. For each plant or seed batch entry, the user can specify their specific genotype or segregation profile (heterozygous/homozygous, single/multiple transgene insertions, resistance or mutation segregation ratio).

The “Plant” and “Seed batch” wizards can be activated by clicking on “Add a plant” or “Add a seed batch” located at the right corner above the plant and seed tables, respectively (**Figures 7A,B**). The only mandatory field here is the personal plant or seed batch identifier. All other information (generation, phenotyping, genotyping, harvest date, etc.) is optional (see Supplementary Figure 1 for the list of available fields) and can

be completed later. Generation stages are entered by the user in the corresponding field, according to the recommendations of the customizable “Laboratory guidelines.” These guidelines can request the use of classical terms, such as T1, T2, F1, F3, as well as other terms such as “unknown” or “Tx” when receiving seeds from another laboratory, for example (see Supplementary Data Sheet 1). Although there is no requirement to track all successive generations, it is possible to associate specific plants with their progeny using the “S” (seed batch) function available for each plant table entry (**Figure 7A**). Reciprocally, mother plants from seed batches can be indicated when using the “Seed batch” wizard.

SeedUSoon will assign a unique ID number to each seed batch (in addition to the identifier entered by the user; **Figure 7B**, “ID” column). If a seed batch is deleted (for instance if no seeds are left), this ID number can never be reallocated to any other seed batch. Similarly, even if two seed batches possess the same personal identifier, they will have two unique ID numbers. This software-generated unique ID number therefore provides an easy and secure way to unambiguously distinguish seed batches. This feature can also be used to improve seed stock organization, if included in the label present on the seed stock tubes. SeedUSoon users can simply enter this ID number in the software search engine, and, with this information alone, directly access the corresponding seed batch and plant line information.

An additional field can only be activated for plants resulting from transgenesis, to permit the recording of the location and sequences of one of several insertion sites in individual plants. This facilitates the work of users who prefer to record a series of plants with independent T-DNA or transposon locations within the same table of a unique “Plant line,” rather than in separate “Plant lines” (an example of this application can be seen in **Figure 7B**).

A “Copy” button is located at the left of each table entry to accelerate the recording of similar seed batches or plants (**Figures 7A,B**). Its activation will open a wizard, and the user will only need to validate or edit the duplicated information. For seed batches, the software will allocate a unique ID number to the new entry.

## Recording Phenotypical Data and Experimental Results

Since phenotypical data are often influenced by plant or seed batches, phenotypical results can be individually recorded for each entry in the “Plant” and/or “Seed batch” tables of SeedUSoon (**Figure 1C-b,c**, Supplementary Figure 1). The user can type a short description of the phenotype in the corresponding wizard field (this text will appear in the table; see the plant example in **Figure 1C-b**). The user can also upload files describing detailed phenotyping results in the same section. Scrolling the mouse over the “phenotyping” section of the table will reveal the presence of the uploaded file.

Results of tissue-expression patterns (such as from GFP-fusion or reporter gene studies) can also be uploaded in this “phenotyping” section of the “Plant” and “Seed batch” tables.

Germination assays, genotyping and sequencing results can also be recorded or uploaded within individual seed or plant

batches. The reference number and pages of the laboratory notebook can be indicated for each result section (phenotyping, germination assay, genotyping, etc.), along with the name of the person who conducted the experiment.

## MTA Tracking

SeedUSoon includes a function to help protect the intellectual property of laboratories, especially related to MTA tracking. An MTA field is included in the plant lines “General information” panel (**Figure 1C-i**). The “MTA details” field allows users to record the recipient of the MTA, its location, or any particular recommendation. When exporting a line, the information regarding the presence of an MTA protecting the line is preserved. However, the “MTA details” field is left empty for confidential reasons.

SeedUSoon’s graphical “tree” representations of plant lines allow users to immediately identify a protected material. Any plant line protected by an MTA will be indicated in red (**Figure 4**), so that tracking its descendants will be straightforward, even long after obtaining and using the original seeds.

## Administrative Mode

The “Admin” icon (**Figure 1C-f**) provides access to the administrative mode (only for users with administrative rights) to be able to customize the user interface, create SeedUSoon user accounts, and specify their rights. All customization choices recorded from the administrative mode (on a single computer) will be effective for all computers that connect to the same database.

The administrative mode is a very simple interface organized in eight tabs, each giving access to a table with editable content (**Figure 8A**). A wizard for generating new entries can be activated by clicking on “New” at the bottom of each table. Existing entries can be separately edited or deleted using the buttons at the right of each table.

## Defining User’s Rights

In the first tab (“SeedUSoon Users”), the administrator can create login accounts and define distinct levels of user’s rights (**Figure 8A**). SeedUSoon users will either be allowed to enter and modify data (“Writer” level), or will only be able to access the data without modifying them (“Reader” level). The final user level (“Administrator”) gives additional access to the administrative mode.

## Scroll-Down Menu Customization

All tabs (aside from the one used to define “SeedUSoon users”) are dedicated to the customization of the user’s module (**Figure 8B**). This allows the administrator to specify the options available in the scroll-down menus presented to the users. The software comes with a set of pre-recorded options for each tab, which can easily be edited by the database manager. For each tab entry, the software will automatically verify and count lines, features or experiments containing the corresponding scroll-down menu option. This will help the manager visualize the relevance of certain fields, in order to only delete unused menu options.

**A**

SeedUSoon Users	Persons	Species	Ecotypes	Mutation methods	Mutation method references	Strains	Resistances	Actions
Login		Profile						
admin			ADMINISTRATOR					<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
DUBOIS			WRITER					<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
DURAND			WRITER					<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
KIM			WRITER					<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
LIN			WRITER					<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
MARTIN			WRITER					<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
SMITH			ADMINISTRATOR					<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
Students			READER					<input type="button" value="View"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>
<input type="button" value="New"/> <input type="button" value="Delete selected row(s)"/>								

**B**

...users	Persons	Species	Ecotypes	Mutation methods	Mutation method references	Strains	Resistances	Actions
	Firstname	Lastname	Shows up in the "persons" menu (User mode)		Number of lines	Number of experiments		
(Other)			<input checked="" type="checkbox"/>	7	0		<input type="button" value="View"/> <input type="button" value="Edit"/>	
Alex	MARTIN		<input type="checkbox"/>	2	4		<input type="button" value="View"/> <input type="button" value="Edit"/>	
Christine	BAKER		<input checked="" type="checkbox"/>	4	18		<input type="button" value="View"/> <input type="button" value="Edit"/>	
Erik	LIN		<input checked="" type="checkbox"/>	0	0		<input type="button" value="View"/> <input type="button" value="Edit"/>	
Isabelle	KIM		<input checked="" type="checkbox"/>	17	0		<input type="button" value="View"/> <input type="button" value="Edit"/>	
Robert	SMITH		<input checked="" type="checkbox"/>	1	2		<input type="button" value="View"/> <input type="button" value="Edit"/>	
Stefan	DURAND		<input checked="" type="checkbox"/>	2	12		<input type="button" value="View"/> <input type="button" value="Edit"/>	

**C**

Database Configuration Panel:

- General
- Keymap
- Appearance
- Miscellaneous
- Labo

Filter (Ctrl+F)

Database User manual

THIS PAGE IS FOR DATABASE ADMINISTRATORS ONLY.  
Please do not modify the following parameters without contacting your database administrator first.

Username (\*) : SeedUSoondatabase

Password (\*) : \*\*\*\*\*

Database URL (\*) : jdbc:mysql://localhost:3306/seedusoon

**FIGURE 8 | Administrative mode and database configuration. (A)** The administrative interface is organized in eight tabs: SeedUSoon users (selected here), Persons, species, Ecotypes, Mutation methods, Method references, Strains and Resistances. **(B)** Persons tab, displaying how many lines or experiments are associated with specific names corresponding to laboratory members. **(C)** Database configuration panel.

## Persons

The “Persons” tab (**Figure 8B**) corresponds to current or past laboratory members who contributed to the generation or the analysis of any plant line recorded in the database. This category should not be confused with the previously mentioned “SeedUSoon users” category. If a “Person” leaves the laboratory and no new entries will be generated under this name, it is possible to deactivate (hide) the name from the scroll-down menus when generating new plant line entries. This limits the length of scroll-down menus in laboratories with a high turnover of members. To

do this, the administrator must uncheck the box “Shows up in the “Persons” Menu (User mode)” in the central column of the “Persons” table (**Figure 8B**). Previous entries containing a reference to this person will still display the name.

## Species and ecotypes

Laboratories can enter their plant models and favorite ecotypes in the “Species” and “Ecotypes” tabs. In the scroll-down menus of the user mode, ecotypes will be specific to each species. For this reason, in the administrative mode, new species must be

recorded before registering ecotypes, so that the correct species can be linked to the new ecotype when creating new entries in the ecotype tab.

#### Mutation methods

The “Mutation methods” tab contains the common short terms used to refer to standard mutagenesis techniques used in the laboratory (such as: “T-DNA,” “CRISPR,” “EMS,” etc.). When specifying a new “Mutation method,” the administrator must first define its “Genetic features” category (i.e., “Transgenesis” or “Endogenous gene mutagenesis”; **Figure 1B** and refer to the dedicated section).

The “Mutation method references” tab can be used to record a precise reference from the literature or a precise protocol registered in the laboratory.

Finally, *Agrobacterium* strains and resistances originated by T-DNA or transposon insertions in the plant genome can respectively be recorded in the “Strains” and “Resistances” tabs.

### Database Configuration

Database connection parameters must be entered at the first startup of the software (**Figure 8C**). The software will then restart to allow users (or administrators) to enter their login and password to access the user (or administrative) mode after this initial configuration.

If some users need to connect to a different database, these connection parameters can be modified by following the path: Tools tab/Options/Labo/Database.

## CONCLUSION AND FUTURE DEVELOPMENTS

SeedUSoon is a new plant line database software, built upon a strong genetic foundation. The software’s ability to track the history of mutations or transgene inheritance, in addition to the possibility to record related seed batches, provides the user with a more clear and organized view of the genomic context of their biological material. SeedUSoon contains novel functions related to MTA tracking and easily distinguishes between GMO and non-GMO plant lines, to facilitate administrative and legal compliance. Exporting data between databases is also greatly simplified by the import/export functions.

Our intention when we started to design SeedUSoon was to improve the management of our own laboratory plant lines and seed collections. Nevertheless, from the start, the software was also meant to be able to adjust to the context and habits of any other plant laboratories conducting basic research. We achieved this goal by developing a customizable user’s module, and by integrating choices for field entries that are respectful of individual user habits. To help managers or PIs standardize entries in their own database or seed collection, a customizable “Laboratory guidelines” document is easily accessible from the software.

Several additional functions were requested during the development of SeedUSoon. The current version of the software

was designed in order to implement most of these requests in the future. For instance, the possibility to connect to different SeedUSoon databases using a simple Login/Logout could be advantageous to access distinct databases dedicated to specific projects. We also took into account future functions that can print labels (with customizable content, including unique ID numbers and plant line names), or export data in a diversity of formats (to generate files necessary for GMO certification, for instance). In collaboration with our Intellectual Property department, we considered the possibility of generating MTAs prefilled with plant line information, which would only require the addition of the recipient identification and the approval signatures. This feature would greatly facilitate and stimulate this procedure, since the signing of MTAs when sending seeds to other research groups is hard to implement in many laboratories.

User feedback (through the project website and the dedicated email address) will be important in helping us decide on the strategy for future SeedUSoon developments. Similarly, the design of the current version was improved by the feedback from users of previous versions of SeedUSoon. In the current configuration, this software has already helped laboratories organize hundreds of plant lines, from their generation to the organization of seed collections. Several plant biology laboratories from our research organization have implemented SeedUSoon in recent years, and it is now available for broader distribution (under the protection of a proprietary license agreement).

The design of this software is intended to help others optimize the tracking of their biological material. Ultimately, SeedUSoon will contribute to a facilitated and improved exchange of information to accompany seed exchange between laboratories.

## AUTHOR CONTRIBUTIONS

HJ, CC, LN, and NP designed the functional aspect of the software. CC and SS programmed the software. HJ drafted the manuscript. All authors tested the software, and have read and approved the final manuscript.

## FUNDING

This project was supported by the GIPSE computational support group of the Commissariat à l’Energie Atomique et aux Énergies Alternatives (CEA) and by a Marie Curie International Reintegration grant to HJ as part of the 7th European Community Framework Programme.

## ACKNOWLEDGMENTS

Members of the LBDP, LGBP, and PCV laboratories (CEA) are gratefully acknowledged for their useful suggestions and testing of the software. We also thank the GIPSE board members for

supporting the development of SeedUSoon, and Brandon Loveall from IMPROVENCE for English proofreading of the software interface and manuscript. We also thank the creators of the FigureJ plug-in for their helpful figure preparation tool (Mutterer and Zinck, 2013).

## REFERENCES

- Baltes, N. J., and Voytas, D. F. (2015). Enabling plant synthetic biology through genome engineering. *Trends Biotechnol.* 33, 120–131. doi: 10.1016/j.tibtech.2014.11.008
- Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313X.1998.00343.x
- Diez, C. M., Roessler, K., and Gaut, B. S. (2014). Epigenetics and plant genome evolution. *Curr. Opin. Plant Biol.* 18, 1–8. doi: 10.1016/j.pbi.2013.11.017
- Exner, V., Hirsch-Hoffmann, M., Gruissem, W., and Hennig, L. (2008). PlantDB – a versatile database for managing plant research. *Plant Methods* 4:1. doi: 10.1186/1746-4811-4-1
- Fauser, F., Schimpl, S., and Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J.* 79, 348–359. doi: 10.1111/tpj.12554
- Gomez, S. K., Javot, H., Deewatthanawong, P., Torres-Jerez, I., Tang, Y., Blancaflor, E. B., et al. (2009). *Medicago truncatula* and *Glomus intraradices* gene expression in cortical cells harboring arbuscules in the arbuscular mycorrhizal symbiosis. *BMC Plant Biol.* 9:10. doi: 10.1186/1471-2229-9-10
- Greco, R., Ouwerkerk, P. B., Sallaud, C., Kohli, A., Colombo, L., Puigdomenech, P., et al. (2001). Transposon insertional mutagenesis in rice. *Plant Physiol.* 125, 1175–1177. doi: 10.1104/pp.125.3.1175
- Hanke, L. A., Botelho, C. S., Braz, F. A., Batista, P. H., Folgueras-Flatschart, A. V., Noda, R. W., et al. (2014). FluxTransgenics: a flexible LIMS-based tool for management of plant transformation experimental data. *Plant Methods* 10, 1–9. doi: 10.1186/1746-4811-10-20
- Hartung, F., and Schiemann, J. (2014). Precise plant breeding using new genome editing techniques: opportunities, safety and regulation in the EU. *Plant J.* 78, 742–752. doi: 10.1111/tpj.12413
- Hellens, R., Mullineaux, P., and Klee, H. (2000). Technical focus: a guide to *Agrobacterium* binary Ti vectors. *Trends Plant Sci.* 5, 446–451. doi: 10.1016/S1360-1385(00)01740-4
- Henikoff, S., Till, B. J., and Comai, L. (2004). TILLING: traditional mutagenesis meets functional genomics. *Plant Physiol.* 135, 630–636. doi: 10.1104/pp.104.041061
- Henry, L., Ramm, K., Zhu, Q.-H., and Upadhyaya, N. M. (2008). RGMIMS: a web-based laboratory information management system for plant functional genomics research. *Mol. Breed.* 22, 151–157. doi: 10.1007/s11032-008-9160-z
- Jung, S., Menda, N., Redmond, S., Buels, R. M., Friesen, M., Bendana, Y., et al. (2011). The Chado Natural Diversity module: a new generic database schema for large-scale phenotyping and genotyping data. *Database (Oxford)* 2011:bar051. doi: 10.1093/database/bar051
- Kim, S.-I., and Gelvin, S. B. (2007). Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the *Arabidopsis* genome generated under non-selective conditions. *Plant J.* 51, 779–791. doi: 10.1111/j.1365-313X.2007.03183.x
- Kim, Y., Schumaker, K. S., and Zhu, J. K. (2006). EMS mutagenesis of *Arabidopsis*. *Methods Mol. Biol.* 323, 101–103.
- Kohl, K. I., and Gremmels, J. (2010). Documentation system for plant transformation service and research. *Plant Methods* 6:4. doi: 10.1186/1746-4811-6-4
- Kohli, A., Melendi, P. G., Abrantes, R., Capell, T., Stoger, E., and Christou, P. (2006). The quest to understand the basis and mechanisms that control expression of introduced transgenes in crop plants. *Plant Signal. Behav.* 1, 185–195. doi: 10.4161/psb.1.4.3195
- Lee, J. M., Davenport, G. F., Marshall, D., Ellis, T. H., Ambrose, M. J., Dicks, J., et al. (2005). GERMINATE. A generic database for integrating genotypic and phenotypic information for plant genetic resource collections. *Plant Physiol.* 139, 619–631. doi: 10.1104/pp.105.065201
- Li, X., Lassner, M., and Zhang, Y. L. (2002). Deleteagene: a fast neutron deletion mutagenesis-based gene knockout system for plants. *Comp. Funct. Genomics* 3, 158–160. doi: 10.1002/cfg.148
- Love, C. G., Andongabo, A. E., Wang, J., Carion, P. W., Rawlings, C. J., and King, G. J. (2012). InterStoreDB: a generic integration resource for genetic and genomic data. *J. Integr. Plant Biol.* 54, 345–355. doi: 10.1111/j.1744-7909.2012.01120.x
- Milc, J., Sala, A., Bergamaschi, S., and Pecchioni, N. (2011). A genotypic and phenotypic information source for marker-assisted selection of cereals: the CEREALAB database. *Database (Oxford)* 2011:baq038. doi: 10.1093/database/baq038
- Mlotshwa, S., Pruss, G. J., Gao, Z., Mgutshini, N. L., Li, J., Chen, X., et al. (2010). Transcriptional silencing induced by *Arabidopsis* T-DNA mutants is associated with 35S promoter siRNAs and requires genes involved in siRNA-mediated chromatin silencing. *Plant J.* 64, 699–704. doi: 10.1111/j.1365-313X.2010.04358.x
- Mutterer, J., and Zinck, E. (2013). Quick-and-clean article figures with FigureJ. *J. Microsc.* 252, 89–91. doi: 10.1111/jmi.12069
- Nieuwland, J., Sornay, E., Marchbank, A., de Graaf, B. H., and Murray, J. A. (2012). Phytotacker, an information management system for easy recording and tracking of plants, seeds and plasmids. *Plant Methods* 8:43. doi: 10.1186/1746-4811-8-43
- Ortega, D., Raynal, M., Laudie, M., Llauro, C., Cooke, R., Devic, M., et al. (2002). Flanking sequence tags in *Arabidopsis thaliana* T-DNA insertion lines: a pilot study. *C. R. Biol.* 325, 773–780. doi: 10.1016/S1631-0691(02)01490-7
- Scott, R., Clark, C., Mathews, A., Yallop, S., Davey, M. R., Lazzeri, P., et al. (2003). Transgenic Plant Monitor (TPM): a relational database for the management and analysis of information in plant tissue culture and transformation laboratories. *Mol. Breed.* 11, 121–125. doi: 10.1023/a:1022406831709
- Svistoonoff, S., Creff, A., Reymond, M., Sigoillot-Claude, C., Ricaud, L., Blanchet, A., et al. (2007). Root tip contact with low-phosphate media reprograms plant root architecture. *Nat. Genet.* 39, 792–796. doi: 10.1038/ng2041
- Wegmüller, S., Svistoonoff, S., Reinhardt, D., Stuurman, J., Amrhein, N., and Bucher, M. (2008). A transgenic dTph1 insertional mutagenesis system for forward genetics in mycorrhizal phosphate transport of *Petunia*. *Plant J.* 54, 1115–1127. doi: 10.1111/j.1365-313X.2008.03474.x
- Yagi, Y., Nakamura, T., and Small, I. (2014). The potential for manipulating RNA with pentatricopeptide repeat proteins. *Plant J.* 78, 772–782. doi: 10.1111/tpj.12377

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00013/full#supplementary-material>



# Estimation of Wheat Plant Density at Early Stages Using High Resolution Imagery

**Shouyang Liu<sup>1\*</sup>, Fred Baret<sup>1</sup>, Bruno Andrieu<sup>2</sup>, Philippe Burger<sup>3</sup> and Matthieu Hemmerlé<sup>4</sup>**

<sup>1</sup> UMR EMMAH, INRA, UAPV, Avignon, France, <sup>2</sup> UMR ECOSYS, INRA, AgroParisTech, Université Paris-Saclay, Thiverval-Grignon, France, <sup>3</sup> UMR AGIR, INRA, Castanet Tolosan, France, <sup>4</sup> Hi-Phen, Avignon, France

Crop density is a key agronomical trait used to manage wheat crops and estimate yield. Visual counting of plants in the field is currently the most common method used. However, it is tedious and time consuming. The main objective of this work is to develop a machine vision based method to automate the density survey of wheat at early stages. RGB images taken with a high resolution RGB camera are classified to identify the green pixels corresponding to the plants. Crop rows are extracted and the connected components (objects) are identified. A neural network is then trained to estimate the number of plants in the objects using the object features. The method was evaluated over three experiments showing contrasted conditions with sowing densities ranging from 100 to 600 seeds·m<sup>-2</sup>. Results demonstrate that the density is accurately estimated with an average relative error of 12%. The pipeline developed here provides an efficient and accurate estimate of wheat plant density at early stages.

**Keywords:** plant density, RGB imagery, neural network, wheat, recursive feature elimination, Hough transform

## OPEN ACCESS

### Edited by:

John Doonan,  
Aberystwyth University, UK

### Reviewed by:

Eric R. Casella,  
Forest Research Agency (Forestry  
Commission), UK

Julia Christine Meitz-Hopkins,  
Stellenbosch University, South Africa

Ankush Prashar,  
Newcastle University, UK

### \*Correspondence:

Shouyang Liu  
shouyang.liu@inra.fr

### Specialty section:

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

**Received:** 20 September 2016

**Accepted:** 20 April 2017

**Published:** 16 May 2017

### Citation:

Liu S, Baret F, Andrieu B, Burger P  
and Hemmerlé M (2017) Estimation  
of Wheat Plant Density at Early  
Stages Using High Resolution  
Imagery. *Front. Plant Sci.* 8:739.  
doi: 10.3389/fpls.2017.00739

## INTRODUCTION

Wheat is one of the main crops cultivated around the world with sowing density usually ranging from 150 to 400 seed·m<sup>-2</sup>. Plant population density may significantly impact the competition among plants as well as with weeds and consequently affect the effective utilization of available resources including light, water, and nutrients (Shrestha and Steward, 2003; Olsen et al., 2006). Crop density appears therefore as one of the important variables that drive the potential yield. This explains why this information is often used for the management of cultural practices (Godwin and Miller, 2003). Plant population density is still investigated most of the time by visually counting the plants in the field over samples corresponding either to a quadrat or to a segment. This is achieved at the stage when the majority of plants have just emerged and before the beginning of tillering (Norman, 1995) which happens few days to few weeks after emergence. This method is time and labor intensive and may be prone to human error.

Some efforts have been dedicated to the development of high-throughput methods for quantifying plant density. This was mainly applied to maize using either capacitive sensors during the harvest (Nichols, 2000; Li et al., 2009) or optical sensors including 2D cameras (Shrestha and Steward, 2003, 2005; Tang and Tian, 2008a,b) and range sensors (Jin and Tang, 2009; Nakarmi and Tang, 2012, 2014; Shi et al., 2013, 2015). However, quantifying the population density of maize is much simpler than that of wheat since maize plants are normally bigger, with larger plant spacing along the row and more evenly distributed. In wheat crops, leaves between neighboring plants overlap rapidly, and tillers will also appear, making the plant identification very difficult

when they have more than three leaves, even using visual counting in the field. Most studies during these early stages report results derived from estimates of the vegetation fraction coverage measured using high resolution imagery (Guo et al., 2013) or based on vegetation indices computed with either multispectral (Sankaran et al., 2015) or hyperspectral (Liu et al., 2008) reflectance measurements. However, none of these investigations specifically addressed the estimation of plant density. Advances in digital photography providing very high resolution images, combined with the development of computer vision systems, offer new opportunities to develop a non-destructive high-throughput method for plant density estimation.

The objective of this study is to develop a system based on high resolution imagery that measures wheat plant population density at early stages. The methods used to acquire the RGB images and the experimental conditions are first presented. Then the pipeline developed to process the images is described. Finally, the method is evaluated with emphasis on its accuracy and on its corresponding domain of validity.

## MATERIALS AND METHODS

### Field Experiments and Measurements

Three experiments were conducted in 2014 in France (**Table 1**): Avignon, Toulouse, and Paris. In Avignon, four sowing densities (100, 200, 300, and 400 seeds·m<sup>-2</sup>) with the same “Apache” cultivar were sampled. In Toulouse, five plant densities (100, 200, 300, 400, and 600 seeds·m<sup>-2</sup>) with two different cultivars, “Apache” and “Caphorn” were considered. In Paris, two cultivars with a single sowing density of 150 seeds·m<sup>-2</sup> were sampled. All measurements were taken around 1.5 Haun stage, when most plants already emerged. A total of 16 plots were therefore available over the three experiments under contrasted conditions in terms of soil, climate, cultivars, and sowing densities. All the plots were at least 10 m length by 2 m width.

In Toulouse and Avignon, images were acquired using an RGB camera fixed on a light moving platform, termed Phenotypette (**Figure 1**). The platform was driven manually at about 0.5 m·s<sup>-1</sup>. For each plot, at least 10 images were collected to be representative of the population. For Paris experiment, the camera was mounted on a monopod to take two pictures with no overlap. In all the cases, the camera was oriented at 45° inclination perpendicular to the row direction and was pointing at the center row from a distance of 1.5 m and with spatial resolution around 0.2 mm (**Figure 1** and **Table 1**). For each plot, 10 images were selected randomly among the whole set of images acquired. The number of plants located in the two central rows was then visually counted over each of the 10 selected images to derive the reference plant density.

### Image Processing

Each image was processed using the pipeline sketched on **Figure 2**. It was mainly programmed using MATLAB and Image Processing Toolbox R2016a (code available on request). To facilitate the application, the corresponding MATLAB functions used are also given in the text.

### Classification of Green Elements

The images display green pixels corresponding to the emerged plants, and brown pixels corresponding to the soil background. The RGB color space was firstly transformed into Lab, to enhance its sensitivity to variations in greenness (Philipp and Rath, 2002). The Otsu automatic thresholding method (Otsu, 1975) was then applied to channel ‘a’ to separate the green from the background pixels (function: *graythresh*). Results show that the proposed method performs well (**Figure 2b**) under the contrasted illumination conditions experienced (**Table 1**). Further, this approach provides a better identification of the green pixels (results not presented for the sake of brevity) as compared to the use of supervised methods (Guo et al., 2013) based on indices such as the excess green (Woebbecke et al., 1995) or more sophisticated indices proposed by Meyer and Neto (2008).

### Geometric Transformation

The perspective effect creates a variation of the spatial resolution within the image: objects close to the lens appear large while distant objects appear small. A transformation was therefore applied to remap the image into an orthoimage where the spatial resolution remains constant. The transformation matrix was calibrated using an image of a chessboard for each camera setup (**Figure 2c**). The chessboard covered the portion of the image that was later used for plant counting. The corners of the squares in the chessboard were identified automatically (function: *detectCheckerboardPoints*). Then the transformation matrix can be derived once the actual dimension of the squares of the chessboard is provided (function: *fitgeotrans*) (**Figure 2c**). The transformation matrix was finally applied to the whole image for given camera setup (function: *tformfwd*) (**Figure 2d**). This allows remapping the image into a homogeneously distributed domain on the soil surface.

### Row Identification and Orientation

The plant density measurement for row crops such as wheat is achieved by counting plants over a number of row segments of given length. Row identification is therefore a mandatory step as sketched in **Figure 2e**. Row identification methods have been explored intensively mostly for the automation of robot navigation in field (Vidović et al., 2016). Montalvo et al. (2012) reviewed the existing methods and found that the Hough transform (Slaughter et al., 2008) is one of the most common and reliable methods. It mainly involves computing the co-distribution of the length ( $\rho$ ) and orientation ( $\theta$ ) of the segments defined by two green pixels (**Figure 3A**). The Hough transform detects dominant lines even in the presence of noise or discontinuous rows. The noise could include objects between rows such as weeds or misclassified background pixels such as stones (Marchant, 1996; Rovira-Más et al., 2005). Although the Hough transform is computationally demanding, its application on edge points of the green objects decreases this constraint. Hence, the ‘Canny Edge Detector’ (Canny, 1986) was consequently used to detect edges prior to the application of the Hough transform. The Hough transform was conducted with orientation  $-90^\circ < \theta < 90^\circ$  with  $0.1^\circ$  angular steps and a radius

**TABLE 1 | Characteristics of the three experimental sites.**

Sites	Latitude	Longitude	Cultivars	Sowing density (seeds·m <sup>-2</sup> )	Reference density (plants·m <sup>-2</sup> )	Illumination conditions	Cameras	Resolution	Focal length	Spatial resolution (mm)
Toulouse	43.5°N	1.5°E	Apache	100, 200, 300, 400, 600	106, 187, 231, 350, 525	Diffuse	Sigma SD14	2640 by 1760	50 mm	0.23
			Caphorn	100, 200, 300, 400, 600	118, 206, 250, 387, 431					
Paris	48.8°N	1.9°E	Premio	150	154	Flash	NIKON D5200	4496 by 3000	0.16	0.13
			Attlass	150	182					
Avignon	43.9°N	4.8°E	Apache	100, 200, 300, 400	54, 129, 232, 425	Direct	Sigma SD14	4608 by 3072		

**FIGURE 1 |** The image acquisition in the field with the Phenotypette.

$-3000 < \rho < 3000$  pixels with 1 pixel steps (function: *hough*) (**Figure 3A**).

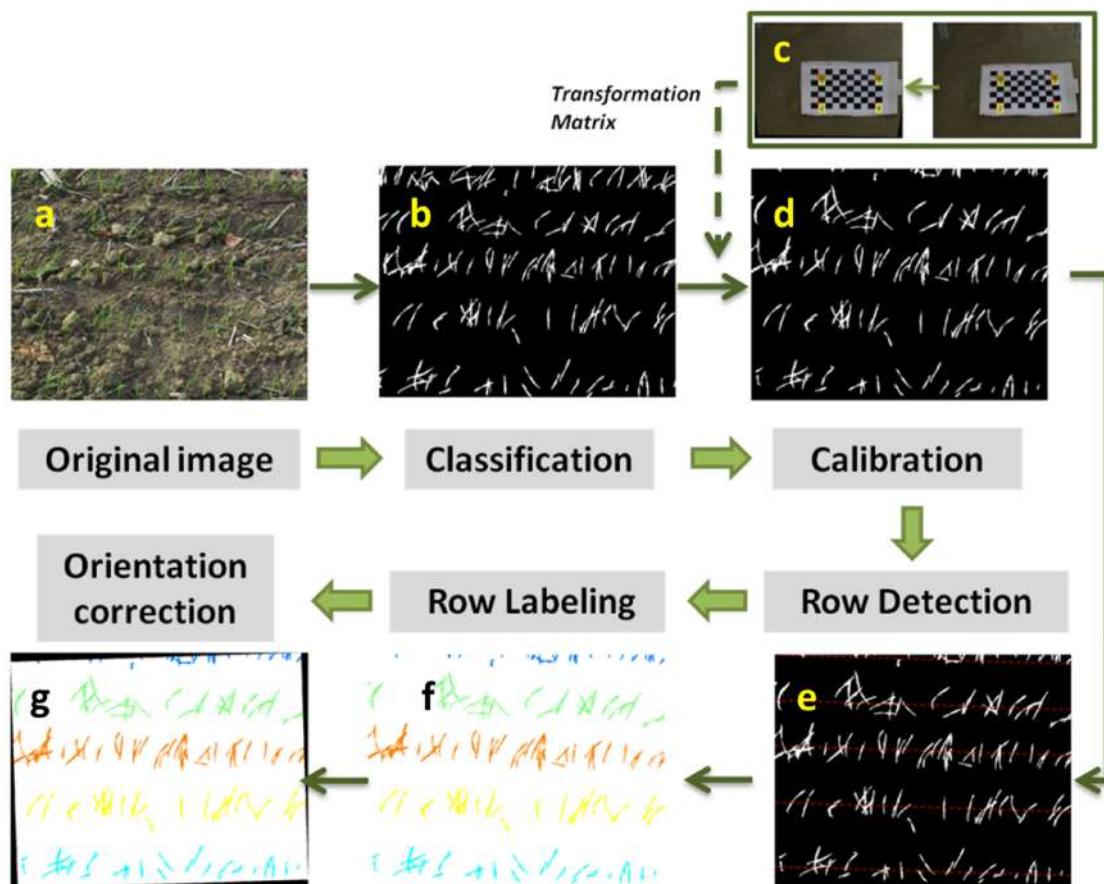
Five main components show up in the image (**Figure 3A**), corresponding to the five rows of the original image (**Figure 2a**). As all rows are expected to be roughly parallel, their orientation could be inferred as the  $\theta$  value,  $\theta_{\text{row}}$  (where  $\theta_{\text{row}} = 90^\circ$  corresponds to the horizontal orientation on the images on **Figure 2f**), that maximizes the variance of  $\rho$ . The positions of the rows are derived from the peaks of frequency for  $\theta = \theta_{\text{row}}$  (**Figure 3B**). Five lines on **Figure 2e** highlight the center of each row. Because of the uncertainty in the orientation of the camera along the row, the row line drawn on the images are not exactly horizontal. This is illustrated in **Figure 2f** where  $\theta_{\text{row}} = -88.2^\circ$ . The images were therefore rotated according to  $\theta_{\text{row}}$  (function: *imrotate*), so that the rows are strictly horizontal in the displayed images **Figure 2g**.

## Object Identification and Feature Extraction

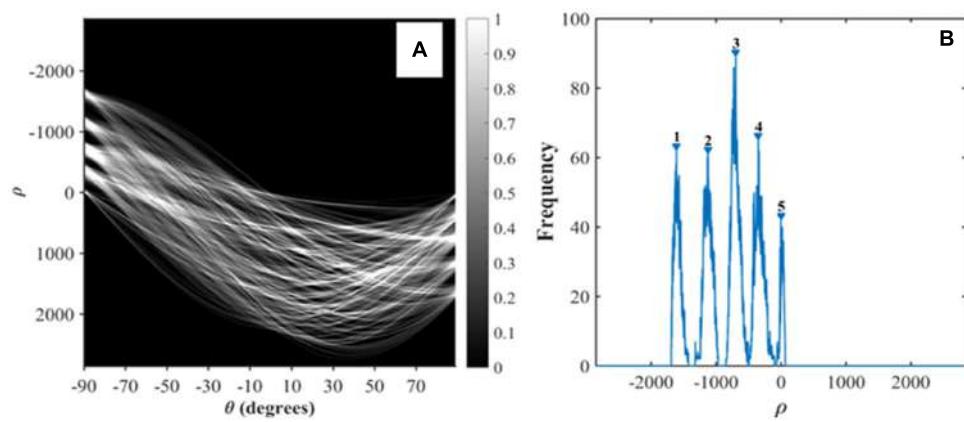
An object in a binary image refers to a set of pixels that form a connected group with the connectivity of eight neighbors. Each object was associated to the closest row line and characterized by 10 main features (function: *bwmorph*) (the top 10 features in **Table 2**). Three additional features were derived from skeletonization of the object: the length, number of branch and end points of the skeleton (function: *regionprops*) (the last three features in **Table 2**). More details on the feature extraction function used can be found in <https://fr.mathworks.com/help/images/>.

## Estimation of the Number of Plants Contained in Each Object

Machine learning methods were used to estimate the number of plants contained in each object from the values of their 13



**FIGURE 2 |** The methodology involving image processing feature extraction. **(a)** Original image. **(b)** Binary image. **(c)** Image of a chessboard to derive the transformation matrix. **(d)** Calibrated image. **(e)** Detecting rows in the image, corresponding to red dashed lines. **(f)** Labeling rows with different colors. **(g)** Correcting row orientation to be horizontal.



**FIGURE 3 |** Hough transform to detect rows. **(A)** Hough transform. **(B)** Identification of the peaks of  $\rho$  corresponding to the rows.

associated features (Table 2). Artificial neural networks (ANNs) have been recognized as one of the most versatile and powerful method to relate a set of variables to one or more variables. ANNs are interconnected neurons characterized by a transfer function.

They combine the input values (the features of the object) to best match the output values (number of plants in our case) over a training database. The training process requires first to define the network architecture (the number of hidden layers and

**TABLE 2 |** The 13 features extracted for each of the connected object.

#	Name	Meaning	Unit
F1	Area	Number of pixels of the connected component (object)	Pixel
F2	FilledArea	Number of pixels of the object with all holes filled	Pixel
F3	ConvexArea	Number of pixels within the associated convex hull	Pixel
F4	Solidity	Ratio of number of pixels in the region to that of the convex hull	Scalar
F5	Extent	Ratio of number of pixels in the region to that of the bounding box	Scalar
F6	EquivDiameter	Diameter of a circle with the same area as the region	Pixel
F7	MajorAxisLength	Length of the major axis of the ellipse equivalent to the region	Pixel
F8	MinorAxisLength	Length of the minor axis of the ellipse equivalent to the region	Pixel
F9	Eccentricity	Eccentricity of the equivalent ellipse to the region	Scalar
F10	Orientation	Orientation of the major axis of the equivalent ellipse	Degree
F11	LengthSkelet	Number of pixels of the skeleton	Pixel
F12	NumEnd	Number of end points of the skeleton	Scalar
F13	NumBranch	Number of branch points of the skeleton	Scalar

nodes per layer and the type of transfer function of each neuron). Then the synaptic weights and biases are tuned to get a good agreement between the number of plants per object estimated from the object's features and the corresponding number of plants per object in the training database. A one-layer feed-forward network with  $k_n$  tangent sigmoid hidden neurons and none linear neuron was used. The number of hidden nodes was varied between  $1 \leq k_n \leq 10$  to select the best architecture. The weights and biases were initialized randomly. The training was achieved independently over each site considering 90% of the data set

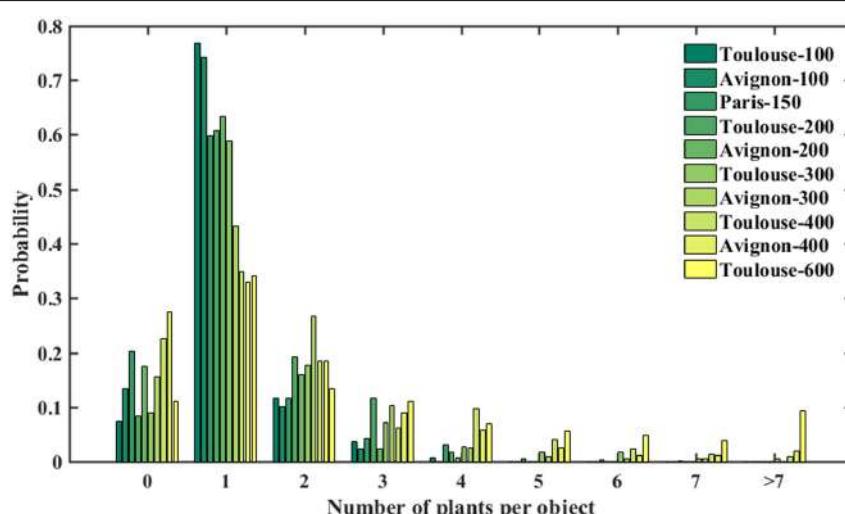
corresponding to a total of the 606 (Toulouse), 347 (Paris), and 476 (Avignon) objects. The remaining 10% objects of each site was used to evaluate the performance of the training. Note that the estimates of number of plants per object were continuous, i.e., representing actually the average probability of getting a discrete number of plants.

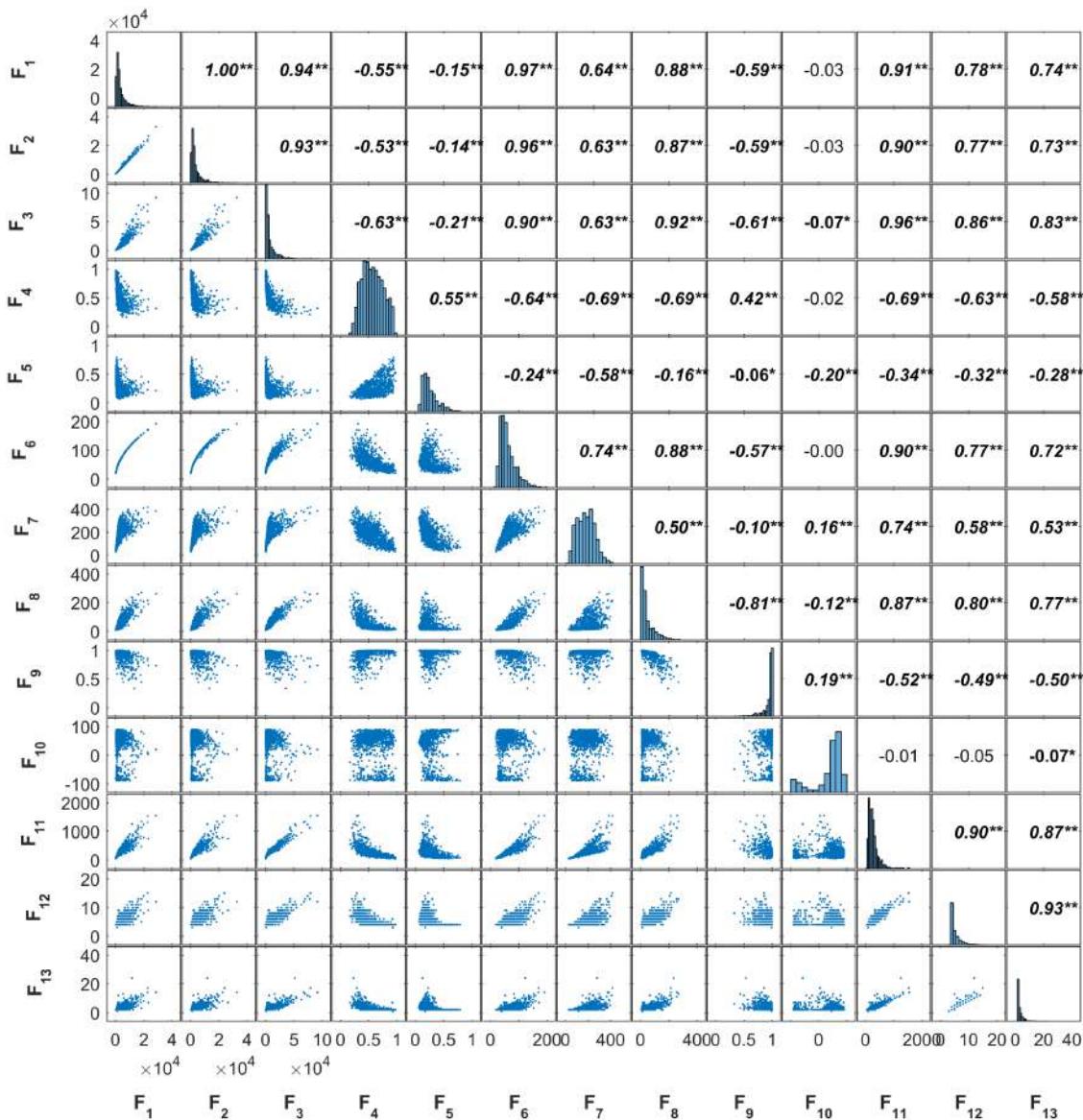
A compact, parsimonious and non-redundant subset of features should contribute to speed up the learning process and improve the generalization of predictive models (Tuv et al., 2009; Kuhn and Johnson, 2013; Louppe, 2014). Guyon et al. (2002) proposed recursive feature elimination (RFE) to select the optimal subset of features. Specific to ANN, the combinations of the absolute values of the weights were used firstly to rank the importance of predictors (features) (Olden and Jackson, 2002; Gevrey et al., 2003). For the subset including  $n$  features, RFE presumes that the subset of the top  $n$  features outperforms the other possible combinations (Guyon et al., 2002; Granitto et al., 2006). Then 13 iterations corresponding to the 13 features need to be computed to select the optimal subset defined as the smallest set providing a  $\text{RMSE}_n$  lower than  $1.02 \text{ RMSE}_{\text{best}}$ , where  $\text{RMSE}_{\text{best}}$  is the minimum RMSE value observed when using the 13 features. To minimize possible overfitting of the training dataset, a cross-validation scheme was used (Seni and Elder, 2010) with the training data set including 90% of the cases and the test data set containing the remaining 10%. The process was repeated five times with a random drawing of the training and test data sets for each trial.

## RESULTS

### Number of Plants per Object and Object Feature Selection

The number of plants per object resulted in a consistently right-skewed distribution over the three experimental sites (Figure 4). For all the plots, objects containing single plants have the highest

**FIGURE 4 |** Number of plants per object over the three sites.



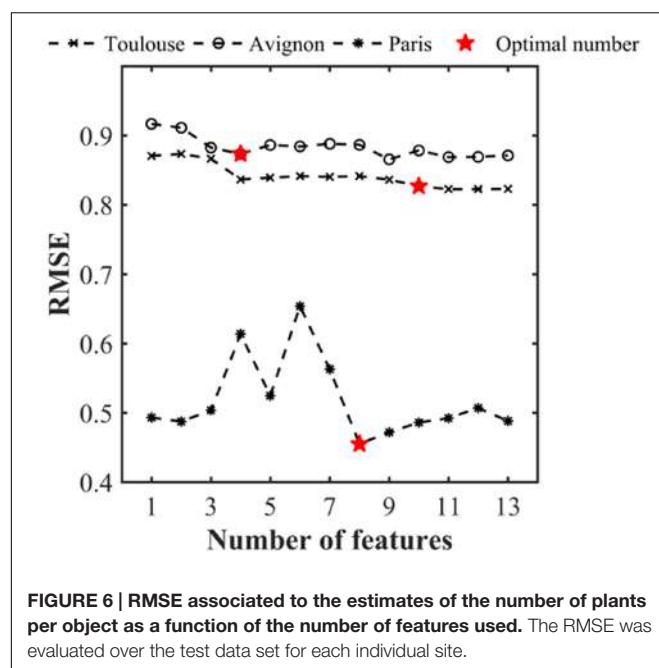
**FIGURE 5 |** The correlation among the 13 objects' features for the Toulouse site (\*\*: 0.01, \*: 0.05). The abbreviation of features refers to names in **Table 2**.

probability of occurrence. However, objects contain generally more plants for high density as compared to the low density conditions. Note that 10–20% of the objects were classified as null, i.e., containing no plants. This corresponds to errors in separating plants from the background: objects such as straw residues, stone, or weeds may show colors difficult to separate in the classification step. Further, due to the variability of the illumination conditions, plants may be misclassified into two disconnected objects. In this case, the larger part is considered as a plant while the smaller remaining part is considered as non-plant, i.e., set to 0.

Most of the 13 features described in **Table 2** are closely related as illustrated by the plot-matrix of the Toulouse site (**Figure 5**). Correlations are particularly high between the four area related

features (F1, F2, F3, F6), between the skeleton derived features (F11, F12, F13), and between the area and skeleton related features. Similar correlations were observed over the Paris and Avignon sites. These strong relationships indicate the presence of redundancy between the 13 features, which may confuse the training of ANN. However, this could be partly overcome by the RFE feature selection algorithm.

The estimation performances of the number of plants per object were evaluated with the RMSE metrics as a function of the number of features used (**Figure 6** and **Table 3**). Note that the RMSE value was calculated based on the visual identification of the number of plants per object in the dataset. **Figure 6** shows that the RMSE decreases consistently when the number of features used increases. However, after using the first four features, the



**FIGURE 6 |** RMSE associated to the estimates of the number of plants per object as a function of the number of features used. The RMSE was evaluated over the test data set for each individual site.

**TABLE 3 |** Performance of the estimation of the number of plants per object over three experiments.

Sites	Training size	n <sub>node</sub>	Number of features	R <sup>2</sup>	RMSE	Bias
Toulouse	606	2	10	0.83	0.83	0.28
Paris	347	2	8	0.79	0.47	0.077
Avignon	476	2	4	0.61	0.87	0.45

improvement in estimation performances is relatively small when including remaining features. The number of features required according to our criterion (1.02. RMSE<sub>best</sub>) varies from 10 (Toulouse) to 4 (Avignon). A more detailed inspection of the main features used across the three sites (**Table 4**) shows the importance of the area related features (F<sub>1</sub>, F<sub>2</sub>, F<sub>3</sub>, F<sub>4</sub>, and F<sub>6</sub>) despite their high inter-correlation (**Figure 5**). The length of the skeleton (F<sub>11</sub>) also appears important particularly for the Avignon site, while the orientation and extent do not help much (**Table 4**).

As expected, the model performs the best for the Paris site (**Table 3**) where the situation is simpler because of the low density inducing limited overlap between plants (**Figure 4**). For sowing density  $\leq 300 \text{ seeds} \cdot \text{m}^{-2}$ , a better accuracy is reached in Toulouse (RMSE = 0.51) and Avignon (RMSE = 0.68) sites. Conversely, the larger number of null objects (**Figure 4**) corresponding to misclassified objects or split plants in the Avignon site, explains the degraded performance (**Table 3**). The bias in the estimation of the number of plants per object appears relatively small, except for the Avignon site. Attention should be paid on the bias since the application of the neural network on a larger number of objects is not likely to improve the estimation of the total number of plants. The bias is mostly due to difficulties associated to the misclassified objects (**Figure 7**). Note that the

**TABLE 4 |** Features selected and the corresponding rank over three sites.

#	Features	Toulouse	Paris	Avignon
F <sub>1</sub>	Area	2	1	3
F <sub>2</sub>	FilledArea	1	3	
F <sub>3</sub>	ConvexArea	4	4	2
F <sub>4</sub>	Solidity	10	7	
F <sub>5</sub>	Extent			
F <sub>6</sub>	EquivDiameter	3	2	4
F <sub>7</sub>	MajorAxisLength			
F <sub>8</sub>	MinorAxisLength	5	5	
F <sub>9</sub>	Eccentricity	8	8	
F <sub>10</sub>	Orientation			
F <sub>11</sub>	LengthSkelet	6	6	1
F <sub>12</sub>	NumEnd	7		
F <sub>13</sub>	NumBranch	9		

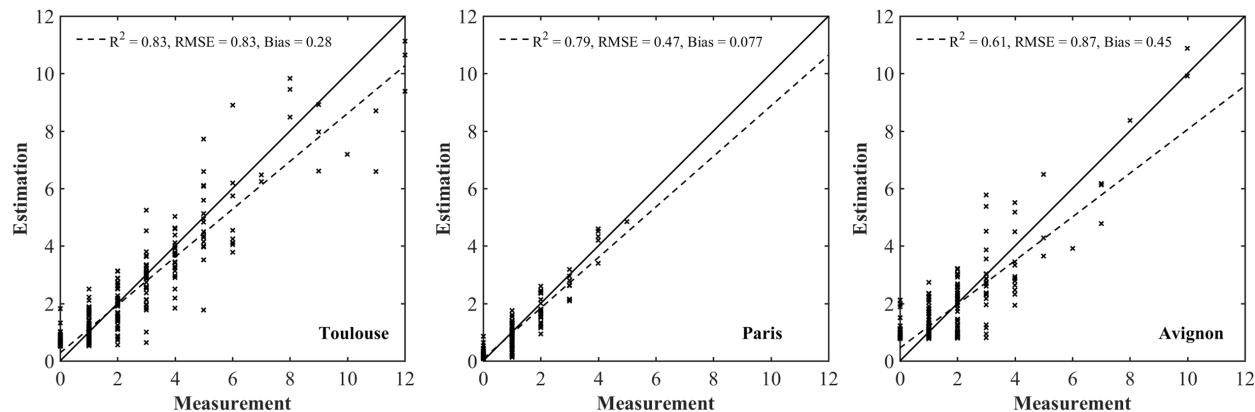
estimation performance degraded for the larger number of plants per objects (**Figure 7**) as a consequence of more ambiguities and smaller samples used in the training process.

## Performance of the Method for Plant Density Estimation

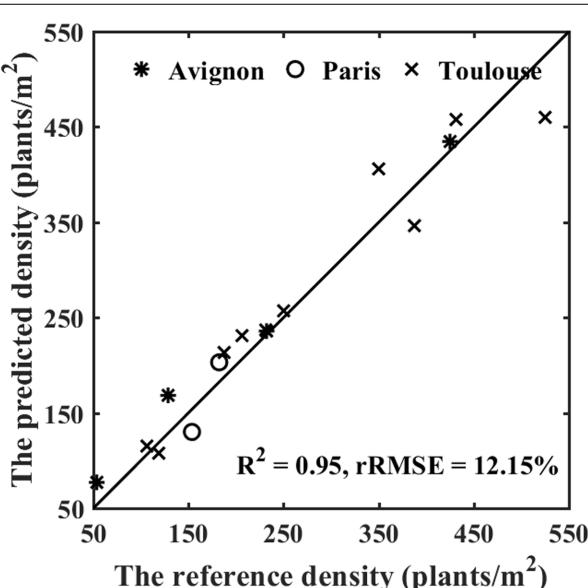
The estimates of plant density were computed by summing the number of plants in all the objects extracted from the row segments identified in the images, divided by the segment area (product of the segment length and the row spacing). The reference density was computed from the visually identified plants. Results show a good agreement between observations and predictions over sowing densities ranging from 100 to 600 plants·m<sup>-2</sup> (**Figure 8**). The performances slightly degrade for densities higher than 350 plants·m<sup>-2</sup>. This may be explained by the difficulty to handle more complex situation when plant spacing decreases, with a higher probability of plant overlap (**Figure 7**). Note that the slight overestimation observed for the low densities in the Avignon site is mainly attributed to the bias in the estimation of the number of plants per object due to the classification problem already outlined.

## DISCUSSION AND CONCLUSION

The method proposed in this study relies on the ability to identify plants or group of plants from RGB images. Image classification is thus a critical step driving the accuracy of the plant density estimation. Wheat plants at emergence have a relatively simple structure and color. The image quality is obviously very important, including the optimal spatial resolution that should be better than 0.4 mm as advised by Jin et al. (2016). Further, the image quality should not be compromised by undesirable effects due to image compression algorithms. As a consequence, when the resolution is between 0.2 and 0.5 mm, it would be preferable to record images in raw format to preserve its quality. A known and fixed white balance should be applied to make the series of images comparable in terms of color. Finally, the view direction was chosen to increase the plant cross section by taking



**FIGURE 7 |** Comparison between the estimated number of plants per object with the value measured over the test dataset for each individual site.



**FIGURE 8 |** Performance of density estimation over the three sites.

images inclined at around 45° zenith angle in a compass direction perpendicular to the row orientation. Note that too inclined views may result in large overlap of plants from adjacent rows which will pose problems for row (and plant) identification.

Plants were separated from the background based on their green color. A unique unsupervised method based on the Lab transform on which automatic thresholding is applied was used with success across a range of illumination conditions. However, the method should be tested under a much larger range of illumination and soil conditions before ensuring that it is actually applicable in all scenarios. Additionally, attention should be paid to weeds that are generally green. Fortunately, weeds were well-controlled in our experiments. Although this is also generally observed during emergence, weed detection algorithm could be integrated in the pipeline in case of significant

infestation. Weeds may be identified by their position relative to the row (Woebbecke et al., 1995). However, for the particular observational configuration proposed (45° perpendicular to the row), the application of these simple algorithms are likely to fail. Additional (vertical) images should be taken, or more refined methods based on the color (Gée et al., 2008) or shape (Swain et al., 2011) should be implemented.

Once the binary images are computed from the original RGB ones, objects containing uncertain number of plants can be easily identified. An ANN method was used in this study to estimate the number of plants from the 13 features of each object. Alternative machine learning techniques were tested including random forest (Breiman, 2001), multilinear regression (Tabachnick et al., 2001) and generalized linear model (Lopatin et al., 2016). The ANN was demonstrated to perform better for the three sites (results not presented in this study for the sake of brevity). The RFE algorithm used to select the minimum subset of features to best estimate the number of plants per object (Granitto et al., 2006) resulted in 4–10 features depending on the data set considered. The features selected are mainly related to the object area and the length of the corresponding skeleton. Conversely, object orientation and extent appear to contribute marginally to the estimation of the number of plants per object. The RFE framework employed here partly accounts for the strong co-dependency between the 13 features considered. The selection process could probably be improved using a recursive scheme similar to the one employed in stepwise regression, or a transformation of the space of the input features.

The wheat population density was estimated with an average of 12% relative error. The error increases with the population density because of the increase of overlap between plants creating larger objects, hence making it more difficult to associate accurately the number of plants they contain. Likewise, a degradation of the performances is also expected when plants are well-developed. Jin and Tang (2009) found that the selection of the optimal growth stage is critical to get accurate estimation of the plant density in maize crops. A timely observation just between Haun stage 1.5–2 corresponding to 1.5–2 phyllochron after emergence appears optimal: plants are enough developed to

be well-identified while the overlap between plants is minimized because of the low number of leaves (between 1 and 2) and their relatively erect orientation. However, in case of heterogeneous emergence, it is frequent to observe a delay of about 1 phyllochron (Jamieson et al., 2008; Hokmalipour, 2011) between the first and the last plant emerged. Observation between Haun stage 1.5 and 2 can thus ensure that the majority has emerged. Since the phyllochron varies between 63 and 150°C·d (McMaster and Wilhem, 1995), the optimal time window of 0.5 phyllochron (between Haun stage 1.5 and 2) can last about 4–8 days under an average 10°C air temperature. This short optimal time window for acquiring the images is thus a strong constraint when operationally deploying the proposed method.

The success of the method relies heavily on the estimation of the number of plants per object. The machine learning technique used in this study was trained independently for each site. This provides the best performances because it takes into account the actual variability of single plant structure that depends on its development stage at the time of observation, on the genotypic variability as well as on possible influence of the environmental conditions, especially wind. Operational deployment of the method therefore requires the model to be re-calibrated over each new experimental site. However, a single training encompassing all the possible situations may be envisioned in near future. This requires a large enough training data set representing the variability of genotypes, development stage and environmental conditions. This single training data base could also include other cereal crop species similar to wheat at emergence such as barley, triticale, or oat.

Several vectors could be used to take the RGB images, depending mostly on the size of the experiment and the

resources available. A monopod and a light rolling platform, the Phenotypette, were used in our study. More sophisticated vectors with higher throughput could be envisioned in the next step, based either on a semi-automatic (Comar et al., 2012) or fully automatic rover (de Solan et al., 2015) or on a UAV platform as recently demonstrated by Jin et al. (2016).

## AUTHOR CONTRIBUTIONS

The experiment and algorithm development were mainly accomplished by SL and FB. SL wrote the manuscript and FB made very significant revisions. BA also read and improved the final manuscript. All authors participated the discussion of experiment design. BA, PB, and MH significantly contributed to the field experiment in Paris, Toulouse, and Avignon, respectively.

## FUNDING

This study was supported by “Programme d’investissement d’Avenir” PHENOME (ANR-11-INBS-012) and Breedwheat (ANR-10-BTR-03) with participation of France Agrimer and “Fonds de Soutien à l’Obtention Végétale”. The grant of the principal author was funded by the Chinese Scholarship Council.

## ACKNOWLEDGMENT

We also thank the people from Paris, Toulouse, and Avignon sites who participated in the field experiments.

## REFERENCES

- Breiman, L. (2001). Random forests. *J. Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 8, 679–698. doi: 10.1109/TPAMI.1986.4767851
- Comar, A., Burger, P., de Solan, B., Baret, F., Daumard, F., and Hanocq, J.-F. (2012). A semi-automatic system for high throughput phenotyping wheat cultivars in-field conditions: description and first results. *Funct. Plant Biol.* 39, 914–924. doi: 10.1071/FP12065
- de Solan, B., Baret, F., Thomas, S., Beauchêne, K., Comar, A., Fournier, A., et al. (2015). “PHENOMOBILE V: a fully automated high throughput phenotyping system,” in *Poster at the Eucarpia*, Montpellier.
- Gée, C., Bossu, J., Jones, G., and Truchetet, F. (2008). Crop/weed discrimination in perspective agronomic images. *Comput. Electron. Agric.* 60, 49–59. doi: 10.1016/j.compag.2007.06.003
- Gevrey, M., Dimopoulos, I., and Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Model.* 160, 249–264. doi: 10.1016/S0304-3800(02)00257-0
- Godwin, R., and Miller, P. (2003). A review of the technologies for mapping within-field variability. *Biosyst. Eng.* 84, 393–407. doi: 10.1016/S1537-5110(02)00283-0
- Granitto, P. M., Furlanello, C., Biasioli, F., and Gasperi, F. (2006). Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometr. Intell. Lab. Syst.* 83, 83–90. doi: 10.1016/j.chemolab.2006.01.007
- Guo, W., Rage, U. K., and Ninomiya, S. (2013). Illumination invariant segmentation of vegetation for time series wheat images based on decision tree model. *Comput. Electron. Agric.* 96, 58–66. doi: 10.1016/j.compag.2013.04.010
- Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422. doi: 10.1023/A:1012487302797
- Hokmalipour, S. (2011). The study of phyllochron and leaf appearance rate in three cultivar of maize (*Zea mays* L.) At Nitrogen Fertilizer Levels. *World Appl. Sci. J.* 12, 850–856.
- Jamieson, P., Brooking, I., Zyskowski, R., and Munro, C. (2008). The vexatious problem of the variation of the phyllochron in wheat. *Field Crops Res.* 108, 163–168. doi: 10.1016/j.fcr.2008.04.011
- Jin, J., and Tang, L. (2009). Corn plant sensing using real-time stereo vision. *J. Field Robot.* 26, 591–608. doi: 10.1002/rob.20293
- Jin, X., Liu, S., Baret, F., Hemerlé, M., and Comar, A. (2016). Estimates of plant density from images acquired from UAV over wheat crops at emergence. *Remote Sens. Environ.*
- Kuhn, M., and Johnson, K. (2013). *Applied Predictive Modeling*. Berlin: Springer. doi: 10.1007/978-1-4614-6849-3
- Li, H., Worley, S., and Wilkerson, J. (2009). Design and optimization of a biomass proximity sensor. *Trans. ASABE* 52, 1441–1452. doi: 10.13031/2013.29124
- Liu, J., Miller, J. R., Haboudane, D., Pattey, E., and Hochheim, K. (2008). Crop fraction estimation from casi hyperspectral data using linear spectral unmixing and vegetation indices. *Can. J. Remote Sens.* 34, S124–S138. doi: 10.5589/m07-062
- Lopatin, J., Dolos, K., Hernández, H., Galleguillos, M., and Fassnacht, F. (2016). Comparing generalized linear models and random forest to model vascular

- plant species richness using LiDAR data in a natural forest in central Chile. *Remote Sens. Environ.* 173, 200–210. doi: 10.1016/j.rse.2015.11.029
- Louppe, G. (2014). Understanding random forests: from theory to practice. *arXiv* 1407, 7502.
- Marchant, J. (1996). Tracking of row structure in three crops using image analysis. *Comput. Electron. Agric.* 15, 161–179. doi: 10.1016/0168-1699(96)00014-2
- McMaster, G. S., and Wilhem, W. W. (1995). Accuracy of equations predicting the phyllochron of wheat. *Crop Sci.* 35, 30–36. doi: 10.2135/cropsci1995.0011183X003500010006x
- Meyer, G. E., and Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Comput. Electron. Agric.* 63, 282–293. doi: 10.1016/j.compag.2008.03.009
- Montalvo, M., Pajares, G., Guerrero, J. M., Romeo, J., Guijarro, M., Ribeiro, A., et al. (2012). Automatic detection of crop rows in maize fields with high weeds pressure. *Expert Syst. Appl.* 39, 11889–11897. doi: 10.1016/j.eswa.2012.02.117
- Nakarmi, A., and Tang, L. (2012). Automatic inter-plant spacing sensing at early growth stages using a 3D vision sensor. *Comput. Electron. Agric.* 82, 23–31. doi: 10.1016/j.compag.2011.12.011
- Nakarmi, A. D., and Tang, L. (2014). Within-row spacing sensing of maize plants using 3D computer vision. *Biosyst. Eng.* 125, 54–64. doi: 10.1016/j.biosystemseng.2014.07.001
- Nichols, S. W. (2000). Method and Apparatus for Counting Crops. US 6073427.
- Norman, D. W. (1995). *The Farming Systems Approach to Development and Appropriate Technology Generation*. Rome: Food Agriculture Organization.
- Olden, J. D., and Jackson, D. A. (2002). Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecol. Model.* 154, 135–150. doi: 10.1016/S0304-3800(02)00064-9
- Olsen, J., Kristensen, L., and Weiner, J. (2006). Influence of sowing density and spatial pattern of spring wheat (*Triticum aestivum*) on the suppression of different weed species. *Weed Biol. Manag.* 6, 165–173. doi: 10.1111/j.1445-6664.2006.00210.x
- Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica* 11, 23–27.
- Philipp, I., and Rath, T. (2002). Improving plant discrimination in image processing by use of different colour space transformations. *Comput. Electron. Agric.* 35, 1–15. doi: 10.1016/S0168-1699(02)00050-9
- Rovira-Más, F., Zhang, Q., Reid, J., and Will, J. (2005). Hough-transform-based vision algorithm for crop row detection of an automated agricultural vehicle. *J. Automobile Eng.* 219, 999–1010. doi: 10.1243/095440705X34667
- Sankaran, S., Khot, L. R., and Carter, A. H. (2015). Field-based crop phenotyping: multispectral aerial imaging for evaluation of winter wheat emergence and spring stand. *Comput. Electron. Agric.* 118, 372–379. doi: 10.1016/j.compag.2015.09.001
- Seni, G., and Elder, J. F. (2010). Ensemble methods in data mining: improving accuracy through combining predictions. *Synth. Lect. Data Mining Knowl. Discov.* 2, 1–126. doi: 10.1186/1471-2105-14-206
- Shi, Y., Wang, N., Taylor, R., and Raun, W. (2015). Improvement of a ground-LiDAR-based corn plant population and spacing measurement system. *Comput. Electron. Agric.* 112, 92–101. doi: 10.1016/j.compag.2014.11.026
- Shi, Y., Wang, N., Taylor, R. K., Raun, W. R., and Hardin, J. A. (2013). Automatic corn plant location and spacing measurement using laser line-scan technique. *Precis. Agric.* 14, 478–494. doi: 10.1007/s11119-013-9311-z
- Shrestha, D. S., and Steward, B. L. (2003). Automatic corn plant population measurement using machine vision. *Trans. ASAE* 46, 559. doi: 10.13031/2013.12945
- Shrestha, D. S., and Steward, B. L. (2005). Shape and size analysis of corn plant canopies for plant population and spacing sensing. *Appl. Eng. Agric.* 21, 295–303. doi: 10.13031/2013.18144
- Slaughter, D., Giles, D., and Downey, D. (2008). Autonomous robotic weed control systems: a review. *Comput. Electron. Agric.* 61, 63–78. doi: 10.1016/j.compag.2007.05.008
- Swain, K. C., Nørremark, M., Jørgensen, R. N., Midtiby, H. S., and Green, O. (2011). Weed identification using an automated active shape matching (AASM) technique. *Biosyst. Eng.* 110, 450–457. doi: 10.1016/j.biosystemseng.2011.09.011
- Tabachnick, B. G., Fidell, L. S., and Osterlind, S. J. (2001). *Using Multivariate Statistics*. Glenview, IL: HarperCollins College Publishers.
- Tang, L., and Tian, L. F. (2008a). Plant identification in mosaicked crop row images for automatic emerged corn plant spacing measurement. *Trans. ASABE* 51, 2181–2191. doi: 10.13031/2013.25381
- Tang, L., and Tian, L. F. (2008b). Real-time crop row image reconstruction for automatic emerged corn plant spacing measurement. *Trans. ASABE* 51, 1079–1087. doi: 10.13031/2013.24510
- Tuv, E., Borisov, A., Runger, G., and Torkkola, K. (2009). Feature selection with ensembles, artificial variables, and redundancy elimination. *J. Mach. Learn. Res.* 10, 1341–1366.
- Vidović, I., Cupec, R., and Hocenski, Ž (2016). Crop row detection by global energy minimization. *Pattern Recognit.* 55, 68–86. doi: 10.1016/j.patcog.2016.01.013
- Woebbecke, D., Meyer, G., Von Bargen, K., and Mortensen, D. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Trans. ASAE* 38, 259–269. doi: 10.13031/2013.27838

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

Copyright © 2017 Liu, Baret, Andrieu, Burger and Hemmerlé. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Evaluation of the SeedCounter, A Mobile Application for Grain Phenotyping

Evgenii Komyshev<sup>1</sup>, Mikhail Genaev<sup>2</sup> and Dmitry Afonnikov<sup>1,2\*</sup>

<sup>1</sup> Laboratory of Evolutionary Bioinformatics and Theoretical Genetics, Department of Systems Biology, Institute of Cytology and Genetics Siberian Branch of Russian Academy of Sciences (SB RAS), Novosibirsk, Russia, <sup>2</sup> Chair of Informational Biology, Novosibirsk State University, Novosibirsk, Russia

## OPEN ACCESS

**Edited by:**

Marcos Egea-Cortines,  
Universidad Politécnica de Cartagena,  
Spain

**Reviewed by:**

Scot Nelson,  
University of Hawaii at Manoa, USA  
Konstantin Kozlov,

Peter the Great St. Petersburg  
Polytechnic University, Russia

**\*Correspondence:**

Dmitry Afonnikov  
ada@bionet.nsc.ru

**Specialty section:**

This article was submitted to  
Technical Advances in Plant Science,  
a section of the journal  
*Frontiers in Plant Science*

**Received:** 29 September 2016

**Accepted:** 15 December 2016

**Published:** 04 January 2017

**Citation:**

Komyshev E, Genaev M and  
Afonnikov D (2017) Evaluation of the  
SeedCounter, A Mobile Application  
for Grain Phenotyping.  
*Front. Plant Sci.* 7:1990.  
doi: 10.3389/fpls.2016.01990

Grain morphometry in cereals is an important step in selecting new high-yielding plants. Manual assessment of parameters such as the number of grains per ear and grain size is laborious. One solution to this problem is image-based analysis that can be performed using a desktop PC. Furthermore, the effectiveness of analysis performed in the field can be improved through the use of mobile devices. In this paper, we propose a method for the automated evaluation of phenotypic parameters of grains using mobile devices running the Android operational system. The experimental results show that this approach is efficient and sufficiently accurate for the large-scale analysis of phenotypic characteristics in wheat grains. Evaluation of our application under six different lighting conditions and three mobile devices demonstrated that the lighting of the paper has significant influence on the accuracy of our method, unlike the smartphone type.

**Keywords:** wheat grain, phenotyping, computer image analysis, mobile devices, Android

## INTRODUCTION

The grains per ear and grain size are important characteristics of cereal yield. Seed counting and morphometry “by eye” is laborious. Therefore, various approaches have been suggested for efficient grain morphometry using image processing techniques (Granitto et al., 2005; Pourreza et al., 2012; Tanabata et al., 2012). Most of these approaches were implemented using desktop PC software for analyzing grain images on a light background obtained using either a digital camera or a scanner (Herridge et al., 2011; Tanabata et al., 2012; Whan et al., 2014). These approaches allow users to estimate a large number of grain morphometric parameters describing shape and color (Bai et al., 2013). They also facilitate methods for identifying the cereal variety using grain images (Wiesnerová and Wiesner, 2008; Chen et al., 2010; Zapotoczny, 2011), determining seed moisture content and predicting semolina yield in durum wheat (Novaro et al., 2001; Tahir et al., 2007). Duan et al. (2011) developed a labor-free engineering solution for high throughput automatic analysis of rice yield-related traits including the number of total spikelets, the number of filled spikelets, the 1000-grain weight, the grain length, and the grain width. Roussel et al. (2016) proposed a detailed analysis of seed shape and size. They used 3D surface reconstruction from the silhouettes of several images obtained by rotation of a seed in front of a digital camera. This method was implemented further in the *phenoSeeder* robotic platform (Jahnke et al., 2016), which was designed for the high-quality measurement of basic seed biometric traits and mass from which seed density

is calculated. Strange et al. (2015) used X-ray computed tomography for the *in situ* determination of grain shape. The engineering facilities for grain morphometry demonstrate high performance and precision; however, they are installed in a limited number of plant research laboratories. There is still a need for low cost, high-throughput methods of grain analysis (Whan et al., 2014).

Large-scale breeding experiments require processing substantial phenotypic data, often in field conditions and thus without access to desktop computers and scanners. In this case, a digital camera is a viable option, but the images must be subsequently copied to a laptop or PC.

Modern mobile devices (smartphones and Internet tablets) contain digital cameras with high resolution. Mobile devices have multicore processors with sufficient computational power for image processing and analysis. These features allow users to capture and process images wherever necessary. A number of applications for mobile devices have been developed for the morphometry of plant organs. Leafsnap (Kumar et al., 2012) is able to identify plant species in real time based on their leaf images: a user takes pictures of a plant leaf using a mobile device and sends the images from the camera to a remote server where they are processed. Leaf Doctor (Pethybridge and Nelson, 2015) is another mobile application that estimates the percentage of disease severity based on leaf images in a semi-automated manner. Mobile devices can also serve as efficient tools to estimate soil-color (Gómez-Robledo et al., 2013).

In this work, we present a mobile application, SeedCounter, for the Android platform that performs automated calculation of morphological parameters of wheat grains using mobile devices in field conditions (without computer facilities). The application estimates the number of grains scattered on a sheet of A4, Letter, Legal, A3, A4, A5, B4, B5, or B6 paper and morphological parameters such as length, width, area, and distance between the geometric center of mass of the grain and the point of intersection of its principal axes.

We conducted several seed counting tests under controlled lighting conditions and daylight to estimate software performance. We demonstrated that the SeedCounter can estimate the number of grains in an image and their size with high accuracy, but performance is dependent on lighting conditions.

## MATERIALS AND METHODS

### Getting Images

The program input is a color image of grains placed arbitrarily on a sheet of white paper (A4, Letter, Legal, A3, A5, B4, B5, or B6). We recommend minimizing any contact between grains. To reduce errors, users should provide the following conditions for image capture: the paper sheet should be placed on a dark background and bright side lighting should be avoided.

The boundaries of the paper sheet on the background should be parallel to the sides of the frame (**Figure 1A**). The fixed size of the paper makes it possible to calculate the scale of the image and evaluate the grain sizes in metric units. The SeedCounter

application receives images directly from the camera of the mobile device.

### Image Processing Algorithm

The algorithm is implemented using the OpenCV image processing library (Howse, 2013; Dawson-Howe, 2014) and consists of several steps.

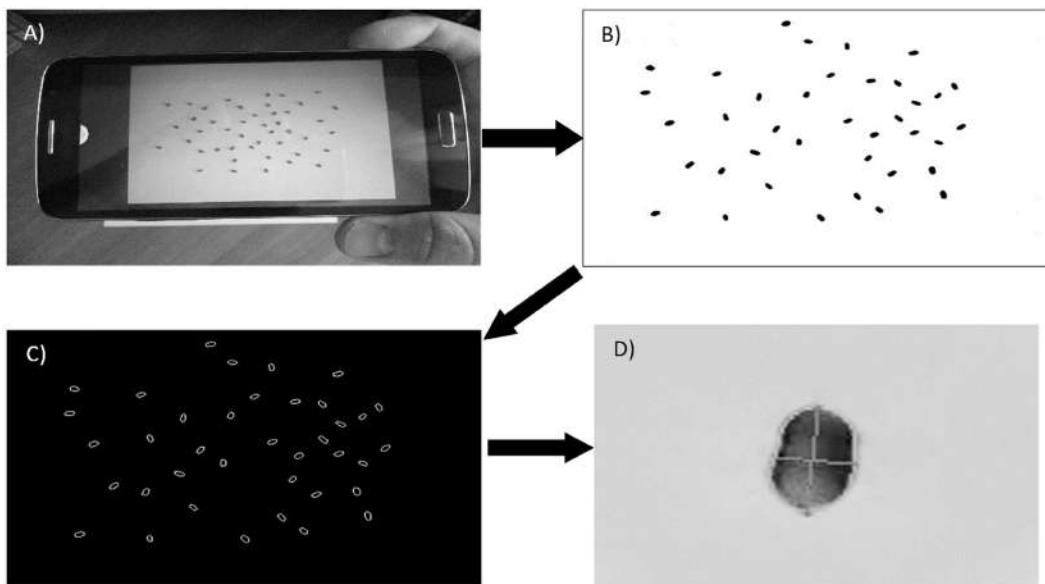
### Paper Sheet Recognition

The paper sheet is recognized as a light area of tetragonal shape on a dark background. For recognition, the original color image (**Figure 1**) is converted to grayscale by the cvtColor() function. To determine the area of the sheet, an adaptive binarization of the entire image is performed by the adaptiveThreshold() function, and the canny() function is used for paper boundary detection. The set of lines close to the sheet boundaries is generated by the houghLinesP() function with the length parameter varying from 20 to 80% of the respective image side. Due to distortions on the image, not all of these lines for the same side are parallel and lines at the adjacent sides are not perpendicular. Therefore, to select lines approximating the paper boundaries, we cluster them with respect to their mutual angle and distance, yielding four clusters of lines corresponding to the paper sides. For each cluster, we reconstruct a sheet boundary line with the smallest distance from the pixels of the cluster lines. The intersections between the sheet boundary lines determine the vertices of the paper tetragonal image. If the paper shape on the image deviates from rectangular, affine transformations convert it to rectangular. This step is performed using the getPerspectiveTransform() function for transformation matrix calculations, and the warpPerspective() function is used to transform the image, making the opposite edges parallel and all angles equal to 90°.

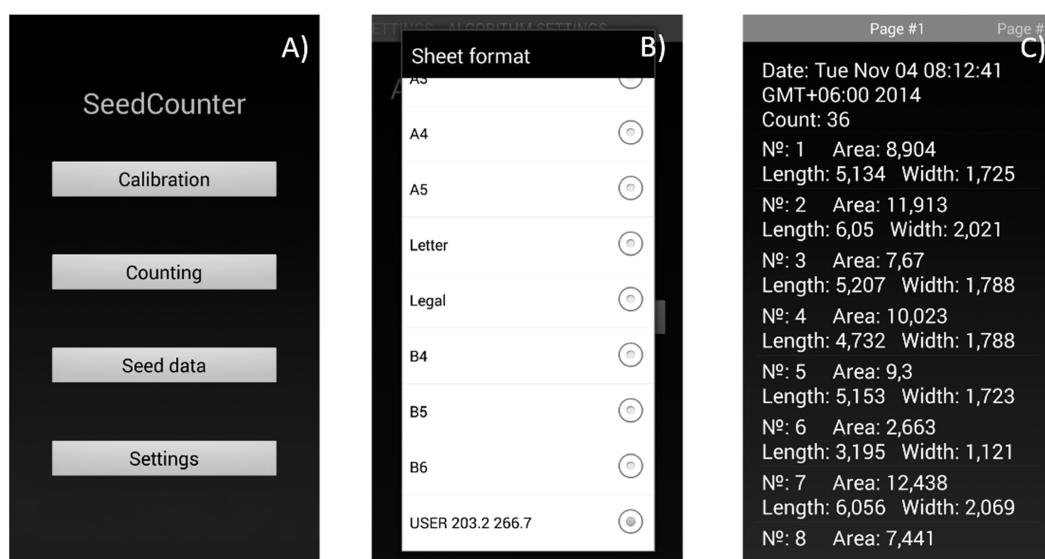
### Grain Identification and Morphometry

Grains are identified as contours by applying the findContours() function to the image fragment corresponding to the paper sheet. We make a further adjustment of the grain boundaries using local Hue Saturation Value (HSV) channel binarization for the neighboring regions of the original image. Local binarization reduces the influence of shadowing during grain boundary determination. It includes converting a local image segment to HSV color space and a subsequent conversion into grayscale based on calibration parameters and color histograms. The resulting channel reflects the degree of conformity of image pixels to the grain color. The local binarization yields more accurate determinations of grain boundaries.

The marked watershed method (Roerdink and Meijster, 2000), as implemented in the watershed() function, is used to resolve the boundaries of seed grains that are in contact with one another. The resulting contours are approximated by grain ellipsoids, allowing for estimates of the size of the major and minor principal axes corresponding to the length and the width of the grain (**Figure 1D**). SeedCounter additionally identifies the grain image area and the distance between the geometric center of mass and the point of intersection of the principal axes.



**FIGURE 1 |** The main steps of grain recognition on a sheet of paper. **(A)** Capturing an image using the camera of the mobile device and paper recognition. **(B)** The image after affine transformation and binarization. **(C)** Grain contours identified on the image. **(D)** Grain image with the major axes shown by crossed lines.



**FIGURE 2 |** The SeedCounter application interface. **(A)** Main menu. **(B)** Selection of the paper size. **(C)** Output screen indicating the results of measurements (grain count and length/width/area for each grain).

## Android Application Interface

The mobile application user can adjust image processing and seed recognition parameters by using the ‘Calibration’ option on the main menu (**Figure 2A**). The user should provide a single seed on the paper, process the image and verify that the algorithm identifies the seed correctly and marks it as a red polygon. The algorithm parameters at this stage are saved automatically. The user can also use the program menu (**Figure 2B**) to define the size of the paper sheet (including user-defined sizes) and the camera

and image resolutions to enable the touching seed separation algorithm and HSV binarization.

Data on the number of counted seeds and seed shape parameters for each seed are stored in XML format and can be displayed using the ‘Seed data’ menu (**Figure 2C**). The user can view the data, delete it, export in tsv format or send it to the SeedCounter web-server. In the last case, the user obtains the data URL that allows the uploading of the data in the web-browser.

## Accuracy Estimation

We considered two types of errors. First, we estimated the accuracy of the grain number identification. Fifty wheat grains of the same variety were poured onto a sheet, and the number of grains was estimated by SeedCounter. After that, one grain was removed from the sheet, the grains were shuffled (no control for the grain separation), and the number of grains was estimated again. This procedure was repeated 40 times. We performed this measurement series using different mobile devices, camera resolutions, and illumination conditions. For each series of grain number estimations, we calculated the mean absolute error (MAE) and the mean absolute percentage error (MAPE) as follows:

$$MAE = \frac{1}{M} \sum_{j=1}^{j=M} |N_j - N'_j|$$

$$MAPE = \frac{100\%}{M} \sum_{j=1}^{j=M} \left( \frac{|N_j - N'_j|}{N_j} \right)$$

where  $j$  is the image number in the experiment,  $N_j$  is the number of grains on the sheet,  $N'_j$  is the number of grains estimated by SeedCounter and  $M = 40$  is the number of images in the experiment. The error in seed grain number estimation increases as the MAE [Eq. (1)] and MAPE [Eq. (2)] values increase. If the MAE and MAPE values are close to 0, the error is low. We additionally estimated the Pearson correlation coefficient,  $r_N$ , between  $N_j$  and  $N'_j$ . The closer  $r_N$  is to unity, the smaller the error in the grain number estimates.

Second, we evaluated the accuracy of the grain length and width estimation. We measured the length and width of 250 grains of five wheat varieties, with each grain placed in a strict order, using a Carl Zeiss Aioscop 2 plus microscope equipped with a digital camera with the AxoCamHRC TV2/3c 0.63 adapter. We placed grains on the paper sheet in the same order and applied the SeedCounter software to estimate their length and width. A series of morphometric measurements of the 250 grains was performed using different mobile devices, camera resolutions and illumination conditions. For each experiment, we calculated MAE [Eq. (1)] separately for length and width and calculated the average values. The same procedure was used to calculate MAPE [Eq. (2)] for the width and length. The Pearson correlation coefficients,  $r_w$  and  $r_l$ , were calculated for these parameters separately.

To compare the accuracy of SeedCounter applications with available software, we compared our results with measurements obtained using the application SmartGrain (Tanabata et al., 2012) running on a personal computer (Intel Core i7, 2400 MHz, 4 Gb RAM) and images from the scanner HP Scanjet 3800 with 600 dpi.

## Experimental Conditions

We evaluated the accuracy of the morphometric parameter estimation of grains using the following three mobile devices running Android OS at maximal camera resolution: the smartphones Samsung Galaxy Grand 2, Sony Ericsson XPERIA

pro mini, and the Internet tablet DNS AirTab m101w. Characteristics for these devices are presented in Supplementary Table S1.

We used the following three types of lighting devices: A 11-W daylight lamp (color temperature 4000 K, luminous flux 900 lm), a 5-W daylight lamp (4000 K, 400 lm), and a 35-W halogen lamp (2700 K, 190 lm). Four types of artificial lighting were used, as follows: a 11-W daylight lamp (L1); a 11-W daylight lamp and two 5-W daylight lamps (L2); a 11-W daylight lamp and four 5-W daylight lamps (L3); and a 11-W daylight lamp, four 5-W daylight lamps, and a halogen lamp (L4). The lamps were set at a height of 60 cm above the sheet of paper. The sheet was placed on a table with a dark top, and the experiments were performed in a dark room. To assess the accuracy of the measurements in the daylight, we also measured the grains without using artificial lighting in cloudy weather indoors and on a clear day outdoors. Details of the experimental conditions are listed in Table 1.

We used two-way ANOVA tests to estimate the influence of device type and lighting conditions on grain number and shape accuracy. We considered device type and lighting to be independent variables and error estimates (MAE and MAPE) to be dependent variables. The Statistica 6.0 software was used to perform this test.

## Wheat Varieties

We used the grains from the following five wheat varieties from the cereal collection of the Chromosome engineering laboratory, Institute of Cytology and Genetics SB RAS: Alen'kaya 1102 II-12, 84/98w 99 II-13, Synthetic 6x x-12, Purple Chance 4480 II-03, and Alcedo n-99. Plants were grown in a field near Novosibirsk in 2014. These varieties have grains with different shapes and sizes. The variety Alcedo is oval in shape and has an average length of ~7 mm and width of ~3.6 mm. The Synthetic variety has an elongated grain shape and an average length of ~8 mm and width of ~2.3 mm. The Alen'kaya variety has smaller dimensions, with an average length of ~5 mm and an average width of ~2.4 mm. The 84/98w and Purple Chance varieties are similar

**TABLE 1 | Light conditions for measuring the accuracy of the wheat grain morphometry.**

Number	Lighting facilities	Luminous flux (lux)	Light temperature
L1	11-W daylight lamp	900 lm	4000K
L2	11-W daylight lamp, 2 × 5-W daylight lamps	1700 lm	4000K
L3	11-W daylight lamp, 4 × 5-W daylight lamps	2500 lm	4000K
L4	11-W daylight lamp, 4 × 5-W daylight lamps, 35-W halogen lamp	2690 lm	4000 and 2700K
L5	Daylight, cloudy day, indoors	(1280 lux)	-
L6	Daylight, sunny day, outdoors	(656000 lux)	-

*Lux units were used to evaluate the light intensity of natural lighting conditions.*

in appearance and have an average length/width of 6.5/2.6 mm and 7/2.9 mm, respectively.

## RESULTS

The SeedCounter mobile application for Android devices is free to download at the Android Play Store<sup>1</sup>). The SeedCounter application requires a minimum of Android API version 15, and Oracle/Sun JDK 6 or 7 is recommended. SeedCounter uses the OpenCV library for image processing. SeedCounter is distributed under the BSD (Berkley Software Distribution) license.

The grain number estimation accuracies for different experiment series are shown in **Table 2**. The table shows that the MAE [Eq. (1)] of the estimate of the number of grains on the sheet is close to 1% and that the MAPE [Eq. (2)] is close to 2%. A more detailed analysis showed that the largest errors in counting the number of grains occur if two or more grains on the paper are in contact and that under poor lighting conditions, the algorithm does not separate most of the grains. If the grains on the sheet are all separated, the seed counting error vanishes.

The accuracy of length and width estimation for the grains by different devices in different conditions is shown in **Table 3**. The table demonstrates that the grain size estimation accuracy was approximately 0.30 mm (average for all series: 0.31 mm) that is approximately 8% of the linear dimensions of the grain (average for all series: 8.03%). The correlation coefficients between the control length and its estimate in all experiments were not lower than 0.79. For the grain width, this parameter

<sup>1</sup><https://play.google.com/store/apps/details?id=org.wheatdb.seedcounter>

**TABLE 2 | Evaluation of the accuracy of wheat grain counting using the SeedCounter mobile application.**

Experiment ID	MAE <sup>a</sup> (mm)	MAPE <sup>a</sup> (%)	r (N <sub>j</sub> , N' <sub>j</sub> ) <sup>a</sup>
Sam_L1	1.425	0.035	0.996
Sam_L2	1.375	0.036	0.994
Sam_L3	0.65	0.015	0.998
Sam_L4	0.975	0.024	0.997
Sam_L5	1.15	0.029	0.992
Sam_L6	0.55	0.017	0.998
Sony_L1	1	0.024	0.995
Sony_L2	0.8	0.019	0.995
Sony_L3	0.675	0.017	0.996
Sony_L4	0.775	0.020	0.997
Sony_L5	0.75	0.018	0.996
Sony_L6	0.775	0.018	0.996
DNS_L1	1.2	0.031	0.997
DNS_L2	0.5	0.012	0.997
DNS_L3	0.125	0.003	0.999
DNS_L4	0.725	0.017	0.998
DNS_L5	1.175	0.030	0.996
DNS_L6	0.775	0.020	0.997

<sup>a</sup>Mean absolute error (MAE), mean absolute percentage error (MAPE), and Pearson correlation coefficient r (N<sub>j</sub>, N'<sub>j</sub>) between the actual number and estimated number of seeds.

**TABLE 3 | The accuracy of estimates of the length and width of wheat grains by SeedCounter mobile application and SmartGrain.**

Experiment ID	MAE <sup>a</sup> (mm)	MAPE <sup>a</sup> (%)	r <sub>l</sub> <sup>a</sup>	r <sub>w</sub> <sup>a</sup>
Sam_L1	0.284	7.453	0.936	0.816
Sam_L2	0.296	7.576	0.928	0.824
Sam_L3	0.283	7.339	0.932	0.822
Sam_L4	0.327	8.306	0.923	0.811
Sam_L5	0.398	9.081	0.797	0.770
Sam_L6	0.349	8.437	0.875	0.769
Sony_L1	0.313	8.277	0.933	0.765
Sony_L2	0.310	8.121	0.931	0.767
Sony_L3	0.298	7.787	0.937	0.777
Sony_L4	0.327	8.418	0.920	0.755
Sony_L5	0.301	7.727	0.913	0.749
Sony_L6	0.338	8.546	0.899	0.730
DNS_L1	0.295	7.852	0.943	0.774
DNS_L2	0.296	7.688	0.935	0.777
DNS_L3	0.287	7.730	0.950	0.779
DNS_L4	0.311	8.229	0.940	0.780
DNS_L5	0.351	8.798	0.890	0.672
DNS_L6	0.346	8.264	0.890	0.787
SmartGrain	0.305	6.973	0.948	0.886

<sup>a</sup>Mean absolute error (MAE), mean absolute percentage error (MAPE) averaged over length and width, and Pearson correlation coefficients for length (r<sub>l</sub>) and width (r<sub>w</sub>) between actual and estimated values.

**TABLE 4 | Significance of the influence of the mobile device type and lighting on errors in estimating grain number and dimensions.**

Error type	Lighting conditions	Device type
Grain counting, MAE	<b>0.004</b>	0.365
Grain counting, MAPE	<b>0.003</b>	0.306
Grain dimensions, MAE	<b>0.036</b>	0.771
Grain dimensions, MAPE	0.094	0.890

ANOVA p-values of two factors are represented. Bold values are significant ( $p < 0.05$ )

was smaller but greater than 0.67. Both correlation coefficients were significant at  $p < 0.01$ . Interestingly, errors for grain length estimates for SeedCounter and SmartGrain are close to each other; however, for grain width SmartGrain demonstrates better performance.

Average values for different devices under the same conditions are shown in Supplementary Table S2. The mobile devices on average demonstrate the best performance in grain size estimation at L3 lighting conditions (two daylight lamps, luminous flux is 2500 lm). The worst performance was obtained at L5 conditions (cloudy day, indoors).

The two-way ANOVA test showed that the lighting conditions significantly influence the estimation of the grain number and the grain length and width (ANOVA  $p$ -value  $< 0.05$ ; **Table 4**). Interestingly, the largest mean MAE [Eq. (1)] for grain counting, 0.458, was obtained for the lighting condition with the lowest luminous flux (L1, 11-W lamp only), whereas the other lighting conditions had lower MAE values: 0.058 for L2, 0.1 for L3,

0.058 for L4, and 0.275 for L5. It should be noted that the seed counting error under conditions without artificial light is smaller than that for the lowest luminous flux but larger than that obtained under all other controlled light conditions. The results shown in **Table 4** demonstrate that device type does not have a significant effect on the grain number/dimension measurements.

**Figure 3** demonstrates the scatterplot of the length and width measurements for 250 seeds obtained by microscope and a Samsung camera using daylight lamps (L3) and sunlight (L6) lighting conditions. This figure demonstrates that with good lighting conditions, the grain size estimates obtained by the mobile device are in agreement with the microscope measurements. However, in sunlight conditions, our software tends to underestimate the grain dimensions for larger grains and overestimate them for smaller grains. This effect is likely due to a shadow effect that introduces systematic bias in the grain size estimation when an image is taken under direct bright sunlight.

We estimated the time used for the analysis of a single image by mobile devices and SmartGrain software at different image resolutions. The results are shown in Supplementary Table S3. The time for low resolution image processing ( $2592 \times 1944$  pixels) is approximately 30 s. For a higher resolution camera (Samsung  $3264 \times 2448$ ), this value is close to 1 min. Interestingly, this is comparable with the time of image processing by SmartGrain (at similar resolutions,  $3510 \times 2550$ ).

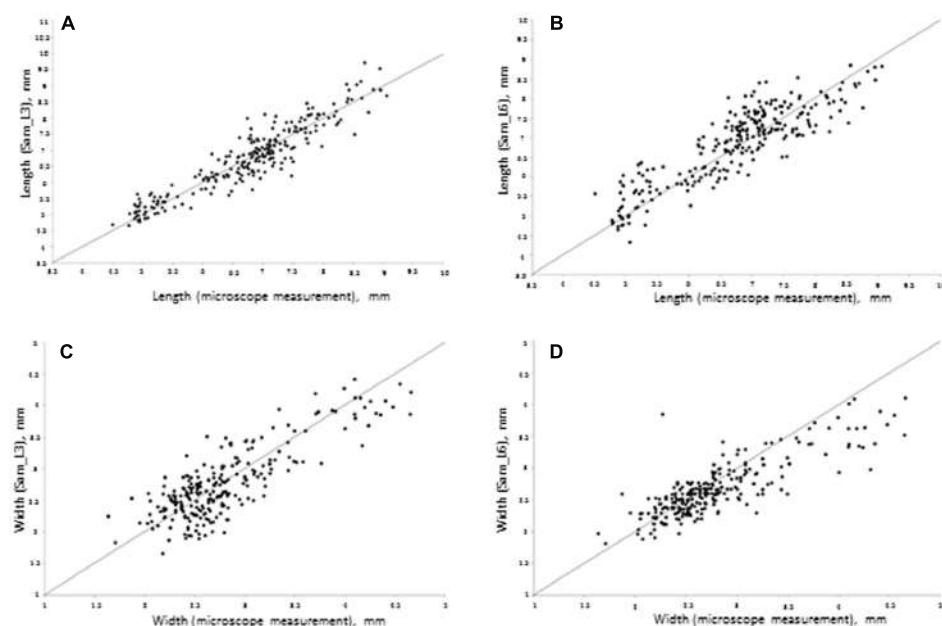
Using the SeedCounter mobile application, we performed wheat grain morphometry of five varieties. For each variety, 50

grains were analyzed, and their length and width were measured. The results are shown in **Figures 4A–C**.

The diagrams in **Figures 4A–C** demonstrate the reliability of discriminating grains from different wheat varieties based on their length and width estimates. The figure shows that the Alcedo cultivar has the thickest grains (average width–3.59 mm) and that the Synthetic cultivar has the longest grains (7.97 mm). The separation of varieties by seed size is clearly demonstrated in **Figure 4C**, where different varieties occupy different plot areas.

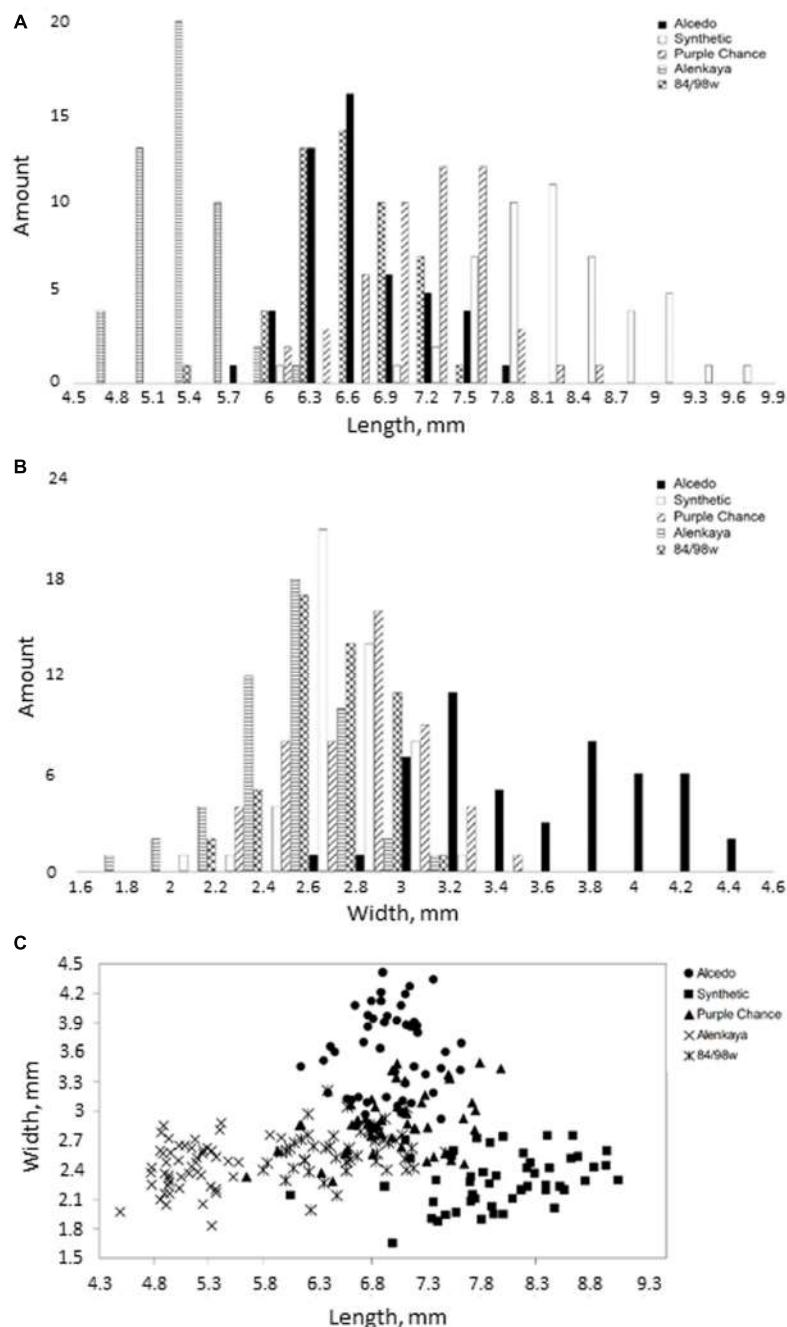
## DISCUSSION

Image processing methods for seed morphometry and classification have been implemented since the 1980s (Sapirstein et al., 1987). Updates of these methods appear constantly, including in recent years (Smykalova et al., 2013; Whan et al., 2014; Miller et al., 2016; Sankaran et al., 2016). New methods use various optical sensing techniques to estimate seed quality and safety (Huang et al., 2015), describe complex seed shapes using 2D images (Williams et al., 2013; Cervantes et al., 2016). Breakthrough 3D imaging technology and robotics (Jahnke et al., 2016; Roussel et al., 2016) or X-ray computed tomography (Strange et al., 2015) implemented for evaluating seed shape in fine detail. However, there is still a need for seed phenotyping using simple and low cost tools (Whan et al., 2014). They can be effectively implemented with high throughput. Despite



**FIGURE 3 | Scatter plot of seed sizes measured by Samsung mobile device (Y-axis) relative to the sizes measured under a microscope (X-axis).**

(A) Seed length at L3 conditions; regression parameters: intercept = 0.25 (Lower 95%: -0.07, Upper: 0.57), slope = 0.96 (Lower 95%: 0.91, Upper 95%: 1.01). (B) Seed length in L6 conditions; regression parameters: intercept = 1.26 (Lower 95%: 0.86, Upper 95%: 1.64), slope = 0.83 (Lower 95%: 0.77, Upper 95%: 0.88). (C) Seed width in L3 conditions; regression parameters: intercept = 0.54 (Lower 95%: 0.33, Upper 95%: 0.73), slope = 0.79 (Lower 95%: 0.73, Upper 95%: 0.87). (D) Seed width in L6 conditions; regression parameters: intercept = 0.90 (Lower 95%: 0.58, Upper 95%: 0.68), slope = 0.64 (Lower 95%: 0.58, Upper 95%: 0.68).



**FIGURE 4 | Distribution of grain width and length for five different wheat varieties. (A,B)** The histograms of the length and width distribution, respectively; **(C)** 2D-scatter plot of grains from different varieties in length (X-axis) and width (Y-axis).

simplicity, they are powerful enough to identify QTL related to seed morphology and size (Gegas et al., 2010; Herridge et al., 2011; Moore et al., 2013; Williams et al., 2013). Mobile devices are valuable tools in this regard. They provide the researcher everything needed for simple phenotyping, including a digital camera, a powerful processor, and Internet access. They can be applied far from the lab, yet provide reasonable precision for phenotypic parameter estimates. Mobile devices are also

convenient for the novel type of plant phenotyping ‘by crowd’ (Rahman et al., 2015).

We suggest a program for grain morphometry using mobile devices. The protocol of the analysis setup is simple and uses a white paper sheet of standard size as a background to convert pixels into the metric scale. To test the accuracy of the program, we performed a series of image analysis experiments using three types of mobile devices and six lighting conditions.

In our work, the mean absolute errors of the length/width estimates are approximately 7–10% and correlation coefficients for length and width between estimated and actual values at ambient lighting are close to 0.93 and 0.77, respectively. Similar analysis performed in a recent work, Miller et al. (2016) reported  $r^2 = 0.996$  for maize kernel length estimated from digital images and their actual values (flatbed document scanner Epson V700, 1200 dpi image, 24-bit color resolution). Sankaran et al. (2016) reported Pearson correlation coefficients between image-based estimates of chickpea seed size and their real values ranging from 0.86 to 0.93 (Canon 70D digital SLR camera, tripod setup, 15–85 mm zoom lens, image resolution set to  $2700 \times 1800$  pixels). Whan et al. (2014) analyzed performance of wheat seed length and width measurements by the following three methods: GrainScan (developed by the authors), SmartGrain (Tanabata et al., 2012), and SeedCount (Next Instruments, 2015). They used an Epson Perfection V330 (Seiko Epson Corporation, Suwa, Japan) scanner to obtain 300 dpi color images. Whan et al. (2014) demonstrated that the average accuracy (Pearson correlation between true parameters and image-based estimates) for GrainScan was very high (0.981–0.996), while the average accuracy for SmartGrain was lower (0.871–0.947), similarly to that of SeedCount at the ambient light conditions (0.731–0.940; Supplementary Table S2). Note, the accuracy for length estimates was higher than for width for all three methods. Our results demonstrate that SeedCounter accuracy and efficiency are comparable with those obtained using desktop PC/scanner/camera devices. Note that we used cameras with moderate resolution and unpretentious lighting conditions for our experiments.

Interference from uncontrolled or uneven lighting is the most basic challenge for smartphone optical sensing (McCracken and Yoon, 2016). We found that the lighting of the paper has significant influence on the accuracy of our method, unlike the smartphone type (Table 4; Supplementary Table S2). We used ANOVA with six different classes of lighting not related directly to luminosity. We chose this approach because our data demonstrated that the influence of luminosity itself on accuracy is not straightforward: images taken at high luminosity under direct sunlight demonstrate increased error in comparison with medium luminosity images and ambient lighting. Under low light conditions (11-W daylight lamp or without artificial lighting), grain number estimation accuracy decreases. Lighting conditions with halogen and daylight lamps (experimental conditions of Sam\_L4, Sony\_L4, and DNS\_L4) caused a small shimmering effect on the images. This effect can complicate the paper recognition process and lead to distorted results. The flicker effect was also present under Sam\_L3 and DNS\_L3 conditions but could be significantly suppressed using the “night shot” technique. The location of light sources and their angle with the paper surface can distort the measurements and degrade sheet recognition conditions. A brighter, diffused light eliminates distortions associated with the appearance of dark spots on the borders of the sheet that can be incorrectly recognized as grains, allows for more efficient separation of touching grains and reduces the likelihood that the grain in the image will merge with the background.

There are several approaches suggested to improve image quality and analysis precision. Some of them require auxiliary/add-on devices (enclosed lighting and imaging attachments) to improve the sensitivity of the smartphone camera (Barbosa et al., 2015). Other methods implement normalization algorithms to reduce lighting inhomogeneity on the image (McCracken et al., 2016). There is still no perfect solution to this problem and further investigation is required to reduce image processing errors from these sources (McCracken and Yoon, 2016).

Mobile applications can significantly accelerate the process of counting the number of grains of wheat in an ear. The time required to calculate approximately 50 grains using a mobile device is approximately 20–55 s, depending on the mobile device and camera resolution. The time required for manually counting the same number of grains may be a little less but mobile devices allow processing a series of images in the background and automatically saving and transmitting data to the server. Increasing the number of grains to 100 increases the running time of the algorithm by 5–10 s. The time required to evaluate the lengths and widths of 50 grains under the microscope is approximately 40–60 min. The mobile application performs this analysis in approximately 1 min.

Thus, the mobile application “SeedCounter” allows for the large-scale measurement of the phenotypic parameters of wheat grains, such as length, width, area, and number of grains per ear, both in “the field” and in the laboratory.

## AUTHOR CONTRIBUTIONS

EK developed algorithms, SeedCounter software, and performed data analysis. MG contributed to algorithm development and data analysis. DA conceived of the study and participated in its design. All authors participated in writing the manuscript as well as read and approved its final version.

## FUNDING

This work was in part supported by the Russian Foundation for Basic Research (project 16-37-00304) and Russian Government Budget project 0324-2015-0003.

## ACKNOWLEDGMENT

Authors are grateful to Tatyana Pshenichnikova for providing wheat grain material, Sodbo Sharapov for providing Samsung Galaxy Tab S2 for technical testing of the application and Sergey Lashin for help in English translation.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01990/full#supplementary-material>

## REFERENCES

- Bai, X. D., Cao, Z. G., Wang, Y., Yu, Z. H., Zhang, X. F., and Li, C. N. (2013). Crop segmentation from images by morphology modeling in the CIE L\*a\*b color space. *Comput. Electron. Agric.* 99, 21–34. doi: 10.1016/j.compag.2013.08.022
- Barbosa, A. I., Gehlot, P., Sidapra, K., Edwards, A. D., and Reis, N. M. (2015). Portable smartphone quantitation of prostate specific antigen (PSA) in a fluoropolymer microfluidic device. *Biosens. Bioelectron.* 70, 5–14. doi: 10.1016/j.bios.2015.03.006
- Cervantes, E., Martín, J. J., and Saadaoui, E. (2016). Updated methods for seed shape analysis. *Scientifica* 2016, 5691825. doi: 10.1155/2016/5691825
- Chen, X., Xun, Y., Li, W., and Zhang, J. (2010). Combining discriminant analysis and neural networks for corn variety identification. *Comput. Electron. Agric.* 71, 48–53. doi: 10.1016/j.compag.2009.09.003
- Dawson-Howe, K. (2014). *A Practical Introduction to Computer Vision with OpenCV*. Chichester: John Wiley & Sons.
- Duan, L., Yang, W., Huang, C., and Liu, Q. (2011). A novel machine-vision-based facility for the automatic evaluation of yield-related traits in rice. *Plant Methods* 7: 44. doi: 10.1186/1746-4811-7-44
- Gegas, V. C., Nazari, A., Griffiths, S., Simmonds, J., Fish, L., Orford, S., et al. (2010). A genetic framework for grain size and shape variation in wheat. *Plant Cell* 22, 1046–1056. doi: 10.1105/tpc.110.074153
- Gómez-Robledo, L., López-Ruiz, N., Melgosa, M., Palma, A. J., Capitan-Vallvey, L. F., and Sánchez-Marañón, M. (2013). Using the mobile phone as Munsell soil-colour sensor: an experiment under controlled illumination conditions. *Comput. Electron. Agric.* 99, 200–208. doi: 10.1016/j.compag.2013.10.002
- Granitto, P. M., Verdes, P. F., and Ceccatto, H. A. (2005). Large-scale investigation of weed seed identification by machine vision. *Comput. Electron. Agric.* 47, 15–24. doi: 10.1016/j.compag.2004.10.003
- Herridge, R. P., Day, R. C., Baldwin, S., and Macknight, R. C. (2011). Rapid analysis of seed size in *Arabidopsis* for mutant and QTL discovery. *Plant Methods* 7: 3. doi: 10.1186/1746-4811-7-3
- Howse, J. (2013). *Android Application Programming with OpenCV*. Birmingham: Packt Publishing Ltd.
- Huang, M., Wang, Q. G., Zhu, Q. B., Qin, J. W., and Huang, G. (2015). Review of seed quality and safety tests using optical sensing technologies. *Seed Sci. Technol.* 43, 337–366. doi: 10.1525/sst.2015.43.3.16
- Jahnke, S., Roussel, J., Hombach, T., Kochs, J., Fischbach, A., Huber, G., et al. (2016). phenoSeeder – a robot system for automated handling and phenotyping of individual seeds. *Plant Physiol.* 172, 1358–1370. doi: 10.1104/pp.16.01122
- Kumar, N., Belhumeur, P. N., Biswas, A., Jacobs, D. W., Kress, W. J., Lopez, I. C., et al. (2012). “Leafsnap: A computer vision system for automatic plant species identification,” in *Proceedings of the 12th European Conference on Computer Vision-ECCV*, eds A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid (Berlin: Springer), 502–516.
- McCracken, K. E., Angus, S. V., Reynolds, K. A., and Yoon, J. Y. (2016). Multimodal imaging and lighting bias correction for improved µPAD-based water quality monitoring via smartphones. *Sci. Rep.* 6: 27529. doi: 10.1038/srep27529
- McCracken, K. E., and Yoon, J.-Y. (2016). Recent approaches for optical smartphone sensing in resource limited settings: a brief review. *Anal. Methods* 8, 6591–6601. doi: 10.1039/C6AY01575A
- Miller, N. D., Haase, N. J., Lee, J., Kaeppler, S. M., de Leon, N., and Spalding, E. P. (2016). A robust, high-throughput method for computing maize ear, cob, and kernel attributes automatically from images. *Plant J.* doi: 10.1111/tpj.13320 [Epub ahead of print].
- Moore, C. R., Gronwall, D. S., Miller, N. D., and Spalding, E. P. (2013). Mapping quantitative trait loci affecting *Arabidopsis thaliana* seed morphology features extracted computationally from images. *G3* 3, 109–118. doi: 10.1534/g3.112.003806
- Next Instruments (2015). *Seedcount*. Condell Park, NSW: Next Instruments.
- Novaro, P., Colucci, F., Venora, G., and D'egidio, M. G. (2001). Image analysis of whole grains: a noninvasive method to predict semolina yield in durum wheat. *Cereal Chem.* 78, 217–221. doi: 10.1094/CCHEM.2001.78.3.217
- Pethybridge, S. J., and Nelson, S. C. (2015). Leaf doctor: a new portable application for quantifying plant disease severity. *Plant Dis.* 99, 1310–1316. doi: 10.1094/PDIS-03-15-0319-RE
- Pourreza, A., Pourreza, H., Abbaspour-Farda, M. H., and Sadriana, H. (2012). Identification of nine Iranian wheat seed varieties by textural analysis with image processing. *Comput. Electron. Agric.* 83, 102–108. doi: 10.1016/j.compag.2012.02.005
- Rahman, M., Blackwell, B., Banerjee, N., and Saraswat, D. (2015). Smartphone-based hierarchical crowdsourcing for weed identification. *Comput. Electron. Agric.* 113, 14–23. doi: 10.1016/j.compag.2014.12.012
- Roerdink, J. B. T. M., and Meijster, A. (2000). The watershed transform: definitions, algorithms and parallelization strategies. *Fundam. Inform.* 41, 187–228. doi: 10.3233/FI-2000-411207
- Roussel, J., Geiger, F., Fischbach, A., Jahnke, S., and Scharr, H. (2016). 3D surface reconstruction of plant seeds by volume carving: performance and accuracies. *Front. Plant. Sci.* 7:745. doi: 10.3389/fpls.2016.00745
- Sankaran, S., Wang, M., and Vandemark, G. J. (2016). Image-based rapid phenotyping of chickpeas seed size. *Eng. Agric. Environ. Food* 9, 50–55. doi: 10.1016/j.eafef.2015.06.001
- Sapirstein, H. D., Neuman, M., Wright, E. H., Shwedyk, E., and Bushuk, W. (1987). An instrumental system for cereal grain classification using digital image analysis. *J. Cereal Sci.* 6, 3–14. doi: 10.1016/S0733-5210(87)80035-8
- Smykalova, I., Grillo, O., Bjelkova, M., Martina, P., and Gianfranco, V. (2013). Phenotypic evaluation of flax seeds by image analysis. *Ind. Crops Prod.* 47, 232–238. doi: 10.1016/j.indcrop.2013.03.001
- Strange, H., Zwiggelaar, R., Sturrock, C., Mooney, S. J., and Doonan, J. H. (2015). Automatic estimation of wheat grain morphometry from computed tomography data. *Funct. Plant Biol.* 42, 452–459. doi: 10.1071/FP14068
- Tahir, A. R., Neethirajan, S., Jayas, D. S., Shahin, M. A., Symons, S. J., and White, N. D. G. (2007). Evaluation of the effect of moisture content on cereal grains by digital image analysis. *Food Res. Int.* 40, 1140–1145. doi: 10.1016/j.foodres.2007.06.009
- Tanabata, T., Shibaya, T., Hori, K., Ebana, K., and Yano, M. (2012). SmartGrain: high-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiol.* 4, 1871–1880. doi: 10.1104/pp.112.205120
- Whan, A. P., Smith, A. B., Cavanagh, C. R., Ral, J. P. F., Shaw, L. M., Howitt, C. A., et al. (2014). GrainScan: a low cost, fast method for grain size and colour measurements. *Plant Methods* 10: 1. doi: 10.1186/1746-4811-10-2310.4225/08/536302C43FC28
- Wiesnerová, D., and Wiesner, I. (2008). Computer image analysis of seed shape and seed color for flax cultivar description. *Comput. Electron. Agric.* 61, 126–135. doi: 10.1016/j.compag.2007.10.001
- Williams, K., Munkvold, J., and Sorrells, M. (2013). Comparison of digital image analysis using elliptic fourier descriptors and major dimensions to phenotype seed shape in hexaploid wheat (*Triticum aestivum* L.). *Euphytica* 190, 99–116. doi: 10.1007/s10681-012-0783-0
- Zapotocny, P. (2011). Discrimination of wheat grain varieties using image analysis and neural networks, Part I, single kernel texture. *J. Cereal Sci.* 54, 60–68. doi: 10.1016/j.jcs.2011.02.012

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

Copyright © 2017 Komyshev, Genaev and Afonnikov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Non-destructive Phenotypic Analysis of Early Stage Tree Seedling Growth Using an Automated Stereovision Imaging Method

**Antonio Montagnoli<sup>1\*</sup>, Mattia Terzaghi<sup>1</sup>, Nicoletta Fulgaro<sup>1</sup>, Borys Stoew<sup>2</sup>, Jan Wipenmyr<sup>2</sup>, Dag Ilver<sup>2</sup>, Cristina Rusu<sup>2</sup>, Gabriella S. Scippa<sup>3</sup> and Donato Chiatante<sup>1</sup>**

<sup>1</sup> Laboratory of Environmental and Applied Botany, Department of Biotechnology and Life Science, University of Insubria, Varese, Italy, <sup>2</sup> Sensor Systems Department, Acreo Swedish ICT, Gothenburg, Sweden, <sup>3</sup> Department of Biosciences and Territory, University of Molise, Pesche, Italy

## OPEN ACCESS

### Edited by:

Marcos Egea-Cortines,  
Universidad Politécnica de Cartagena,  
Spain

### Reviewed by:

Risto Sievänen,  
The Finnish Forest Research Institute,  
Finland

Yonghui Liu,  
Aberystwyth University, UK

### \*Correspondence:

Antonio Montagnoli  
[antonio.montagnoli@uninsubria.it](mailto:antonio.montagnoli@uninsubria.it)

### Specialty section:

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

**Received:** 08 March 2016

**Accepted:** 18 October 2016

**Published:** 28 October 2016

### Citation:

Montagnoli A, Terzaghi M, Fulgaro N,  
Stoew B, Wipenmyr J, Ilver D,  
Rusu C, Scippa GS and Chiatante D  
(2016) Non-destructive Phenotypic  
Analysis of Early Stage Tree Seedling  
Growth Using an Automated  
Stereovision Imaging Method.  
*Front. Plant Sci.* 7:1644.  
doi: 10.3389/fpls.2016.01644

A plant phenotyping approach was applied to evaluate growth rate of containerized tree seedlings during the precultivation phase following seed germination. A simple and affordable stereo optical system was used to collect stereoscopic red–green–blue (RGB) images of seedlings at regular intervals of time. Comparative analysis of these images by means of a newly developed software enabled us to calculate (a) the increments of seedlings height and (b) the percentage greenness of seedling leaves. Comparison of these parameters with destructive biomass measurements showed that the height traits can be used to estimate seedling growth for needle-leaved plant species whereas the greenness trait can be used for broad-leaved plant species. Despite the need to adjust for plant type, growth stage and light conditions this new, cheap, rapid, and sustainable phenotyping approach can be used to study large-scale genome variations due to genome variability and interaction with environmental factors.

**Keywords:** plant phenotype, biomass, seedlings, *Picea abies* L., *Pinus sylvestris* L., *Fagus sylvatica* L., *Quercus ilex* L., RGB image analysis

## INTRODUCTION

Worldwide, an estimated two billion ha of forests are degraded (Minnemayer et al., 2011; Stanturf et al., 2014). In addition to the continuing anthropogenic alterations of global ecosystems (Foley et al., 2005; Kareiva et al., 2007; Ellis et al., 2013), the anticipated effects of global climate change are expected to lead to further deforestation and forest degradation in the future (Steffen et al., 2007; Malhi et al., 2008; Zalasiewicz et al., 2010; Stanturf et al., 2014). Recently, restoration of degraded land has received increasing attention due to its potential to reconcile agricultural development and forest conservation (Robertson and Swinton, 2005; Kissinger et al., 2012). Among the many techniques and tools available for restoration strategies (Stanturf et al., 2014), container seedlings may be the most cost-effective when the planting season is to be extended or adverse sites are to be planted (Brissette et al., 1991; Luoranen et al., 2005, 2006; Stanturf et al., 2014). Container seedlings are produced to meet desired characteristics for outplanting under specified conditions (Brissette et al., 1991; Landis et al., 2010). This requires the artificial production of high-quality forest planting stock material (Wang et al., 2007; Cole et al., 2011) able to successfully survive and grow after outplanting (Wilson and Jacobs, 2006). To achieve this, there is an urgent need to improve the phenotypic assessment of containerised tree seedlings.

The phenotype of a plant is the result of a complex interaction between morphological, ontogenetical, physiological, and biochemical factors (Gratani, 2014). A thorough knowledge of the phenotypic variation occurring spontaneously in nature or after induction by non-intrinsic factors such as environmental stressors is essential for a better understanding of all events taking place in the life of a plant (Grant-Downton and Dickinson, 2006; Kuromori et al., 2009). For the purposes of this paper, we refer to phenotyping as a method to measure plant growth using non-invasive technologies that have become increasingly available in recent years (Fiorani and Schurr, 2013). Unfortunately, measurements of relative growth rate on a mass basis still depend on destructive and time-consuming approaches (Walter et al., 2007; Fiorani and Schurr, 2013; Humplík et al., 2015) with the result of limiting the possibility to examine (1) a large number of samples enabling metadata analysis, and (2) the same sample repeatedly over time (Furbank and Tester, 2011; Busemeyer et al., 2013; Rahaman et al., 2015). To overcome these constraints and to increase the usefulness of phenotype investigation, new approaches based upon the use of technologically advanced equipment that do not affect the samples under examination have been attempted (Tsaftaris and Noutsos, 2009; Walter et al., 2015). Among these, the one based on a non-destructive image analysis seems to achieve a good reliability for rapid phenotyping measurements of a number of plant traits (Li et al., 2014; Humplík et al., 2015). The reliability of this approach was demonstrated in shoot growth rate analyses in which increments measured as differences of digital area showed a high degree of correlation with those obtained by traditional fresh or dry weights measurements (Humplík et al., 2015; Rahaman et al., 2015; and reference herein). A further improvement of this approach is likely to contribute to a better understanding of the principles governing plant biomass distribution in all organs during the lifespan of a plant, a factor of primary importance for phenotype determination. Similarly, other investigations based on measurements of morphometric parameters (i.e., leaf area, stem height, number of tillers, and inflorescence architecture) of plant growth in controlled and natural conditions could benefit from adopting this non-destructive approach (Busemeyer et al., 2013; Fiorani and Schurr, 2013; Rahaman et al., 2015). Moreover, in recent years a large body of literature is rapidly accumulating, mainly for *Arabidopsis* and agricultural plant species, demonstrating how non-destructive analysis of plant phenotype supports other omics approaches to plant science (Edwards and Batley, 2004; Kuromori et al., 2009). However, despite the undeniable merits of this non-destructive method, it cannot be ignored that a number of biases affect these measurements due to overlapping, twisting, curling, and circadian movement of plant organs during image acquisition, especially when 2D color red-green-blue (RGB) image is taken from a single direction (top view) (Lati et al., 2013; Tessmer et al., 2013; Humplík et al., 2015). Indeed, it is difficult to reliably separate overlapped plant canopies into individual plants and the development and implementation of these methods is limited to early growth stages of a specific plant (Jin and Tang, 2009). To overcome these biases the utilization of a stereo vision system has advantages over conventional 2D machine vision-based plant

sensing systems (Jin and Tang, 2009; Piron et al., 2009; Lati et al., 2013). Even though stereo vision system appear promising for estimation of plant growth parameters and development of models, development and implementation of these methods is still limited in terms of species and plant developmental stage (Lati et al., 2013).

We have specifically developed a simple and flexible optical system together with its associated imaging and processing software able to compare acquired images and to obtain, rapidly and efficiently, measurements of height and greenness of young containerised seedlings during the precultivation period. Unlike most of the commercially available solutions for plant phenotyping which are costly and require a large space (Granier et al., 2006; Tsaftaris and Noutsos, 2009), our system is low cost and has the dimension of a bench instrument. In particular, the small size characteristic makes the system easily transportable and combinable with other equipment as well as with high potential to be straightforward integrated in mass-industry. In the present paper, we describe this in-house developed optical system together with the results obtained from a growth kinetics study on tree seedlings grown in a growth chamber, from seed germination to 5-weeks-old plants. Plant biomass is defined as the total mass of all the above- and below-ground parts at a given point in a plant's life (Roberts et al., 1993; Humplík et al., 2015; Wang and Ruan, 2016). The rationale for testing the functioning of our system with this important parameter is twofold: (a) its considerable influence on the plant phenome, and (b) its great variability in response to environmental factors (Coleman et al., 1994; Di Iorio et al., 2011; Montagnoli et al., 2012, 2014; Chiatante et al., 2015). In order to widen the implementation and development of stereo vision method, seedling analysis was performed with four different species characterized by different canopy geometries and development, two broad-leaved (*Fagus sylvatica* L., *Quercus ilex* L.) and two needle-leaved (*Picea abies* L., *Pinus sylvestris* L.). Since different species and types of plants are characterized by differences in architectural organization (Barthélémy and Caraglio, 2007; Díaz et al., 2016), the effectiveness of our in-house built optical system and its corresponding software was characterized by using both broad-leaved and needle-leaved. We present also a comparison between the data obtained by automated imaging analysis with those obtained with the traditional destructive method.

## MATERIALS AND METHODS

### Plant Material and Growth Chamber Characteristics

Seeds of four tree species (*Fagus sylvatica* L., *Quercus ilex* L., *Picea abies* L., and *Pinus sylvestris* L.) were provided by the National Forest Service (National Centre for Study and Conservation of Forest Biodiversity-Peri, Italy) and sorted for uniform size. Seeds of *F. sylvatica* were first hydrated by soaking for 24 h in tap water; then seeds were surface sterilized with 3,5% household bleach for 2 min, and rinsed four times with sterile water to remove all traces of bleach. Afterward, seeds were treated with "Teldor" fungicide (3 ml in 1 l of sterile water per 10 min) and placed

under a hood for 3 h to improve fungicide adherence to the seed coat. Finally, seeds were subjected to cold stratification in perlite at 4°C for 2 months. Seeds of *Q. ilex* were hydrated by soaking them for 24 h in tap water and sown without further pretreatment. *P. sylvestris* and *P. abies* seeds were directly sown directly without any pretreatment. A total of 104 seeds were sown in four different mini-plug plastic container trays (QPD 104 VW – 104 cells; 33 mm × 33 mm × 45 mm; 40 mm/height; 27 cc) (QuickPot by HerkuPlast-Kubern, Germany), containing sterile stabilized peat growing medium Preforma VECO3 (Jiffy® Products). The temperature and humidity settings in the growth chamber are detailed in **Table 1**. The trays were placed on a steel table with a 50 mm-high edge in order to fill it up with water. The mini-plugs had drainage holes in their base, allowing watering from underneath. Watering operations were made every 3 days during germination and every 2 days during growth period to maintain constant water content in each tray. Seed germination was 78% for *Q. ilex*, 66% for *F. sylvatica*, 78% for *P. sylvestris*, and 96% for *P. abies*. Plants were grown under fluorescent light (FLUORA T8), yielding approximately  $120 \mu\text{mol m}^{-2} \text{s}^{-1}$  (Light Meter sensor – HD2302.0 – Delta Ohm, Italy) at tray height. Each plant species was grown independently in the same chamber until the harvest date. A single growth chamber was used to allow for a strict control of environmental factors (uniform conditions) and seedling development (coetaneous cohort).

## Experimental Design

For each species, four trays were grown for a total of 416 seedlings (104 seedlings per tray). To investigate the kinetics of plant growth, half a tray was considered for destructive analysis and the other half for non-destructive image analysis. The first sampling point was 14, 15, and 21 days after germination (a.g.) depending on the plant species. Following samplings were carried out at intervals of no less than 6 days and not more than 12 days depending on the plant species, for four sampling points and 4 weeks of growth period.

## Measurement of Shoot Height and Plant Biomass

At each sampling date, plant height of seedlings for non-destructive analysis ( $n = 52$ ) was measured manually with a wooden measuring stick from the base of the seedling to the highest leaf. Furthermore, five seedlings per tray (20 seedlings in total per species) were randomly collected at each sampling

**TABLE 1 |** Growth chamber settings (number of dark/light hours, relative temperatures, and humidity) for each species.

Plant species	Photoperiod (h day/night)	Temperature (°C)	Relative humidity (%)
<i>Pinus sylvestris</i> L.	16/8	21–26	80 (germination), 55–70 (growth)
<i>Picea abies</i> L.	16/8	21–26	80 (germination), 55–70 (growth)
<i>Fagus sylvatica</i> L.	16/8	21–22	70
<i>Quercus ilex</i> L.	16/8	21–22	70

point. Leaves, shoots, and roots from each seedling were oven dried (52 h at 75°C) and weighed in order to measure total plant biomass.

## Optical System

The optical data acquisition system consists of two digital color cameras equipped with identical lenses from Edmund Optics: 1/1.8" CMOS, 1280 × 1024 pixels, sensor area 6.79 × 5.43 mm, 5 mm fixed focal length lens, field-of-view of 65.5° (UI-1240SE: USB 2.0 uEye industrial camera from IDS Imaging<sup>1</sup>). A rugged USB cable is used for both data transmission and supplying the current to the camera electronics.

The cameras are mounted next to each other as close as possible (~5.5 cm) for stereographic imaging technique for the image color extraction of plant-green and for plant height estimation.

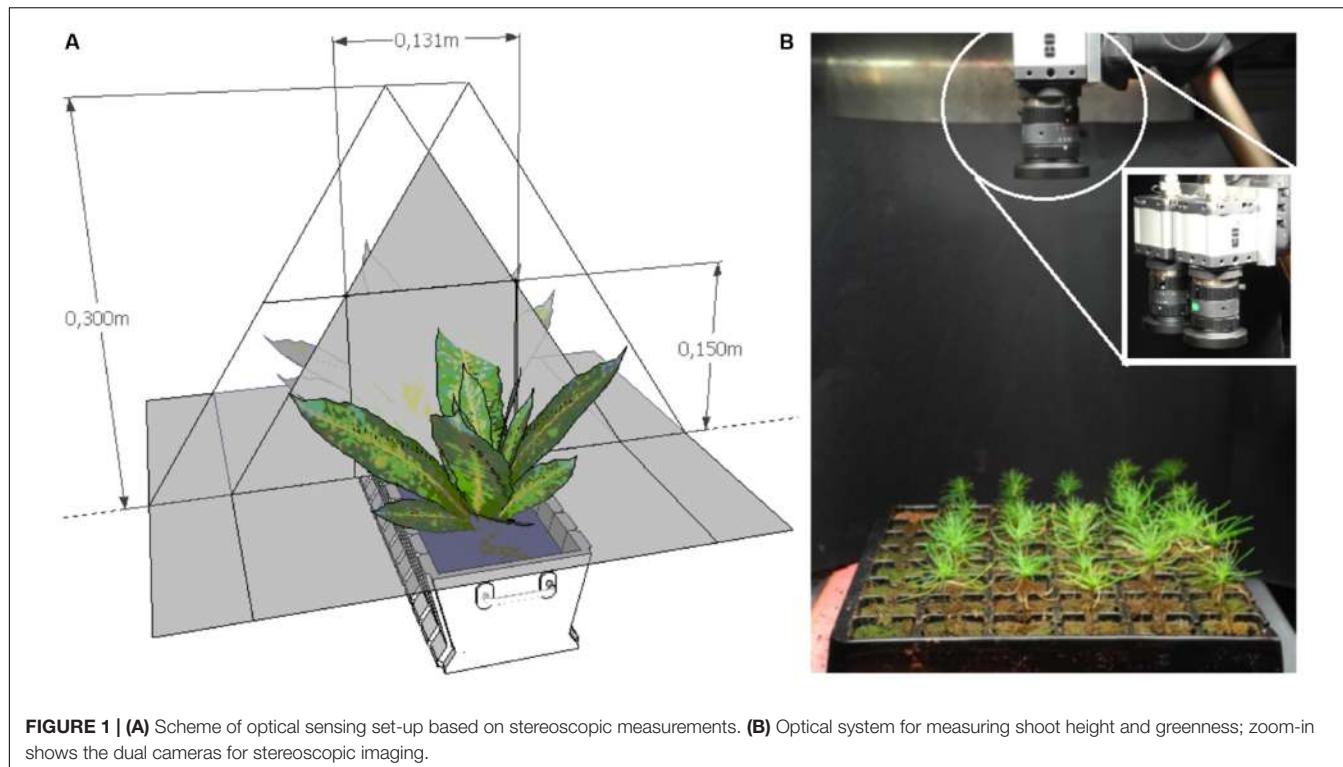
## Image Capture

Shoot stereoscopic images were taken at the same time as the destructive sampling. The trays were manually moved into the image capture cabinets where one stereoscopic image – top view – of each experimental half tray was taken. The tested optical sensing system is based on image acquisition and data processing using in-house developed algorithms derived from hue-saturation-value (HSV) analysis of the image data. Shoot height sensing is based on analysis of reflected light by using a stereoscopic imaging system (**Figure 1**). Total leaf area or green biomass sensing is based on analysis of reflected light using the percentage of green ground coverage by foliage when observed from above. The same hardware is used for extraction of plant greenness and stereoscopic analysis. The depth of focus of the image is a combination of sensor size, focal length and aperture of the lens, and the distance between camera and object. This system can measure various leave colors (e.g., green, red–brown) and different seedling heights (e.g., 4–5 cm, 15–20 cm). The green pixel selection is sensitive to the light source; the proper configuration is also controlled by the .ini file for the respective camera. In particular, parameters in .ini file such as timing (pixel clock, frame rate, exposure time), master and color gain (red, blue, and green), were adjusted given their effect upon green pixel selection. In fact, by setting these parameters it was possible to level out the image colors recorded by the two cameras as well as to distinguish more clearly leaves within the background frame. A long enough sequence of these images can be used to provide a time-series of plant growth – averaged either over the entire scene, or for individual plants. The achieved resolution of the height map is about 1mm that is adequate to follow plant development. After image capture, all images were analyzed using uEyeDualcam and HeightMap software products (Acreeo Swedish ICT).

## Software for Data Acquisition with Optical System

The control of the cameras is carried out using a vendor-supplied software library, uEye (from IDS GmbH). This library

<sup>1</sup><https://en.ids-imaging.com/store/ui-1240se.html>



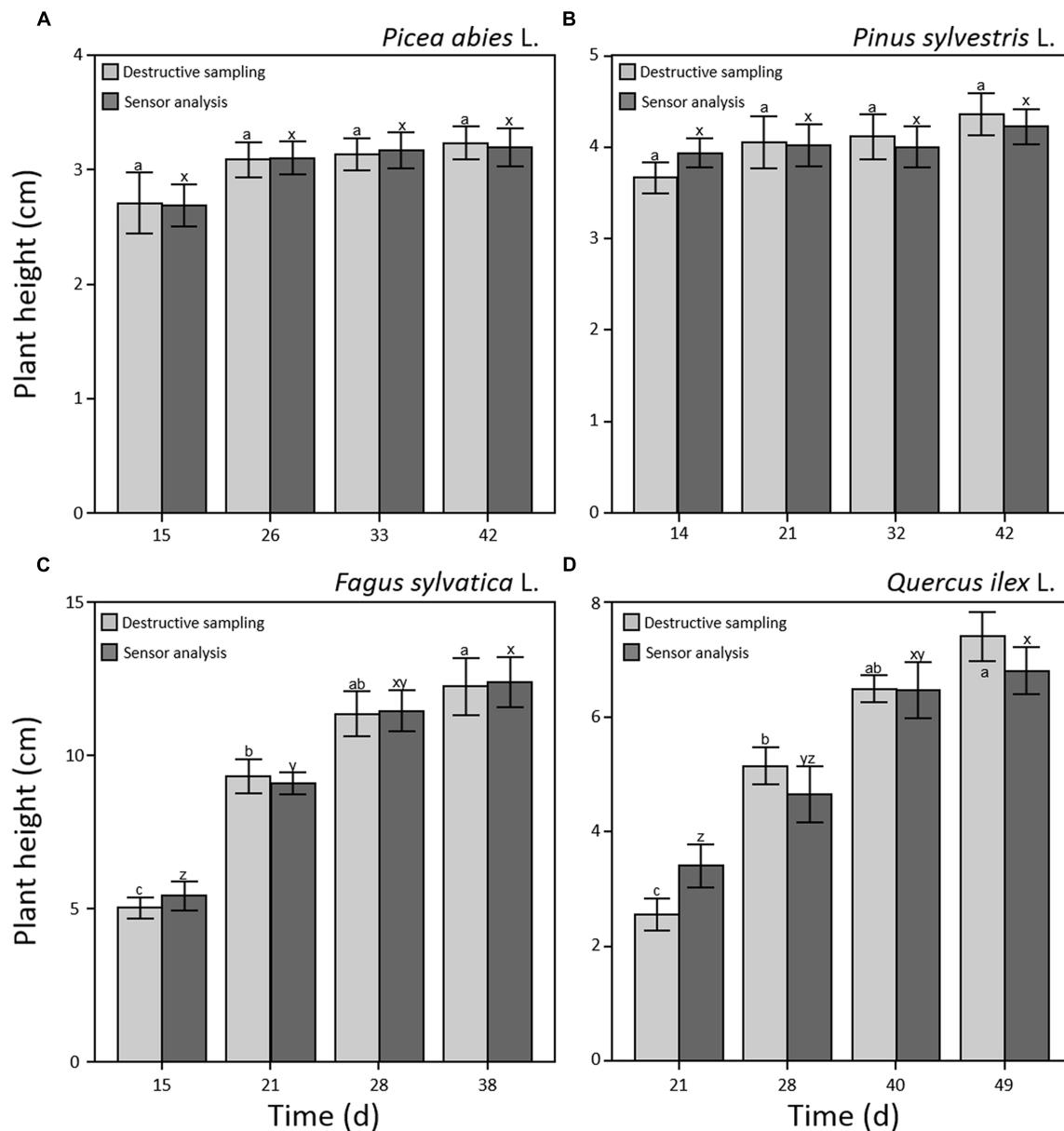
is linked to a graphical user interface (GUI) developed in-house in Microsoft Visual C++. From now on, our developed GUI software executable is referred to as uEyeDualCam.

This uEyeDualCam software has been designed to both functioning for the configuration of individual parameters for each camera as well as the extraction of the “green-only” information for each picture taken. In particular, the automatic setup of individual parameters configuration can be performed by means of special initialization (.ini) files for each camera. The .ini files can be edited by hand providing individual setups for the different light sources because, for example, the green-pixel selection is sensitive to the characteristics of the light source. Moreover, the “green-only” information can be extracted and saved in the PNG format, as a picture. The extraction of the pixels with the relevant shades of plant color is accomplished by converting the color information from RGB format into the HSV format. Both formats are commonly used in image processing and the conversion algorithm is free. This step is relatively simple but time-consuming as each image contains approximately 1.3 Mpixels for our system. Afterward, the uEyeDualCam software allows editing the selection of useful HSV color information corresponding to the “green-only,” whatever color characterize leaves of the species analyzed (e.g., bright green, reddish green, or orange). The shades of useful color form a cylindrical segment in the HSV color space. The uEyeDualCam software selects only pixels within that segment, whereby the non-plant pixels are replaced by the black color. At this point, the uEyeDualCam software provides the percentage of plant pixels for the currently processed image.

A separate set of processing tools (HeightMap software) was developed for the function of height-mapping of each stereoscopic image pair. A discussion of the basic principles of stereoscopic analysis is available from Ensenso and IDS Imaging Development Systems GmbH (2012). In particular, the HeightMap software recalculated greenness using “green-only” information in order to create a plant height map (cm) of the tray conferring a value to the pixel of selected images. Thus, the main innovation in our work is the removal of the soil background and keeping only the plant information within each image. This improves the processing speed and the ability of the tool to match/correlate the relevant image pixels without interference. The current revision of the HeightMap software allows for computing the height distribution of each image at the pixel level, within a selectable sub-set of the scene. The HeightMap software provides a pixel map for the entire scene that can be saved in the monochromatic PNG format.

## Statistical Analysis

Morphological measurements were square root or log transformed to ensure normal distributions and equal variances for the use of parametric statistics. Analysis of variance (one-way ANOVA) was carried out to test the effect of time on plant height, greenness, and biomass. *Post hoc* Bonferroni tests were conducted to detect significant differences between sampling days. An independent samples *t*-test was applied to test the significance of differences between plant height obtained by destructive sampling and plant height obtained by sensor analysis for each sampling date. Analyses of parametric methods were applied at a 95% significance level.



**FIGURE 2 |** Plant height (cm) during early seedling development measured by optical sensors (dark gray) and destructive sampling (light gray) for *Picea abies* (A), *Pinus sylvestris* (B), *Fagus sylvatica* (C), and *Quercus ilex* (D). Data refer to each sampling date after germination and are represented as means ( $n = 52$ )  $\pm 1$  SE. Lowercase letters indicate statistically significant differences ( $p < 0.05$ ) between each sampling date (a, b, and c) and between sensor and destructive analysis (x, y, and z).

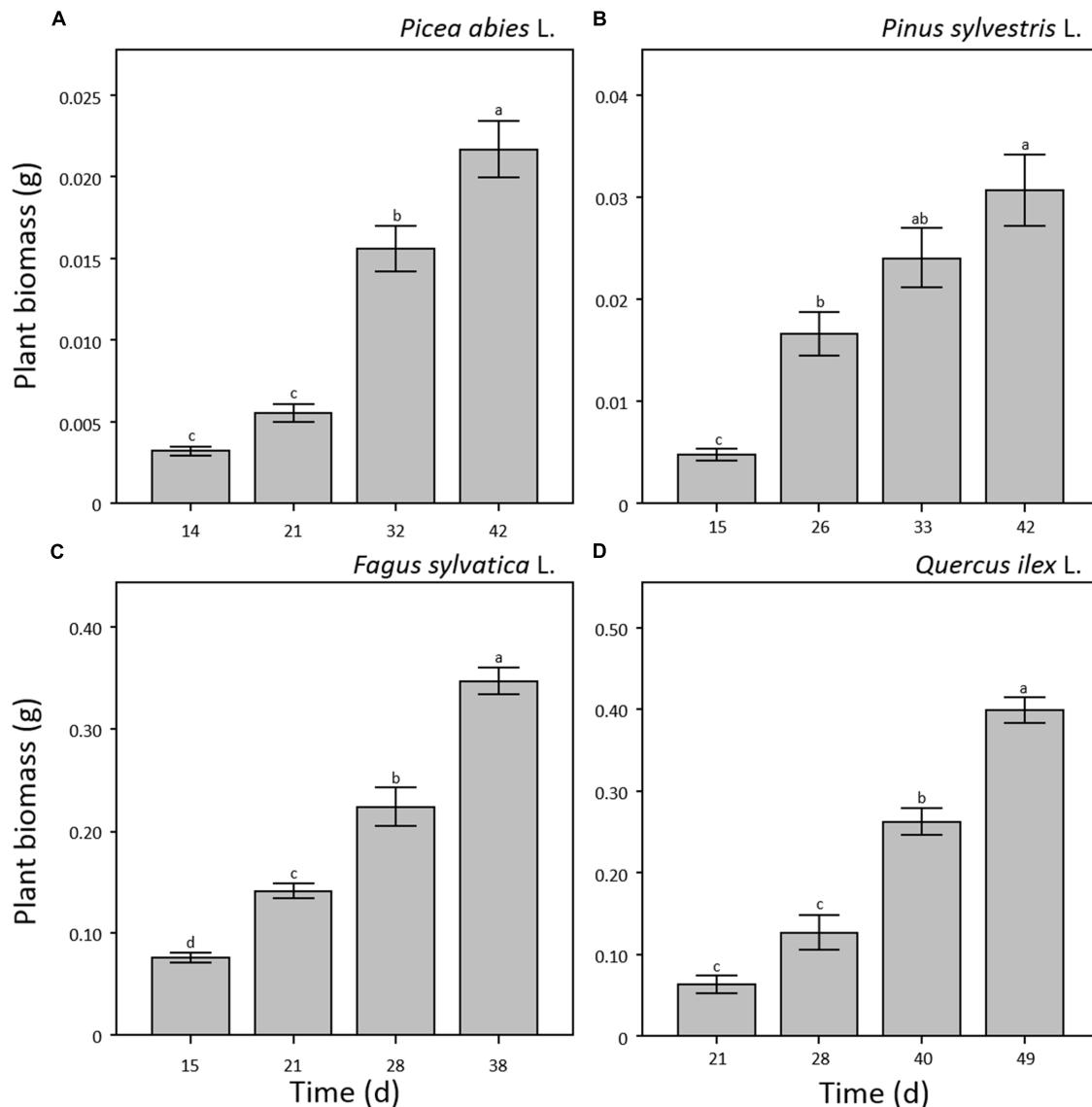
Data of plant height and greenness were related to seedling biomass and allometric equations were obtained by regression analysis. Significant equations were used to develop a regression growth model for each species based on the variation of plant height or greenness over time. In order to test the performance of applied models, the relative root mean squared error (RMSE%) and the relative model bias (BIAS%) were calculated by comparing biomass values predicted from plant height or greenness model with actual biomass values in the range of measured values. Statistical analysis was carried out

using statistical software package SPSS 17.0 (SPSS Inc, Chicago IL, USA).

## RESULTS AND DISCUSSION

### Shoot Height and Plant Biomass

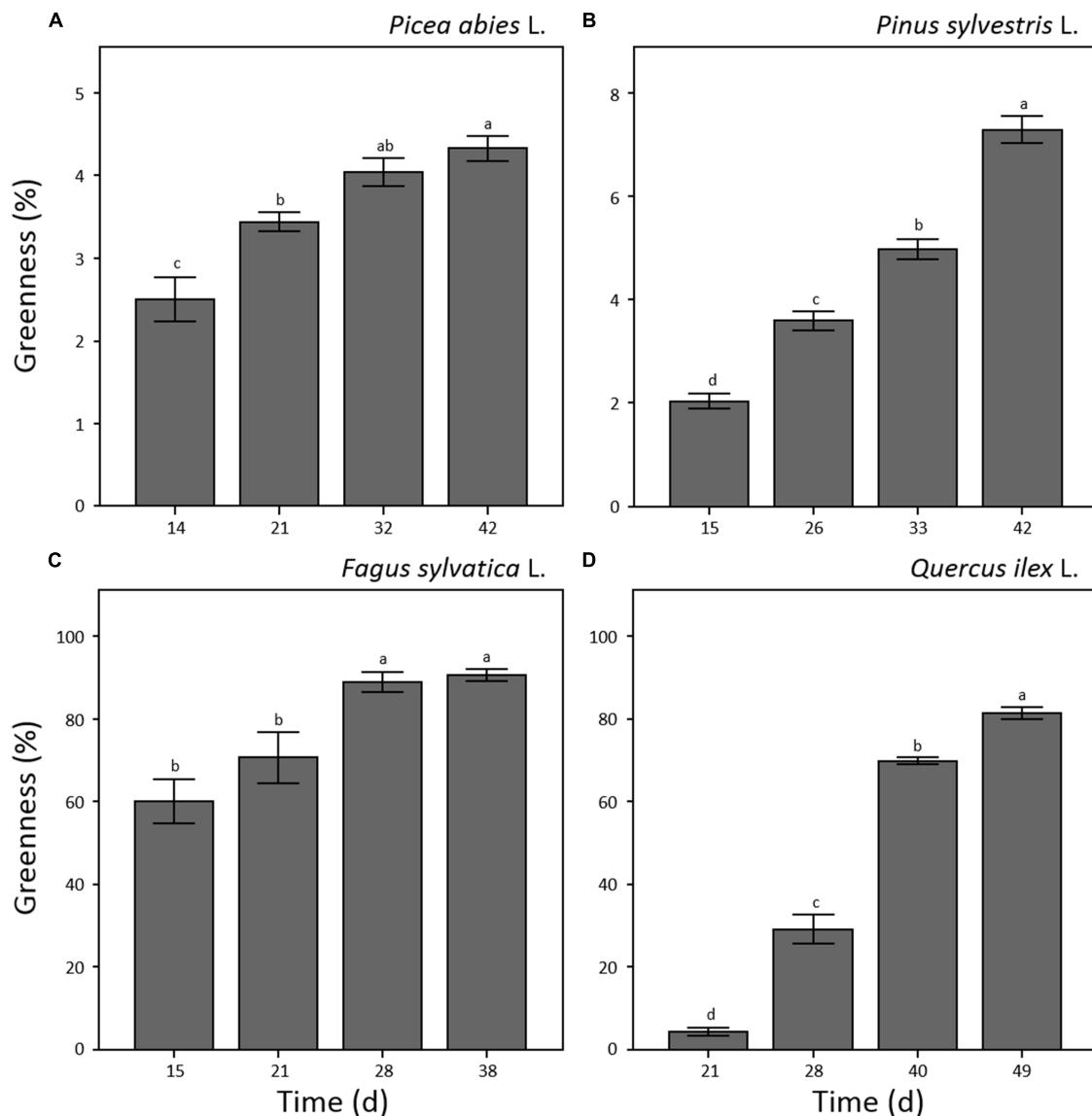
Results on plant height did not show significant differences between manual and software measurements for all four species and sampling points (Figure 2) demonstrating that the



**FIGURE 3 |** Plant biomass (g) measured by destructive sampling for seedlings of *Picea abies* (A), *Pinus sylvestris* (B), *Fagus sylvatica* (C), and *Quercus ilex* (D). Data refer to each sampling date after germination and are represented as means ( $n = 5$ )  $\pm$  1 SE. Lowercase letters indicate statistically significant differences ( $p < 0.05$ ) between each sampling date.

combination of optical sensors and software analysis constitutes a valuable alternative to destructive methods. Shoot height throughout the experiment showed different patterns for needle- and broad-leaved species (Figure 2). In the case of both needle-leaved species, no significant increment of plant height was detected after the emergence of cotyledons ( $p = 0.240$  and  $p = 0.256$  for *P. abies* and *P. sylvestris*, respectively; Figures 2A,B) as internode elongation did not occur during the consecutive emissions of new leaves at this early developmental stage. Therefore, seedlings reached almost maximum height at the first sampling point (day 14th and 15th a.g., respectively), with a slight not significant increment detectable at the last sampling point (day 42nd a.g.; Figures 2A,B). Our results fall within the range

of the rates measured by other researchers for pine species (Jarvis and Jarvis, 1964; Grime and Hunt, 1975; Grotkopp et al., 2002). Moreover, our findings are in line with those of other authors (Evans, 1972; Causton and Venus, 1981; Hunt, 1982; Grotkopp et al., 2002) who showed that growth of *Pines* typically increases sharply between 2 and 4 weeks after seedling emergence and then declines over time. This is probably due to the invaders habit of *Pinus* species characterized by high growth rate, small seed mass, and short generation time (Grotkopp et al., 2002; de Chantal et al., 2003). Broad-leaved species showed a different growth pattern. Plant height increased significantly ( $p < 0.001$ ) throughout the experiment that reached a maximum value of 13 and 7 cm for *F. sylvatica* and *Q. ilex*, respectively, at the

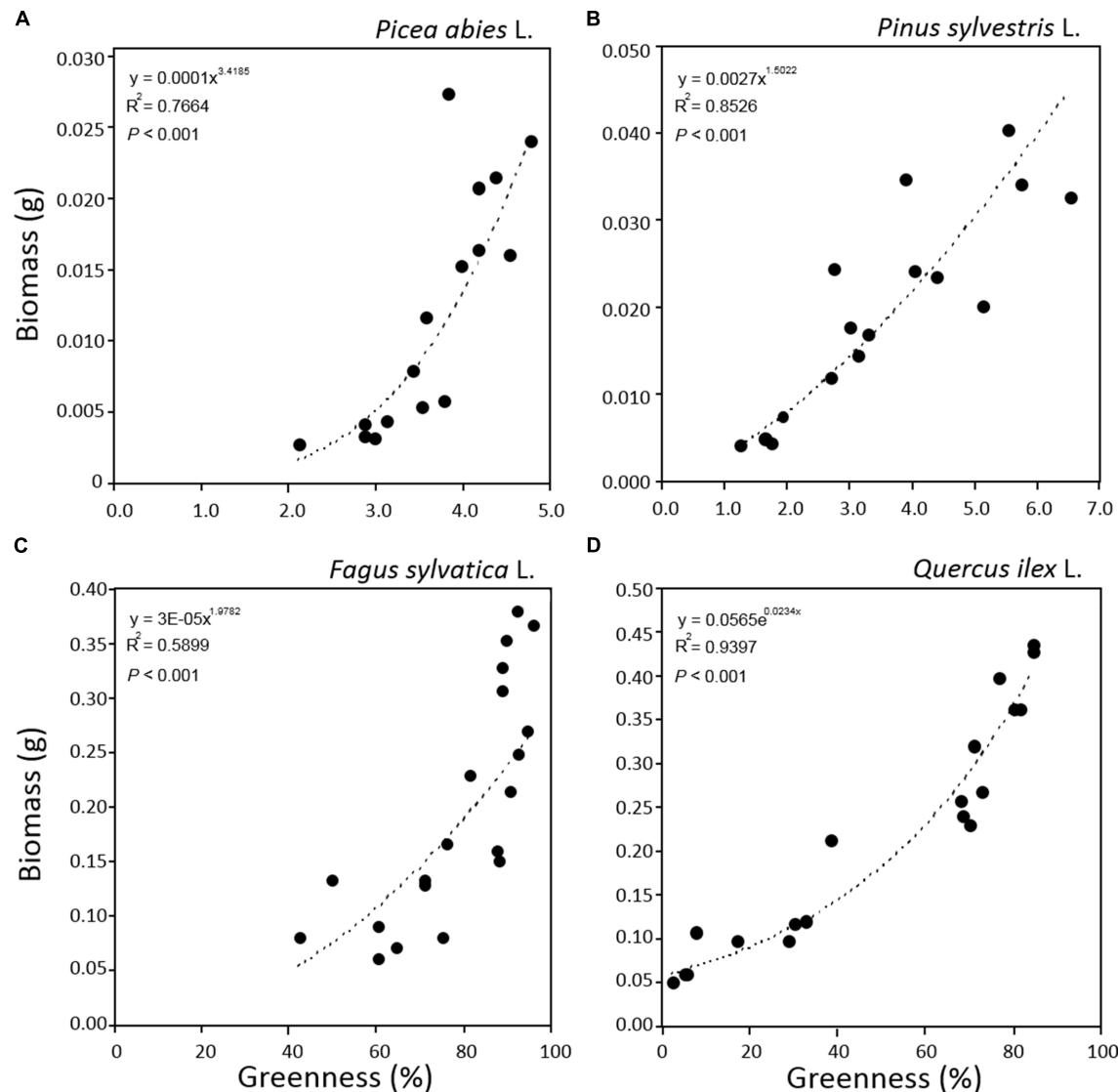


**FIGURE 4 |** Plant greenness (%) measured by optical system analysis for seedlings of *Picea abies* (A), *Pinus sylvestris* (B), *Fagus sylvatica* (C) and *Quercus ilex* (D). Data refer to each sampling date after germination and are represented as means ( $n = 52$ )  $\pm$  1 SE. Lowercase letters indicate statistically significant difference ( $p < 0.05$ ) between each sampling date.

third sampling point (day 28th and 40th a.g.), without further increment until the end of the experiment (Figures 2C,D). As plant growth at this stage of seed development is still depending on endogenous factors (Bentsink and Koornneef, 2008; Baskin and Baskin, 2014), the observed pattern is probably attributable to species-specific growth habits. Despite the importance of early seedling development, studies evaluating this process remain scarce or absent (Walter et al., 2007) as in the case of *F. sylvatica* and *Q. ilex*. Concerning plant biomass development, all four species showed a significant power function increase ( $p < 0.001$ ) throughout the experiment (Figure 3). Moreover, the two broad-leaved species (Figures 3C,D) showed a 10-fold higher total biomass than needle-leaved species (Figures 3A,B).

## Shoot Greenness

Shoot greenness of the seedlings showed significant variation throughout the experiment ( $p < 0.001$ ) with different patterns for each of the considered species (Figure 4). In the case of *F. sylvatica*, the maximum value was reached at the third sampling point (day 28 a.g.; Figure 4C) remaining stable until the end of the experiment. *P. abies*, *P. sylvestris*, and *Q. ilex* (Figures 4A,B,D) showed a continuous increase in greenness throughout the experiment reaching maximum values at the last sampling point (day 42, 42 and 49 a.g., respectively). In general, broad-leaved species showed 10–20 time fold higher values of greenness than needle-leaved species (Figures 4C,D). Seedling leaves of *F. sylvatica* covered almost 80% of the trays at day



**FIGURE 5 | Relationships between greenness and plant biomass for seedlings of *Picea abies* (A), *Pinus sylvestris* (B), *Fagus sylvatica* (C), and *Quercus ilex* (D).** Data refer to all sampling dates.

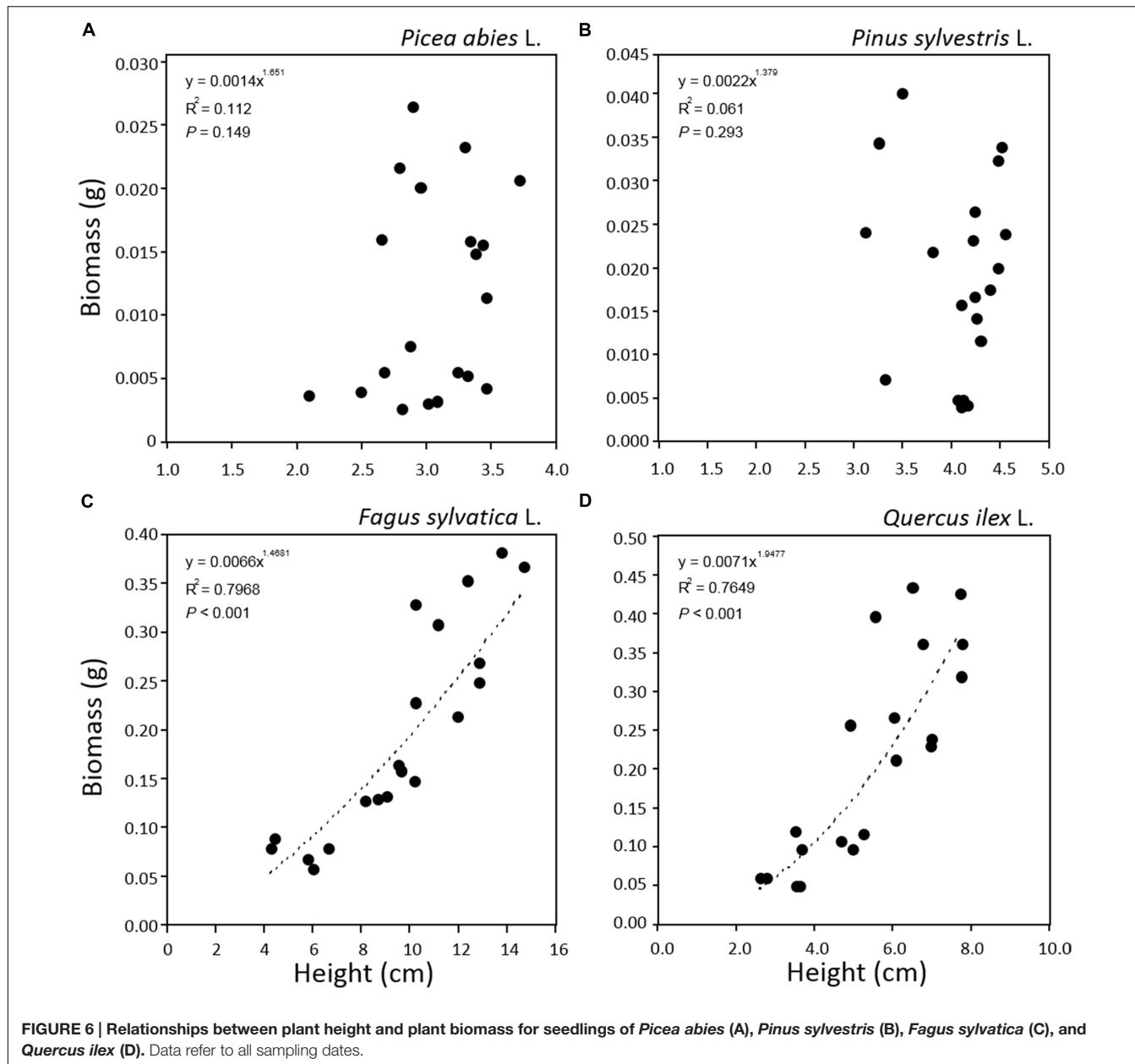
21 a.g. while *Q. ilex* reached 80% tray coverage at day 49 a.g. (Figures 4C,D). On the other hand, *P. abies* and *P. sylvestris* covered less than 7% of the total tray area at 42 days a.g. (Figures 4A,B).

## Regression Model

In order to test our non-destructive measurement method as a tool for monitoring tree seedling growth, patterns of tray greenness and seedling height obtained by software analysis, were related to seedling biomass data obtained by classical destructive analysis. The power function was selected as best fit for all the relationships. This might be explained by Richards (1959) which demonstrated that in the allometric relationship between two correlated growth characteristics, throughout plant development,

if the known growth characteristic is conforming to one curve type, any other unknown characteristic increasing allometrically with it will have the same family of growth curve. In our case, plant height and greenness growth characteristics were related to plant biomass which developed in time conforming to power function curve.

The relationship between tray greenness and seedling biomass showed a good correlation for all species until the tray was almost fully covered (Figures 5). However, in the case of *F. sylvatica* almost the whole tray was covered in less than 1 month but, as its biomass continued to increase after full coverage (Figures 3C and 4C), a lower coefficient of determination (Figure 5C) was observed. This might be due to deviation in the segmented plant's area occurring when different plant



**FIGURE 6 |** Relationships between plant height and plant biomass for seedlings of *Picea abies* (A), *Pinus sylvestris* (B), *Fagus sylvatica* (C), and *Quercus ilex* (D). Data refer to all sampling dates.

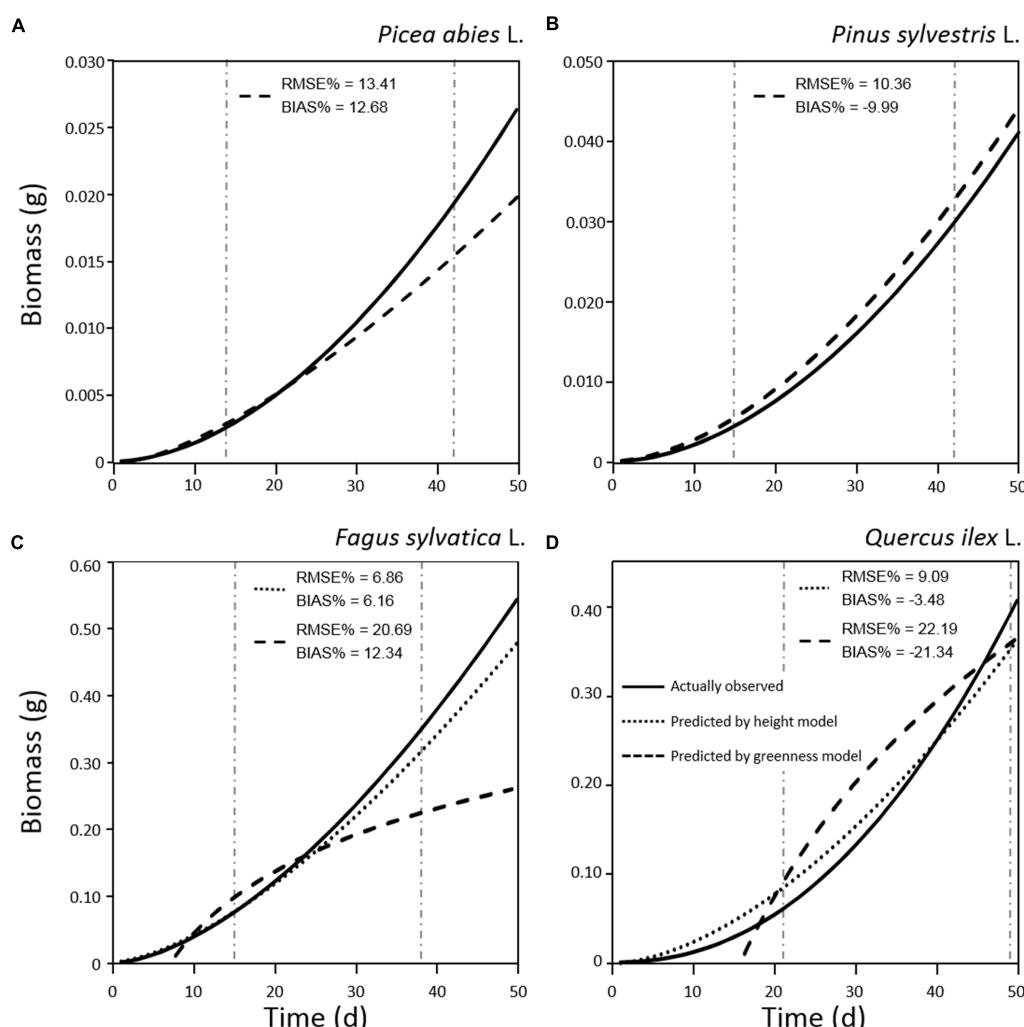
leaves overlap (Jin and Tang, 2009; Lati et al., 2013). Alternative approaches are offered by stereovision-based models which allow plant characterization using 3D spatial properties. A weak relationship between seedling height and biomass was found in the case of the two needle-leaved species (Figures 6A,B) while a strong relationship was found for the two broad-leaved species (Figures 6C,D). Indeed, both needle-leaved species (*P. abies* L. and *P. sylvestris* L.) did not significantly increase plant height during the growth period (Figures 2A,B) despite the continuous increment of seedling biomass. Regression growth models, derived from the relationship of both height and greenness with biomass and their variation with time (Table 2), were compared with allometric equations obtained by destructive sampling (Figure 7). Therefore, the greenness regression growth

model showed the best and the only fit for *P. abies* and *P. sylvestris* (Figures 7A,B). Moreover, the RMSE values showed the best fit to the plant height regression growth model for *F. sylvatica* and *Q. ilex* (Figures 7C,D). As highlighted by previous studies (Downie et al., 2012; Humplík et al., 2015; Walter et al., 2015), image analysis is a robust method to record three-dimensional information. Our study confirms this, showing a good fit to models with destructive data. Furthermore, in the present work is clearly demonstrated how crucial the choice of the parameter to analyze is, depending on the species under investigation and the growing conditions. In particular, for our system, the possibility to choose plant greenness or plant height parameters enhances the analysis over the other already existing systems. In particular, the results obtained by stereo optical system were

**TABLE 2 | Regression growth model obtained by actually measured data and by predicted data based on height and greenness measured by optical sensors.**

Species		Equation	$R^2$
<i>Picea abies</i> L.	Actually measured	Biomass = 0.00002 × Time <sup>1.8344</sup>	0.93
	Predicted based on greenness	Biomass = 0.00004 × Time <sup>1.5931</sup>	0.77
<i>Pinus sylvestris</i> L.	Actually measured	Biomass = 0.00003 × Time <sup>1.8447</sup>	0.88
	Predicted based on greenness	Biomass = 0.00004 × Time <sup>1.7889</sup>	0.93
<i>Fagus sylvatica</i> L.	Actually measured	Biomass = 0.0009 × Time <sup>1.6366</sup>	0.95
	Predicted based on greenness	Biomass = 0.1384 × ln(Time) – 0.2741	0.67
<i>Quercus ilex</i> L.	Predicted based on plant height	Biomass = 0.0012 × Time <sup>1.5317</sup>	0.77
	Actually measured	Biomass = 0.00007 × Time <sup>2.2141</sup>	0.92
	Predicted based on greenness	Biomass = 0.3118 × ln(Time) – 0.884	0.95
	Predicted based on plant height	Biomass = 0.0005 × Time <sup>1.6824</sup>	0.69

Models estimate biomass values at a certain time (day).



**FIGURE 7 | Regression models of biomass growth for seedlings of *Picea abies* (A), *Pinus sylvestris* (B), *Fagus sylvatica* (C), and *Quercus ilex* (D), calculated by destructive sampling (actually observed; solid line) and by plant height (dotted line) and greenness (dashed line) models. RMSE% and BIAS% values were calculated comparing values predicted by plant height or greenness models with actual observations in the range of measurements (vertical dot-dashed lines).**

highly comparable with results obtained with direct destructive methods. This highlighted how the 'choose' function strongly reduced problems often occurring in image-based phenotyping such as overlapping, twisting, curling, and circadian movement. Indeed, in the case of broad-leaved species (*F. sylvatica* and *Q. ilex*) that fully cover a tray, the plant height parameter works better than greenness because new leaves, although related to an increase of plant height, overlap with other leaves and quickly cover the entire tray. On the other hand, the greenness parameter works better than plant height for needle-leaved species (*P. abies* and *P. sylvestris*) because new leaves do not overlap while plant height remains almost constant during the early growth period.

## CONCLUSION

Automatic plant phenotyping is under rapid development due to its potential for comparative phenotyping of a large number of samples in an easy, rapid and non-destructive manner. Therefore, technological effort is being put into the development of a low cost and more accessible phenotyping stereo vision system. Here we present a simple and flexible system that is less inexpensive compared to most of the solutions available today on the market and does not require any specific skills to be run. The data collected refer to a comparative analysis of a number of morphological traits obtained from containerised tree seedlings at their precultivation stage. Results suggest that at present the system is reliable, allowing for straightforward control and adjustment of various plant and light source parameters. Therefore, the system and models developed provided a strong agreement between the actual and estimated growth parameters for plants with interconnected canopies. Although the possibility to adapt the system to other growing conditions, conclusion of the present work are specie-specific and focus on containerized early stage of seedlings. Further implementation in both software and hardware can be done for improving the characterization

efficiency of bigger plants, different species and light conditions. Finally, the phenotyping approach used to measure the growth of young seedlings it might also be of support to different omics investigations.

## AUTHOR CONTRIBUTIONS

AM make substantial contributions to the study concept and design, to data collection process and relative interpretation. AM writes the article and dealt with manuscript process, improvements and revisions. MT participate to all works aspects such as concept and design, lab work, software development, data collection and analysis. NF contribute to the experiment design, seedlings growth and data collection. BS, JW, DI, and CR make substantial contribution to the software development and hardware settings for image acquisition and analysis. BS and CR equally contribute in drafting and revising the article concerning the optical sensors and image data acquisition parts. GS supervise the research and contribute to all works aspects. DC conceive and supervise the research in all aspects. Give important intellectual content in outlining the article and revising it critically.

## FUNDING

The European Commission within the Seventh Framework Programme through Project ZEPHYR (grant number 308313) supported this work.

## ACKNOWLEDGMENTS

We are grateful to Dr. Barbara Baesso and Rosaria Santamaria for helping in seedlings production, to National Forest Service (National Centre for Study and Conservation of Forest Biodiversity- Peri, IT) for providing seeds.

## REFERENCES

- Barthélemy, D., and Caraglio, Y. (2007). Plant architecture: a dynamic, multilevel and comprehensive approach to plant form, structure and ontogeny. *Ann. Bot.* 99, 375–407. doi: 10.1093/aob/mcl260
- Baskin, C. C., and Baskin, J. M. (2014). *Seeds, Ecology, Biogeography and Evolution of Dormancy and Germination*, 2nd Edn. San Diego, CA: Academic/Elsevier.
- Bentsink, L., and Koornneef, M. (2008). Seed dormancy, and germination. *Arabidopsis Book* 6:e0119. doi: 10.1199/tab.0119
- Brissette, J. C., Barnett, J. P., and Landis, T. D. (1991). "Container seedlings," in *Forest Regeneration Manual*, eds M. L. Duryea and P. M. Dougherty (Boston, MA: Kluwer Academic Publishers), 117–141.
- Busemeyer, L., Mentrup, D., Möller, K., Wunder, E., Alheit, K., Hahn, V., et al. (2013). Breedvision—a multi-sensor platform for non-destructive field-based phenotyping in plant breeding. *Sensors* 13, 2830–2847. doi: 10.3390/s130302830
- Causton, D. R., and Venus, J. C. (1981). *The Biometry of Plant Growth*. London: Edward Arnold.
- Chiatante, D., Tognetti, R., Scippa, G. S., Congiu, T., Baesso, B., Terzaghi, M., et al. (2015). Interspecific variation in functional traits of oak seedlings (*Quercus ilex*, *Quercus trojana*, *Quercus virginiana*) grown under artificial drought and fire conditions. *J. Plant. Res.* 128, 595–611. doi: 10.1007/s10265-015-0729-4
- Cole, R. J., Holl, K. D., Keene, C., and Zahawi, R. A. (2011). Direct seeding of late-successional trees to restore tropical montane forest. *For. Ecol. Manag.* 261, 1590–1597. doi: 10.1016/j.foreco.2010.06.038
- Coleman, J. S., McConaughay, K. D. M., and Ackerly, D. D. (1994). Interpreting phenotypic variation in plants. *Trends Ecol. Evol.* 9, 187–191. doi: 10.1016/0169-5347(94)90087-6
- de Chantal, M., Leinonen, K., Kuuluvainen, T., and Cescatti, A. (2003). Early response of *Pinus sylvestris* and *Picea abies* seedlings to an experimental canopy gap in a boreal spruce forest. *For. Ecol. Manag.* 176, 321–336. doi: 10.1016/S0378-1127(02)00273-6
- Di Iorio, A., Montagnoli, A., Scippa, G. S., and Chiatante, D. (2011). Fine root growth of *Quercus pubescens* seedlings after drought stress and fire disturbance. *Environ. Exp. Bot.* 74, 272–279. doi: 10.1016/j.envexpbot.2011.06.009
- Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., Dray, S., et al. (2016). The global spectrum of plant form and function. *Nature* 529, 167–177. doi: 10.1038/nature16489
- Downie, H., Holden, N., Otten, W., Spiers, A. J., Valentine, T. A., and Dupuy, L. X. (2012). Transparent soil for imaging the rhizosphere. *PLoS ONE* 7:e44276. doi: 10.1371/journal.pone.0044276
- Edwards, D., and Batley, J. (2004). Plant bioinformatics: from genome to phenotype. *Trends Biotechnol.* 22, 5. doi: 10.1016/j.tibtech.2004.03.002

- Ellis, E. C., Kaplan, J. O., Fuller, D. Q., Vavrus, S., Klein Goldewijk, K., and Verburg, P. H. (2013). Used planet: a global history. *Proc. Natl. Acad. Sci.* 110, 7978–7985. doi: 10.1073/pnas.1217241110
- Ensenso and Ids Imaging Development Systems GmbH (2012). *Obtaining Depth Information from Stereo Images. Whitepaper*. Available at: <https://en.ids-imaging.com/whitepaper.html>
- Evans, G. C. (1972). *The Quantitative Analysis of Plant Growth*. Berkeley, CA: University of California Press.
- Fiorani, F., and Schurr, U. (2013). Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 64, 267–291. doi: 10.1146/annurev-arplant-050312-120137
- Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., et al. (2005). Global consequences of land use. *Science* 309, 570–574. doi: 10.1126/science.1111772
- Furbank, R. T., and Tester, M. (2011). Phenomics—technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 12. doi: 10.1016/j.tplants.2011.09.005
- Granier, C., Aguirreababal, L., Chenu, K., Cookson, S. J., Dauzat, M., Hamard, P., et al. (2006). PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytol.* 169, 623–635. doi: 10.1111/j.1469-8137.2005.01609.x
- Grant-Downton, R. T., and Dickinson, H. G. (2006). Epigenetics and its implications for plant biology 2. The “epigenetic epiphany”: epigenetics, evolution and beyond. *Ann. Bot.* 97, 11–27. doi: 10.1093/aob/mcj001
- Gratani, L. (2014). Plant phenotypic plasticity in response to environmental factors. *Adv. Bot.* 2014, 208747. doi: 10.1155/2014/208747
- Grime, J. P., and Hunt, R. (1975). Relative growth rate: its range and adaptive significance in a local flora. *J. Ecol.* 63, 393–422. doi: 10.2307/2258728
- Grotkopp, E., Rejmánek, M., and Rost, T. L. (2002). Toward a causal explanation of plant invasiveness: seedling growth and life-history strategies of 29 pine (*Pinus*) species. *Am. Nat.* 159, 396–419. doi: 10.1086/338995
- Humprik, J. F., Lazár, D., Husičková, A., and Spíchal, L. (2015). Automated phenotyping of plant shoots using imaging methods for analysis of plant stress responses—a review. *Plant Methods* 11, 29. doi: 10.1186/s13007-015-0072-8
- Hunt, R. (1982). *Plant Growth Curves: The Functional Approach to Plant Growth Analysis*. East Kilbride: Thomson Litho.
- Jarvis, P. G., and Jarvis, M. S. (1964). Growth rates of woody plants. *Physiol. Plant.* 17, 654–666. doi: 10.1111/j.1399-3054.1964.tb08192.x
- Jin, J., and Tang, L. (2009). Corn plant sensing using real-time stereo vision. *J. Field Robot.* 26, 591–608. doi: 10.1002/rob.20293
- Kareiva, P., Watts, S., McDonald, R., and Boucher, T. (2007). Domesticated nature: shaping landscapes and ecosystems for human welfare. *Science* 316, 1866–1869. doi: 10.1126/science.1140170
- Kissinger, G., Herold, M., and De Sy, V. (2012). *Drivers of Deforestation and Forest Degradation: A Synthesis Report for REDD+ Policymakers*. Vancouver: Lexeme Consulting.
- Kuromori, T., Takahashi, S., Kondou, Y., Shinozaki, K., and Matsui, M. (2009). Phenome analysis in plant species using loss-of-function and gain-of-function mutants. *Plant Cell Physiol.* 50, 1215–1231. doi: 10.1093/pcp/pcp078
- Landis, T. D., Dumroese, R. K., and Haase, D. L. (2010). *Seedling Processing, Storage and Outplanting. The Container Tree Nursery Manual*. Washington, DC: USDA Forest Service. Agriculture Handbook, 674.
- Lati, R. N., Sagi Filin, S., and Eizenberg, H. (2013). Estimating plant growth parameters using an energy minimization-based stereovision model. *Comput. Electron. Agric.* 98, 260–271. doi: 10.1016/j.compag.2013.07.012
- Li, L., Zhang, Q., and Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors* 14, 20078–20111. doi: 10.3390/s141120078
- Luoranen, J., Rikala, R., Konttinen, K., and Smolander, H. (2005). Extending the planting period of dormant and growing Norway spruce container seedlings to early summer. *Silva Fenn.* 39, 481. doi: 10.14214/sf.361
- Luoranen, J., Rikala, R., Konttinen, K., and Smolander, H. (2006). Summer planting of *Picea abies* container-grown seedlings: effects of planting date on survival, height growth and root egress. *For. Ecol. Manag.* 237, 534–544. doi: 10.1016/j.foreco.2006.09.073
- Malhi, J. Y., Roberts, T., Betts, R. A., Killeen, T. J., Li, W., and Nobre, C. A. (2008). Climate change, deforestation, and the fate of the Amazon. *Science* 319, 169–172. doi: 10.1126/science.1146961
- Minnemayer, S., Laestadius, L., and Sizer, N. (2011). *A World of Opportunity*. Washington, DC: World Resource Institute.
- Montagnoli, A., Di Iorio, A., Terzaghi, M., Trupiano, D., Scippa, G. S., and Chiatante, D. (2014). Influence of soil temperature and water content on fine-root seasonal growth of European beech natural forest in Southern Alps, Italy. *Eur. J. For. Res.* 133, 957–968. doi: 10.1007/s10342-014-0814-6
- Montagnoli, A., Terzaghi, M., Di Iorio, A., Scippa, G. S., and Chiatante, D. (2012). Fine-root morphological and growth traits in a Turkey-oak stand in relation to seasonal changes in soil moisture in the Southern Apennines, Italy. *Ecol. Res.* 27, 1015–1025. doi: 10.1007/s11284-012-0981-1
- Piron, A., Leemans, V., Lebeau, F., and Destain, M.-F. (2009). Improving in-row weed detection in multispectral stereoscopic images. *Comput. Electron. Agric.* 69, 73–79. doi: 10.1016/j.compag.2009.07.001
- Rahaman, M. M., Chen, D., Gillani, Z., Klukas, C., and Chen, M. (2015). Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front. Plant Sci.* 6:619. doi: 10.3389/fpls.2015.00619
- Richards, F. J. (1959). A flexible growth function for empirical use. *J. Exp. Bot.* 10, 290–301. doi: 10.1093/jxb/10.2.290
- Roberts, M. J., Long, S. P., Tieszen, L. L., and Beadle, C. L. (1993). “Measurement of plant biomass and net primary production of herbaceous vegetation,” in *Photosynthesis and Production in a Changing Environment*, eds D. O. Hall, J. M. O. Scurlock, H. R. Bolhar-Nordenkampf, R. C. Leegood, and S. P. Long (Netherlands: Springer).
- Robertson, G. P., and Swinton, S. M. (2005). Reconciling agricultural productivity and environmental integrity: a grand challenge for agriculture. *Front. Ecol. Environ.* 3:38–46. doi: 10.1890/1540-9295(2005)003[0038:RAPAEI]2.0.CO;2
- Stanturf, J. A., Palik, B. J., and Dumroese, R. K. (2014). Contemporary forest restoration: a review emphasizing function. *For. Ecol. Manag.* 331, 292–323. doi: 10.1016/j.foreco.2014.07.029
- Steffen, W., Crutzen, P. J., and McNeill, J. R. (2007). The Anthropocene: are humans now overwhelming the great forces of nature. *Ambio* 36, 614–621. doi: 10.1579/0044-7447(2007)36[614:TAAHNO]2.0.CO;2
- Tessmer, O. L., Jiao, Y., Cruz, J. A., Kramer, D. M., and Chen, J. (2013). Functional approach to high-throughput plant growth analysis. *BMC Syst. Biol.* 7(Suppl. 6):S17. doi: 10.1186/1752-0509-7-S6-S17
- Tsaftaris, S. A., and Noutsos, C. (2009). “Plant phenotyping with low cost digital cameras and image analytics,” in *Information Technologies in Environmental Engineering*, eds I. N. Athanasiadis, P. A. Mitkas, A. E. Rizzoli, and J. M. Gomez (Berlin: Springer), 238–251.
- Walter, A., Liebisch, F., and Hund, A. (2015). Plant phenotyping: from bean weighing to image analysis. *Plant Methods* 11, 14. doi: 10.1186/s13007-015-0056-8
- Walter, A., Scharr, H., Gilmer, F., Zierer, R., Nagel, K. A., Ernst, M., et al. (2007). Dynamics of seedling growth acclimation towards altered light conditions can be quantified via GROWSCREEN: a setup and procedure designed for rapid optical phenotyping of different plant species. *New Phytol.* 174, 447–455. doi: 10.1111/j.1469-8137.2007.02002.x
- Wang, F., Yu, C. L., and Liu, D. (2007). Effects of plant growth regulators on drought resistance of shrub seedlings. *For. Sci. Tech.* 32, 56–60.
- Wang, L., and Ruan, Y.-L. (2016). Shoot–root carbon allocation, sugar signalling and their coupling with nitrogen uptake and assimilation. *Funct. Plant Biol.* 43, 105–113. doi: 10.1071/FP15249
- Wilson, C. B., and Jacobs, F. D. (2006). Quality assessment of temperate zone deciduous hardwood seedlings. *New For.* 31, 417–433. doi: 10.1007/s11056-005-0878-8
- Zalasiewicz, J., Williams, M., Steffen, W., and Crutzen, P. (2010). The new world of the Anthropocene. *Environ. Sci. Technol.* 44, 2228–2231. doi: 10.1021/es903118j
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2016 Montagnoli, Terzaghi, Fulgaro, Stoew, Wipenmyr, Ilver, Rusu, Scippa and Chiatante. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Exploring Relationships between Canopy Architecture, Light Distribution, and Photosynthesis in Contrasting Rice Genotypes Using 3D Canopy Reconstruction

Alexandra J. Burgess<sup>1,2</sup>, Renata Retkute<sup>3</sup>, Tiara Herman<sup>4</sup> and Erik H. Murchie<sup>1\*</sup>

<sup>1</sup> Division of Plant and Crop Sciences, School of Biosciences, University of Nottingham, Loughborough, UK, <sup>2</sup> Crops For the Future, Semenyih, Malaysia, <sup>3</sup> School of Life Sciences, The University of Warwick, Coventry, UK, <sup>4</sup> School of Biosciences, University of Nottingham Malaysia Campus, Semenyih, Malaysia

## OPEN ACCESS

### Edited by:

John Doonan,  
Aberystwyth University, UK

### Reviewed by:

Eric Ober,  
National Institute of Agricultural  
Botany, UK  
Yuhui Chen,  
Samuel Roberts Noble Foundation,  
USA

### \*Correspondence:

Erik H. Murchie  
erik.murchie@nottingham.ac.uk

### Specialty section:

This article was submitted to  
Technical Advances in Plant Science,  
a section of the journal  
*Frontiers in Plant Science*

**Received:** 23 November 2016

**Accepted:** 20 April 2017

**Published:** 17 May 2017

### Citation:

Burgess AJ, Retkute R, Herman T and  
Murchie EH (2017) Exploring  
Relationships between Canopy  
Architecture, Light Distribution, and  
Photosynthesis in Contrasting Rice  
Genotypes Using 3D Canopy  
Reconstruction.

*Front. Plant Sci.* 8:734.  
doi: 10.3389/fpls.2017.00734

The arrangement of leaf material is critical in determining the light environment, and subsequently the photosynthetic productivity of complex crop canopies. However, links between specific canopy architectural traits and photosynthetic productivity across a wide genetic background are poorly understood for field grown crops. The architecture of five genetically diverse rice varieties—four parental founders of a multi-parent advanced generation intercross (MAGIC) population plus a high yielding Philippine variety (IR64)—was captured at two different growth stages using a method for digital plant reconstruction based on stereocameras. Ray tracing was employed to explore the effects of canopy architecture on the resulting light environment in high-resolution, whilst gas exchange measurements were combined with an empirical model of photosynthesis to calculate an estimated carbon gain and total light interception. To further test the impact of different dynamic light patterns on photosynthetic properties, an empirical model of photosynthetic acclimation was employed to predict the optimal light-saturated photosynthesis rate ( $P_{max}$ ) throughout canopy depth, hypothesizing that light is the sole determinant of productivity in these conditions. First, we show that a plant type with steeper leaf angles allows more efficient penetration of light into lower canopy layers and this, in turn, leads to a greater photosynthetic potential. Second the predicted optimal  $P_{max}$  responds in a manner that is consistent with fractional interception and leaf area index across this germplasm. However, measured  $P_{max}$ , especially in lower layers, was consistently higher than the optimal  $P_{max}$  indicating factors other than light determine photosynthesis profiles. Lastly, varieties with more upright architecture exhibit higher maximum quantum yield of photosynthesis indicating a canopy-level impact on photosynthetic efficiency.

**Keywords:** 3D reconstruction, canopy architecture, crop productivity, light environment, MAGIC population, photosynthesis, rice (*Oryza* spp.)

## INTRODUCTION

The rate of photosynthesis of a given stand of crops is dependent on a multitude of factors including weather, temperature, leaf age, and plant development. Photosynthesis, in turn, is closely linked to potential yield (Murchie et al., 2009; Zhu et al., 2010). However, the complex arrangement of overlapping leaves of different ages and in different states of photosynthesis means that assessing canopy level photosynthesis from individual leaf activity is difficult and time consuming. For an accurate prediction of canopy photosynthesis from leaf measurements, it is necessary to have data on multiple leaf characteristics including physical orientation, positioning and physiological characteristics, such as photosynthetic acclimation and nutrient status (Burgess et al., 2015, 2016). However, predicted productivity tends to be higher than that measured in the field (Zhu et al., 2010). The cause of this disparity is unclear, but may arise from suboptimal photosynthetic responses to dynamic environmental changes partly caused by architectural traits (Zhu et al., 2010; Burgess et al., 2015).

In the absence of methods for whole canopy measurements, such as in Song et al. (2016), predictions require knowledge of the architectural characteristics and its effect on canopy light distribution. Photosynthetic rate is highly sensitive to light intensity, and, in turn, the light intensity within crop canopies has high spatio-temporal variability, and is dependent upon features such as leaf angle, size and shape, leaf number, and the arrangement of this material in three-dimensional space. These findings have led to the concept of an “idealized plant type” or “ideotype.” For example, the International Rice Research Institute (IRRI) proposed that upright leaves, large panicles and fewer tillers would represent the ideal structure for rice (Dingkuhn et al., 1991; Virk et al., 2004). Erect leaf morphology is a characteristic that repeatedly arises within the concept of an ideotype. This is due to the increased light penetration to deeper canopy layers leading to uniformity of light within the canopy setting and maximal net photosynthesis (Clendon and Millen, 1979; Hodanova, 1979; Turitzin and Drake, 1981; Setter et al., 1995; Normile, 1999). Within dense canopies, steeper leaf angles potentially lead to an improvement in whole day carbon gain by enhancing light absorption at low solar angles (Falster and Westoby, 2003). Erect leaf stature is also associated with reduced susceptibility to photoinhibition and reduced risk of overheating (King, 1997; Murchie et al., 1999; Werner et al., 2001; Falster and Westoby, 2003; Burgess et al., 2015). As such, the erect ideotype is predicted to be most effective in low latitudes but it has also been found to be productive in high latitudes (Reynolds and Pfeiffer, 2000; Peng et al., 2008 and references within). However, despite this, there is still variation in crop morphology and the erect ideotype is not widespread in many species. As such, there may still be potential for yield improvement by alteration of canopy architectural characteristics (Reynolds et al., 2000; Khush, 2005; Khan et al., 2015; Rötter et al., 2015).

There is currently no method for producing accurate high-resolution 3D architectural reconstructions of entire field grown crop canopies via imaging techniques for modeling purposes. This is largely due to problems of occlusion at high leaf densities

i.e., of being unable to produce images of leaves deep within the canopy using the most common optical techniques. Being able to do so would be highly advantageous for testing hypothesis about canopy structure within fundamental or applied research. However, advances in hardware and image processing have led to new methods for capturing and evaluating plant architecture. These methods have been used for numerous purposes including both plants grown in pots and those grown under field conditions (e.g., Falster and Westoby, 2003; Godin and Sinoquet, 2005; Watanabe et al., 2005; Quan et al., 2006; Sinoquet et al., 2007; Zheng et al., 2008; Burgess et al., 2015). Whilst previous studies have attempted to look at the relationship between canopy architecture and the light environment (e.g., Zheng et al., 2008; Song et al., 2013), these have been restricted due to the relatively inaccurate manual reconstruction and modeling techniques used and the limited genetic variation and architectural types studied. Architectural traits are inherently linked to the resulting light environment and since photosynthetic rate is strongly light-dependent it therefore follows that photosynthetic rate will be dependent upon architecture.

To overcome the limitations of previous studies we used a new approach for high resolution 3D reconstruction of crop plants (Pound et al., 2014; Burgess et al., 2015) to investigate fundamental structure-function canopy properties. This is not a high throughput technique but rather uses individual plants extracted from field grown plots to generate highly accurate representations that can then be used to populate a canopy *in silico* for ray-tracing and photosynthesis modeling. The parental lines used for the creation of multi-parent advanced generation inter-cross (MAGIC) populations in rice (Bandillo et al., 2013) were selected for analysis within this study. These lines have a well-researched genetic background and contain desirable traits for yield, grain quality, and biotic and abiotic stress resistance (more details on each line are given in **Supplementary Table S1**). Furthermore, the contrasting origin of each line means that they are cultivated in diverse habitats with different stressors and constraints. The initial phase of this study involved a preliminary small-scale screening experiment to assess differences in terms of architectural and physiological features for 15 of the lines (referred to here as M1–M15 in **Supplementary Table S1**). Four of these lines, Shan-Huang Zhan-2 (SHZ-2), IR4630-22-2-5-1-3, WAB 56-125, and Inia Tacuari (referred to here as M2, M4, M11, and M13, respectively), plus the Philippine high-yielding variety IR64 were chosen for an in depth physiological study. These lines were chosen due to their differences in a number of features including leaf area index (LAI; leaf area per unit ground area), chlorophyll a:b ratios (a reliable indicator of shade acclimation state, reflecting the proportion of chlorophyll in light harvesting complexes), chlorophyll content and physical appearance. The aims are to: (1) assess the method for image based reconstruction on genetically variable rice plants grown in simulated field environment (see materials and methods); (2) test the hypothesis that there are common links between canopy architecture and photosynthetic traits across genetically diverse rice cultivars (such as leaf angle, light distribution, and photosynthetic capacity) and; (3) test the hypothesis that canopy-induced dynamic light properties are associated with

the acclimation status of leaves in genetically diverse cultivars. The latter uses a new empirical acclimation model which predicts the optimal  $P_{max}$  (if light were the sole determinant; Retkute et al., 2015). Acclimation is a process whereby leaves adjust their photosynthetic capacity, dark respiration and light compensation point according to long term changes in the light environment. However, the ability to acclimate optimally in fluctuating conditions has not been fully tested (Anderson et al., 1995; Murchie and Horton, 1997, 1998; Yano and Terashima, 2001; Walters, 2005; Athanasiou et al., 2010; Retkute et al., 2015).

## MATERIALS AND METHODS

### Plant Material and Growth

The preliminary screening used 15 of the possible 16 parental lines from a MAGIC rice population (Bandillo et al., 2013; details given in **Supplementary Table S1** with results of the screening in **Supplementary Table S2**). Seeds were sown into module trays containing Levington Module compost [N (96 ppm), P (49 ppm), K (159 ppm)] mixed with 30% sand by volume in the FutureCrop Glasshouse facilities, University of Nottingham Sutton Bonington Campus ( $52^{\circ}49'59''$  N,  $1^{\circ}14'50''$  W), UK on the 7th May 2015. The FutureCrop Glasshouse is a south-facing glasshouse designed and built by CambridgeHOK (Brough, UK) for the growth of crop stands within a controlled environment. It consists of a concrete tank  $5 \times 5 \times 1.25$  m positioned at ground level. The tank is filled entirely with a sandy loam soil, extracted from local fields, and sieved through a fine mesh. The seedlings were transplanted into microplots (containing  $5 \times 5$  plants with  $10 \times 10$  cm spacing between adjacent plants;  $100$  plants  $m^{-2}$ ) within soil beds 7 days after root establishment. For the preliminary screen, key measurements were made 55–60 days after transplanting (DAT), corresponding to a vegetative growth phase (**Supplementary Table S2**). Ten centimeters of spacing is consistent with rice field planting guidelines ([www.irri.org](http://www.irri.org)). Following the preliminary screening, four lines; Shan-Huang Zhan-2 (SHZ-2), IR4630-22-2-5-1-3, WAB 56–125, and Inia Tacuari (referred to here as M2, M4, M11 and M13, respectively), were selected for the in depth study as well as the popular Philippine variety IR64, from IRRI. Selection was made largely on the basis of contrasting architecture including leaf area index (LAI; leaf area per unit ground area), chlorophyll a:b ratios and content plus physical appearance. This selection also represents rice from diverse origins (**Supplementary Table S1**) and genetic backgrounds (M2, M4 and IR64 of indica and M11 plus M13 of japonica). The seeds were sown into module trays on the 15th October 2015 and transplanted into replicate microplots of  $6 \times 6$  plants (10 cm spacing as above) using a completely randomized design. Plots were arranged in a  $3 \times 4$  design that minimized edge effects and plants on edge of plots were not used in this study. The glasshouse conditions were kept consistent for both the screening and the in depth study. Irrigation was supplied using drip irrigation for 15 min, twice daily. Sodium (Son T Agro, Philips) lamps provided additional lighting whenever the photosynthetically active radiation (PAR) fell below  $300 \mu\text{mol m}^{-2} \text{ s}^{-1}$  and a 12 h photoperiod (07:00–19:00) was maintained using blackout blinds. A temperature of

$28 \pm 3^{\circ}\text{C}$  and relative humidity (RH) of 50–60% was maintained throughout. Nutrient composition of plots was measured by sampling soil at leaf 3, during the vegetative growth stage. Consequently Yara Milla complex fertilizer (applied at rate equivalent to  $50 \text{ kg ha}^{-1}$  of N plus micronutrients) was applied to the plots, 80 days after transplanting (DAT).

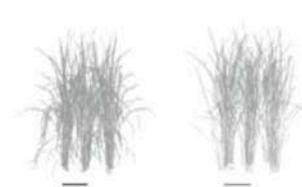
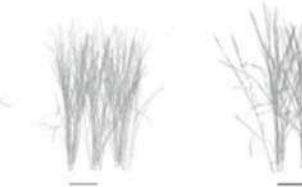
### Physiological Measurements: In Depth Study

In depth measurements were made at two different growth stages: 45 and 85 DAT, which correspond to an early (prior to full canopy development) and late (full canopy development prior to flowering) vegetative phase. Here, we refer to these stages as growth stage 1 (GS1) and growth stage 2 (GS2), terms used in this study only. Five replicate measurements of plant height per plot were taken weekly, from four DAT. Five replicate measurements per plot were taken for tiller numbers at each of the growth stages. Three replicate plants per line were taken for leaf width, leaf area, fresh, and dry weight measurements at each growth stage. Individual plant dry weight and area was analyzed by passing material through a leaf area meter (LI3000C, Licor, Nebraska) and drying in an oven at  $80^{\circ}\text{C}$  for 48 h or until no more weight loss was noted. Measured LAI (leaf area per unit ground area:  $\text{m}^2 \text{ m}^{-2}$ ) was calculated as the total area (leaf + stem) divided by the area of ground each plant covered (distance between rows  $\times$  distance within rows) and averaged across the replicate plants. A Walz MiniPam fluorometer was used to measure dark-adapted values of Fv/Fm in the glasshouse at mid-day. Leaves were dark adapted using clips (DLC-08; Walz) for at least 20 min and Fo and Fm were measured by applying a saturating pulse (0.8 s,  $6,000 \mu\text{mol m}^{-2} \text{ s}^{-1}$ ). Five replicate measurements on different leaves were taken per plot. Chlorophyll a and b content and ratios were determined through chlorophyll assays corresponding to GS2. Frozen leaf samples of known area were ground in 80% acetone, centrifuged for 5 min at 1,600 g, and the absorbance (at 663.6 and 646.6 nm) of the supernatant was measured using a spectrophotometer according to the method of Porra et al. (1989).

### Imaging and Ray Tracing

3D analysis of a plant from each plot (i.e., three replicate plants per line which accounts for any within-genotype variability caused by environment) was made according to the protocol of Pound et al. (2014) based on stereo-imaging in the in-depth analysis (GS1 and GS2). Briefly, plants were removed carefully from the central part of the plots (with roots and soil). They were positioned on a calibration target and turntable. SLR cameras were placed at three positions and 45–60 images recorded as the plant was carefully rotated. Automated reconstruction of a 3-D point cloud and conversion of this to a 3D canopy representation made up of 2D flat leaves took place using existing software described in Pound et al. (2014). These reconstructions were duplicated and rotated to form a  $3 \times 3$  canopy grid (with set 10 cm spacing between plants), with the same leaf area index (LAI) as the measured plants (see **Table 1**). The LAI of each reconstructed canopy was calculated as the area of mesh inside the ray tracing boundaries divided by the ground

**TABLE 1 |** Canopy reconstructions and description.

	Line							
	M2	M4	M11	M13	IR64			
<b>GS1</b>								
<b>Plant Height (cm)</b>	36.60±3.12	47.07±0.47	52.20±3.14	59.53±1.40	49.4±1.06	<b>54.30</b>	<b>0.001</b>	3.250
<b>No. of Tillers</b>	4.47±0.18	4.47±0.18	3.07±0.18	2.67±0.07	4.93±0.24	<b>3.92</b>	<b>&lt;0.001</b>	0.249
<b>No. of Leaves</b>	22.33±1.20	18.67±1.20	15.67±2.4	10.33±0.67	25.67±0.67	<b>18.53</b>	<b>&lt;0.001</b>	1.955
<b>Leaf Width (cm)</b>	0.90±0.03	0.92±0.02	1.05±0.00	1.08±0.03	0.87±0.02	<b>0.963</b>	<b>&lt;0.001</b>	0.032
<b>Dry Weight (g)</b>	1.22±0.13	1.31±0.07	1.69±0.07	1.20±0.07	1.52±0.30	<b>1.39</b>	<b>0.207</b>	0.223
<b>Measured LAI</b>	2.19±0.18	2.22±0.26	2.45±0.13	1.86±0.48	2.82±0.09	<b>2.31</b>	<b>0.206</b>	0.378
<b>Reconstruction LAI</b>	2.21	2.22	2.47	1.86	2.83			
<b>GS2</b>								
<b>Plant Height (cm)</b>	54.87±3.89	65.00±0.92	69.27±1.85	70.13±1.70	67.07±1.88	<b>65.30</b>	<b>0.005</b>	3.210
<b>No. of Tillers</b>	5.40±0.31	4.87±0.52	4.13±0.47	2.53±0.18	5.67±0.29	<b>4.52</b>	<b>&lt;0.001</b>	0.528
<b>No. of Leaves</b>	28.67±2.19	27.33±2.6	21.33±3.33	14.00±2.52	51.00±3.06	<b>28.50</b>	<b>&lt;0.001</b>	3.920
<b>Leaf Width (cm)</b>	0.97±0.03	1.05±0.03	1.18±0.04	1.38±0.02	1.02±0.04	<b>1.12</b>	<b>&lt;0.001</b>	0.049
<b>Dry Weight (g)</b>	3.13±0.56	4.70±0.46	3.47±0.80	5.2±0.59	6.60±0.40	<b>4.62</b>	<b>0.011</b>	0.817
<b>Measured LAI</b>	4.72±0.83	5.46±0.68	4.12±0.59	3.67±0.34	8.75±3.11	<b>5.35</b>	<b>0.206</b>	2.125
<b>Reconstruction LAI</b>	4.75	5.44	4.18	3.68	8.69			

The means of three plots are shown with standard errors of the mean. P-values correspond to ANOVA. Plants were imaged and reconstructed as a single plant according to the protocol of Pound et al. (2014). These were then duplicated and rotated and arranged on a 3 × 3 canopy grid. Rotating the plants enabled a similar leaf area index (LAI) to be achieved. Measured LAI was calculated as the total area (leaf + stem), using a leaf area meter (LI3000C, Licor, Nebraska), divided by the area of ground each plant covered (distance between rows × distance within rows). The reconstructed LAI was calculated as mesh area inside the designated ray tracing boundaries (see Section Imaging and Ray Tracing). Following imaging and measurement of leaf area, dry weights were calculated. The resting plant height (the plant height in the natural position, i.e., taking into account leaf curling), number of tillers, number of leaves, and leaf width of five plants per plot was calculated. Growth stage 1 corresponds to 45 DAT and 2 at 85 DAT. M2, M4, M11, and M13 refer to Shan-Huang Zhan-2 (SHZ-2), IR4630-22-2-5-1-3, 157 WAB 56-125, and Inia Tacuari, respectively. Growth stage 1 corresponds to 45 DAT and 2 at 85 DAT.

area. A forward ray-tracing algorithm, fastTracer (fastTracer version 3; PICB, Shanghai, China from Song et al., 2013), was used to calculate diurnal change in total light per unit leaf area throughout the canopies. Latitude was set at 14.2 (for the Philippines), atmospheric transmittance 0.5, light scattering 7.5%, light transmittance 7.5%, days 344 (GS1 10th December), and 21 (GS2 21st January). The diurnal course of light intensities over a whole canopy was recorded at 1 min intervals. The aim was to study the effect of canopy architecture on the resultant light environment and the impact on whole canopy photosynthesis thus the same parameters for ray tracing were used for each of the canopies, despite the diverse origin of each of the lines (see Supplementary Table S1).

## Gas Exchange

Photosynthesis-light response curves (LRC) and Photosynthesis vs. Ci (leaf internal CO<sub>2</sub> concentration; ACi) curves were taken via infra-red gas exchange (IRGA). Leaves were not dark-adapted prior to measurements. LRCs were taken at GS1 and 2 whereas ACi curves were taken at GS1 only. Leaf gas exchange

measurements (LRC and ACi) were taken with a LI-COR 6400XT infra-red gas-exchange analyser (LI-COR, Nebraska). The block temperature was maintained at 30°C using a flow rate of 500 ml min<sup>-1</sup> and ambient humidity. For light response curves, light was provided by a combination of in-built red and blue LEDs. Illumination occurred over a series of 12 photosynthetically active radiation values (low to high), between 0 and 2,000 μmol m<sup>-2</sup> s<sup>-1</sup>, with a minimum of 2 min and maximum of 3 min at each light level at two different canopy heights; top (center of flag leaf) and bottom (25% of full canopy height). Therefore, the positions were not affected by canopy height. Separate induction curves showed that this was sufficient to fully induce leaves. For the A-Ci curves; leaves were exposed to 1,500 μmol m<sup>-2</sup> s<sup>-1</sup> throughout. They were placed in the chamber at 400 p.p.m. CO<sub>2</sub> for a maximum of 2 min and then CO<sub>2</sub> was reduced stepwise to 50 p.p.m. CO<sub>2</sub> was then increased to 1500 p.p.m. again in a stepwise manner. Two replicates were taken per layer per treatment plot for both sets of measurements apart from LRCs for GS2, which has five replicates overall for each of the five varieties.

## Statistical Analysis

Analysis of variance (ANOVA) was carried out using GenStat for Windows, 17th Edition (VSN International Ltd.). Data was checked to see if it met the assumption of constant variance and normal distribution of residuals. A correlation matrix was used to investigate the relationships between different physiological traits.

## Modeling

All modeling was carried out using Mathematica (Wolfram).

$$F_{NRH}(L, \phi, \theta, P_{max}, \alpha) = \frac{\phi L + (1 + \alpha) P_{max} - \sqrt{(\phi L + (1 + \alpha) P_{max})^2 - 4\theta\phi L (1 + \alpha) P_{max}}}{2\theta} - \alpha P_{max} \quad (5)$$

Cumulative leaf area index (cLAI; leaf area per unit ground area as a function of depth) was calculated from each of the canopy reconstructions. cLAI was not measured in this study but previous work has validated this method using manual measurements of leaf area (Pound et al., 2014). Leaves are represented here as a series of small 2D triangles. For each depth ( $d$ ; distance from the highest point of the canopy), all triangles with centers lying above  $d$  were found (Equation 1).

$$d_i = \max_{j=1,2,3; 1 \leq i \leq n} z_i^j - (z_i^1 + z_i^2 + z_i^3) / 3 \quad (1)$$

The sum of the areas of these triangles was calculated and divided by the ground area. The cumulative LAI as a function of depth through the canopy was calculated using Equation (2).

$$cLAI(d) = \frac{\sum_{i=1}^n I(d_i \leq d) S_i}{(\max_{1 \leq i \leq n} x_i - \min_{1 \leq i \leq n} x_i)(\max_{1 \leq i \leq n} y_i - \min_{1 \leq i \leq n} y_i)} \quad (2)$$

where  $I(A) = 1$  if condition  $A$  is satisfied and  $S_i$  is the area of a triangle  $i$ .

The light extinction coefficient of the canopy was calculated using the 3D structural data and the light distribution obtained from ray tracing. In order to calculate fractional interception (FI) within a canopy as a function of depth at time  $t$ , all triangles lying above depth,  $d$ , were identified (Equation 1). Their contribution to intercepted light was then calculated by multiplying PPFD received per unit surface area (ray tracing output) by the area of triangle. The light intercepted was summed for all triangles above the set  $d$ , and divided by light intercepted by ground area according to Equation (3).

$$F(d, t) = \frac{\sum_{i=1}^n I(d_i \leq d) S_i L_i(t)}{L_0(t) * \text{ground area}} \quad (3)$$

where  $L_0(t)$  is light received on a horizontal surface with a ground area  $(\max_{1 \leq i \leq n} x_i - \min_{1 \leq i \leq n} x_i)(\max_{1 \leq i \leq n} y_i - \min_{1 \leq i \leq n} y_i)$ , and  $L_i(t)$  is light intercepted by a triangle  $i$ .

The light extinction coefficient,  $k$ , was calculated by fitting (by least squares) the function,

$$f(x) = a(1 - e^{-kx}) \quad (4)$$

to the set of points  $\{cLAI(d), F(d, t)\}$  calculated by varying depth from 0 to the height at total cLAI with step  $\Delta d = 1$  mm,  $a$  in Equation (4) is a fitted parameter.

The response of photosynthesis to light irradiance,  $L$ , was calculated using a non-rectangular hyperbola given by Equation (5):

Values for  $P_{max}$  were determined from leaf gas exchange measurements (see Section Gas Exchange). The value of  $\alpha$  was obtained by fitting a line of best fit between all measured  $P_{max}$  and  $Rd$ -values. All other parameters (e.g.,  $P_{max}$ ,  $\Phi$ , and  $\theta$ ) were estimated from the light response curves for three canopy layers using the Mathematica command *FindFit*.

As each canopy was divided into two layers, and each triangle from the digital plant reconstruction was assigned to a particular layer,  $m$ , according to the triangle center (i.e. with triangle center between upper and lower limit of a layer depth). Carbon gain per unit canopy area was calculated as daily carbon assimilation over a whole canopy divided by the total surface area of the canopy according to Equation (6).

$$C = \frac{\sum_{i=1}^n P_i}{\sum_{i=1}^n S_i} \quad (6)$$

Total canopy light interception per unit leaf area over whole day was calculated according to Equation (7).

$$TL_{LA} = \frac{\sum_{i=1}^n S_i \int_6^{18} L_i(t) dt}{\sum_{i=1}^n S_i} \quad (7)$$

where  $S_i$  is the area of triangle  $i$ .

An empirical model of acclimation was employed to predict the distribution of optimal  $P_{max}$ -values throughout each of the canopies. Details of the model can be found in Retkute et al. (2015). The model can be used to predict the maximum photosynthetic capacity,  $P_{max}^{opt}$ , as the  $P_{max}$  that represents maximal carbon gain at a single point within the canopy, based on the light pattern that point has experienced (i.e., using the light pattern output from ray tracing). This was predicted across 250 canopy points, thus leading to distribution of  $P_{max}^{opt}$ -values throughout each of the canopies. The canopy locations were chosen as a subset of triangles that were of similar size (i.e., area) and constitute a representative sample distribution throughout canopy depth.

Carbon gain,  $C$  ( $\text{mol m}^{-2}$ ) was calculated over a given time period (e.g., daily)  $t \in [6, 18]$  (Equation 8).

$$C(L(t), P_{max}) = \int_6^{18} P(L(t), P_{max}) dt \quad (8)$$

Experimental data indicates that the response of photosynthesis to a change in irradiance is not instantaneous and thus to incorporate this into the model Retkute et al. (2015) introduced a time-weighted average for light (Equation 9).

$$L_\tau(t) = \frac{1}{\tau} \int_{-\infty}^t L(t') e^{-\frac{|t-t'|}{\tau}} dt' \quad (9)$$

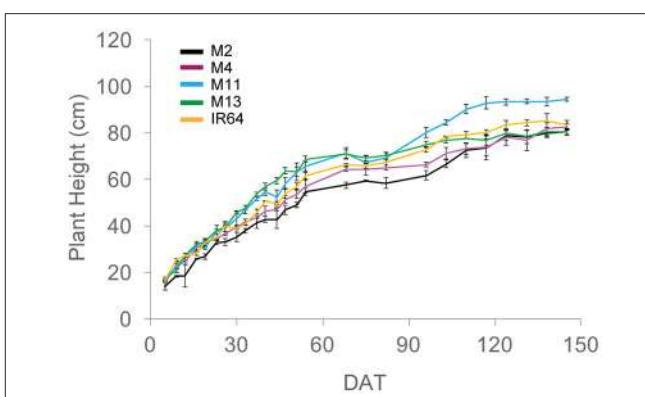
This effectively accounts for photosynthetic induction state, which is hard to quantify *in situ* as it varies according to the light history of the leaf. The more time recently spent in high light, the faster the induction response, thus the time-weighted average effectively acts as a “fading memory” of the recent light pattern using an exponentially decaying weight. If  $\tau = 0$  then a plant will able to instantaneously respond to a change in irradiance, whereas if  $\tau > 0$  the time-weighted average light pattern will relax over the timescale  $\tau$ . Within this study,  $\tau$  was fixed at 0.2 (unless otherwise stated) in agreement with previous studies and fit with past experimental data (Pearcy and Seemann, 1990; Retkute et al., 2015) and measurements of induction state in rice leaves. The time-weighted average only applies to the transition from low to high light; from high to low, response is instantaneous and does not use the weighted average (see **Supplementary Figure S1**). The model was parameterised using the convexity and dark respiration values taken from the fitted LRCs. A moving average of the  $P_{max}$  throughout canopy height was fitted using the Mathematica command *MovingAverage* to give an approximate relationship between canopy height and optimal  $P_{max}$  based on the light environment.

## RESULTS

### Architectural Features

#### Manual Measurements

A summary of the key architectural features is given in **Table 1** (see **Supplementary Table S2** for the initial screening experiment). Similarities can be seen between the key architectural features: the initial screening experiment and the in-depth study (**Table 1** and **Supplementary Table S2**) however the variation seen between the lines was reduced in the second, in depth experiment. For the rest of the paper, only data from the in-depth study will be considered. Plant height varied between lines in both growth stages ( $P = 0.001$  for GS1 and  $P = 0.005$  for GS2), with M2 the shortest and M13 the tallest of the five lines. The change in plant height over the course of the experiment is given in **Figure 1**. One-hundred and fifty DAT is full maturity and just before harvest and the increase in height after 90 DAT likely corresponds to stem elongation. Height is a relevant architectural trait since upland cultivars can be taller than lowland, thought to be a trait associated with weed competition. Here, M11 has aerobic adaptation and M13 is NERICA i.e., derived partly from *Oryza glaberrima*. Since plant height infers greater stem and leaf sheath extension it may be an important trait in determining partitioning, available leaf area and productivity in a given environment. Leaf blade width differed between the lines at each growth stage ( $P < 0.001$  GS1 and 2) with M11 and M13 exhibiting the widest leaf

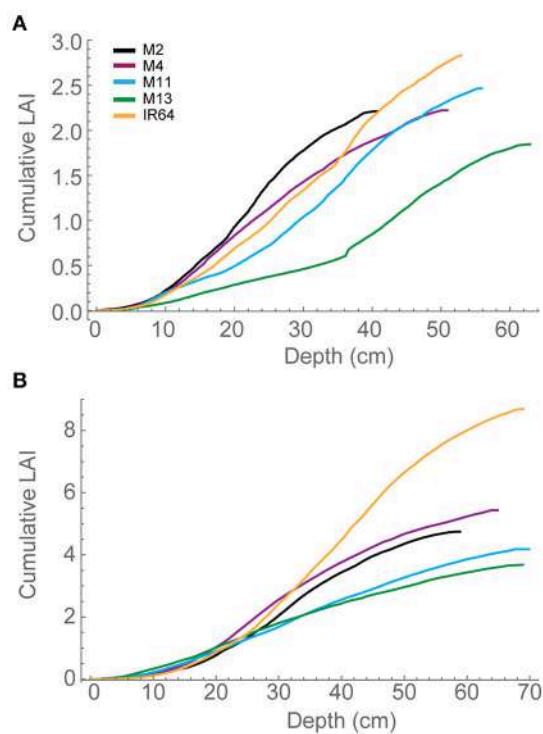


**FIGURE 1 |** Plant height over the course of the experiment, calculated as the average of five measurements per plot. The means of three plots are shown with standard errors of the mean. M2, M4, M11, and M13 refer to Shan-Huang Zhan-2 (SHZ-2), IR4630-22-2-5-1-3, 157 WAB 56-125, and Inia Tacuari, respectively.

blades (**Table 1**). Leaf number and tiller number also differed significantly between the lines ( $P < 0.001$  both growth stages) with M13 containing the fewest number of leaves and IR64 the greatest, however there was no significant difference in leaf area index (LAI) at either growth stage (**Table 1**). Dry matter was not significantly different between lines (**Table 1**) indicating that modeled photosynthesis was not a reliable predictor of biomass production in this case. This could be caused by a number of factors including lack of inclusion of partitioning of biomass to roots or measuring photosynthesis at a limited number of stages.

#### Modeled Data

Each plant within the *in silico* canopy was rotated around the vertical axis such that the LAI inside the ray tracing boundaries was consistent with measured data (**Table 1**; see Section Materials and Methods). Previous papers have validated the modeling using measured data of LAI and extinction coefficients (Burgess et al., 2015). Cumulative leaf area index (cLAI) was calculated through canopy depth (i.e., from top-down; see Section Modeling) for each of the canopies at each growth stage (see **Figures 2A,B**). A curve was deliberately not fitted because the reconstruction and modeling approach used within this study permits the actual relationship between LAI and depth in the canopy to be depicted, without the need for curve fitting. Generally, a sigmoidal response was seen for most genotypes with a more rapid accumulation of leaf area toward the center of the canopy. At GS1, M2, and M13 show the greatest difference among lines in terms of the position of accumulation of LAI according to depth (distance from the top of the canopy) with the latter accumulating more biomass in the bottom half of the canopy (**Figure 2A**). At GS2 (**Figure 2B**) this pattern is not pronounced with other lines showing a similar increase in cLAI up to ~20 cm depth. From here on, differences are shown with M11 and M13 exhibiting least accumulation of leaf material and IR64 exhibiting the greatest. This variation is consistent with total measured LAI-values, with IR64 exhibiting a much higher overall



**FIGURE 2 |** Modeled cLAI, the area of leaf material (or mesh area) per unit ground as a function of depth through the canopy (i.e., distance from the top) at 12:00 h for (A) GS1 and (B) GS2. M2, M4, M11, and M13 refer to Shan-Huang Zhan-2 (SHZ-2), IR4630-22-2-5-1- 3, 157 WAB 56-125, and Inia Tacuari, respectively.

LAI compared to the other lines (Table 1), although according to ANOVA on the measured leaf area, this is not significant.

These distinctive patterns are partly as a result of architecture and arrangement, specifically angles of the leaves, within each canopy. This technique allows automatic and rapid calculation of leaf angle of every triangle in the reconstruction. Leaf angle distributions were calculated (Burgess et al., 2015) for each canopy and averaged at each canopy depth (see Section Modeling; Figures 3A,B), where a leaf inclination angle toward 0 indicates a more horizontal leaf and an inclination angle of 90 indicates a more vertical leaf. M2, M4, and IR64 lines exhibited a trend toward more horizontal leaves at base of canopy at both growth stages 1 and 2, with M11 and M13 more vertical stature.

## Light Environment

### Modeled Data

To explore interactions between depth and light interception, modeled fractional interception (FI) was calculated as a function of depth (Figures 4A,B). This enables the interception to be calculated at a resolution of 1 mm throughout the canopy. Generally, the pattern was similar to that of modeled LAI. At GS1 (Figure 4A), M2, and M4 are achieving ~60% of interception within the top 25 cm of the canopy. This can be compared to M13, which exhibits a near linear relationship between FI and canopy depth. By GS2 (Figure 4B), the lines exhibit a more

similar interception within the top 20 cm of the canopy but a greater variation in the bottom layers in the canopy. M2, M4, and IR64 achieve the greatest FI and M11 and M13 the lowest.

We hypothesize that leaf angle will be related to vertical FI and LAI distribution: we note that toward the top of the canopy, leaves tend to be more horizontal (i.e., angles approaching 0) for those lines with a higher LAI (Figures 2, 3), and this contributes to a higher interception of light (Figure 4). In the lines studied here, erectness does not seem to be associated with a higher LAI.

## Photosynthesis

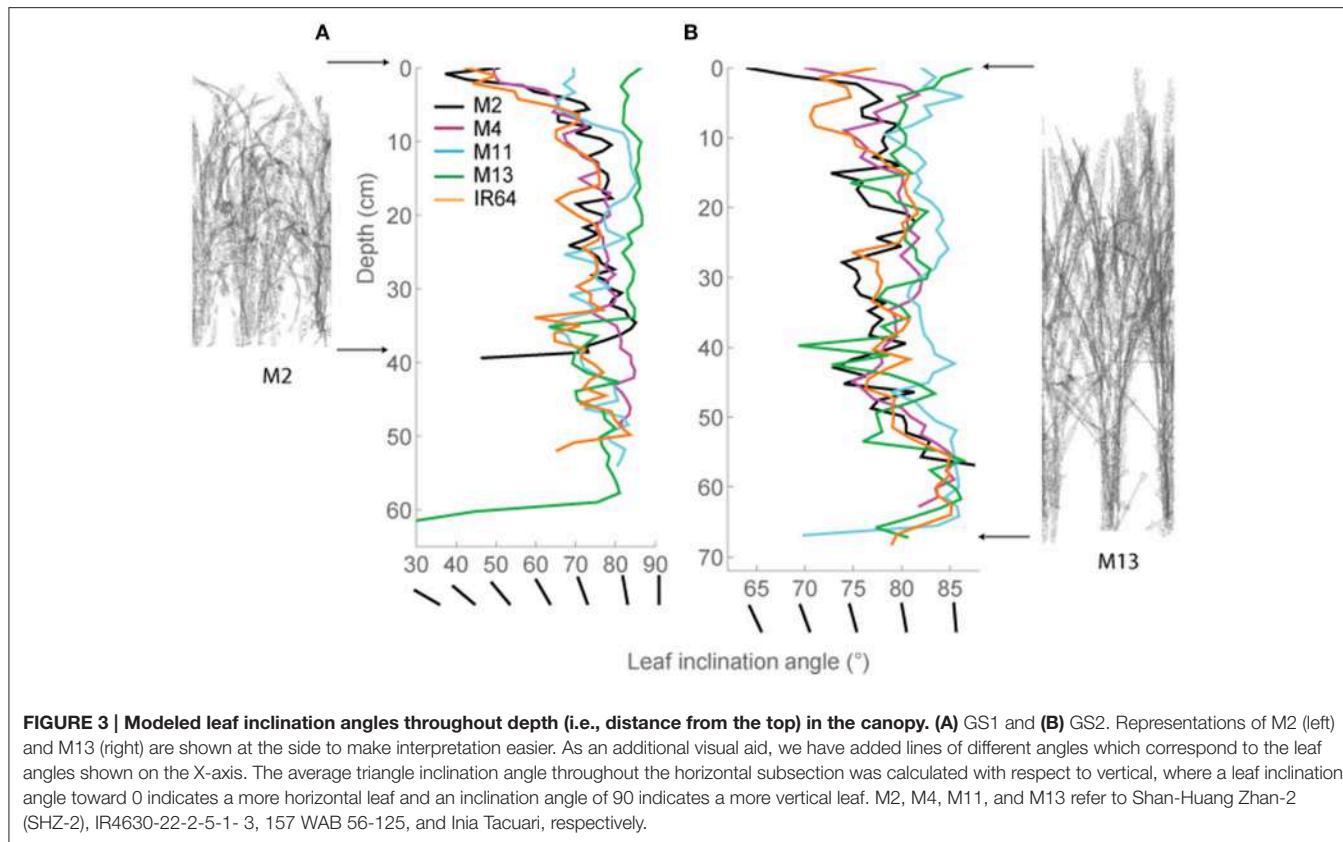
### Measured Data

There were no significant differences between any of the ACi curve parameters ( $V_{max}$ , J, and TPU) at either growth stage (see Table 2). There was a significant difference in Chlorophyll a content ( $P = 0.034$ ) and total chlorophyll content ( $P = 0.041$ ) between the lines with M11 and M13 containing the highest levels and Chl a:b ratios showing little change (Table 3). The dark-adapted Fv/Fm measurement measured at the top of the canopy also shows significant differences between the lines at both growth stages under two different weather conditions, full sun and cloudy with supplementary lights, ( $P < 0.002$  for all) with the lowest Fv/Fm-value found in M2 (Table 4). This is in agreement with previous work on canopy architecture and susceptibility of plants to photoinhibition, whereby erect architectures are less susceptible to high light and have a higher Fv/Fm in accordance with Burgess et al. (2015). Lowered Fv/Fm are seen under high irradiance in healthy rice and wheat plants in the field and represent a decline in maximum photosystem II quantum yield, caused either by damage to reaction centers or another form of sustained quenching (Murchie et al., 1999; Burgess et al., 2015).

We assessed photosynthesis at different canopy layers and compared it to patterns of LAI accumulation above.  $P_{max}$  for the top layer varied between species for GS1 ( $P < 0.001$ ), with M13 having a higher  $P_{max}$  than M4, but not GS2 ( $P = 0.053$ ; Table 2). There was no significant difference in  $P_{max}$  for the bottom layer at either growth stage ( $P = 0.062$  for GS1 and  $P = 0.321$  for GS2). There were no apparent consistencies between canopy structure and distribution of  $P_{max}$  except that the highest  $P_{max}$ , and the largest decline in  $P_{max}$  for the top layer between GS1 and 2 is shown by M13; the line with the lowest cumulative LAI (Figure 5).

### Modeled Data

An empirical model of photosynthesis was employed to calculate the total canopy carbon gain per unit leaf area and per unit ground area (see Section Materials and Methods); results are presented in Table 2. For GS1, M13 exhibits the highest carbon gain per unit leaf area followed by M2 and M4, respectively, with IR64 showing the lowest value. For carbon gain per unit ground area, M13 remains the highest, followed by M2 and M11. This can be attributed to the higher  $P_{max}$  for that line, despite the reduced LAI. At GS2, all canopies show a reduced carbon gain per unit leaf area and increased carbon gain per unit ground area. This is presumably due to an increase in LAI of all canopies and a concurrent increase in proportion of shaded leaves. Per unit leaf area M11 and M13 show the highest



values of carbon gain and per unit ground area M11 is the highest, followed by M2 and M13. However, we saw only weak correlations between  $P_{max}$  and carbon gain per unit leaf area and ground area (**Supplementary Figure S2**).

Canopy structures result in dynamic fluctuations in light from solar movement. The different architectures studied here are likely to generate different characteristics of fluctuations, in addition to the light interception shown above (Burgess et al., 2015). The most appropriate approach is a functional analysis of this variation in dynamic light via the impact that it has on the predicted distribution of a modeled optimal  $P_{max}$ . This was calculated using an empirical model of acclimation (see Section Modeling; Retkute et al., 2015). The model takes into account the fluctuating light over a full day within the canopy and provides an optimal  $P_{max}$ ; the value of  $P_{max}$  that is optimized in terms of carbon gain for that particular light pattern, if light were the sole determinant, using the frequency and duration of high light periods. This differs from previous models that use integrated light over the whole day (e.g., Stegemann et al., 1999). Thus, the optimal  $P_{max}$  provides a means of analyzing both the frequency and duration of high light events in the canopy.

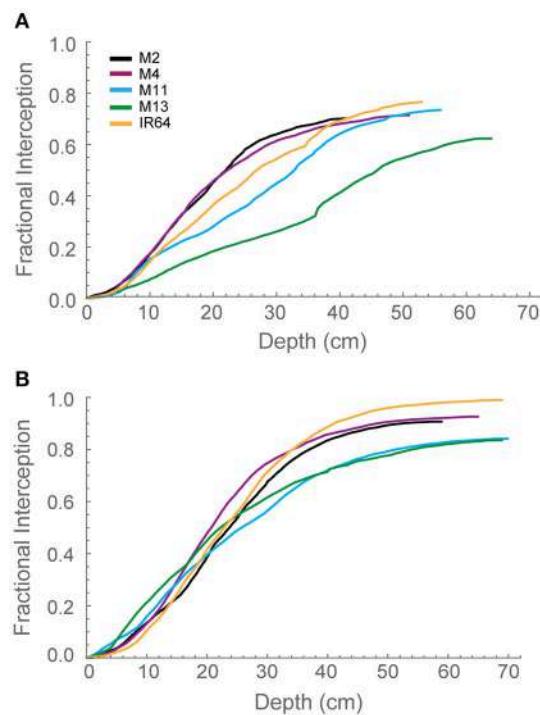
The distribution in optimal  $P_{max}$  for each of the canopies is given in **Figure 5**. This shows distinctive differences between the lines. At GS1, M4, M11, and IR64 show similar patterns for distribution of optimal photosynthetic capacity. These rank in the same order as FI and LAI for depths of 15–35 cm, with lower FI and LAI leading to higher optimal  $P_{max}$ , as

one would expect. M13 with its upright leaves and more open canopy shows a similar pattern for reduction in optimal  $P_{max}$  throughout but a greater value achieved at all canopy layers (depths) and a plateau in optimal  $P_{max}$  toward the top of the canopy. By GS2, differences between each of the canopies are less obvious. All canopies exhibit similar steep gradients within the top section of the canopy followed by a shallower gradient at the bottom of the canopy. IR64 has the lowest predicted optimal  $P_{max}$ -values of all canopies with the bottom ~40 cm under  $5 \mu\text{mol m}^{-2} \text{s}^{-1}$ . However, the ranking is still persistent, this time at lower canopy regions >40 cm. This indicates that optimal  $P_{max}$  can be consistently related to these features of canopy architecture. However, the relationship with leaf angle is less obvious. Measured  $P_{max}$ -values in the lower regions of the canopy were higher than the predicted optimal  $P_{max}$ .

## DISCUSSION

### Canopy Reconstructions

Plant canopies often consist of an assemblage of structurally diverse plants with particular spatial distributions of photosynthetic material. The way in which these photosynthetic surfaces intercept light energy and assimilate CO<sub>2</sub> is the basis for whole canopy photosynthesis, and thus the arrangement of plant material that optimizes light interception will inherently lead to increased productivity. If all incident light is absorbed (FI = 1) then whole canopy photosynthesis is a result of the efficiency of



**FIGURE 4 |** Modeled fractional interception as a function of depth in the canopy at 12:00 h for (A) GS1 and (B) GS2, using ray tracing data. Curves were calculated with step  $\Delta d = 1$  mm. M2, M4, M11, and M13 refer to Shan-Huang Zhan-2 (SHZ-2), IR4630-22-2-5-1-3, 157 WAB 56-125, and Inia Tacuari, respectively.

distribution of light across a particular LAI. The architectures of five diverse rice cultivars at two different growth stages were captured using a low-tech method for high-resolution canopy reconstruction. This reconstruction method has previously been shown to provide an accurate representation of the plants with replication of leaf area between 1 and 4% of that of measured data and accurate capture of leaf angles (Pound et al., 2014; Burgess et al., 2015). In combination with ray tracing using fastTracer3, the reconstruction method provides an accurate depiction of the light gradients found within real life canopies in field settings (Burgess et al., 2015). The structural differences (i.e., CLAI and leaf angle distributions) between diverse rice lines and their relationship to whole canopy photosynthesis can be explored in more depth using this modeling approach than would be possible using manual methods under field conditions.

## The Relationship between Canopy Architecture and Photosynthesis

To investigate the relationships between architectural features and photosynthetic traits, a correlation matrix was produced for manually measured data. Significant correlations (both positive and negative, given in bold) relating to canopy architectural features are given in Table 5. Among the factors that influence photosynthesis [here associated with  $P_{max}$  for the top (T) and bottom (B) canopy layers] are: tiller number, plant height, leaf number, and leaf width. However, these relationships are only

**TABLE 2 |** Parameters taken from ACi curve fitting at GS1 (45 DAT) using Sharkey et al. (2007) (fitting at 30°C).

Line	Layer	$V_{cmax}$	J	TPU
M2	Top	140.5 ± 13.4	187.6 ± 11.1	13.1 ± 0.7
M4		145.9 ± 18.0	202.7 ± 9.0	13.7 ± 0.6
M11		135.6 ± 12.0	195.8 ± 16.3	12.9 ± 1.0
M13		143.4 ± 12.3	186.9 ± 12.3	12.4 ± 0.5
IR64		134.8 ± 12.7	181.3 ± 9.2	12.0 ± 0.6
Mean		140	190.9	12.82
P		0.982	0.847	0.695
SED		22.23	20.45	1.21
M2	Bottom	120.4 ± 8.0	173.1 ± 9.1	11.5 ± 0.8
M4		131.4 ± 19.9	180.2 ± 11.8	11.8 ± 0.5
M11		127.3 ± 10.8	201.6 ± 24.9	13.0 ± 0.8
M13		141.2 ± 17.0	182.0 ± 6.9	11.6 ± 0.5
IR64		126.1 ± 15.7	166.3 ± 11.0	11.4 ± 0.9
Mean		129.3	180.6	11.83
P		0.905	0.606	0.551
SED		22.05	22.07	1.05

The means of six independent curves are shown with standard errors of the mean. P-value corresponds to ANOVA. M2, M4, M11, and M13 refer to Shan-Huang Zhan-2 (SHZ-2), IR4630-22-2-5-1-3, 157 WAB 56-125, and Inia Tacuari, respectively.

**TABLE 3 |** Chlorophyll content and chlorophyll a:b ratio at GS2 (85 DAT), top of canopy.

Line	Chl a ( $\mu\text{g}/\text{cm}^2$ )	Chl b ( $\mu\text{g}/\text{cm}^2$ )	Chl a+b ( $\mu\text{g}/\text{cm}^2$ )	Chl a:b
M2	36.10 ± 2.40	8.46 ± 0.55	44.56 ± 2.92	4.27 ± 0.08
M4	36.53 ± 2.71	8.93 ± 0.85	45.46 ± 3.43	4.19 ± 0.19
M11	45.67 ± 3.78	10.30 ± 0.80	55.98 ± 4.57	4.42 ± 0.07
M13	53.69 ± 2.61	11.70 ± 0.50	65.40 ± 3.08	4.58 ± 0.08
IR64	39.01 ± 1.71	9.19 ± 0.39	48.20 ± 2.06	4.25 ± 0.09
Mean	42.2	9.72	51.9	4.344
P	0.034	0.126	0.041	0.356
SED	5.28	1.20	6.41	0.20

The means of three plots are shown with standard errors of the mean. P-value corresponds to ANOVA. M2, M4, M11, and M13 refer to Shan-Huang Zhan-2 (SHZ-2), IR4630-22-2-5-1-3, 157 WAB 56-125, and Inia Tacuari, respectively.

significant at the first growth stage, not the second, indicating (i) the architecture at certain developmental stages (smaller plants) are more critical in determining photosynthesis characteristics, (ii) beyond a certain developmental stage, or a certain amount of leaf area, the levels of light inside the canopy are below a certain threshold so as to not significantly influence photosynthetic characteristics in particular acclimation to light intensity or (iii) photosynthetic performance is determined by factors other than architectural traits. Given the data concerning optimal  $P_{max}$  it seems possible that all of these suggestions could be contributing, as we explain below.

There is a positive correlation, although weak, between plant height and photosynthesis during GS1, which may be initially

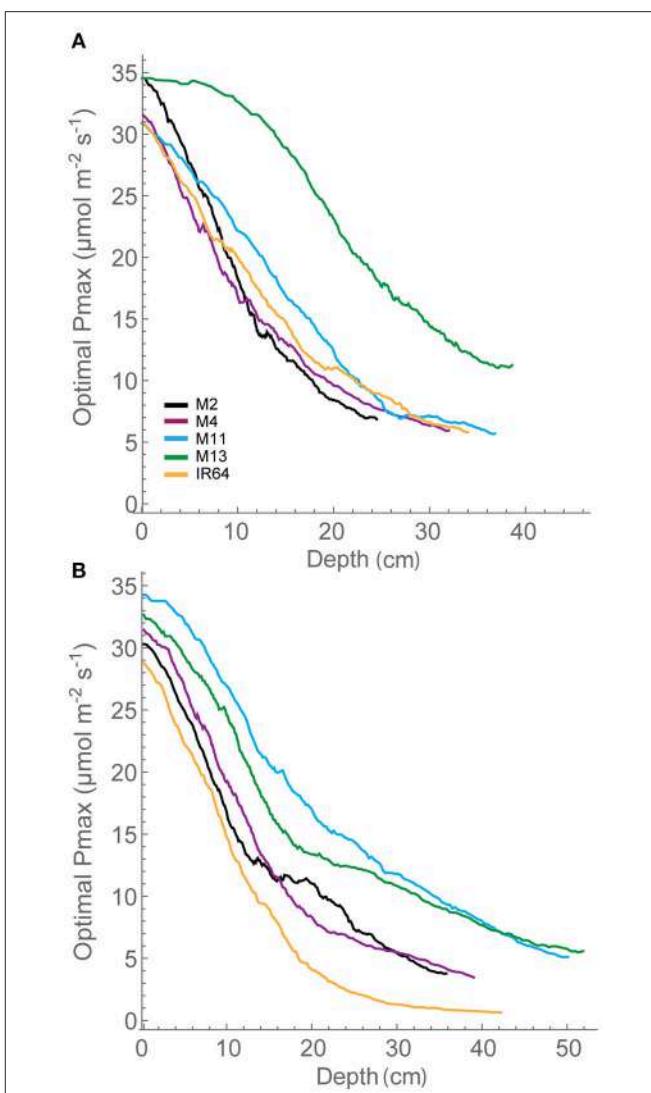
**TABLE 4 | Maximum quantum yield of PSII (Fv/Fm) measured after 20 min dark adaptation.**

Line	GS1		GS2	
	Full sun	Clouds + sup lights	Full sun	Clouds + sup lights
M2	0.748 ± 0.009	0.780 ± 0.010	0.788 ± 0.005	0.801 ± 0.004
M4	0.785 ± 0.004	0.805 ± 0.003	0.803 ± 0.007	0.830 ± 0.006
M11	0.813 ± 0.001	0.828 ± 0.004	0.810 ± 0.007	0.838 ± 0.006
M13	0.814 ± 0.013	0.848 ± 0.009	0.841 ± 0.007	0.846 ± 0.004
IR64	0.792 ± 0.007	0.816 ± 0.003	0.816 ± 0.003	0.826 ± 0.003
Mean	0.791	0.819	0.812	0.828
P	0.001	0.002	0.001	<0.001
SED	0.0115	0.0090	0.0084	0.0067

Five measurements were taken per plot. The means of three plots are shown with standard errors of the mean. Growth stage 1 corresponds to 45 DAT and 2 at 85 DAT. M2, M4, M11, and M13 refer to Shan-Huang Zhan-2 (SHZ-2), IR4630-22-2-5-1-3, 157 WAB 56-125, and Inia Tacuari, respectively.

contrary to what would be expected. Whilst extra height may provide an advantage during competition with shorter neighbors (such as weeds in Upland cultivars), it is also possible that height may increase self-shading over a greater surface area of the canopy, thus could intuitively reduce canopy productivity (diffuse light notwithstanding). Alternatively, plant height could be linked closely with leaf angles, with taller plants containing more elongated and erect leaves (as seen within our two tallest study lines: M11 and M13), which can lead to greater penetration of light throughout the canopy especially at mid-day, despite the greater height. Conversely, increased photosynthetic potential could provide plants with the means to achieve greater height. There is increasing evidence that tall plants provide greater sinks for photosynthate (i.e., within the stems) that can reduce limitations based on source-sink processes. This can lead to higher photosynthetic rates, at the leaf level, within taller crops. Therefore, the positive correlation between plant height and photosynthesis at GS1 could be a result of stem sink development during this stage.

To explore how canopy architecture influences photosynthesis and light interception at the whole canopy level, a line of best fit between measured LAI and modeled data were made (**Supplementary Figure S2**). Total canopy light interception is negatively correlated to measured LAI at both growth stages ( $R^2 = 0.981$  and  $0.967$  for GS1 and GS2, respectively). Similarly, there is also a negative correlation between measured LAI and carbon gain per unit leaf area ( $R^2 = 0.775$  and  $0.914$  for GS1 and GS2, respectively). Thus across the five rice lines, an increase in leaf area leads to a decrease in total light intercepted and in carbon gain per unit leaf area, possibly representing the “dilution effect” (Field and Mooney, 1983), although this does not translate to a significant decrease in measured  $P_{max}$  (**Table 6**), nor does it translate into an effect on carbon gain per unit ground area, with no clear relationship at either growth stage ( $R^2 = 0.311$  and  $0.091$  for GS1 and GS2, respectively).



**FIGURE 5 | Whole canopy acclimation model output.** The acclimation model was run at 250 locations throughout canopy depth to predict the optimal  $P_{max}$  at each location throughout canopy depth (i.e., from the top of the canopy) dependent upon the light environment that it experienced, calculated via ray tracing. A moving average has been fitted to the data. **(A)** GS1 and **(B)** GS2. M2, M4, M11, and M13 refer to Shan-Huang Zhan-2 (SHZ-2), IR4630-22-2-5-1-3, 157 WAB 56-125, and Inia Tacuari, respectively.

This lack of a relationship may be due to a high canopy density, high nutrient accumulation within the canopy leading to a large proportion of shaded leaves with a high respiratory burden (see below; Reich et al., 1998). It might be expected that leaf angle, canopy light interception and LAI distribution are closely related: indeed this was shown in **Figures 2–4** at depths between 10 and 30 cm (e.g., where M11 and M13 have lowest LAI and F but highest leaf angle). The conclusion is that a more upright leaf angle permits a greater light penetration but a greater LAI accumulation at GS2 lessens this effect. This is consistent with previous work (e.g., Song et al., 2013).

The dynamic light pattern cast by canopies presents a complex problem: how do leaves determine the optimal properties of a

**TABLE 5 |** The relationship between measured canopy architectural traits and photosynthesis: the sample correlation coefficient value taken from the correlation matrix output for select canopy architectural and physiological traits.

Growth Stage		Tiller number	Plant height	Leaf area	Leaf number	Leaf width (B)	Leaf Width (T)
1	Plant Height	<b>-0.638*</b>	–				
	Leaf Area	0.412	-0.221	–			
	Leaf Number	<b>0.890*</b>	<b>-0.629*</b>	<b>0.521*</b>	–		
	Leaf Width (B)	-0.240	<b>0.601*</b>	-0.045	-0.420	–	
	Leaf Width (T)	<b>-0.907*</b>	<b>0.635*</b>	-0.358	<b>-0.813*</b>	0.445	–
	$P_{max}$ (B)	<b>-0.574*</b>	<b>0.601*</b>	0.112	-0.425	0.513	<b>0.519*</b>
	$P_{max}$ (T)	<b>-0.721*</b>	<b>0.730*</b>	-0.189	<b>-0.624*</b>	<b>0.626*</b>	<b>0.737*</b>
	Fv/Fm Sun	<b>-0.561*</b>	<b>0.830*</b>	0.066	<b>-0.585*</b>	0.480	<b>0.555*</b>
	Fv/Fm Cloudy	<b>-0.755*</b>	<b>0.881*</b>	-0.150	<b>-0.692*</b>	<b>0.589*</b>	<b>0.675*</b>
	Fl/Height	0.101	<b>-0.713*</b>	-0.012	0.158	<b>-0.578*</b>	-0.2145
2	Leaf Area	0.389	-0.053	–			
	Leaf Number	<b>0.689*</b>	-0.166	<b>0.663*</b>	–		
	Leaf Width (B)	<b>-0.615*</b>	0.408	-0.307	<b>-0.627*</b>	–	
	Leaf Width (T)	<b>-0.819*</b>	<b>0.673*</b>	-0.413	<b>-0.652*</b>	<b>0.683*</b>	–
	$P_{max}$ (B)	<b>-0.524*</b>	<b>0.587*</b>	-0.357	-0.645	<b>0.706*</b>	<b>0.825*</b>
	$P_{max}$ (T)	-0.311	0.453	-0.088	0.165	-0.065	0.176
	Fv/Fm Sun	<b>-0.736*</b>	0.486	-0.040	-0.195	0.294	<b>0.694*</b>
	Fv/Fm Cloudy	<b>-0.709*</b>	<b>0.661*</b>	-0.236	-0.382	0.441	<b>0.734*</b>
	Chl a	<b>-0.752*</b>	<b>0.5645*</b>	-0.251	-0.470	0.401	<b>0.689*</b>
	Chl b	<b>-0.692*</b>	<b>0.619*</b>	-0.290	-0.453	0.329	<b>0.643*</b>
	Total Chl	<b>-0.7467*</b>	<b>0.5772*</b>	-0.2587	-0.4702	0.3913	<b>0.6855*</b>
	Fl/height	-0.2241	<b>-0.5415*</b>	-0.1975	-0.3912	0.2445	-0.0507

Growth stage 1 corresponds to 45 DAT and 2 at 85 DAT. (T) corresponds to measurements from the top canopy layer and (B) from the bottom canopy layer. Fl/Height refers to fractional interception as a function of height throughout the canopy. Significant correlations are given in bold, \*indicates  $P < 0.05$ .

Correlations based on plot means. Dry weight was not significantly correlated to any trait and so is not shown.

light response curve for a given time period? We used a model that predicts the optimal  $P_{max}$  based on ray tracing throughout the canopy depth. The optimal  $P_{max}$  distribution (**Figure 5**) follows a similar pattern (in terms of ranking responses among lines) to LAI and FI at the first growth stage. The ranking similarity is not so clear in the second, see above comment regarding  $P_{max}$  measurements. The differences between each of the lines, particularly at the first growth stage, indicate that whilst the quantity of leaf material (i.e., the LAI) may be similar, the arrangement of this material in 3-dimensional space can lead to dramatic changes in carbon assimilation in different canopy layers.

The greater potential optimal  $P_{max}$  at the bottom of the canopy in M13 at GS1 relative to the other varieties can be linked to the low accumulation of leaf material with canopy depth (as seen with cLAI; **Figures 2A,B**) and the reduced FI of light (**Figure 4**) but an increased total light intercepted over the whole canopy (**Table 6**). This suggests that architecture which enables greater light penetration to lower canopy layers leads to a greater assimilation of carbon at lower canopy layers, which contributes to overall canopy photosynthesis. This is seen as an increased carbon gain per unit leaf area relative to the other lines (**Table 6**). However, when assessing the carbon assimilation per unit ground area, M13 ranks in the middle of the five varieties, indicating

that despite the open canopy and greater light penetration, the reduced LAI of the variety leads to reduced productivity on a per land area basis. This indicates a small level of consistency between diverse canopy architectural traits and the long-term responses of photosynthesis to the light environment in this study. It shows that the architectural traits measured and modeled in this study are having a consistent impact on the light dynamics within the canopy, albeit over a limited number of genotypes. However, it is not possible to conclude whether it is possible to predict acclimation state from the distribution of FI and LAI within the canopy without detailed direct photosynthetic analysis of a wider range of genotypes.

When predicting optimal  $P_{max}$  we assumed that light dynamics are the sole factor determining photosynthetic capacity and that canopy nitrogen profiles correlate with canopy photosynthesis profiles. However, nitrogen profiles are frequently suboptimal with respect to photosynthesis (Hikosaka, 2016). The optimal  $P_{max}$  measurement is therefore a novel and potentially useful method for indicating photosynthetic nitrogen use efficiency in crop canopies, clearly shown here for all lines, even M13 with its more efficient light penetration. It needs to be pointed out that the use of the “time weighted average” or  $\tau$  that was fixed at 0.2 was chosen to represent the time taken for photosynthetic induction, but we do not know whether

**TABLE 6 | Gas exchange and modeling results at each growth stage.**

Growth stage	Line	$P_{max}$ top layer ( $\mu\text{mol m}^{-2} \text{s}^{-1}$ )	$P_{max}$ bottom layer ( $\mu\text{mol m}^{-2} \text{s}^{-1}$ )	Carbon gain per unit leaf area ( $\text{mol m}^{-2} \text{d}^{-1}$ )	Carbon gain per unit ground area ( $\text{mol m}^{-2} \text{d}^{-1}$ )	Total light interception ( $\text{mol m}^{-2} \text{d}^{-1}$ )
1	M2	24.29 ± 1.61	20.23 ± 1.69	0.241	0.532	11.98
	M4	22.67 ± 1.76	17.91 ± 1.82	0.220	0.489	12.25
	M11	29.99 ± 2.37	24.52 ± 2.53	0.204	0.504	11.16
	M13	38.65 ± 2.82	27.34 ± 3.54	0.432	0.798	13.34
	IR64	25.96 ± 1.63	20.24 ± 1.70	0.169	0.480	10.18
	Mean	28.31	22.1			
2	<i>P</i>	<0.001	0.062			
	SED	2.96	3.34			
	M2	20.15 ± 0.77	14.14 ± 1.82	0.174	0.827	7.08
	M4	22.67 ± 1.78	14.78 ± 1.87	0.121	0.661	6.28
	M11	26.83 ± 2.72	17.69 ± 1.63	0.232	0.968	7.72
	M13	23.53 ± 1.11	18.83 ± 1.34	0.236	0.828	8.57
	IR64	26.35 ± 1.02	15.90 ± 1715	0.082	0.714	4.08
	Mean	23.91	16.33			
	<i>P</i>	0.053	0.321			
	SED	2.32	2.39			

Measured  $P_{max}$  for the top and bottom layer was calculated from light response curve fitting; the means of six (GS1) or five (GS2) measurements are shown with standard errors of the mean. *P*-value corresponds to ANOVA. An empirical model of photosynthesis was employed to calculate carbon gain per unit leaf area and ground area using light levels predicted by ray tracing for 10th December (GS1) and 21st January (GS2), respectively (see Section Imaging and Ray Tracing). Total light interception over the course of the day was also calculated. Growth stage 1 corresponds to 45 DAT and 2 at 85 DAT. M2, M4, M11, and M13 refer to Shan-Huang Zhan-2 (SHZ-2), IR4630-22-2-5-1-3, 157 WAB 56–125, and Inia Tacuari, respectively.

acclimation status according to canopy position will have an effect on this.

The leaf inclination angle is critical in determining the flux of solar radiation per unit leaf area (Ehleringer and Werk, 1986; Ezcurra et al., 1991; Falster and Westoby, 2003). Plants containing steep leaf inclination angles tend to have a decreased light capture when the sun is directly overhead (i.e., during midday hours or during summer) but increases light capture at lower solar angles (i.e., start/end of the day or during seasonal changes in the higher latitude regions). This feature has a number of practical applications including the decrease in susceptibility to photoinhibition (Ryel et al., 1993; Valladares and Pugnaire, 1999; Werner et al., 2001; Burgess et al., 2015); reduced risk of overheating due to reduction in mid-day heat loads (King, 1997); and minimized water-use relative to carbon gain (Cowan et al., 1982). This architecture feature, combined with a relatively open canopy, has been adopted within our studied line; M13, and contributes to its inherent heat tolerance and higher Fv/Fm-values (Figure 3, Table 4). The erect leaf stature and higher Fv/Fm is also present in our studied line M11 (Figure 3, Table 4). This may suggest a relationship between erectness, maximum quantum yield, and latitude of origin of the lines with M11 and M13 originating in locations closer to the equator [Latin America including equatorial regions and WARDA (now AfricaRice), Western Africa, respectively] relative to the other lines. Such characteristics are in line with previous work to predict the optimal leaf angle according to latitude (Herbert, 2003; Baldocchi, 2005) and work in

*Arabidopsis thaliana* (Hopkins et al., 2008). Correlations between architectural traits and latitude have also been seen within tree species, with a linear decrease in petiole length with an increase in latitude and change in leaf arrangement (King and Maindonald, 1999). The differences in Fv/Fm between the varieties may also be linked to the genetic background of the lines M11 and M13 with the *japonica* background and M2, M4 and IR64 with the *indica* background. This is in agreement with previous work on rice with higher Fv/Fm-values found in *japonica* cultivars relative to *indica* (Kasajima et al., 2011). Differences in Fv/Fm between the two groups are also mirrored in the capacity for non-photochemical quenching (NPQ) for energy dissipation, with much higher NPQ-values found in *japonica* lines (Kasajima et al., 2011).

Rice cultivation areas are highly diverse and are affected in differing ways by fluctuations in environmental conditions. Thus, the origin of each of the parental founders may also indicate why these specific architectural traits are present and how they interact with leaf photosynthetic properties. The five lines selected for this study have diverse origins including China (M2), South East Asia (International Rice Research Institute; M4 and IR64), Africa (M13), and Latin America (M11). The rapid maturation and early flowering of M13 relates to the short-growing seasons of upland rice production in Western Africa whilst stable yields under low nitrogen inputs enables relatively high yields under low-input upland systems (Gridley et al., 2002). Whilst there is little data relating to canopy architecture in divergent rice lines grown across the world, there has been some

work done studying architectural differences between key African and Australian savannah tree taxa (Moncrieff et al., 2014). They found distinct differences between the two sets of taxa in key architectural traits including plant height and canopy area, and attributed the differences not to disparities in the environmental conditions in which the trees grew, but rather in the differing evolutionary history of African vs. Australian savannas. This may indicate that when assessing regional differences in rice architecture, we must take into account not only the biotic and abiotic differences between areas but also the biogeography, interactions with other species and historic cultivation practices.

Structure function relationships in terms of canopy architecture are complex and affected by growing environment. Many factors, in addition to the ideotype principle, will shape the commercial breeder's decision making process. There may be negative linkages with a particular trait (Rasmusson, 1991). Erect leaved ideotypes do not necessarily perform (Breseghezzo and Coelho, 2013) and architecture "performance" depends on location and environmental factors, inputs, and agronomy (Hammer et al., 2009). The erect ideotype means that a higher LAI and hence higher canopy photosynthesis could be supported but this also requires a high fertilizer (especially nitrogen) input which raises cost and reduces sustainability.

This is the first high-resolution study that has been used to attempt to assess the link between canopy architecture and photosynthesis characteristics. One of the drawbacks of this study was the inability to grow the lines in the location they originated, or under a range of different environments. This poses a challenge as canopy architecture is determined by a combination of the genetics of plant but also the conditions in which the plant was grown, including climate, weather patterns, soil type and the competitive presence of neighboring plants. Thus, the architecture adopted by the genotypes in this study may not be totally representative to that when grown elsewhere due to differences in growing conditions. In this study, we used the latitude of the Philippines as a fastTracer3 parameter as a standard to compare the different lines, which will be a different light environment to those in which the plants were grown or in which the lines traditionally grow or have originated. However, the conditions we used were enough to expose significant differences in architecture between lines which are genetically different in origin.

Other factors relating to whole canopy photosynthesis must also be taken into account such as: the angular relationship between the photosynthetic leaf surfaces and the sun; environmental conditions (i.e., wind speed, temperature, CO<sub>2</sub> concentration); soil properties; the photosynthetic pathway used and; the presence of other biotic or abiotic stressors (Baldocchi and Amthor, 2001). This highlights the need for more in depth studies of canopy photosynthesis and architecture within the range of different environmental conditions in which a plant is likely to be exposed to. Also for more realistic modeling; i.e., modeling mimicking the weather conditions or more realistic representations of the plant stands in general (such as incorporating canopy movement due to wind: Burgess et al., 2016). These high-resolution

studies will be critical in determining the exact relationships between canopy architectural features, photosynthesis, the light environment and productivity of our cropping systems and will provide the framework necessary for any future improvements.

Use of the parental founders of an elite MAGIC population of rice leads to possibilities for future studies using a wider number of crosses and their progeny into the genetic control of specific architectural features or breeding attempts to produce an "optimal plant type." Whilst the genetic control of certain architectural traits is relatively understood (e.g., Wang and Li, 2006; Busov et al., 2008; Neeraja et al., 2009; Pearce et al., 2011), the interactions between genotype, phenotype, management, and the environment are less well-known. These relationships are confounded further by the variability in weather patterns and the relatively unknown effects of climate change on our agricultural systems. However, combining high-resolution studies and crop simulations with new breeding methods and genetic modeling provides a promising future for accelerating the discovery and creation of new idealized plant types. Multi-parent populations provide an attractive background for study when combined with high-throughput SNP genotyping (Bandillo et al., 2013).

## AUTHOR CONTRIBUTIONS

TH performed the initial screening experiment with the assistance of AB; AB performed the second in depth study; AB performed all imaging and reconstruction plus most of the modeling work with the assistance of RR; AB and EM wrote the article with the contributions of the other authors.

## FUNDING

AB is supported by the CFF-UNMC Doctoral Training Programme (CFF-UNMC DTP) under BiomassPLUS Programme BioP1-006 and the University of Nottingham, School of Biosciences. This work was also supported by the Biotechnology and Biological Sciences Research Council (grant number BB/JOO3999/1).

## ACKNOWLEDGMENTS

We would like to thank the glasshouse team at Sutton Bonington. We are grateful to Dr. Hei Leung and Dr. Marietta Baraoidan at the International Rice Research Institute for provision of the rice lines.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00734/full#supplementary-material>

**Supplementary Figure S1 | Example of a time-weighted light pattern at  $\tau = 0.2$  (black line) relative to a non-weighted line (i.e.,  $\tau = 0$ ).** The time weighted average (Equation 9) is an exponentially decaying weight

used to represent the fact that photosynthesis is not able to respond instantaneously to a change in irradiance levels. If  $\tau = 0$  then a plant will able to instantaneously respond to a change in irradiance, whereas if  $\tau > 0$  the time-weighted average light pattern will relax over the timescale  $\tau$ . Within this study,  $\tau$  was fixed at 0.2.

#### Supplementary Figure S2 | Correlations for different parameters for the two growth stages.

#### Supplementary Table S1 | Agronomic details on the 16 Parental Lines used to develop the *indica* and *japonica* MAGIC Populations.

Data for MAGIC

lines taken from Bandillo et al. (2013). The four MAGIC lines plus IR64 selected for in depth study are given in bold.

**Supplementary Table S2 | Physiological characteristics of the 15 parental MAGIC lines + IR64 used in the initial screening.** All measurements, apart from harvest dry weight and seed dry weight, were taken 55–60 days after transplanting (DAT), corresponding to the vegetative growth stage. SPAD and leaf discs for chlorophyll samples were taken on the last full expanded leaf. The means of three plots are shown with standard errors of the mean. The bold lines are those selected for use in the in depth study due to their contrasting physiological features.

## REFERENCES

- Anderson, J. M., Chow, W. S., and Park, Y. I. (1995). The grand design of photosynthesis: acclimation of the photosynthetic apparatus to environmental cues. *Photosynth. Res.* 46, 129–139. doi: 10.1007/BF00020423
- Athanasiou, K., Dyson, B. C., Webster, R. E., and Johnson, G. N. (2010). Dynamic acclimation of photosynthesis increases plant fitness in changing environments. *Plant Physiol.* 152, 366–373. doi: 10.1104/pp.109.149351
- Baldocchi, D. (2005). Environmental science: the carbon cycle under stress. *Nature* 437, 483–484. doi: 10.1038/437483a
- Baldocchi, D., and Amthor, J. (2001). Canopy photosynthesis: history, measurements and models," in *Terrestrial Global Productivity*, eds J. Roy, B. Saugier, and H. A. Mooney (San Diego, CA: Academic Press), 9–31. doi: 10.1016/B978-012505290-0/50003-X
- Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., et al. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* 6:11. doi: 10.1186/1939-8433-6-11
- Breseghezzo, F., and Coelho, A. S. G. (2013). Traditional and modern plant breeding methods with examples in rice (*Oryza sativa* L.). *J. Agric. Food Chem.* 61, 8277–8286. doi: 10.1021/jf305531j
- Burgess, A. J., Retkute, R., Pound, M. P., Foulkes, J., Preston, S. P., Jensen, O. E., et al. (2015). High-resolution three-dimensional structural data quantify the impact of photoinhibition on long-term carbon gain in wheat canopies in the field. *Plant Physiol.* 169, 1192–1204. doi: 10.1104/pp.15.00722
- Burgess, A. J., Retkute, R., Preston, S. P., Jensen, O. E., Pound, M. P., Pridmore, T. P., et al. (2016). The 4-dimensional plant: effects of wind-induced canopy movement on light fluctuations and photosynthesis. *Front. Plant Sci.* 7:1392. doi: 10.3389/fpls.2016.01392
- Busov, V. B., Brunner, A. M., and Strauss, S. H. (2008). Genes for control of plant stature and form. *New Phytol.* 177, 589–607. doi: 10.1111/j.1469-8137.2007.02324.x
- Clendon, J. H. M., and Millen, G. G. M. (1979). Leaf angle: an adaptive feature of sun and shade leaves. *Bot. Gaz.* 140, 437. doi: 10.1086/337110
- Cowan, I. R., Lange, O. L., Nobel, P. S., Osmond, C. B., and Ziegler, H. (1982). Regulation of water use in relation to carbon gain in higher plants. *Physiol. Plant Ecol.* 12, 589–613. doi: 10.1007/978-3-642-68150-9\_18
- Dingkuhn, M., Penning De Vries, F., Datta, S. K. D., and Van Laar, H. (1991). "Concepts for a new plant type for direct seeded flooded tropical rice," in *Direct Seeded Flooded Rice in the Tropics* (Manila: IRRI), 17–38.
- Ehleringer, J., and Werk, K. (1986). "Modifications of solar-radiation absorption patterns and implications for carbon gain at the leaf level," in *On the Economy of Plant Form and Function, Proceedings of the Sixth Maria Moors Cabot Symposium, Evolutionary Constraints on Primary Productivity: Adaptive Patterns of Energy Capture in Plants, Harvard Forest, August 1983*, ed T. Givnish (Cambridge; London; New York, NY; New Rochelle, NY; Melbourne: Cambridge University Press), 57–82.
- Ezcurra, E., Montana, C., and Arizaga, S. (1991). Architecture, light interception, and distribution of Larrea species in the Monte Desert, Argentina. *Ecology* 72, 23–34. doi: 10.2307/1938899
- Falster, D. S., and Westoby, M. (2003). Leaf size and angle vary widely across species: what consequences for light interception? *New Phytol.* 158, 509–525. doi: 10.1046/j.1469-8137.2003.00765.x
- Field, C. B., and Mooney, H. A. (1983). Leaf age and seasonal effects on light, water, and nitrogen use efficiency in a California shrub. *Oecologia* 56, 348–355. doi: 10.1007/BF00379711
- Godin, C., and Sinoquet, H. (2005). Functional-structural plant modelling. *New Phytol.* 166, 705–708. doi: 10.1111/j.1469-8137.2005.01445.x
- Gridley, H., Jones, M., and Wopereis-Pura, M. (2002). "Development of new rice for Africa (NERICA) and participatory varietal selection," in *Breeding Rainfed Rice for Drought-Prone Environments: Integrating Conventional and Participatory Plant Breeding in South and Southeast Asia: Proceedings of a DFID Plant Sciences Research Programme/IRRI Conference* (Wallingford, UK), 12–15.
- Hammer, G. L., Dong, Z., McLean, G., Doherty, A., Messina, C., Schussler, J., et al. (2009). Can changes in canopy and/or root system architecture explain historical maize yield trends in the U.S. Corn Belt? *Crop Sci.* 49, 299. doi: 10.2135/cropsci2008.03.0152
- Herbert, T. J. (2003). A latitudinal cline in leaf inclination of *Dryas octopetala* and implications for maximization of whole plant photosynthesis. *Photosynthetica* 41, 631–633. doi: 10.1023/B:PHOT.0000027531.32053.87
- Hikosaka, K. (2016). Optimality of nitrogen distribution among leaves in plant canopies. *J. Plant Res.* 129, 299–311. doi: 10.1007/s10265-016-0824-1
- Hodanova, D. (1979). Sugar beet canopy photosynthesis as limited by leaf age and irradiance. Estimation by models. *Photosynthetica* 13, 376–385.
- Hopkins, R., Schmitt, J., and Stinchcombe, J. R. (2008). A latitudinal cline and response to vernalization in leaf angle and morphology in *Arabidopsis thaliana* (Brassicaceae). *New Phytol.* 179, 155–164. doi: 10.1111/j.1469-8137.2008.02447.x
- Kasajima, I., Ebana, K., Yamamoto, T., Takahara, K., Yano, M., Kawai-Yamada, M., et al. (2011). Molecular distinction in genetic regulation of nonphotochemical quenching in rice. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13835–13840. doi: 10.1073/pnas.1104809108
- Khan, M. H., Dar, Z. A., and Dar, S. A. (2015). Breeding strategies for improving rice yield—a review. *Agric. Sci.* 6, 467–478. doi: 10.4236/as.2015.65046
- Khush, G. S. (2005). What it will take to Feed 5.0 Billion Rice consumers in 2030. *Plant Mol. Biol.* 59, 1–6. doi: 10.1007/s11103-005-2159-5
- King, D. A. (1997). The functional significance of leaf angle in eucalyptus. *Aust. J. Bot.* 45, 619. doi: 10.1071/BT96063
- King, D. A., and Maingdonald, J. H. (1999). Tree architecture in relation to leaf dimensions and tree stature in temperate and tropical rain forests. *J. Ecol.* 87, 1012–1024. doi: 10.1046/j.1365-2745.1999.00417.x
- Moncrieff, G. R., Lehmann, C. E. R., Schnitzler, J., Gambiza, J., Hiernaux, P., Ryan, C. M., et al. (2014). Contrasting architecture of key African and Australian savanna tree taxa drives intercontinental structural divergence. *Glob. Ecol. Biogeogr.* 23, 1235–1244. doi: 10.1111/geb.12205
- Murchie, E. H., Chen, Y. Z., Hubbard, S., Peng, S. B., and Horton, P. (1999). Interactions between senescence and leaf orientation determine *in situ* patterns of photosynthesis and photoinhibition in field-grown rice. *Plant Physiol.* 119, 553–563. doi: 10.1104/pp.119.2.553
- Murcie, E. H., and Horton, P. (1997). Acclimation of photosynthesis to irradiance and spectral quality in British plant species: chlorophyll content, photosynthetic capacity and habitat preference. *Plant Cell Environ.* 20, 438–448. doi: 10.1046/j.1365-3040.1997.d01-95.x
- Murcie, E. H., and Horton, P. (1998). Contrasting patterns of photosynthetic acclimation to the light environment are dependent on the differential expression of the responses to altered irradiance and spectral quality. *Plant Cell Environ.* 21, 139–148. doi: 10.1046/j.1365-3040.1998.00262.x

- Murchie, E. H., Pinto, M., and Horton, P. (2009). Agriculture and the new challenges for photosynthesis research. *New Phytol.* 181, 532–552. doi: 10.1111/j.1469-8137.2008.02705.x
- Neeraja, C. N., Vemireddy, L. R., Malathi, S., and Siddiq, E. A. (2009). Identification of alternate dwarfing gene sources to widely used Dee-Gee-Woo-Gen allele of sd1 gene by molecular and biochemical assays in rice (*Oryza sativa* L.). *Electron. J. Biotechnol.* 12, 1–11. doi: 10.2225/vol12-issue3-fulltext-11
- Normile, D. (1999). Crossing rice strains to keep Asia's rice bowls brimming. *Science* 283:313. doi: 10.1126/science.283.5400.313
- Pearce, S., Saville, R., Vaughan, S. P., Chandler, P. M., Wilhelm, E. P., Sparks, C. A., et al. (2011). Molecular characterization of Rht-1 dwarfing genes in hexaploid wheat. *Plant Physiol.* 157, 1820–1831. doi: 10.1104/pp.111.183657
- Pearcy, R. W., and Seemann, J. R. (1990). Photosynthetic induction state of leaves in a soybean canopy in relation to light regulation of ribulose-1,5-bisphosphate carboxylase and stomatal conductance. *Plant Physiol.* 94, 628–633. doi: 10.1104/pp.94.2.628
- Peng, S., Khush, G. S., Virk, P., Tang, Q., and Zou, Y. (2008). Progress in ideotype breeding to increase rice yield potential. *Field Crops Res.* 108, 32–38. doi: 10.1016/j.fcr.2008.04.001
- Porra, R., Thompson, W., and Kriedman, P. (1989). Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim. Biophys. Acta* 975, 384–394. doi: 10.1016/S0005-2728(89)80347-0
- Pound, M. P., French, A. P., Murchie, E. H., and Pridmore, T. P. (2014). Automated recovery of three-dimensional models of plant shoots from multiple color images. *Plant Physiol.* 166, 1688–1698. doi: 10.1104/pp.114.248971
- Quan, L., Tan, P., Zeng, G., Yuan, L., Wang, J., and Kang, S. B. (2006). Image-based plant modeling. *ACM Trans. Graph.* 25, 599. doi: 10.1145/1141911.1141929
- Rasmussen, D. C. (1991). A plant breeder's experience with ideotype breeding. *Field Crop. Res.* 26, 191–200. doi: 10.1016/0378-4290(91)90035-T
- Reich, P. B., Walters, M. B., Ellsworth, D. S., Vose, J. M., Volin, J. C., Gresham, C., et al. (1998). Relationships of leaf dark respiration to leaf nitrogen, specific leaf area and leaf life-span: a test across biomes and functional groups. *Oecologia* 114, 471–482. doi: 10.1007/s004420050471
- Retkute, R., Smith-Unna, S. E., Smith, R. W., Burgess, A. J., Jensen, O. E., Johnson, G. N., et al. (2015). Exploiting heterogeneous environments: does photosynthetic acclimation optimize carbon gain in fluctuating light? *J. Exp. Bot.* 66, 2437–2447. doi: 10.1093/jxb/erv055
- Reynolds, M. P., and Pfeiffer, W. H. (2000). "Applying physiological strategies to improve yield potential," in *Durum Wheat Improvement in the Mediterranean Region: New challenges*, eds C. Royo, M. M. Nachit, N. Di Fonzo, and J. L. Araus (CIHEAM), 95–103.
- Reynolds, M. P., van Ginkel, M., and Ribaut, J. M. (2000). Avenues for genetic modification of radiation use efficiency in wheat. *J. Exp. Bot.* 51, 459–473. doi: 10.1093/jexbot/51.suppl\_1.459
- Rötter, R. P., Tao, F., Höhn, J. G., and Palosuo, T. (2015). Use of crop simulation modelling to aid ideotype design of future cereal cultivars. *J. Exp. Bot.* 66, 3463–3476. doi: 10.1093/jxb/erv098
- Ryel, R. J., Beyschlag, W., and Caldwell, M. (1993). Foliage orientation and carbon gain in two tussock grasses as assessed with a new whole-plant gas-exchange model. *Funct. Ecol.* 7, 115–124. doi: 10.2307/2389874
- Setter, T. L., Conocono, E. A., Egglestone, J. A., and Kropff, M. J. (1995). Possibility of increasing yield potential of rice by reducing panicle height in the canopy. I. effects of panicles on light interception and canopy photosynthesis. *Aust. J. Plant Physiol.* 22, 441–451.
- Sharkey, T. D., Bernacchi, C. J., Farquhar, G. D., and Singsaas, E. L. (2007). Fitting photosynthetic carbon dioxide response curves for C<sub>3</sub> leaves. *Plant. Cell Environ.* 30, 1035–1040. doi: 10.1111/j.1365-3040.2007.01710.x
- Sinoquet, H., Stephan, J., Sonohat, G., Lauri, P. É., and Monney, P. (2007). Simple equations to estimate light interception by isolated trees from canopy structure features: assessment with three-dimensional digitized apple trees. *New Phytol.* 175, 94–106. doi: 10.1111/j.1469-8137.2007.02088.x
- Song, Q., Xiao, H., Xiao, X., and Zhu, X. G. (2016). A new canopy photosynthesis and transpiration measurement system (CAPTS) for canopy gas exchange research. *Agric. For. Meteorol.* 217, 101–107. doi: 10.1016/j.agrformet.2015.11.020
- Song, Q., Zhang, G., and Zhu, X.-G. (2013). Optimal crop canopy architecture to maximise canopy photosynthetic CO<sub>2</sub> uptake under elevated CO<sub>2</sub> – a theoretical study using a mechanistic model of canopy photosynthesis. *Funct. Plant Biol.* 40, 109–124. doi: 10.1071/FP12056
- Stegemann, J., Timm, H.-C., and Küppers, M. (1999). Simulation of photosynthetic plasticity in response to highly fluctuating light: an empirical model integrating dynamic photosynthetic induction and capacity. *Trees* 14, 145–160. doi: 10.1007/s004680050219
- Turitzin, S. N., and Drake, B. G. (1981). The effect of a seasonal change in canopy structure on the photosynthetic efficiency of a salt marsh. *Oecologia* 48, 79–84. doi: 10.1007/BF00346991
- Valladares, F., and Pugnaire, F. (1999). Tradeoffs between irradiance capture and avoidance in semi-arid environments assessed with a crown architecture model. *Ann. Bot.* 83, 459–469. doi: 10.1006/anbo.1998.0843
- Virk, P., Khush, G., and Peng, S. (2004). Breeding to enhance yield potential of rice at IRRI: the ideotype approach. *Int. Rice Res. Notes* 29, 5–9.
- Walters, R. G. (2005). Towards an understanding of photosynthetic acclimation. *J. Exp. Bot.* 56, 435–447. doi: 10.1093/jxb/eri060
- Wang, Y., and Li, J. (2006). Genes controlling plant architecture. *Curr. Opin. Biotechnol.* 17, 123–129. doi: 10.1016/j.copbio.2006.02.004
- Watanabe, T., Hanan, J. S., Room, P. M., Hasegawa, T., Nakagawa, H., and Takahashi, W. (2005). Rice morphogenesis and plant architecture: Measurement, specification and the reconstruction of structural development by 3D architectural modelling. *Ann. Bot.* 95, 1131–1143. doi: 10.1093/aob/mci136
- Werner, C., Ryel, R. J., Correia, O., and Beyschlag, W. (2001). Effects of photoinhibition on whole-plant carbon gain assessed with a photosynthesis model. *Plant Cell Environ.* 24, 27–40. doi: 10.1046/j.1365-3040.2001.00651.x
- Yano, S., and Terashima, I. (2001). Separate localization of light signal perception for sun or shade type chloroplast and palisade tissue differentiation in *Chenopodium album*. *Plant Cell Physiol.* 42, 1303–1310. doi: 10.1093/pcp/pce183
- Zheng, B., Guo, Y., Ma, Y., Shi, L., Deng, Q., and Li, B. (2008). Comparison of architecture among different cultivars of hybrid rice using a spatial light model based on 3-D digitising. *Funct. Plant Biol.* 35, 900–910. doi: 10.1071/FP08060
- Zhu, X.-G., Long, S. P., and Ort, D. R. (2010). Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.* 61, 235–261. doi: 10.1146/annurev-arplant-042809-112206

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

Copyright © 2017 Burgess, Retkute, Herman and Murchie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# UAV-Based Thermal Imaging for High-Throughput Field Phenotyping of Black Poplar Response to Drought

Riccardo Ludovisi<sup>1†</sup>, Flavia Tauro<sup>1†</sup>, Riccardo Salvati<sup>1</sup>, Sacha Khoury<sup>2</sup>, Giuseppe Scarascia Mugnozza<sup>1</sup> and Antoine Harfouche<sup>1\*</sup>

<sup>1</sup> Department for Innovation in Biological, Agro-food and Forest Systems, University of Tuscia, Viterbo, Italy, <sup>2</sup> Department of Plant Sciences, University of Cambridge, Cambridge, United Kingdom

## OPEN ACCESS

### Edited by:

John Doonan,  
Aberystwyth University,  
United Kingdom

### Reviewed by:

Véronique Jorge,  
Institut National de Recherche  
Agronomique, France  
Stephen Hunt,  
Queen's University, Canada

### \*Correspondence:

Antoine Harfouche  
aharfouche@unitus.it

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

### Specialty section:

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

Received: 07 January 2017

Accepted: 13 September 2017

Published: 27 September 2017

### Citation:

Ludovisi R, Tauro F, Salvati R, Khoury S, Scarascia Mugnozza G and Harfouche A (2017) UAV-Based Thermal Imaging for High-Throughput Field Phenotyping of Black Poplar Response to Drought. *Front. Plant Sci.* 8:1681. doi: 10.3389/fpls.2017.01681

Poplars are fast-growing, high-yielding forest tree species, whose cultivation as second-generation biofuel crops is of increasing interest and can efficiently meet emission reduction goals. Yet, breeding elite poplar trees for drought resistance remains a major challenge. Worldwide breeding programs are largely focused on intra/interspecific hybridization, whereby *Populus nigra* L. is a fundamental parental pool. While high-throughput genotyping has resulted in unprecedented capabilities to rapidly decode complex genetic architecture of plant stress resistance, linking genomics to phenomics is hindered by technically challenging phenotyping. Relying on unmanned aerial vehicle (UAV)-based remote sensing and imaging techniques, high-throughput field phenotyping (HTFP) aims at enabling highly precise and efficient, non-destructive screening of genotype performance in large populations. To efficiently support forest-tree breeding programs, ground-truthing observations should be complemented with standardized HTFP. In this study, we develop a high-resolution (leaf level) HTFP approach to investigate the response to drought of a full-sib  $F_2$  partially inbred population (termed here 'POP6'), whose  $F_1$  was obtained from an intraspecific *P. nigra* controlled cross between genotypes with highly divergent phenotypes. We assessed the effects of two water treatments (well-watered and moderate drought) on a population of 4603 trees (503 genotypes) hosted in two adjacent experimental plots (1.67 ha) by conducting low-elevation (25 m) flights with an aerial drone and capturing 7836 thermal infrared (TIR) images. TIR images were undistorted, georeferenced, and orthorectified to obtain radiometric mosaics. Canopy temperature ( $T_c$ ) was extracted using two independent semi-automated segmentation techniques, eCognition- and Matlab-based, to avoid the mixed-pixel problem. Overall, results showed that the UAV platform-based thermal imaging enables to effectively assess genotype variability under drought stress conditions.  $T_c$  derived from aerial thermal imagery presented a good correlation with ground-truth stomatal conductance ( $g_s$ ) in both segmentation techniques. Interestingly, the HTFP approach was instrumental to detect drought-tolerant response in 25% of the population. This study shows the potential of UAV-based thermal imaging for field phenomics of poplar and other tree species. This is anticipated to have tremendous implications for accelerating forest tree genetic improvement against abiotic stress.

**Keywords:** UAV remote sensing, high-throughput field phenotyping (HTFP), phenomics, poplar thermal imagery, image processing, stomatal conductance, drought

## INTRODUCTION

Fast growing *Populus* clones are among the most common lignocellulosic feedstocks for second-generation bioenergy production in Europe (Amichev et al., 2010; Sannigrahi et al., 2010; Sabatti et al., 2014; Djomo et al., 2015). A recent report of the International Poplar Commission indicates that the total area of short rotation coppice (SRC) *Populus* across Europe is about 23,502 ha (FAO, 2016). However, the current surface area planted with SRC *Populus* in Europe was estimated at about 45,000 ha by Alasia Franco Vivai (Franco Alasia, Personal Communication), taking into account the establishment of few thousands ha in recent years in Poland. Compared to alternative bioenergy crops, SRC plantations offer versatile year-round coppice cycles and high yield to input ratio (Sims and Venturi, 2004). To alleviate the conflict between food and biofuel production (Rockwood et al., 2004; Edrissi and Abhilash, 2016), SRC plantations are typically grown on marginal lands, which are inadequate for high productivity crop growth (Herr and Carlson, 2013), thus generating an income without the need for a land remediation period (Paulson et al., 2003).

Short rotation coppice *Populus* clones are selected from worldwide breeding programs based on interspecific hybridization, whereby black poplar (*Populus nigra* L.) is a fundamental parental pool (Stanton et al., 2013); being widely and naturally spread in Europe, typically associated with riparian ecosystems, and characterized by large phenotypic and genetic variability (van der Schoot et al., 2000; Richardson et al., 2014; DeWoody et al., 2015). Moreover, *P. nigra* has been thoroughly studied due to its numerous adaptive characteristics, including easy clonal propagation, good coppicing ability, resistance to pathogens and parasites (Benetka et al., 2012), prolonged growing season (Rohde et al., 2011), and high plasticity in response to environmental conditions (Chamaillard et al., 2011). Breeding strategies based on recurrent selection and testing are frequently implemented for gradual population improvement (Neale and Kremer, 2011). Acceleration of the *Populus* domestication is also expected through recurrent intraspecific crossing and higher order species mixes, yet to be supplemented by genomic selection, association genetics, and genetic engineering (Harfouche et al., 2012). While first-generation hybridization ( $F_1$ ) is adopted to obtain heterosis for growth rate (Stettler et al., 1988), advanced generation breeding, such as  $F_2$  hybridization, among closely related *Populus* species, has proved to be an efficient strategy toward genetic improvement (Stanton et al., 2010).

Field-grown trees are routinely exposed to environmental stress and are likely to experience unprecedented rises in temperature and increases in the frequency and severity of summer drought episodes in the future (Lindner et al., 2014; IPCC, 2014). The physiological responses to drought are complex and traits vary in their importance depending on intensity, duration, and timing of the drought (Bréda and Badeau, 2008; Tardieu and Tuberrosa, 2010; Harfouche et al., 2014). These traits present as reduced leaf size, decreased leaf growth rate, lowered stomatal aperture and density, reduced stomatal conductance ( $g_s$ ), and altered patterns of root development (Tardieu and Tuberrosa, 2010). Furthermore, inside the leaf, prolonged drought

periods reduce  $\text{CO}_2$  assimilation rates and the extra energy dissipation, with a consequent increase in reactive oxygen species production, leading to leaf senescence and yield loss (Pintó-Marijuan and Munné-Bosch, 2014). Physiological and molecular studies on drought tolerance in *Populus* have shed light on the considerable divergence in response to water deficit between different genotypes (Marron et al., 2002; Monclús et al., 2006; Street et al., 2006; Huang et al., 2009; Regier et al., 2009; Cocozza et al., 2010; Viger et al., 2016).

Plant tolerance to abiotic stresses is an ambiguous concept, even after distinguishing different strategies such as avoidance, tolerance, and escape (Levitt, 1972). Depending on their genetically dictated molecular and physiological attributes, plants budget their water in very different ways, along a continuum that ranges from the water-conserving or risk-aversion behavior displayed by isohydric plants to the risk-taking behavior displayed by anisohydric plants (Tardieu and Simonneau, 1998; Moshelion et al., 2014; Sade and Moshelion, 2014; Attia et al., 2015).

In reduced water availability conditions, the relationship between  $g_s$  and leaf temperature has been utilized as a valid indicator of trees' response (Chaves et al., 2003; Bréda et al., 2006; Jiménez-Bello et al., 2011; Costa et al., 2013; Rebetzke et al., 2013; Struthers et al., 2015). Therefore, relating the leaf temperature of individuals to the average value of a population exposed to similar environmental conditions may be indicative of the trees' state of stress.

A major obstacle to a more effective dissection of plant response to drought is the difficulty in properly phenotyping in a high-throughput fashion. To relieve a phenotyping bottleneck, phenotypic traits should be turned into quantifiable, objective, and consistent measures. Furthermore, automated and high-throughput phenotyping (HTP) on large-scale plant populations is expected to increase the probability of detecting crucial traits and, thus, identifying effective genotype-phenotype relationships (Goggin et al., 2015). HTP envisions a suite of strategies to speed up the phenotyping process and maximize the number of studied plants per experiment (Goggin et al., 2015). These methods enable automated, non-destructive, and non-invasive screening of high dimensionality populations, and thus, allowing the same plants and their responses to changing environmental factors to be monitored throughout their life cycle (Fahlgren et al., 2015a). To facilitate data interpretation, HTP platforms often involve observations in controlled conditions, such as growth chambers and greenhouses. However, plant performance in highly controlled conditions is poorly correlated with breeders' target commercial and real-world environments (White et al., 2012; Araus and Cairns, 2014; Deery et al., 2014; Ghanem et al., 2015; Izawa, 2015; Poland, 2015). With regards to the specific case of drought, phenotyping under controlled conditions is highly challenging. In fact, the declining soil moisture content and increasing mechanical impedance typical of droughts are difficult to replicate in pots that are much smaller than the volume of soil available in the field (Cairns et al., 2011; Passioura, 2012). This fact may result in fast plant response through stomatal closure, which may, in turn, mask slower adaptive processes.

High-throughput phenotyping approaches seek to gather remote information; however, close-range (proximal) sensing is frequently required to provide the adequate resolution to decipher phenotypic traits (White et al., 2012). Proximal HTP combines robotic technology and imaging to enable high-dimensional phenotype screening and capture. Unmanned aerial vehicles (UAVs) equipped with cameras and sensors are proximal remote sensing that bridge the gap between time consuming ground-based measurements and satellite/airborne observations (Gago et al., 2015). Compared to traditional ground-based techniques, UAVs enable rapid and non-destructive measurements. They also offer much quicker turnaround times than satellites at competitive costs (Berni et al., 2009). In terms of spatial resolution, different from satellites, UAVs allow the acquisition of images whose pixels are significantly smaller than objects of interest, thus minimizing the bias effect due to background intensity (mixed-pixel problem) (Jones and Sirault, 2014). In contrast to aircrafts, such as manned helicopters, UAVs can safely hover at low altitudes in the proximity of plants, thus allowing high resolutions at low costs (White et al., 2012). In light of such advantages, UAVs are expected to open new avenues in field-based phenotyping of multiple stress traits and large populations rapidly, precisely, and accurately.

With regards to the evaluation of drought response, conventional ground-based methods typically require  $g_s$  measurements, which involve time consuming contact with leaves (Maes and Steppe, 2012; Costa et al., 2013). Alternatively, based on the relationship between  $g_s$  and leaf temperature, UAVs have been furnished with thermometers and thermal infrared (TIR) cameras to capture images of large-scale populations (Berni et al., 2009). TIR cameras show great potential in phenotyping as thermal images contain spatially distributed information about the energy emitted from body surfaces, such as plant leaves. Thermal images can be used to detect the state of stress due to drought and indirectly estimate  $g_s$  through the leaf energy balance equation (Jones, 1992, 1999). Since canopy temperature ( $T_c$ ) has long been recognized as a measure of plant water status (Jones et al., 2009), UAVs mounting sensors have been used for mapping drought response of agricultural crops at 20–40 cm spatial resolution (Sepulcre-Canto et al., 2006, 2007; Zarco-Tejada et al., 2012).

Although imaging has revolutionized plant phenotyping through the early and quantitative detection of plant traits in objective terms (Dhondt et al., 2013; Goggin et al., 2015), massive image data handling and processing remains the rate-limiting step in HTP (Fahlgren et al., 2015b). Image post-processing may include several steps, such as calibration and undistortion (Berni et al., 2009; Zarco-Tejada et al., 2013; Araus and Cairns, 2014), which should be automated in data management pipelines to boost HTP. Within data post-processing, the identification of objects of interest in images against other objects and the background (segmentation) (Jähne, 2005) is often the most critical step (Dhondt et al., 2013). In phenotyping studies, segmentation algorithms have been developed for plant three-dimensional measurements (Chéné et al., 2012) and to estimate crown architecture parameters (Díaz-Varela et al., 2015). For

imaging to substantially contribute to HTP, standardized semi-automated image processing tools should be introduced and complemented with ground-truthing through well-established point measurement sensors (Fahlgren et al., 2015b).

Here, we developed a novel methodology for field HTP (HTFP) of drought stress in a *P. nigra* F<sub>2</sub> partially inbred population using TIR images recorded from a UAV platform. The objectives of this study were to establish a field-scale HTP procedure to rapidly and precisely examine drought stress in trees; to provide objective image-based tools and statistical protocols to quantify phenotypic traits of moderately stressed and non-stressed trees; and to use HTFP to identify promising drought-tolerant genotypes for potential use within *Populus* breeding programs.

## MATERIALS AND METHODS

The workflow of the developed methodology is illustrated in **Figure 1**. The HTFP method involves the following four steps: (1) UAV-based thermography to capture tree  $T_c$ ; (2) semi-automatic georeferencing, orthorectification, and mosaicking of TIR images; (3) tree canopy identification against bare soil through two independent image segmentation approaches; and (4) ground-truthing validation of UAV-based thermal data with  $g_s$  data.

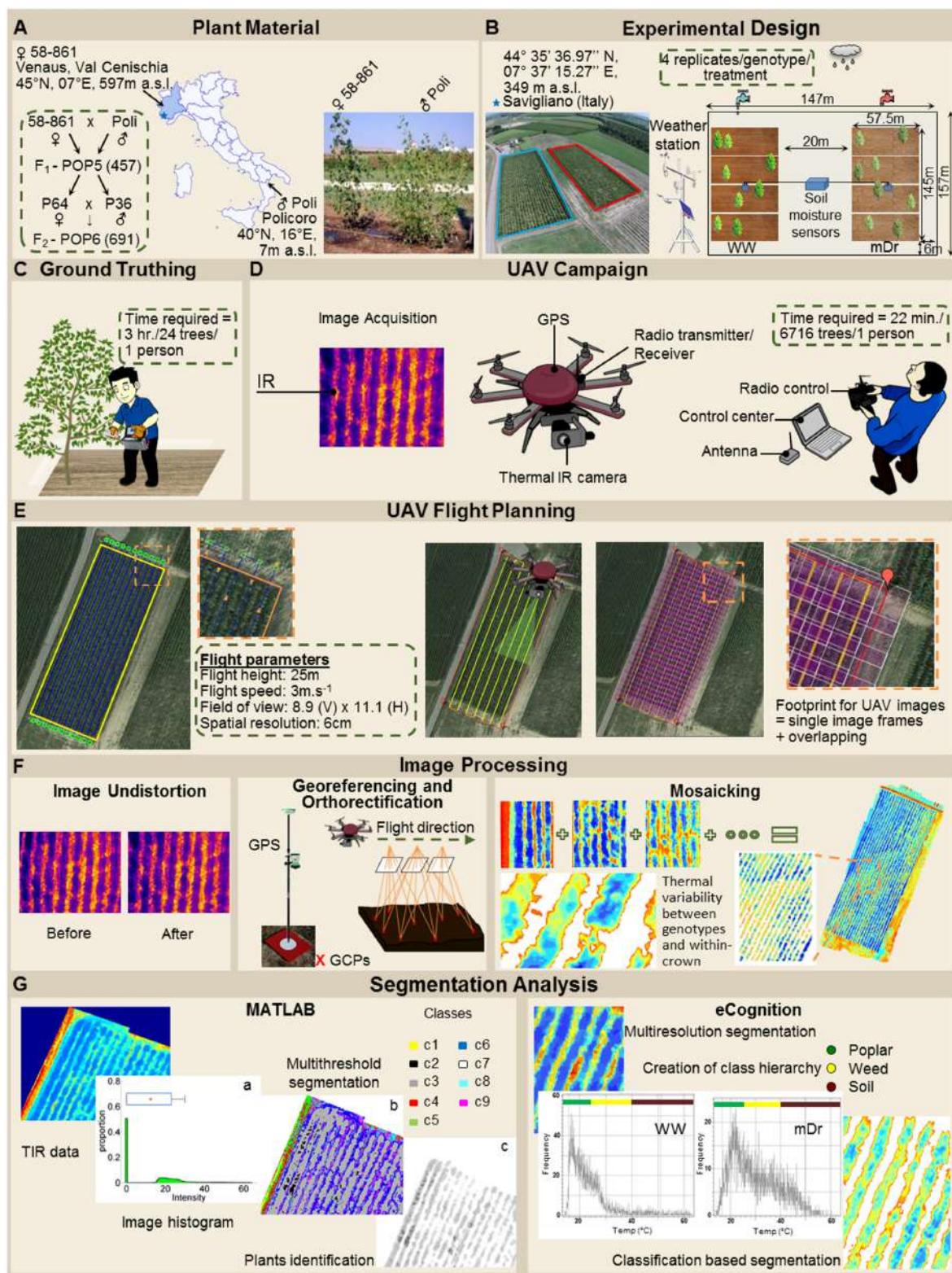
## Field Experiments

### Plant Material and Experimental Design

POP6 is a full-sib F<sub>2</sub> partially inbred population consisting of 691 genotypes obtained from an intraspecific controlled cross between two *P. nigra* parents, P64 and P36. Parents have been randomly selected from an F<sub>1</sub> breeding population (POP5) of 457 genotypes, obtained from an intraspecific *P. nigra* controlled cross between genotypes 58-861 and Poli (**Figure 1A**). Such genotypes originated from natural populations in divergent environmental conditions: 58-861 is from cold/wet climates typical of Val Cenischia (Northern Italy; 45°09'N, 07°01'E, 597 m above sea level), whereas Poli originates from warm/dry climates typical of Policoro (Southern Italy; 40°09'N, 16°41'E, 7 m above sea level). These genotypes are characterized by contrasting responses to water stress (Regier et al., 2009; Cocozza et al., 2010).

In January 2012, F<sub>2</sub> genotypes were used to establish a stool bed and produce 1-year old material. Such material was harvested in January 2013 to produce hardwood cuttings. A large number of hardwood cuttings (20 cm long) was obtained per each genotype, and cuttings whose diameter was within 3–4 cm and with a large number of intact buds were labeled and stored at +4°C.

The experimental field-scale plots were located in Savigliano (Northern Italy; 44°35'36.97"N, 07°37'15.27"E, 349 m above sea level), and were established in April 2013. Two adjacent plots (1.67 ha of total extension) were developed to expose POP6 genotypes to two different water treatments (see Section "Water Treatment") (**Figure 1B**). Each plot featured a completely randomized block design with four blocks. A single hardwood cutting per genotype was randomly assigned to each block to minimize variability attributable to eventually uncontrollable



**FIGURE 1 |** Workflow of the high-throughput field phenotyping (HTFP) methodology. **(A)** Plant material includes intraspecific *Populus nigra* full-sib *F*<sub>2</sub> partially inbred population (POP6) obtained from the controlled cross between two *P. nigra* parents, P64 and P36. These parents were selected from an *F*<sub>1</sub> breeding population (POP5) of 457 genotypes, obtained from an intraspecific *P. nigra* controlled cross between genotypes 58-861 and Poli. **(B)** Two adjacent plots were developed in

(Continued)

**FIGURE 1 |** Continued

Savigliano (Italy) to host POP6 genotypes exposed to well-watered (WW) and moderate drought (mDr) stress conditions. In a plot, WW conditions were maintained, whereby water lost during the day through tree evapotranspiration ( $ET_c$ , mm) was daily restored via drip irrigation. In the other plot, mDr conditions were maintained by withholding irrigation. In each plot, soil water content was daily monitored through a time domain reflectometry SM150 Soil Moisture Sensor installed at 50 cm underneath the soil surface. **(C)** Leaf stomatal conductance  $g_s$  was collected using a dynamic diffusion porometer. Measurements were taken on three biological replicates per treatment on each parental genotype (24 trees). **(D)** An unmanned FlyNovex® multi-copter was integrated with a FLIR A35 TIR (thermal infrared) camera. The UAV campaign allowed for capturing TIR images of both experimental plots. **(E)** The flight mission was planned using the open source autopilot Mission Planner. The UAV was flown in the autonomous mode at a nominal speed of 3 m/s. **(F)** Fish-eye undistortion, image orthorectification, georeferencing, and mosaicking were performed using 16 ground control points (GCPs) captured with a global positioning system (GPS). **(G)** Canopy identification was achieved through two alternative automatic image segmentation approaches (an in-house algorithm in Matlab and eCognition).

environmental factors (such as soil composition and fertility). Differently, for P64, P36, 58-861, and Poli, four cuttings were replicated per each block to ensure larger data availability. Cuttings were planted at a distance of 2.5 m × 1 m, between and within rows, respectively. In addition, the border effect, i.e., trees planted at external locations of plots display better growth conditions, was minimized by planting a double border row of *P. nigra* cv. Jean Pourtet around the sides of each plot. Therefore, a total of 6716 trees were planted in the experimental plots.

In March 2013, multiple sprouts were thinned to a single stem per stamp, choosing the most vigorous. During the first growing season in 2013, drought treatment was not applied to ensure a homogeneous root system development and, thus, minimize effects on shoot growth due to different cutting dimensions. In the same growing season, both plots were regularly irrigated with a drip irrigation system during dry periods, and weeds were controlled using mechanized and chemical practices. In February 2014, experimental plots were coppiced, and for the second growing season in 2014, plots were managed similarly to 2013. In February 2015, trees were again coppiced and, in March 2015, re-sprouts were thinned to a single stem. Finally, due to a mortality of 12.7% of the original trees planted in 2013, 4603 plants (503 genotypes) were available at the beginning of the drought experiment in July 2015.

### Weather and Soil Condition Analysis

Weather data at the experimental setup were obtained from a meteorological station managed by Agency for Protection of Environment – Piemonte (ARPA-Piemonte<sup>1</sup>) and located at 10 km from the experimental plots. Observations gathered from 1994 to 2013 indicate an annual mean temperature of 11.8°C and a total annual rainfall of 740 mm at the site. According to Köppen-Geiger classification (Kottek et al., 2006), climate was classified as Cfb, that is, warm temperate, fully humid and with warm summers. During the experiment in 2015, a 2 km distant meteorological station (Delta-T Devices Ltd., United Kingdom) was also used to record hourly air temperature ( $T_{air}$ ), air humidity, and precipitation.

Soil samples were collected from the middle of plots to quantify soil texture and to estimate field capacity and permanent wilting point. These samples were taken from topsoil down to a maximum depth of 0.5 m. The soil type was a sandy loam (50.5% sand, 35.5% loam, and 14.0% clay), using the United States Department of Agriculture soil taxonomy. Soil field

capacity and permanent wilting point were estimated to 34 and 9.5%, respectively. Soil texture was estimated using gravimetric analysis, whereas soil field capacity and permanent wilting point were evaluated using the pressure-based method (Richards and Weaver, 1944).

### Water Treatment

Response to drought stress was investigated by exposing POP6 to different water regimes for a period of 26 days, from 2nd July 2015 (day of the year; DOY 183) to 28th July 2015 (DOY 209). In one plot, well-watered (WW) conditions were maintained, whereby water lost during the day through tree evapotranspiration ( $ET_c$ , mm) was daily restored via drip irrigation, see **Figure 1B**. Water provisioned through drip irrigation was estimated based on site-specific reference evapotranspiration ( $ET_0$ , mm) and on *Populus* crop coefficient ( $k_c$ ).  $ET_0$  was found according to FAO-56 Penman-Monteith equation (Allen et al., 1998), and  $k_c$  was set to 0.84 in July and to 1.21 in August (Guidi et al., 2008). In the other plot, moderate drought (mDr) conditions were established by exposing plants only to natural rainfall. Drought was imposed by withholding irrigation from DOY 183, and monitoring the progressive reduction of soil moisture until a pre-wilting (i.e., sub-lethal) level of soil water content was achieved.

In each plot, soil water content was daily monitored through a time domain reflectometry SM150 Soil Moisture Sensor (Delta-T Devices Ltd., United Kingdom) installed at 50 cm underneath the soil surface. Four measures per day were recorded using a DL6 Data Logger (Delta-T Devices Ltd., United Kingdom), and the daily average soil water content was estimated. Such system allowed controllability over the experiment, by ensuring that WW and mDr conditions were maintained in the plots. During the drought experiment, daily average  $T_{air}$  (°C), daily rainfall (mm), daily water deficit (mm), and soil water content expressed as percentage of water field capacity (%FC) in WW and mDr plots were reported (Supplementary Figures S1A–C). Soil permanent wilting point is reported as a percentage of the water field capacity (%FC = 27.9%). Daily water deficit was calculated as the difference between precipitation and  $ET_c$ , according to Cantero-Martínez et al. (2007).

### Data Acquisition and Processing

#### UAV Campaign

An unmanned FlyNovex® multi-copter (FlyTop, Italy) was integrated with a FLIR A35 TIR camera (FLIR Systems, United States), see **Figure 1D**. FlyNovex® is a versatile and

<sup>1</sup><http://www.arpa.piemonte.it>

powerful (24V-6S motors) hexacopter (120 cm diagonal size) with a highly resistant carbon fiber frame and offering a 7 kg take-off weight. Its maximum transmission distance is 2 km and its maximum flight time is 20 min. The FLIR A35 was equipped with a 9 mm f1.25 lens. Its thermal sensitivity is less than 0.05°C at 30°C and the camera enables measurements in the range -25°C to +135°C. The image sensor is a Focal Plane Array (FPA) based on uncooled microbolometers with a spectral response in the range 7.5–13 μm. The camera field of view is equal to 48° (horizontally) × 39° (vertically), its resolution to 320 pixels × 256 pixels, and its spatial resolution to 2.78 mrad. The camera captures images at an acquisition frequency of 60 Hz.

Pictures are stored as 14 bit digital raw images. The camera is radiometrically calibrated, and its high accuracy and pixel-to-pixel sensitivity circumvent the need for ground infrared calibration targets and temperature correction during post-processing (Deery et al., 2016; Gómez-Candón et al., 2016). The camera is controlled by an embedded computer (Pico PC with a Cortex 9 processor) that stores raw images on an internal micro SDD memory card for the entire duration of the flight.

A total of 16 highly visible targets were positioned along the borders of the experimental plots (8 targets per plot were located at the corners and in the middle of each side) and used as ground control points (GCPs). The targets were used for georeferencing thermal images. A Real-Time Kinematic global positioning system (GPS; Leica Geosystems, Switzerland) GS08plus model with an accuracy of 3 mm was used for capturing GCP locations.

### Flight Plan and Thermal Imaging

The flight mission was planned using the open source autopilot Mission Planner (APM Mission Planner, United States). The UAV was flown in the autonomous mode (GPS-waypoint navigation) at a nominal speed of 3 m/s. Experimental plots were scanned during two 11-min flights conducted on 28th July 2015 under stable cloudless and low-wind conditions. To ensure similar solar illumination angles and consistent proportions of sunlit and shaded leaves (Deery et al., 2014), flights were performed at 13:41 local time above WW and at 14:30 local time above mDr.

To capture the experimental plots in a single flight pass while maximizing image resolution, flight plan consisted of transects parallel to the plant rows, see **Figure 1E**. A ground station processed the UAV safety manual control and sent telemetry data (position, attitude, and status data) through a radio link at 2.4 GHz to a laptop computer. This communication link also allowed operation of the onboard TIR camera.

Thermal images were recorded at an elevation of 25 m from ground level, thus yielding an 8.9 (vertical) m × 11.1 (horizontal) m and a pixel size of 6 cm × 6 cm. Such a resolution is within POP6 average leaf area (46.47 cm<sup>2</sup> or 1.3 pixels, unpublished data). Flying at such an elevation minimizes image distortions due to atmospheric effects. The selected flight plan enabled capture of thousands of high quality single images presenting 30% overlap and 50% sidelap. Since images captured during take-off, landing, and flight maneuvers were discarded from further processing,

acquisition of images took a few minutes per experimental plot.

### Stomatal Conductance Ground-truthing

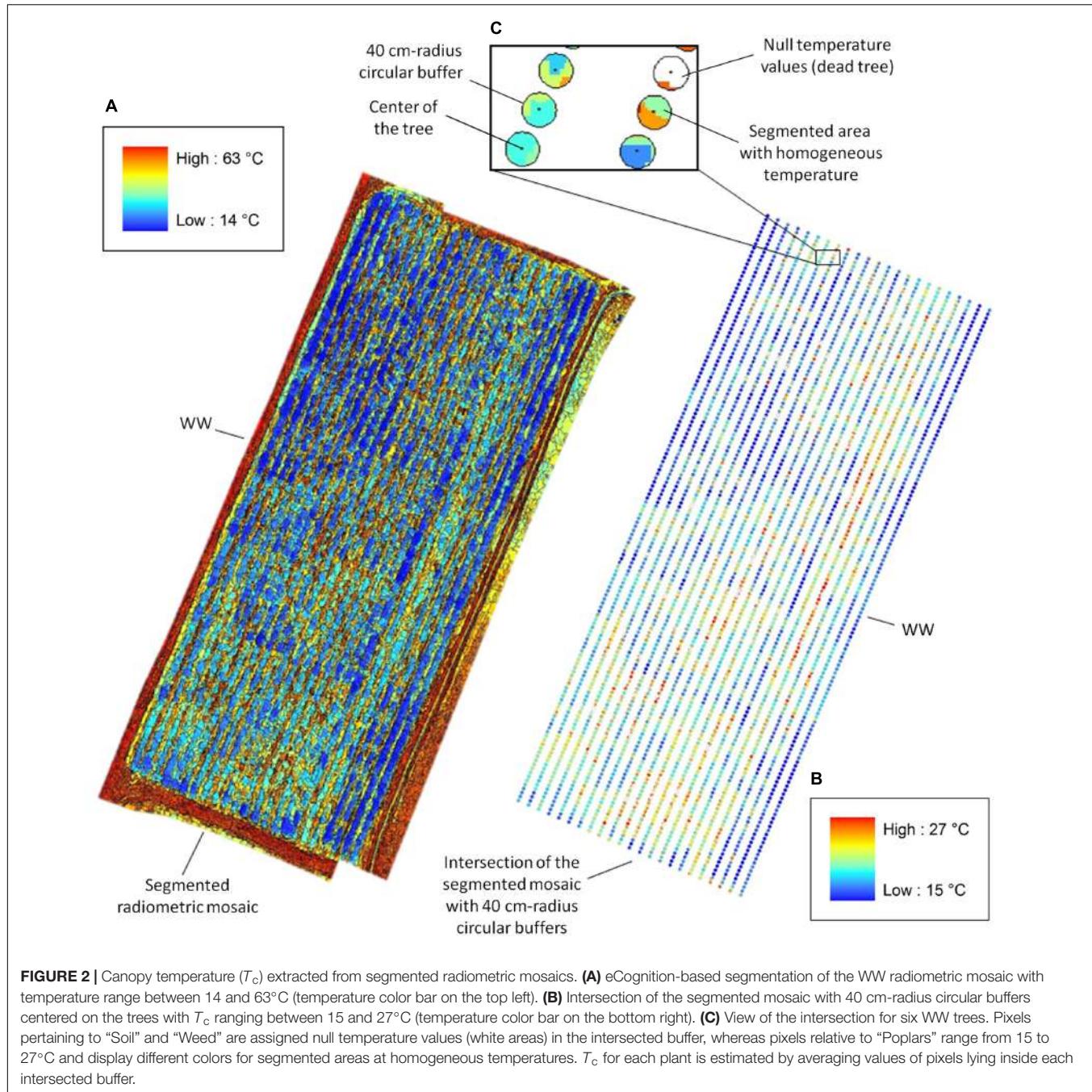
To validate the UAV-based approach, we studied the relationship between ground-based midday stomatal conductance ( $g_s$ , mmol m<sup>-2</sup> s<sup>-1</sup>) and  $T_c$  collected from the UAV for selected genotypes, see **Figure 1C**. At the same time of UAV flights, we collected abaxial  $g_s$  data using a dynamic diffusion porometer (AP4, Delta-T devices Ltd., United Kingdom). Measurements were taken on three biological replicates per two water treatments on each of the four parental genotypes ( $3 \times 2 \times 4 = 24$  trees). On each tree, two  $g_s$  technical replicates were taken on the first sunlit fully expanded leaf from the canopy top. For each plot, before measurements, the porometer was calibrated according to the experimental site  $T_{air}$  and humidity. Instrument calibration,  $g_s$  measurements, and moving within and between plots required approximately 3 h.

### Thermal Image Processing

We collected a total of 7836 thermal images (“.fff” files) during the UAV campaign. Given the high overlap and sidelap, one frame every 20 (392 images) was converted to radiometric “.jpg” through IRT Analyzer (Grayess, United States), and fish-eye undistorted through Adobe Photoshop CC (Adobe Systems, United States) (lens adjustment tool) by setting the camera focal length to 9 mm, see **Figure 1F**. Image mosaicking was performed with the Image Composite Editor software (Microsoft Corporation, United States), and radiometric mosaics were then converted to grid data with Surfer (Golden Software, United States). Mosaics were orthorectified and georeferenced with ArcGIS 9.2 (ESRI, United States). Orthomosaics were georeferenced by manually matching the surveyed 16 GCPs (**Figure 1F**).

Mosaics were processed to remove bare soil pixels (Jones and Sirault, 2014), and used to estimate average  $T_c$  (°C) for each tree. In particular, radiometric mosaics were combined with the position of the tree centers and spacing in the experimental plots (**Figure 2**). Firstly, radiometric mosaics were segmented to automatically identify regions depicting trees. Canopy identification was achieved through two independent semi-automatic image segmentation approaches (**Figure 1G**). We utilized the eCognition commercial software (eCognition Developer 9, Tremble Inc., United States) that is commonly adopted in image-based analysis for environmental applications. Further, we in-house developed a second segmentation algorithm in a Matlab environment (Matlab R2014a, The Mathworks Inc., United States).

In the eCognition segmentation, parameters such as canopy shape (set to 0.1), compactness (set to 0.5), and scale (set to 10) were used to identify trees. To address bare soil pixel removal, two pixel classes were introduced. One class included pixels at temperatures ranging from 15 to 27°C and from 14 to 28°C for WW and mDr, respectively (named “Poplar” in **Figure 1G**). Conversely, the second class consisted of pixels at temperatures lower and higher than the above ranges (named “Weed” and “Soil” in **Figure 1G**). We visually ascertained in the radiometric



mosaics that the first class of temperatures was related to plants, and the second to soil and weed. Segmented areas "Weed" and "Soil" were discarded in the subsequent temperature extraction analysis.

In the alternative segmentation approach, Otsu's threshold selection method was implemented on the entire radiometric mosaics (Otsu, 1979). Mosaic histograms were thresholded to obtain nine level-segmented images. The effectiveness of the segmentation was assessed through Otsu's objective criterion ( $N = 0.99$  in case of eight thresholds). Based on visual inspection of the mosaics, pixels at temperatures ranging from 15 to 27°C

(WW) and from 14 to 28°C (mDr) were retained for estimating average  $T_c$ .

After segmentation, pixels relative to soil and weed were assigned null temperature values in the radiometric mosaics. Then, 40 cm-radius circular buffers centered on the trees were intersected with the segmented mosaics (Figure 2). Since neighboring plants were spaced by 1 m within a row, such conservative buffer was selected to guarantee the precise identification of each plant canopy. Finally,  $T_c$  for each tree was calculated by averaging pixel values lying inside each intersected buffer.

## Statistical Analysis

Phenotypic variance of a total of 503 genotypes was investigated using R software (R v.3.1.3, R Foundation for Statistical Computing, Austria). Due to natural replicates mortality since plot establishment, only genotypes with at least three replicates in both WW and mDr were retained in the analysis. Two-way ANOVA (ANalysis Of VAriance) inferential statistic procedure was used to describe the effects of genotype, treatment, and their interaction on total phenotypic variance observed in POP6. To respect ANOVA assumptions, the Box-Cox procedure (Box and Cox, 1964) was performed on the additive model to yield optimal data transformation, and Bartlett's test was used to test the homogeneity of variance. When statistically significant differences among blocks were found, block effect adjustment was conducted to minimize the influence of competition among neighboring trees on genotype-specific response to drought conditions. Adjustment was performed according to Dillen et al. (2009), and it was repeated separately for each experimental plot.

The generalized linear mixed model was built to test differences among genotypes within each treatment and between the two treatments, and differences due to the genotype by treatment interaction ( $G \times T$ ).  $G \times T$  aims at testing the consistency in the relative performance of genotypes grown in different conditions (White et al., 2007). Statistical significance was considered for  $p$ -values  $\leq 0.05$ . To capture the proportion of total phenotypic variance due to genetic variation, broad-sense heritability ( $H^2$ ) was estimated for  $T_c$  in each treatment according to Sehgal et al. (2015):

$$H^2 = \sigma_G^2 / [\sigma_G^2 + (\sigma_\epsilon^2 / r)] \quad (1)$$

Here,  $\sigma_G^2$  and  $\sigma_\epsilon^2$  are the genetic and residual variance components, respectively, and  $r$  is the average number of replicates for a genotype within treatment (i.e., 3.5).

Moreover, to provide a better estimation,  $H^2$  was calculated over combined treatments, taking into account the  $G \times T$  effects on total variance:

$$H^2 = \sigma_G^2 / [\sigma_G^2 + (\sigma_{G \times T}^2 / n) + (\sigma_\epsilon^2 / nr)] \quad (2)$$

Here,  $\sigma_{G \times T}^2$  is the  $G \times T$  variance component,  $n$  is the number of treatments, and  $r$  is the average number of replicates for a genotype in the experiment (i.e., 7.3).

Variance components were obtained using the restricted maximum likelihood (REML) procedure, with treatment as fixed effect, and genotype and  $G \times T$  as random effects.

Finally, to evaluate the robustness of the independent segmentation procedures, differences between  $T_c$  datasets obtained with eCognition and Matlab, and the precision of both procedures was tested using the Student's  $t$ -test and the Spearman's rank correlation test. Statistical significance of the correlation coefficient ( $\rho$ ) was considered for  $p$ -values  $\leq 0.05$ .

## Stress Susceptibility Index

*Populus nigra* genotype response to drought was dissected by computing the Stress Susceptibility Index (SSI) on UAV-based  $T_c$ . Such an index has proved to be an efficient tool to classify plants according to their tolerance or sensitivity to water stress (Sánchez

et al., 2015). SSI was calculated based on genotypic mean  $T_c$  according to Fischer and Maurer (1978):

$$SSI = [1 - (T_{cmDr}/T_{cWW})] / [1 - (\bar{T}_{cmDr}/\bar{T}_{cWW})] \quad (3)$$

Here,  $T_{cmDr}$  and  $T_{cWW}$  correspond to genotypic means under mDr and WW conditions, respectively, and  $\bar{T}_{cmDr}$  and  $\bar{T}_{cWW}$  correspond to POP6 means in mDr and WW conditions, respectively.

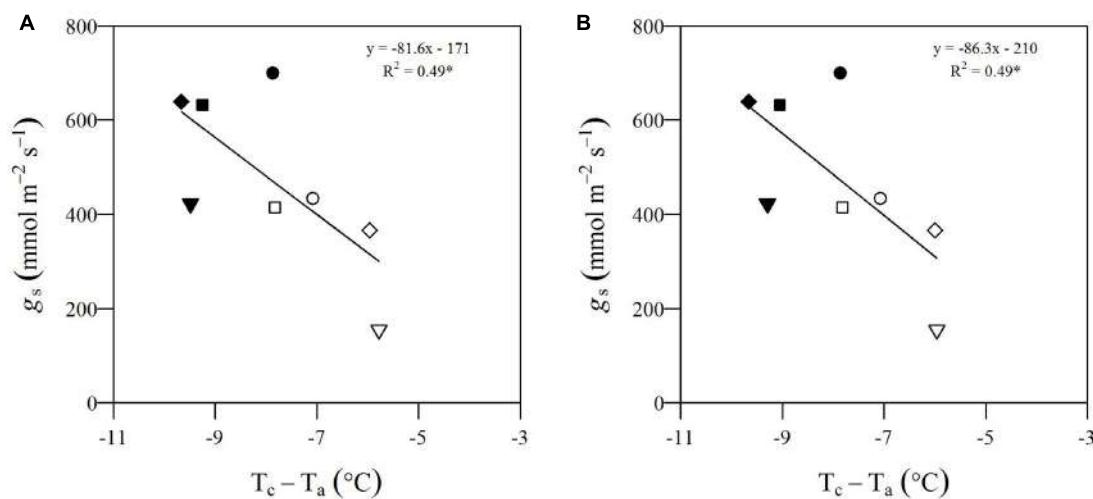
Negative SSI values correspond to a decrease in the genotypic mean  $T_{cmDr}$  with respect to  $T_{cWW}$  together with an increase in  $\bar{T}_{cmDr}$  with respect to  $\bar{T}_{cWW}$ . SSI values equal to 0 indicate consistent genotype means in mDr and WW conditions, regardless to POP6 mean responses. SSI values comprised between 0 and 1 indicate that  $T_c$  increase in mDr with respect to WW in the genotype is lower than  $T_c$  increase observed in the population. SSI values equal to 1 correspond to consistent deviations between mDr and WW conditions for both genotype and population means. SSI values greater than 1 indicate that  $T_c$  increase in mDr with respect to WW in the genotype is greater than  $T_c$  increase observed in the population. Therefore, genotypes whose index lies between 0 and 1 suggest an improved response to drought as compared to the overall population behavior.

## RESULTS

### UAV-Based HTFP for Detecting Response to Drought

The UAV-based methodology allowed collection of high-throughput thermal data on experimental plots covering an area of 1.67 ha in only 22 min. While ground-truthing required 3 h to obtain  $g_s$  data on 24 trees, by performing only two low-elevation UAV flights, we were able to capture the response to drought of 6716 trees with a spatial resolution of 6 cm  $\times$  6 cm, that is, at POP6 sub-leaf definition. The methodology was simple to implement in the field: TIR camera required no infrared target-based calibration; standard GCPs were located in the experimental plots; and the UAV was autonomously navigated based on pre-planned mission (Figure 1). Data processing was almost fully automated: images were undistorted, mosaicked, and orthorectified with commercial user-friendly software. The supervision of an expert was mandatory to georeference the orthomosaics and to segment images. Image segmentation allowed for addressing the mixed-pixel problem and automatically identifying tree canopies through two independent algorithms. Both methodologies required an expert user to set parameters and classes based on visual inspection of the orthomosaics, and their implementation was computationally inexpensive. The extraction of tree  $T_c$  from radiometric orthomosaics for both treatments was executed within 1 h, and computational time devoted to statistical analyses was on the order of a few minutes.

To validate UAV-based data, we utilized  $T_c$  of selected genotypes to recover the well-known inverse correlation between  $g_s$  and the difference between  $T_c$  and  $T_{air}$  (Farooq et al., 2009;



**FIGURE 3 |** Relationship between  $g_s$  (mmol  $m^{-2} s^{-1}$ ) and the difference between  $T_c$  and  $T_a$  ( $^{\circ}C$ ) for parental genotypes. **(A)** eCognition-based data and **(B)** Matlab-based data. For Poli (circle), 58-861 (triangle), P36 (square), and P64 (diamond),  $T_c$  and  $g_s$  are obtained by averaging measurements taken on biological replicates. Both sets of data are fitted with a linear function ( $R^2 = 0.49$ ). Data relative to mDr conditions are displayed as white symbols, whereas WW data are shown as black symbols. During the UAV flights,  $T_a$  was equal to  $28.75^{\circ}C$ . Statistically significant ( $p$ -value  $\leq 0.05$ ) regressions are indicated with the symbol \*.

Costa et al., 2013; Virlet et al., 2014). **Figure 3** illustrates the relationship between  $g_s$  and the difference between  $T_c$  and the air temperature at the time of the UAV flight ( $T_a$ ) for parental genotypes. They were chosen due to specific morphological and ecological traits, which were expected to lead to divergent drought responses. In **Figure 3A**, we show eCognition-based data, and, in **Figure 3B**, we report Matlab-based data. In both graphs, data relative to mDr conditions are found at high values on the X-axis and with  $g_s$  ranging from 154–434 mmol  $m^{-2} s^{-1}$  on the Y-axis, whereas lower WW data correspond to higher  $g_s$  values (ranging from 422 to 700 mmol  $m^{-2} s^{-1}$ ). Both data sets were fitted with a linear regression, statistically significant at an  $R^2$  of 0.49 ( $p$ -value  $< 0.05$ ), suggesting that UAV-based thermal data accurately captured plant response to drought conditions.  $T_c$  and  $g_s$  mean values, with relative standard errors, are shown in **Table 1**.

Remarkably, both segmentation approaches led to consistent relationships. To further assess the robustness of the independent segmentation methodologies, we computed the Student's *t*-test on POP6 genotypic mean  $T_c$ , separately for mDr and WW conditions. Notably, differences in data obtained with eCognition and Matlab were not statistically significant for both treatments

(WW:  $p$ -value = 0.82 and mDr:  $p$ -value = 0.36). Moreover, on data sets obtained with eCognition and Matlab, we also evaluated the Spearman's rank correlation test, and  $p$ -values of 0.98 ( $p$ -value  $< 0.001$ ) were found for both WW and mDr conditions. Based on such strongly significant correlation, both segmentation methods showed consistent  $T_c$  estimations and, therefore, could be interchangeably utilized to extract  $T_c$  from thermal images. In the succeeding figures, we report results for eCognition data; genotypic mean  $T_c$  along with standard error and SSI values for both eCognition and Matlab are provided in the Supplementary Tables S1, S2.

## POP6 Response to Drought

We report results for 503 genotypes (out of the original 691, due to tree mortality) for eCognition and Matlab-based segmentations. We only retained genotypes with at least three survived replicates in both mDr and WW conditions.

To show within-population and within-genotype variability in  $T_c$ , **Figure 4** displays genotypic mean  $T_c$  in WW (A) and mDr (B) conditions as obtained after segmentation with eCognition. In addition, in **Figure 4C**, we compared the genotypic response to drought by reporting the relative increase in  $T_c$  mDr with respect

**TABLE 1 |** Phenotypic data for  $T_c$  and  $g_s$  for poplar parental genotypes.

Genotype	$g_s$ (mmol $m^{-2} s^{-1}$ )		eCognition-based $T_c - T_a$ ( $^{\circ}C$ )		Matlab-based $T_c - T_a$ ( $^{\circ}C$ )	
	WW (Mean $\pm$ SE)	mDr (Mean $\pm$ SE)	WW (Mean $\pm$ SE)	mDr (Mean $\pm$ SE)	WW (Mean $\pm$ SE)	mDr (Mean $\pm$ SE)
Poli	700 $\pm$ 60	434 $\pm$ 16	-7.86 $\pm$ 1.33	-7.08 $\pm$ 0.73	-7.86 $\pm$ 1.33	-7.08 $\pm$ 0.70
58-861	422 $\pm$ 32	154 $\pm$ 6	-9.48 $\pm$ 1.25	-5.77 $\pm$ 0.36	-9.29 $\pm$ 1.44	-5.95 $\pm$ 0.18
P36	631 $\pm$ 95	414 $\pm$ 215	-9.25 $\pm$ 0.80	-7.82 $\pm$ 1.50	-9.05 $\pm$ 0.99	-7.82 $\pm$ 1.49
P64	639 $\pm$ 120	366 $\pm$ 45	-9.66 $\pm$ 0.69	-5.96 $\pm$ 1.21	-9.66 $\pm$ 0.69	-5.99 $\pm$ 1.17

Mean values with standard errors (Mean  $\pm$  SE) within each drought treatment for  $g_s$  (mmol  $m^{-2} s^{-1}$ ) and for the difference between  $T_c$  and  $T_a$  ( $^{\circ}C$ ).

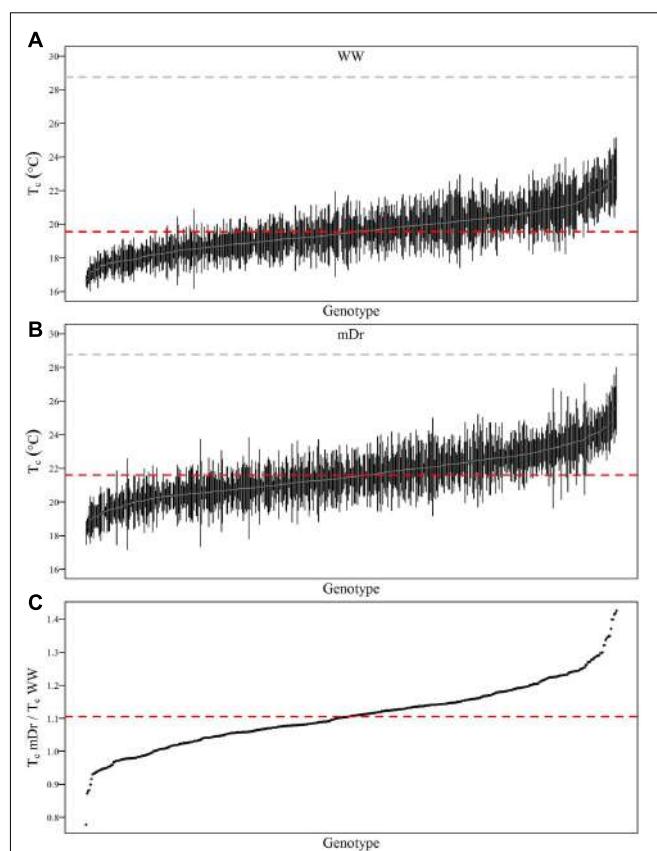
to  $T_c$ WW ( $T_c$ mDr/ $T_c$ WW). Genotypes were ordered based on increasing  $T_c$  (**Figures 4A,B**) and  $T_c$ mDr/ $T_c$ WW (**Figure 4C**). Dashed red lines show POP6 mean  $T_c$  (**Figures 4A,B**) and  $T_c$ mDr/ $T_c$ WW (**Figure 4C**).

In **Figure 4A**, a percentage of 49.3% of genotypes lies above the average  $T_c$  of 19.55°C. For Matlab, 48.4% of genotypes were found above a consistent average  $T_c$  of 19.55°C. Further, the standard error of each genotypic mean ranges from 0.05 to 2.86 for eCognition and from 0.05 to 3.24 for Matlab. In **Figure 4B**, a percentage of 46.3% of genotypes lies above the average  $T_c$  of 21.60°C. For Matlab, 47.4% of genotypes were found above a similar average  $T_c$  of 21.55°C. Further, the standard error of each genotypic mean ranges from 0.08 to 3.56 for eCognition and from 0.16 to 3.17 for Matlab. Genotypic mean  $T_c$  in mDr was on average warmer than in WW conditions. In WW settings, in eCognition, average  $T_c$  ranged from 16.62 to 23.33°C (Matlab:  $T_c$  ranged from 16.62 to 23.06°C). With regards to mDr settings, in eCognition, average  $T_c$  spanned from 18.16 to 26.00°C (average  $T_c$  in mDr ranged from 18.14 to 24.87°C for Matlab).

In **Figure 4C**, the average ratio is equal to 1.11 for both eCognition and Matlab. The percentage of genotypes lying above such value (up to a maximum value of 1.41) is equal to 49.7 and 48.6% for eCognition and Matlab, respectively. A total of 37.38% (Matlab: 38.49%) of genotypes are found between 1 and 1.11, whereas only 12.92% (Matlab: 12.90%) of genotypes have ratios ranging from 0.79 to 1.

To inspect the frequency distribution of genotypic response to drought, in **Figure 5** we report histograms for genotypic mean  $T_c$  obtained with eCognition for WW (A) and mDr (B) conditions. Such frequency distribution is expected to provide insights into POP6 average response to treatments and also the population response with respect to 58-861 and Poli. Both graphs display an approximately symmetric distribution (skewness lower than 0.4) with a slightly platykurtic shape (kurtosis almost equal to 0). Similar data distributions were found for Matlab (**Table 2**). Further, we show average  $T_c$  for 58-861 and Poli genotypes. In WW, parental genotypes present a similar  $T_c$  (58-861 equal to 19.84 and 19.80°C; Poli equal to 19.76 and 19.75°C in eCognition and Matlab, respectively). Conversely, a greater difference in  $T_c$  between parental genotypes was found in mDr than in WW (58-861 equal to 20.88 and 20.93°C; Poli equal to 21.78 and 21.66°C in eCognition and Matlab, respectively). Different from parental genotypes, POP6 showed a large range of variation within WW and mDr treatments for both segmentation procedures. Indeed, for eCognition, average  $T_c$  ranged from 16.62 to 23.33°C in WW, and from 18.20 to 26.01°C in mDr. For Matlab, a very similar range of variation (with respect to eCognition) was found in WW (from 16.62 to 23.05°C), whereas a narrower range was observed in mDr (from 18.16 to 24.84°C).

To quantitatively assess the effect of the treatments on POP6, a two-way ANOVA was used to analyze  $T_c$  differences among genotypes, between treatments, and due to G × T interaction (**Table 3**). Notably, differences among genotypes were not statistically significant. On the other hand, differences between treatments were statistically significant. Finally, G × T interaction was not statistically significant for both eCognition- and Matlab-based  $T_c$ .

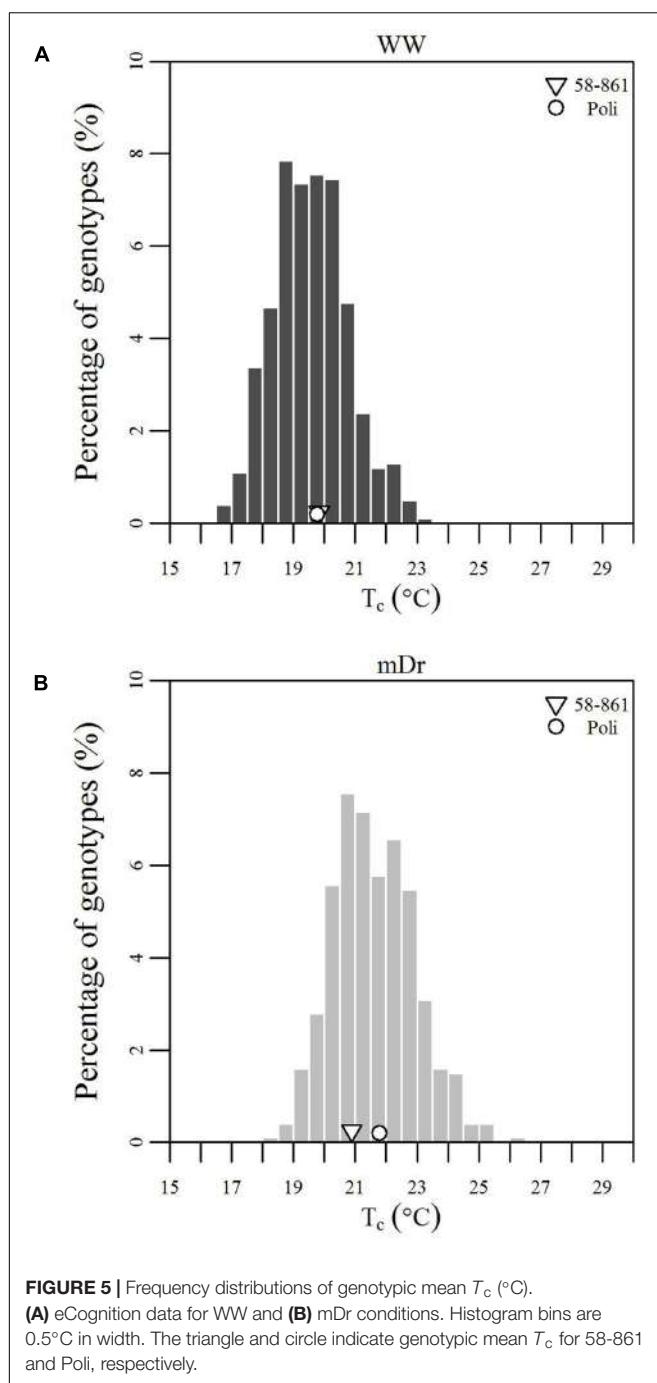


**FIGURE 4 |** Average  $T_c$  (°C) for each genotype in WW and mDr. **(A)** Genotypic mean  $T_c$  (gray circles) and standard error (black bars) in WW conditions. **(B)** Genotypic mean  $T_c$  (gray circles) and standard error (black bars) in mDr conditions. **(C)** Ratio of  $T_c$ mDr to  $T_c$ WW (black circles). Genotypes in **(A,B)** are ordered based on increasing  $T_c$ . Genotypes in **(C)** are ordered based on increasing ratios. The dashed gray line and dashed red lines in **(A,B)** show  $T_c$  during the unmanned aerial vehicle (UAV) flights and population averaged  $T_c$ , respectively. The dashed red line in **(C)** displays population averaged  $T_c$ mDr/ $T_c$ WW. A total number of 503 genotypes is shown in **(A–C)**.

To address the relationship between genetic and environmental sources of variance,  $H^2$  of the  $T_c$  was estimated. Within treatment analysis resulted in very low  $H^2$  for both eCognition (WW: 0.09, mDr: <0.01) and Matlab-based  $T_c$  (WW: 0.10, mDr: <0.01). Similarly, very low  $H^2$  (0.015 in eCognition and <0.01 in Matlab) was observed for combined treatments.

## Selection of Putative Drought-Tolerant Genotypes

Although the genotypic effect did not show a statistically significant influence on trait variance, we further analyzed  $T_c$  data to identify genotypes with improved drought tolerance. Genotype performance in response to drought is illustrated in **Figure 6**. Herein, the difference between single genotypic mean  $T_c$ mDr and POP6  $\bar{T}_c$ mDr is plotted against the difference between genotypic mean  $T_c$ mDr and  $T_c$ WW conditions. Data show a positive correlation, where highly stressed individuals



are located at the tails of the distribution and correspond to a less frequent behavior. Genotypes are mostly evenly distributed above and below the  $X$ -axis; however, stressed conditions lead to a majority of data points lying in the first and fourth quadrants.

From a drought-response perspective, the fourth quadrant in the biplot in **Figure 6** provides the most relevant information. Genotypes lying in the first and fourth quadrants [436 (86.68%) in eCognition and 438 (86.88%) in Matlab] present a  $T_c\text{mDr}$  greater than  $T_c\text{WW}$ . While genotypes in the first quadrant have a  $T_c\text{mDr}$  greater than  $\bar{T}_c\text{mDr}$ , those in the fourth quadrant show a  $T_c\text{mDr}$

smaller than  $\bar{T}_c\text{mDr}$ . A few genotypes display greater  $T_c\text{WW}$  than  $T_c\text{mDr}$  [data points in the second and third quadrants, 67 (13.32%) out of the total number of genotypes in eCognition and 66 (13.12%) in Matlab].

Due to mDr treatment, in genotypes found in the fourth quadrant ( $T_c\text{mDr} > T_c\text{WW}$  and  $T_c\text{mDr} < \bar{T}_c\text{mDr}$ ),  $T_c$  increased less than POP6  $\bar{T}_c\text{mDr}$ . This suggested the onset of acclimation mechanisms and, therefore, an improved response to drought stress with respect to the overall population. Among such drought-tolerant genotypes, 25% of the total number of tested ones (25.65% in eCognition and 25.79% in Matlab) presented an SSI ranging from 0 to 1, that is, the relative increase in  $T_c$  from WW to mDr conditions was lower than the increase observed on average in the population. Genotypes 58-861 and Poli were also located in the fourth quadrant; however, none of them had an SSI comprised between 0 and 1.

## DISCUSSION

In this study, we developed a UAV-based HTFP approach to investigate the response of *P. nigra* to mDr conditions in the field. We assessed the effects of two water treatments on the observed phenotypic variance of an  $F_2$  partially inbred population of 503 genotypes. Notably, we remotely captured high resolution images, whereby image pixels were several orders of magnitude smaller than a single tree crown, and smaller than POP6 average leaf area. Such a resolution could be essential to uncover physiological differences within single crowns. A large leaf-to-leaf variability has been observed for  $g_s$  and leaf temperature in wheat (*Triticum aestivum*), which has led to low values of  $H^2$  when estimated on a single-leaf basis (Rebetzke et al., 2001, 2013). Gonzalez-Dugo et al. (2012) detected in mildly stressed almond (*Prunus dulcis*) that few areas within the crown had substantial stomatal closure while, in the rest of the crown, the stomata were still open and this increased heterogeneity of the  $T_c$ . It has also been proposed that in cotton (*Gossypium* sp.) the variability of leaf temperature may provide important information about the degree of stomatal closure (Fuchs, 1990). In apple trees (*Malus pumila*) grown under drought conditions, the spatial variability of leaf temperature and  $g_s$  was higher for the whole crown than for the top crown (Ngao et al., 2017). By contrast, our UAV-based thermal imaging provides an ideal approach for the collection of the large number of individual leaf temperatures, which are necessary for methods based on temperature frequency distributions simultaneously in one image, rather than point-wise approaches for investigating tree response to drought. We also expect  $g_s$  to be consistently homogeneous in the upper canopy, that is, where  $g_s$  measurements were conducted herein.

High image resolution enabled the extraction of biologically meaningful data; in fact, plant leaves were attributed several image pixels, thus allowing accurate estimations of  $T_c$ . Such a rapid and non-invasive UAV-based procedure is expected to highly benefit phenotypic-based assisted-mass selection in early generation screening in breeding programs for bioenergy purposes. UAVs could, indeed, be adopted to capture high resolution images over sparsely vegetated environments, such

**TABLE 2 |** Frequency distribution shape parameters.

Segmentation Procedure	Treatment	Mean (°C)	Median (°C)	Kurtosis	Skewness
eCognition-based $T_c$	WW	19.55	19.53	-0.13	0.27
	mDr	21.60	21.49	-0.01	0.35
Matlab-based $T_c$	WW	19.55	19.48	-0.13	0.28
	mDr	21.55	21.48	-0.30	0.18

$T_c$  Mean (°C), median (°C), kurtosis, and skewness values for histograms in **Figure 5**.

**TABLE 3 |** Two-way ANalysis Of VAriance (ANOVA) on eCognition and Matlab-based  $T_c$  (°C) in POP6.

Segmentation procedure	Source of variance	Degrees of freedom	Sums of squares	Mean square	F-ratio	p-value
eCognition-based $T_c$ *	Genotype	498	0.01564	0.000031	1.056	0.211
	Treatment	1	0.02108	0.021077	708.437	<0.001
	G × T	498	0.01542	0.000031	1.041	0.276
	Error	2634	0.07836	0.000030		
	Total	3631	0.13050			
Matlab-based $T_c$ **	Genotype	498	0.0751	0.00015	1.037	0.295
	Treatment	1	0.1038	0.10380	713.445	<0.001
	G × T	498	0.0765	0.00015	1.056	0.209
	Error	2642	0.3844	0.00015		
	Total	3639	0.6398			

\*To respect ANOVA assumptions, individual eCognition-based  $T_c$  raw data were transformed by  $\sim -1$ . \*\*To respect ANOVA assumptions, individual Matlab-based  $T_c$  raw data were transformed by  $\sim -0.5$ . Effects of genotype, treatment (well-watered, WW and moderate drought, mDr), and two-factor interaction (G × T) on  $T_c$ .

as orchards (Sepulcre-Canto et al., 2006; Sepulcre-Canto et al., 2007), or over very extended areas, such as natural forests (Torresan et al., 2016). While UAV-based phenotyping approaches have been already tested in agriculture, screening high dimensionality populations is a remarkable bottleneck in forestry applications and this is the first time that these measures have been applied to forest species.

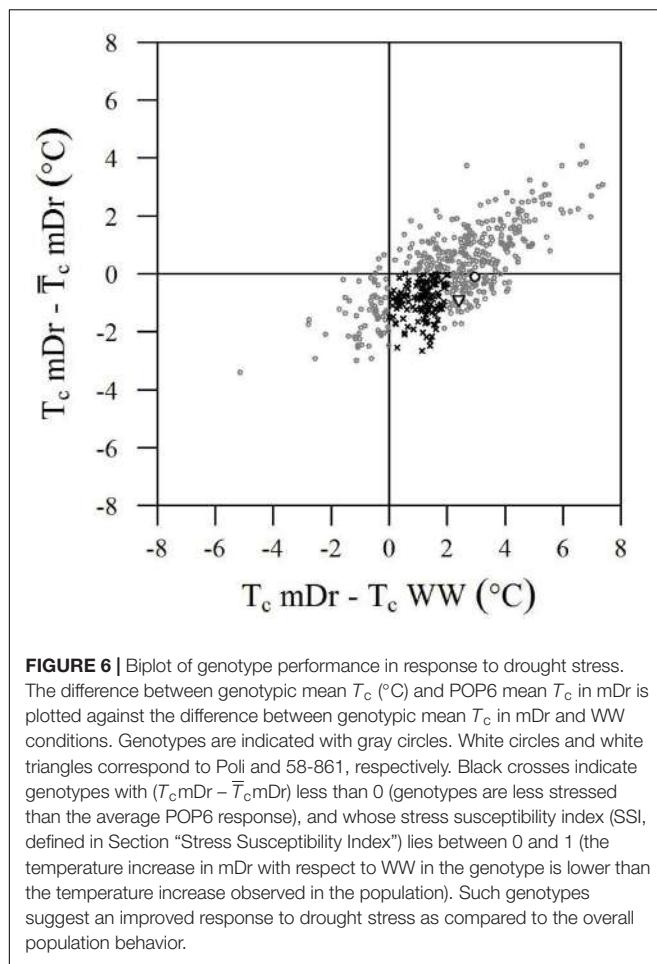
Image processing at high speeds is a central challenge in the field of HTFP. Thermal imaging also allows leaves to be distinguished from the background. If done manually, however, the necessary image processing can be rather labor-intensive and may also be dependent on subjective image interpretation. Our images also identified thousands of tree canopies against background soil and weed through two independent semi-automated segmentation procedures. Segmentation was utilized to reduce the mixed-pixel problem by extracting contours of areas at consistent temperatures. Then, areas relative to trees were identified based on visual inspection of orthomosaics and retained for data processing. Alternative approaches, which are based on manually drawing tree canopies (Virlet et al., 2014), would be extremely time-consuming and may lead to user-biased results. In this study, we developed and standardized a semi-automated image-based analysis procedure and directly applied it on thermal orthomosaics. Without relying on RGB images, we evaluated the sensitivity of the method to image segmentation by experimenting with two independent algorithms. Remarkably, both segmentations led to statistically similar tree responses to drought, thus supporting the robustness of the methodology.

The level of stress induced by the treatment was fully captured through thermal images. This was indicated by the inverse linear

relationship between  $g_s$  and  $T_c$ . An increase in the difference of  $T_c$  with respect to  $T_a$  corresponded to a decrease in transpiration flux and, therefore, to a decrease in the ratio of actual to potential transpiration (Farooq et al., 2009; Virlet et al., 2014). Similar linear regressions with slightly higher  $R^2$  have also been observed for orange (*Citrus sinensis*) ( $R^2 = 0.70\text{--}0.78$ ) (Zarco-Tejada et al., 2012; Ballester et al., 2013a), persimmon (*Diospyros kaki*) ( $R^2 = 0.46$ ) (Ballester et al., 2013b), and almond (*Prunus dulcis*) ( $R^2 = 0.59\text{--}0.66$ ) (Gonzalez-Dugo et al., 2012), using UAV, small airplanes, and ground screening techniques. However, comparable results for forest tree species through UAV-based phenomics are still undocumented.

As already pointed out in a previous greenhouse study on early effects of drought on *P. nigra*, Poli tended to quickly respond to stress by closing stomata due to the fact that Poli is adapted to dry/hot climatic conditions. However, 58-861 reacted more slowly to drought as it is considered better adapted to cool and moist climates (Cocozza et al., 2010). This behavior could be explained by proved geographical and environmental gradients of  $g_s$ , with higher  $g_s$  values observed in northern proveniences of *Populus* sp. (McKown et al., 2013; Kaluthota et al., 2015). This motivates lower  $T_c$  values for 58-861 than Poli as observed in **Figure 5B**.

Due to drought stress, POP6  $T_c$  increased and tended to the temperature of the environment as similarly seen in Mahan and Upchurch (1988). During the UAV flights,  $T_a$  was equal to 28.75°C; in WW conditions, the average difference between  $T_c$  and  $T_a$  was equal to  $-9.19^\circ\text{C}$ . Similar differences (from 10 to 15°C) between air and leaf temperatures have already been demonstrated to be plausible (Jackson et al., 1981). As expected, such a difference decreased to  $-7.14^\circ\text{C}$  in case of



mDr conditions, when transpiration flux was reduced. These findings are consistent with previous drought studies (Zarco-Tejada et al., 2012; Ballester et al., 2013a,b). Also, in agreement with Ballester et al. (2013a,b), and Costa et al. (2013), the treatment resulted in a difference of  $2^{\circ}\text{C}$  between WW and mDr POP6 average  $T_c$ . Figures 5A,B demonstrated that drought induced high phenotypic variability (large ranges of variation in frequency distributions). In fact, genotypic mean  $T_c$  followed a Gaussian distribution with a high degree of transgressive segregation in both treatments (thermal response of POP6 was extreme as compared to parental response). The high variability in  $F_2$  populations may be due to the transgressive segregation as was previously observed by Wu and Stettler (1994) and Rae et al. (2009) for growth-related traits in similar populations. This suggests that thermal response to drought is a quantitative trait controlled by several genes (polygenic trait) (White et al., 2007), and that  $T_c$  is under complex but repeatable genetic control (Rebetzke et al., 2013). In our study, the lack of statistically significant  $G \times T$  interaction suggests that POP6 response is consistent between both treatments (i.e., stressed genotypes increase their  $T_c$ , and genotypes with higher  $T_c$ WW with respect to other genotypes also display higher  $T_c$ mDr). Such consistency may facilitate the selection of genotypes of interest.

Poplar  $T_c$  response to water stress was further explored by investigating  $H^2$ . A very low  $H^2$  was observed for POP6, due to a large residual error and a modest genetic influence on the phenotypic variance. Generally,  $H^2$  of a trait varies across different populations and environments (Griffiths et al., 2000), and it is overestimated when  $G \times T$  interaction is significant (White et al., 2007). To the best of our knowledge,  $H^2$  estimation of  $T_c$  in forest tree species is still not reported. Yet, it was observed in the range from 0.05 to 0.91 in *T. aestivum*, with higher values of  $H^2$  obtained for multi-year and multi-environment trials (Rebetzke et al., 2013). Such a large variation supports the urge to conduct experiments in different environments and water limitation conditions. Moreover, it is noted that recurrent selection is optimal for improving traits with low  $H^2$  (Hopkins et al., 2009). Therefore, recurrent selection of the herein identified putative drought-tolerant genotypes would be a promising approach to accumulate favorable alleles in future crosses of POP6.

In addition, our technique aids in identifying genotypes that appear to be risk-takers versus those that are risk-averse; such an observation is known to occur in response to drought (Sade et al., 2012; Moshelion et al., 2014; Attia et al., 2015). In our experiment, Poli exhibits a risk-averse strategy by limiting transpiration and allowing  $T_c$  to increase, as opposed to 58-861, which appears to be a risk-taker genotype. In fact, Poli considerably increases its  $T_c$  from WW to mDr conditions. This response supports the fact that Poli could have a sensing mechanism that detects reduced water availability, and thereby closes stomata to avoid stress due to drought. Closed stomata at high light intensities could lead to photo-damage, and thus drought stress may become photo-oxidative stress, ultimately leading to biomass yield loss. In 58-861, a lower  $T_c$  increment ( $1.05^{\circ}\text{C}$  in eCognition and  $1.12^{\circ}\text{C}$  in Matlab) than Poli was observed between treatments. This suggests that 58-861 could employ a less conservative strategy, and thus, being a putative risk-taker, which developed an anisohydric survival strategy to drought. This adaptive choice may be beneficial in moderate stress conditions; however, it may not provide any advantages in case of prolonged and more intense drought conditions. The likelihood of anisohydric genotypes to succumb earlier to drought would increase, and that of isohydric genotypes to more likely die of carbon starvation is a function of drought intensity and duration.

Our findings show that, among trees exposed to mDr conditions, 63 (13.32%) out of 503 genotypes were found in the third quadrant in Figure 6, thus indicating greater  $T_c$ WW than  $T_c$ mDr. This behavior may be attributed to lower-density canopies (with short and sparse branches and with a small canopy surface area covered with leaves) in WW than in mDr. Conversely, 129 genotypes (25%) were located in the fourth quadrant and displayed an SSI comprised between 0 and 1. Even though  $g_s$  measurements are required to validate this response, these POP6 genotypes could be considered more risk-takers, which supposedly maintain high  $g_s$  and lower  $T_c$ . However, whether or not these risk-taker genotypes could be considered drought tolerant would depend on the period and severity of drought, and on the rate at which plants recover from drought exposure (Sade et al., 2012). For example, in Alvarez et al. (2007),

anisohydric purple lovegrass (*Eragrostis spectabilis*) maintained higher  $g_s$  and CO<sub>2</sub> assimilation, and showed better performance than isohydric miscanthus (*Miscanthus sinensis*) plants under optimal and mild-to-moderate drought conditions, but little difference was noted when both plants were subjected to severe drought. A greater research effort is currently needed to better understand the physiological mechanisms involved when plants tend to be risk-takers in stressed and risk-averse in non-stressed conditions.

Even if genotypic effect was not statistically significant, risk-taking genotypes accounting for 25% of the tested ones support that crossing genotypes with divergent morphological and ecological traits, such as 58-861 and Poli, resulted in augmented phenotypic variability (see also the large ranges of variation in **Figure 5**) in the early F<sub>2</sub> POP6 generation. Higher-order crossings with differential breeding techniques based on selection are expected to lead to higher  $H^2$  and improved genetic gains. Promising genotypes were capable of controlling stomatal closure to raise their  $T_c$  by less than 2°C with respect to non-stressed conditions (SSI ranging from 0 to 1). Interestingly, in Rizhsky et al. (2002, 2004), upon similar drought conditions, leaf temperature increased by 2–5°C. This supports our finding that trees whose  $T_c$  increases by less than 2°C showed a better response within POP6.

The extensive HTP executed in this study was done in drought trials under natural field conditions. The results of this study prove that this methodology enabled high-throughput data analysis in the field upon fast and non-invasive acquisition of thermal images. Notably, this approach allowed efficient and precise phenotyping of large population of individuals at the same time, thus minimizing the influence of variable meteorological conditions on control (in this case WW) and treated (in this case mDr) trees. Image resolution was remarkably higher than in previous studies (Berni et al., 2009) and highly sufficient for accurately characterizing tree response even at crown and leaf levels. We indicate that flight elevation was also standardized to 25 m to ensure high image definition while guaranteeing that UAV rotor downwash did not disturb  $T_c$ . Finally, the degree of automation employed in the developed HTP method is highly desirable and will be widely adopted in field-based phenomics for forest trees genetics and genomics research. Since  $T_c$  is strongly related to  $g_s$ , photosynthetic rate (Way and Oren, 2010), and leaf water potential (Testi et al., 2008), UAV-based thermal sensing may be an efficient, robust, and high-throughput tool for indirectly screening breeding populations, for selecting drought-tolerant, as well as risk-takers and risk-averse genotypes. This methodology is also promising toward monitoring the dynamics of stomatal movement in response to environmental stresses through rapid and repeated surveys. Improved knowledge of stress response will also be enabled through the synergistic integration of UAV remote sensing with more direct ground-based measurements (i.e., leaf fluorescence, leaf gas exchange, leaf water potential, leaf morphology, tree sap flow, and biomass production). Coupled with ground-penetrating radar, field-based phenomics will accelerate screening for drought tolerance by multi-scale analysis of root, crown, and leaf

traits. Currently, UAVs have been outfitted with multispectral, hyperspectral, thermal, RGB, and near-infrared cameras, which are useful to evaluate plant response to stress (Costa et al., 2013; Shi et al., 2016). Future technological ameliorations will afford unprecedented measurements such as chlorophyll fluorescence imaging and 3D mapping using light detection and ranging (LIDAR) sensors, directly from drones, thus opening novel avenues in high-throughput stress phenotyping in forest trees. Promising examples of this vision can also include UAV integrated platforms equipped with multiple sensors for simultaneous data collection.

In breeding programs, this new aerial screening method may sensibly alleviate the effect of environmental factors such as varying exposure to sunlight according to time of the day, local climatic conditions, background radiation due to soil vegetative cover and soil water content, being precise, automated and quick. In fact, utilizing highly sensitive radiometrically calibrated TIR cameras may speed up tree screening, thus enabling repeated observations over longer periods of time and larger scale areas. Furthermore, augmenting breeders' visual definition by selecting low flight altitudes may provide novel insights on the response to drought stress at the single leaf scale.

## CONCLUSION

In this study, we developed a high-resolution and high-throughput UAV-based phenomics method to investigate drought in the field. We applied this screening method to precisely and efficiently assess the response to drought of a *P. nigra* F<sub>2</sub> population consisting of 503 genotypes planted on an area of 1.67 ha. We captured thermal images of stressed and non-stressed trees from an elevation of 25 m. We reconstructed thermal mosaics and extracted the average  $T_c$  by using two independent image segmentation techniques. We statistically analyzed genotypic temperatures, and identified putative drought-tolerant genotypes. This newly developed approach enabled high resolution thermal orthomosaics from quick UAV-based acquisitions and simultaneous screening of a significant number of individuals.

Two segmentation techniques for accurately analyzing TIR images, one developed in-house using Matlab, and another relying on commercial software eCognition, were successfully implemented to eliminate the mixed-pixel problem. They both led to consistent results, indicating that it is possible to use HTFP-based thermography for the screening of tolerance to drought stress in forest trees. However, considering the complexity of drought tolerance, we suggest it can only act as an accessory means in an active breeding program for drought by contributing significantly to phenotyping of tree response to water stress. Future studies will aim at extending the methodology to rapidly generate environmentally nuanced temporal measurements of physiological differences to contribute to predictive environmental stress response models. Through the approach developed here, candidate genes for drought stress responses can also be identified when this HTFP is combined with advanced genomics approaches.

Based on our analysis, a good correlation was found between UAV-based parental genotypic mean  $T_c$  and ground-truth  $g_s$  measurements. Genotypic mean temperatures exhibited a Gaussian distribution centered about parental behavior. Furthermore, the statistically significant differences observed between treatments were attributed to environmental conditions. Finally, based on SSI values, 25% of the population exhibited increase in temperature under mDr conditions by less than 2°C, and, thus, can be regarded as candidate drought-tolerant genotypes.

The use of UAV for field-based tree phenotyping under drought conditions is novel, but is expected to become an important tool for improving efficiency in forest-tree breeding for climate change. To date, no studies have been carried out attempting to use UAV-based HTFP for forest tree phenotyping in managed stress trials in which specific and well-defined conditions are imposed, and effectively deploy such platform in a breeding program. Thanks to its high resolution aerial imagery, accurate data processing, and relatively simple implementation, this HTFP shows promise as a precise and efficient tool for use in phenomics studies.

## AUTHOR CONTRIBUTIONS

RL performed the ground-truthing, analyzed experimental data, and executed phenotypic and statistical data analyses. FT executed Matlab segmentation and, together with RL, developed results and wrote the manuscript. RS planned the airborne campaign, reconstructed thermal mosaics, and executed eCognition segmentation. SK contributed to develop the workflow and write the manuscript. GSM contributed to the original concept of the project. AH conceived the project and its

components, designed and supervised the study, and revised the manuscript. All authors discussed the results, read and approved the manuscript.

## FUNDING

This research was partially supported by the Regione Lazio/Lazio Innova under the grant “AgroEnVision -*Unmanned Aerial Vehicles as Mobile Multi-Sensor Platforms for Innovative and Sustainable Management of Agro-Environmental Ecosystems*” (FILAS-RU-2014-1191), grants from the European Community’s Seventh Framework Program (WATBIO FP7-311929) and the Brain Gain Program (Rientro dei cervelli) of the Italian Ministry of Education, University, and Research (AH).

## ACKNOWLEDGMENTS

We wish to thank Vincenzo Mantovano for providing the UAV remote sensing platform. We are very grateful to Francesco Fabbrini for his help in developing the experimental design, to Paolo Latini and Chiara Evangelista for their assistance during the airborne campaign and for collecting ground-truth data, and to Maurizio Sabatti for providing plant material. All authors thank Alasia Franco Vivai for plantation management and field maintenance.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.01681/full#supplementary-material>

## REFERENCES

- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*. FAO Irrigation and Drainage Paper No. 56. Rome: FAO.
- Alvarez, E., Scheiber, S. M., Beeson, R. C. Jr., and Sandrock, D. R. (2007). Drought tolerance responses of purple Lovegrass and ‘Adagio’ maiden grass. *HortScience* 42, 1695–1699.
- Amichev, B. Y., Johnston, M., and Van Rees, K. C. N. (2010). Hybrid poplar growth in bioenergy production systems: biomass prediction with a simple process-based model (3PG). *Biomass Bioenerg.* 34, 687–702. doi: 10.1016/j.biombioe.2010.01.012
- Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/j.tplants.2013.09.008
- Attia, Z., Domec, J.-C., Oren, R., Way, D. A., and Moshelion, M. (2015). Growth and physiological responses of isohydric and anisohydric poplars to drought. *J. Exp. Bot.* 66, 4373–4381. doi: 10.1093/jxb/erv195
- Ballester, C., Castel, J., Jiménez-Bello, M. A., Castel, J. R., and Intrigliolo, D. S. (2013a). Thermographic measurement of canopy temperature is a useful tool for predicting water deficit effects on fruit weight in citrus trees. *Agric. Water Manage.* 122, 1–6. doi: 10.1016/j.agwat.2013.02.005
- Ballester, C., Jiménez-Bello, M. A., Castel, J. R., and Intrigliolo, D. S. (2013b). Usefulness of thermography for plant water stress detection in citrus and persimmon trees. *Agric. For. Meteorol.* 168, 120–129. doi: 10.1016/j.agrformet.2012.08.005
- Benetka, V., Novotná, K., and Štochlová, P. (2012). Wild populations as a source of germplasm for black poplar (*Populus nigra* L.) breeding programmes. *Tree Genet. Genomes* 8, 1073–1084. doi: 10.1007/s11295-012-0487-6
- Berni, J. A. J., Zarco-Tejada, P. J., Suárez, L., and Fereres, E. (2009). Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *IEEE Trans. Geosci. Remote Sens.* 47, 722–738. doi: 10.1109/TGRS.2008.2010457
- Box, G. E. P., and Cox, D. R. (1964). An analysis of transformations. *J. Roy. Stat. Soc. B* 26, 211–252.
- Bréda, N., and Badeau, V. (2008). Forest tree responses to extreme drought and some biotic events: towards a selection according to hazard tolerance? *C. R. Geosci.* 340, 651–662. doi: 10.1016/j.crte.2008.08.003
- Bréda, N., Huc, R., Gramier, A., and Dreyer, E. (2006). Temperate forest trees and stands under severe drought: a review of ecophysiological responses, adaptation processes and long-term consequences. *Ann. For. Sci.* 63, 625–644. doi: 10.1051/forest:2006042
- Cairns, J. E., Impa, S. M., O’Toole, J. C., Jagadish, S. V. K., and Price, A. H. (2011). Influence of the soil physical environment on rice (*Oryza sativa* L.) response to drought stress and its implications for drought research. *Field Crop Res.* 121, 303–310. doi: 10.1016/j.fcr.2011.01.012
- Cantero-Martínez, C., Angás, P., and Lampurlánés, J. (2007). Long-term yield and water use efficiency under various tillage systems in Mediterranean rainfed conditions. *Ann. Appl. Biol.* 150, 293–305. doi: 10.1111/j.1744-7348.2007.00142.x
- Chamaillard, S., Fichot, R., Vincent-Barbaroux, C., Bastien, C., Depierreux, C., Dreyer, E., et al. (2011). Variations in bulk leaf carbon isotope discrimination,

- growth and related leaf traits among three *Populus nigra* L. populations. *Tree Physiol.* 31, 1076–1087. doi: 10.1093/treephys/tpo89
- Chaves, M. M., Maroco, J. P., and Pereira, J. S. (2003). Understanding plant responses to drought – from genes to the whole plant. *Funct. Plant Biol.* 30, 239–264. doi: 10.1071/FP02076
- Chéné, Y., Rousseau, D., Lucidarme, P., Bertheloot, J., Caffier, V., Morel, P., et al. (2012). On the use of depth camera for 3D phenotyping of entire plants. *Comput. Electron. Agric.* 82, 122–127. doi: 10.1016/j.compag.2011.12.007
- Cocozza, C., Cherubini, P., Regier, N., Saurer, M., Frey, B., and Tognetti, R. (2010). Early effects of water deficit on two parental clones of *Populus nigra* grown under different environmental conditions. *Funct. Plant Biol.* 37, 244–254. doi: 10.1071/FP09156
- Costa, J. M., Grant, O. M., and Chaves, M. M. (2013). Thermography to explore plant–environment interactions. *J. Exp. Bot.* 64, 3937–3949. doi: 10.1093/jxb/ert029
- Deery, D., Berni, J. A. J., Jones, H., Sirault, X., and Furbank, R. (2014). Proximal remote sensing buggies and potential applications for field-based phenotyping. *Agronomy* 5, 349–379. doi: 10.3390/agronomy4030349
- Deery, D. M., Rebetzke, G. J., Berni, J. A. J., James, R. A., Condon, A. G., Bovill, W. D., et al. (2016). Methodology for high-throughput field phenotyping of canopy temperature using airborne thermography. *Front. Plant Sci.* 7:1808. doi: 10.3389/fpls.2016.01808
- DeWoody, J. D., Trewin, H. T., and Taylor, G. T. (2015). Genetic and morphological differentiation in *Populus nigra* L.: isolation by colonization or isolation by adaptation? *Mol. Ecol.* 24, 2641–2655. doi: 10.1111/mec.13192
- Dhondt, S., Wuyts, N., and Inzé, D. (2013). Cell to whole-plant phenotyping: the best is yet to come. *Trends Plant Sci.* 18, 428–439. doi: 10.1016/j.tplants.2013.04.008
- Díaz-Varela, R. A., de la Rosa, R., León, L., and Zarco-Tejada, P. J. (2015). High-resolution airborne UAV imagery to assess olive tree crown parameters using 3D photo reconstruction: application in breeding trials. *Remote Sens.* 7, 4213–4232. doi: 10.3390/rs70404213
- Dillen, S. Y., Marron, N., Sabatti, M., Ceulemans, R., and Bastien, C. (2009). Relationships among productivity determinants in two hybrid poplar families grown during three years at two contrasting sites. *Tree Physiol.* 29, 975–987. doi: 10.1093/treephys/tpp036
- Djomo, S. N., Ac, A., Zenone, T., De Groote, T., Bergante, S., Facciotti, G., et al. (2015). Energy performances of intensive and extensive short rotation cropping systems for woody biomass production in the EU. *Renew. Sustain. Energy Rev.* 41, 845–854. doi: 10.1016/j.rser.2014.08.058
- Edrisi, S. A., and Abhilash, P. C. (2016). Exploring marginal and degraded lands for biomass and bioenergy production: an Indian scenario. *Renew. Sustain. Energy Rev.* 54, 1537–1551. doi: 10.1016/j.rser.2015.10.050
- Fahlgren, N., Feldman, M., Gehan, M. A., Wilson, M. S., Shyu, C., Bryant, D. W., et al. (2015a). A versatile phenotyping system and analytics platform reveals diverse temporal responses to water availability in *Setaria*. *Mol. Plant* 8, 1520–1535. doi: 10.1016/j.molp.2015.06.005
- Fahlgren, N., Gehan, M. A., and Baxter, I. (2015b). Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curr. Opin. Plant Biol.* 24, 93–99. doi: 10.1016/j.pbi.2015.02.006
- FAO (2016). “Poplars and other fast-growing trees - Renewable resources for future green economies - Synthesis of country progress reports,” in *25th Session of the International Poplar Commission, 13–16 September 2016*, Berlin.
- Farooq, M., Wahid, A., Kobayashi, N., Fujita, D., and Basra, S. M. A. (2009). Plant drought stress: effects, mechanisms and management. *Agron. Sustain. Dev.* 29, 185–212. doi: 10.1051/agro:2008021
- Fischer, R., and Maurer, R. (1978). Drought resistance in spring wheat cultivars. I. Grain yield responses. *Crop Pasture Sci.* 29, 897–912. doi: 10.1071/AR9780897
- Fuchs, M. (1990). Infrared measurement of canopy temperature and detection of plant water stress. *Theor. Appl. Climatol.* 42, 253–261. doi: 10.1007/BF00865986
- Gago, J., Douthe, C., Coopman, R. E., Gallego, P. P., Ribas-Carbo, M., Flexas, J., et al. (2015). UAVs challenge to assess water stress for sustainable agriculture. *Agric. Water Manage.* 153, 9–19. doi: 10.1016/j.agwat.2015.01.020
- Ghanem, M. E., Marrou, H., and Sinclair, T. R. (2015). Physiological phenotyping of plants for crop improvement. *Trends Plant Sci.* 20, 139–144. doi: 10.1016/j.tplants.2014.11.006
- Goggin, F. L., Lorence, A., and Topp, C. N. (2015). Applying high-throughput phenotyping to plant-insect interactions: picturing more resistant crops. *Curr. Opin. Insect Sci.* 9, 69–76. doi: 10.1016/j.cois.2015.03.002
- Gómez-Candón, D., Virlet, N., Labbé, S., Jolivot, A., and Regnard, J.-L. (2016). Field phenotyping of water stress at tree scale by UAV-sensed imagery: new insights for thermal acquisition and calibration. *Precis. Agric.* 17, 786–800. doi: 10.1007/s11119-016-9449-6
- Gonzalez-Dugo, V., Zarco-Tejada, P., Berni, J. A. J., Suárez, L., Goldhamer, D., and Fereres, E. (2012). Almond tree canopy temperature reveals intra-crown variability that is water stress-dependent. *Agric. For. Meteorol.* 15, 156–165. doi: 10.1016/j.agrformet.2011.11.004
- Griffiths, A. J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C., and Gelbart, W. M. (eds.) (2000). “Quantifying heritability,” in *An Introduction to Genetic Analysis*, 7th Edn (New York, NY: W. H. Freeman).
- Guidi, W., Piccioni, E., and Bonari, E. (2008). Evapotranspiration and crop coefficient of poplar and willow short-rotation coppice used as vegetation filter. *Bioresour. Technol.* 99, 4832–4840. doi: 10.1016/j.biortech.2007.09.055
- Harfouche, A., Meilan, R., and Altman, A. (2014). Molecular and physiological responses to abiotic stress in forest trees and their relevance to tree improvement. *Tree Physiol.* 34, 1181–1198. doi: 10.1093/treephys/tpu012
- Harfouche, A., Meilan, R., Kirst, M., Morgante, M., Boerjan, W., Sabatti, M., et al. (2012). Accelerating the domestication of forest trees in a changing world. *Trends Plant Sci.* 17, 64–72. doi: 10.1016/j.tplants.2011.11.005
- Herr, J. R., and Carlson, J. E. (2013). “Traditional breeding, Genomics-assisted breeding, and biotechnology modification of forest trees and short rotation woody crops,” in *Wood-Based Energy in the Northern Forests*, eds M. Jacobson and D. Ciolkosz (New York, NY: Springer Science + Business Media), 69–89.
- Hopkins, A. A., Saha, M. C., and Wang, Z. Y. (2009). “Breeding, genetics, and cultivars,” in *Tall Fescue for the Twenty-first Century. Agronomy Monographs*, eds H. A. Fribourg, D. B. Hannaway, and C. P. West (Madison, WI: American Society of Agronomy), 339–366.
- Huang, X., Xiao, X., Zhang, S., Korpelainen, H., and Li, C. (2009). Leaf morphological and physiological responses to drought and shade in two *Populus cathayana* populations. *Biol. Plant.* 53, 588–592. doi: 10.1007/s10535-009-0107-y
- 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,” in *Core Writing Team*, eds R. K. Pachauri and L. A. Meyer (Geneva: IPCC), 151.
- Izawa, T. (2015). Deciphering and prediction of plant dynamics under field conditions. *Curr. Opin. Plant Biol.* 24, 87–92. doi: 10.1016/j.pbi.2015.02.003
- Jackson, R. D., Idso, S. B., Reginato, R. J., and Pinter, P. J. Jr. (1981). Canopy temperature as a crop water stress indicator. *Water Resour. Res.* 17, 1133–1138. doi: 10.1029/WR017i004p01133
- Jähne, B. (2005). *Digital Image Processing*. Berlin: Springer.
- Jiménez-Bello, M. A., Ballester, C., Caste, J. R., and Intrigliolo, D. S. (2011). Development and validation of an automatic thermal imaging process for assessing plant water status. *Agric. Water Manage.* 98, 1497–1504. doi: 10.1016/j.agwat.2011.05.002
- Jones, H. G. (1992). *Plants and Microclimate*, 2nd Edn. Cambridge: Cambridge University Press, 423.
- Jones, H. G. (1999). Use of thermography for quantitative studies of spatial and temporal variation of stomatal conductance over leaf surfaces. *Plant Cell Environ.* 22, 1043–1055. doi: 10.1046/j.1365-3040.1999.00468.x
- Jones, H. G., Serraj, R., Loveys, B. R., Xiong, L., Wheaton, A., and Price, A. H. (2009). Thermal infrared imaging of crop canopies for the remote diagnosis and quantification of plant responses to water stress in the field. *Funct. Plant Biol.* 36, 978–989. doi: 10.1071/FP09123
- Jones, H. G., and Sirault, X. R. R. (2014). Scaling of thermal images at different spatial resolution: the mixed pixel problem. *Agronomy* 4, 380–396. doi: 10.3390/agronomy4030380
- Kaluthota, S., Pearce, D. W., Evans, L. M., Letts, M. G., Whitham, T. G., and Rood, S. B. (2015). Higher photosynthetic capacity from higher latitude: foliar characteristics and gas exchange of southern, central and northern populations of *Populus angustifolia*. *Tree Physiol.* 35, 936–948. doi: 10.1093/treephys/tpv069

- Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorol. Z.* 15, 259–263. doi: 10.1127/0941-2948/2006/0130
- Levitt, J. (1972). *Responses of Plants to Environmental Stresses*. New York, NY: Academic Press, 698.
- Lindner, M., Fitzgerald, J. B., Zimmermann, N. E., Reyer, C., Delzon, S., van der Maaten, E., et al. (2014). Climate change and European forests: what do we know, what are the uncertainties, and what are the implications for forest management? *J. Environ. Manage.* 146, 69–83. doi: 10.1016/j.jenvman.2014.07.030
- Maes, W. H., and Steppe, K. (2012). Estimating evapotranspiration and drought stress with ground-based thermal remote sensing in agriculture: a review. *J. Exp. Bot.* 67, 4671–4712. doi: 10.1093/jxb/ers165
- Mahan, J. R., and Upchurch, D. R. (1988). Maintenance of constant leaf temperature by plant. 1. Hypothesis limited homeothermy. *Environ. Exp. Bot.* 28, 351–357. doi: 10.1016/0098-8472(88)90059-7
- Marron, N., Delay, D., Petit, J. M., Dreyer, E., Kahlem, G., Delmotte, F. M., et al. (2002). Physiological traits of two *Populus × euramericana* clones, Luisa avanzo and dorskamp, during water stress and re-watering cycle. *Tree Physiol.* 22, 849–858. doi: 10.1093/treephys/22.12.849
- McKown, A. D., Guy, R. D., Klápková, J., Geraldes, A., Friedmann, M., Cronk, Q. C. B., et al. (2013). Geographical and environmental gradients shape phenotypic trait variation and genetic structure in *Populus trichocarpa*. *New Phytol.* 201, 1263–1276. doi: 10.1111/nph.12601
- Monclús, R., Dreyer, E., Villar, M., Delmotte, F. M., Delay, D., Petit, J. M., et al. (2006). Impact of drought on productivity and water use efficiency in 29 genotypes of *Populus deltoides* × *Populus nigra*. *New Phytol.* 169, 765–777. doi: 10.1111/j.1469-8137.2005.01630.x
- Moshelion, M., Halperin, O., Wallach, R., Oren, R., and Way, D. A. (2014). Role of aquaporins in determining transpiration and photosynthesis in water-stressed plants: crop water-use efficiency, growth and yield. *Plant Cell Environ.* 38, 1785–1793. doi: 10.1111/pce.12410
- Neale, D. B., and Kremer, A. (2011). Forest tree genomics: growing resources and applications. *Nat. Rev. Genet.* 12, 111–122. doi: 10.1038/nrg2931
- Ngao, J., Boris, A., and Saudreau, M. (2017). Intra-crown spatial variability of leaf temperature and stomatal conductance enhanced by drought in apple tree as assessed by the RATP model. *Agric. For. Meteorol.* 23, 340–354. doi: 10.1016/j.agrformet.2017.02.036
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* 9, 62–66. doi: 10.1109/TSMC.1979.4310076
- Passioura, J. B. (2012). Phenotyping for drought tolerance in grain crops: when is it useful to breeders? *Funct. Plant Biol.* 39, 851–859. doi: 10.1071/FP12079
- Paulson, M., Bardos, P., Harmsen, J., Wilczek, J., Barton, M., and Edwards, D. (2003). The practical use of short rotation coppice in land restoration. *Land Contam. Reclam.* 11, 323–338. doi: 10.2462/09670513.624
- Pintó-Marijuan, M., and Munné-Bosch, S. (2014). Photo-oxidative stress markers as a measure of abiotic stress-induced leaf senescence: advantages and limitations. *J. Exp. Bot.* 65, 3845–3857. doi: 10.1093/jxb/eru086
- Poland, J. (2015). Breeding-assisted genomics. *Curr. Opin. Plant Biol.* 24, 119–124. doi: 10.1016/j.pbi.2015.02.009
- Rae, A. M., Street, N. R., Robinson, K. M., Harris, N., and Taylor, G. (2009). Five QTL hotspots for yield in short rotation coppice bioenergy poplar: the poplar biomass loci. *BMC Plant Biol.* 9:23. doi: 10.1186/1471-2229-9-23
- Rebetzke, G. J., Condon, A. G., Richards, R. A., and Read, J. J. (2001). Phenotypic variation and sampling for leaf conductance in wheat (*Triticum aestivum* L.) breeding populations. *Euphytica* 121, 335–341. doi: 10.1023/A:1012035720423
- Rebetzke, G. J., Rattey, A. R., Farquhar, G. D., Richards, R. A., and Condon, A. G. (2013). Genomic regions for canopy temperature and their genetic association with stomatal conductance and grain yield in wheat. *Funct. Plant Biol.* 40, 14–33. doi: 10.1071/FP12184
- Regier, N., Streb, S., Cocozza, C., Schaub, M., Cherubini, P., Zeeman, S. C., et al. (2009). Drought tolerance of two black poplar (*Populus nigra* L.) clones: contribution of carbohydrates and oxidative stress defence. *Plant Cell Environ.* 32, 1724–1736. doi: 10.1111/j.1365-3040.2009.02030.x
- Richards, L. R., and Weaver, L. A. (1944). Moisture retention by some irrigated soils as related to soil moisture tension. *J. Agric. Res.* 69, 215–235.
- Richardson, J., Isebrands, J. G., and Ball, J. B. (2014). “Ecology and physiology of poplars and willows,” in *Poplars and Willows: Trees for Society and the Environment*, eds J. G. Isebrands and J. Richardson (Rome: The Food and Agriculture Organization of United Nations and CABI), 92–123. doi: 10.1079/9781780641089.0092
- Rizhsky, L., Liang, H., and Mittler, R. (2002). The combined effect of drought stress and heat shock on gene expression in tobacco. *Plant Physiol.* 130, 1143–1151. doi: 10.1104/pp.006858
- Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S., and Mittler, R. (2004). When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. *Plant Physiol.* 134, 1683–1696. doi: 10.1104/pp.103.033431
- Rockwood, D. L., Naidu, C. V., Carter, D. R., Rahmani, M., Spriggs, T. A., Lin, C., et al. (2004). Short-rotation woody crops and phytoremediation: opportunities for agroforestry? *Agroforest. Syst.* 61, 51–63. doi: 10.1023/B:AGFO.0000028989.72186
- Rohde, A., Bastien, C., and Boerjan, W. (2011). Temperature signals contribute to the timing of photoperiodic growth cessation and bud set in poplar. *Tree Physiol.* 31, 472–482. doi: 10.1093/treephys/tp038
- Sabatti, M., Fabbrini, F., Harfouche, A., Beritognolo, I., Mareschi, L., Carlini, M., et al. (2014). Evaluation of biomass production potential and heating value of hybrid poplar genotypes in a short-rotation culture in Italy. *Ind. Crop Prod.* 61, 62–73. doi: 10.1016/j.indcrop.2014.06.043
- Sade, N., Gebremedhin, A., and Moshelion, M. (2012). Risk-taking plants. Anisohydric behavior as a stress-resistance trait. *Plant Signal. Behav.* 7, 767–770. doi: 10.4161/psb.20505
- Sade, N., and Moshelion, M. (2014). The dynamic isohydric-anisohydric behavior of plants upon fruit development: taking a risk for the next generation. *Tree Physiol.* 34, 1199–1202. doi: 10.1093/treephys/tpu070
- Sánchez, E., Scordia, D., Lino, G., Aries, A., Cosentino, S. L., and Nogués, S. (2015). Salinity and water stress effects on biomass production in different *Arundo donax* L. clones. *Bioenerg. Res.* 8, 1461–1479. doi: 10.1007/s12155-015-9652-8
- Sannigrahi, P., Ragauskas, A. J., and Tuskan, G. (2010). Poplar as a feedstock for biofuels: a review of compositional characteristics. *Biofuels Bioprod. Bioref.* 4, 209–226. doi: 10.1002/bbb.206
- Sehgal, D., Skot, L., Singh, R., Srivastava, R. K., Das, S. P., Taunk, J., et al. (2015). Exploring potential of pearl millet germplasm association panel for association mapping of drought tolerance traits. *PLOS ONE* 10:e0122165. doi: 10.1371/journal.pone.0122165
- Sepulcre-Canto, G., Zarco-Tejada, P., Jimenez-Munoz, J., Sobrino, J., de Miguel, E., and Villalobos, F. J. (2006). Detection of water stress in an olive orchard with thermal remote sensing imagery. *Agric. For. Meteorol.* 136, 31–44. doi: 10.1016/j.agrformet.2006.01.008
- Sepulcre-Canto, G., Zarco-Tejada, P., Jimenez-Munoz, J., Sobrino, J., Soriano, M., Fereres, E., et al. (2007). Monitoring yield and fruit quality parameters in open-canopy tree crops under water stress. Implications for ASTER. *Remote Sens. Environ.* 107, 455–470. doi: 10.1016/j.rse.2006.09.014
- Shi, Y., Thomasson, J. A., Murray, S. C., Pugh, N. A., Rooney, W. L., Shahian, S., et al. (2016). Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PLOS ONE* 11:e0159781. doi: 10.1371/journal.pone.0159781
- Sims, R. E. H., and Venturi, P. (2004). All-year round harvesting of short rotation coppice *Eucalyptus* compared with the delivered costs of biomass from more conventional short season, harvesting systems. *Biomass Bioenerg.* 26, 27–37. doi: 10.1016/S0961-9534(03)00081-3
- Stanton, B. J., Neale, D. B., and Li, S. (2010). “Populus breeding: from the classical to the genomic approach,” in *Genetics and Genomics of Populus (Plant Genetics and Genomics: Crop and Models)*, eds S. Jansson, R. Bhalerao, and A. Groover (New York, NY: Springer Science + Business Media), 309–348. doi: 10.1007/978-1-4419-1541-2\_14
- Stanton, B. J., Serapiglia, M. J., and Smart, L. B. (2013). “The domestication and conservation of *Populus* and *Salix* genetic resources” in *Poplars and Willows: Trees for Society and the Environment*, eds J. G. Isebrands and J. Richardson (Rome: The Food and Agriculture Organization of United Nations and CABI), 124–199.
- Stettler, R. F., Fenn, R. C., Heilman, P. E., and Stanton, B. J. (1988). *Populus trichocarpa* × *Populus deltoides* hybrids for short rotation culture: variation patterns and 4-year field performance. *Can. J. For. Res.* 18, 745–753. doi: 10.1139/x88-114

- Street, N. R., Skogstrom, O., Sjodin, A., Tucker, J., Rodriguez-Acosta, M., Nilsson, P., et al. (2006). The genetics and genomics of the drought response in *Populus*. *Plant J.* 48, 321–341. doi: 10.1111/j.1365-313X.2006.02864.x
- Struthers, R., Ivanova, A., Tits, L., Swennen, R., and Coppin, P. (2015). Thermal infrared imaging of the temporal variability in stomatal conductance for fruit trees. *Int. J. Appl. Earth Obs.* 39, 9–17. doi: 10.1016/j.jag.2015.02.006
- Tardieu, F., and Simonneau, T. (1998). Variability among species of stomatal control under fluctuating soil water status and evaporative demand: modelling isohydric and anisohydric behaviours. *J. Exp. Bot.* 49, 419–432. doi: 10.1093/jxb/49.Special\_Issue.419
- Tardieu, F., and Tuberrosa, R. (2010). Dissection and modelling of abiotic stress tolerance in plants. *Curr. Opin. Plant Biol.* 13, 206–212. doi: 10.1016/j.pbi.2009.12.012
- Testi, L., Goldhamer, D. A., Iniesta, F., and Salinas, M. (2008). Crop water stress index is a sensitive water stress indicator in pistachio trees. *Irrig. Sci.* 26, 395–405. doi: 10.1007/s00271-008-0104-5
- Torresan, C., Berton, A., Carotenuto, F., Di Gennaro, S. F., Gioli, B., Matese, A., et al. (2016). Forestry applications of UAVs in Europe: a review. *Int. J. Remote Sens.* 38, 2427–2447. doi: 10.1080/01431161.2016.1252477
- van der Schoot, J., Pospišková, M., Vosman, B., and Smulders, M. J. M. (2000). Development and characterization of microsatellite markers in black poplar (*Populus nigra* L.). *Theor. Appl. Genet.* 101, 317–322. doi: 10.1007/s001220051485
- Viger, M., Smith, H. K., Cohen, D., Dewoody, J., Trewin, H., Steenackers, M., et al. (2016). Adaptive mechanisms and genomic plasticity for drought tolerance identified in European black poplar (*Populus nigra* L.). *Tree Physiol.* 36, 909–928. doi: 10.1093/treephys/tpw017
- Virlet, N., Lebourgues, V., Martinez, S., Costes, E., Labbé, S., and Regnard, J. (2014). Stress indicators based on airborne thermal imagery for field phenotyping a heterogeneous tree population for response to water constraints. *J. Exp. Bot.* 65, 5429–5442. doi: 10.1093/jxb/eru309
- Way, D. A., and Oren, R. (2010). Differential responses to changes in growth temperature between trees from different functional groups and biomes: a review and synthesis of data. *Tree Physiol.* 30, 669–688. doi: 10.1093/treephys/tpq015
- White, J. F., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., et al. (2012). Field-based phenomics for plant genetics research. *Field Crop Res.* 133, 101–112. doi: 10.1016/j.fcr.2012.04.003
- White, T. L., Adams, W. T., and Neale, D. B. (eds). (2007). “Quantitative genetics – polygenic traits, heritabilities and genetic correlations,” in *Forest Genetics* (Wallingford: CABI), 113–148. doi: 10.1079/9781845932855.0113
- Wu, R., and Stettler, R. F. (1994). Quantitative genetics of growth and development in *Populus*. I. A three-generation comparison of tree architecture during the first 2 years of growth. *Theor. Appl. Genet.* 89, 1046–1054. doi: 10.1007/BF00224537
- Zarco-Tejada, P. J., González-Dugo, V., and Berni, J. A. J. (2012). Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera. *Remote Sens. Environ.* 117, 322–337. doi: 10.1016/j.rse.2011.10.007
- Zarco-Tejada, P. J., Suárez, L., and González-Dugo, V. (2013). Spatial resolution effects on chlorophyll fluorescence retrieval in a heterogeneous canopy using hyperspectral imagery and radiative transfer simulation. *IEEE Geosci. Remote Sens. Lett.* 10, 937–941. doi: 10.1109/LGRS.2013.2252877

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

Copyright © 2017 Ludovisi, Tauro, Salvati, Khoury, Scarascia Mugnozza and Harfouche. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Non-destructive Phenotyping of Lettuce Plants in Early Stages of Development with Optical Sensors

Ivan Simko<sup>1\*</sup>, Ryan J. Hayes<sup>1</sup> and Robert T. Furbank<sup>2,3</sup>

<sup>1</sup> U.S. Department of Agriculture, Agricultural Research Service, Crop Improvement and Protection Research Unit, Salinas, CA, USA, <sup>2</sup> High Resolution Plant Phenomics Centre, Australian Plant Phenomics Facility, Commonwealth Scientific and Industrial Research Organisation Agriculture and Food, Canberra, ACT, Australia, <sup>3</sup> Australian Research Council Centre of Excellence for Translational Photosynthesis, Plant Science Division, Research School of Biology, Australian National University, Acton, ACT, Australia

## OPEN ACCESS

### Edited by:

Marcos Egea-Cortines,  
Universidad Politécnica de Cartagena,  
Spain

### Reviewed by:

Francisco M. Padilla,  
University of Almería, Spain  
Maria Isabel Gil,  
Consejo Superior de Investigaciones  
Científicas, Spain

### \*Correspondence:

Ivan Simko  
ivan.simko@ars.usda.gov

### Specialty section:

This article was submitted to  
Technical Advances in Plant Science,  
a section of the journal  
*Frontiers in Plant Science*

**Received:** 10 November 2016

**Accepted:** 14 December 2016

**Published:** 27 December 2016

### Citation:

Simko I, Hayes RJ and Furbank RT (2016) Non-destructive Phenotyping of Lettuce Plants in Early Stages of Development with Optical Sensors. *Front. Plant Sci.* 7:1985.  
doi: 10.3389/fpls.2016.01985

Rapid development of plants is important for the production of ‘baby-leaf’ lettuce that is harvested when plants reach the four- to eight-leaf stage of growth. However, environmental factors, such as high or low temperature, or elevated concentrations of salt, inhibit lettuce growth. Therefore, non-destructive evaluations of plants can provide valuable information to breeders and growers. The objective of the present study was to test the feasibility of using non-destructive phenotyping with optical sensors for the evaluations of lettuce plants in early stages of development. We performed the series of experiments to determine if hyperspectral imaging and chlorophyll fluorescence imaging can determine phenotypic changes manifested on lettuce plants subjected to the extreme temperature and salinity stress treatments. Our results indicate that top view optical sensors alone can accurately determine plant size to approximately 7 g fresh weight. Hyperspectral imaging analysis was able to detect changes in the total chlorophyll (RCC) and anthocyanin (RAC) content, while chlorophyll fluorescence imaging revealed photoinhibition and reduction of plant growth caused by the extreme growing temperatures (3 and 39°C) and salinity (100 mM NaCl). Though no significant correlation was found between  $F_v/F_m$  and decrease in plant growth due to stress when comparisons were made across multiple accessions, our results indicate that lettuce plants have a high adaptability to both low (3°C) and high (39°C) temperatures, with no permanent damage to photosynthetic apparatus and fast recovery of plants after moving them to the optimal (21°C) temperature. We have also detected a strong relationship between visual rating of the green- and red-leaf color intensity and RCC and RAC, respectively. Differences in RAC among accessions suggest that the selection for intense red color may be easier to perform at somewhat lower than the optimal temperature. This study serves as a proof of concept that optical sensors can be successfully used as tools for breeders when evaluating young lettuce plants. Moreover, we were able to identify the locus for light green leaf color (*qLG4*), and position this locus on the molecular linkage map of lettuce, which shows that these techniques have sufficient resolution to be used in a genetic context in lettuce.

**Keywords:** temperature stress, elevated salinity, relative chlorophyll content, relative anthocyanin content, photosynthesis, relative growth rate, visual rating of color intensity

## INTRODUCTION

Lettuce is an economically valuable, leafy vegetable that is harvested when the plant 'head' reaches maturity, or in the early stages of development when plant leaves are cut for 'baby-leaf' or 'spring-mix' bagged salad (Simko et al., 2014). Plants intended for baby-leaf production are grown in the extremely high densities (7.4 million seeds per hectare) until they reach four- to eight-leaf stage and then their leaves are cut. Because baby-leaf lettuces may be cut repeatedly for multiple harvests (Grahn et al., 2015), they need to regrow rapidly, and consistently produce leaves with the shape, color, texture, and taste attractive to consumers. Environmental conditions, such as temperature or soil salinity, however, can severely affect plant growth. Lettuce is rather sensitive to the elevated levels of salt ( $\text{NaCl}/\text{CaCl}_2$ ) in soil or water. When the decline in yield was compared across several vegetable species, lettuce was classified into the 'most sensitive' group (Shannon and Grieve, 1999). The increased sensitivity of plants is accentuated during the early stages of development when elevated concentrations of salt inhibit lettuce growth (Shannon et al., 1983; Xu and Mou, 2015). Similarly, the photosynthesis (Seginer et al., 1991) and the growth (Thompson et al., 1998) of lettuce plants are significantly reduced at both suboptimal and supraoptimal temperatures.

Simple optical sensors, such as photographic cameras, have been used for non-destructive plant phenotyping for a long time (Taubenhaus et al., 1929), but the recent technological progress in the development of digital and spectral cameras together with strides in analytical software made these tools more appealing to plant scientists. Phenotyping with optical sensors can accurately be performed on individual plants, plant organs, or a group of plants in laboratory, growth-chamber, greenhouse, or field conditions (Fiorani and Schurr, 2013; Araus and Cairns, 2014). These sensors can be used individually and operated manually or integrated into a high-throughput, fully automated, imaging systems (White et al., 2012). The objective of the present study was to test feasibility of using non-destructive phenotyping with optical sensors for the evaluations of lettuce plants in early stages of development. We performed the series of experiments to determine if hyperspectral imaging and chlorophyll fluorescence imaging can determine phenotypic changes manifested on lettuce plants that were subjected to the extreme temperature and salinity stress treatments. We also analyzed the relationship between data obtained from visual observations of plant color (a consumer's perspective) and non-destructive phenotyping with optical sensors. Hyperspectral imaging devices used in our study collect data on electromagnetic radiation reflected by a plant for each pixel of the image. These spectral data can be combined to identify specific characteristics, such as internal structure, chemical composition, physiological status, or a damage that may not be evident in the visible spectrum (Simko et al., 2015a, 2017). In difference from the devices that use hyperspectral imaging, instruments based on chlorophyll fluorescence detect only the light that is re-emitted by chlorophyll after the light of the defined wavelength is directed to a plant. Chlorophyll fluorescence analysis thus provides information about the efficiency of photosynthesis (Maxwell and Johnson,

2000) and can detect photosynthetic response of a tissue to environmental stress (Murata et al., 2007). Both the hyperspectral imaging and the chlorophyll fluorescence imaging are routinely used to analyze plant performance (Furbank and Tester, 2011), but to our knowledge they have previously not been used to analyze the early stages of lettuce development described in the present work. Automated, high-throughput phenotyping with optical sensors is well suited for fast evaluations of a large number of plants, such as mapping populations. In this study, we present the application of hyperspectral imaging for rapid, non-destructive determination of chlorophyll content in plants, and use of these data for mapping of the underlying locus.

## MATERIALS AND METHODS

### Plant Material

The following lettuce accessions were used in one or more experiments: Annapolis (Ann), Balady Banha (BB), Bibb (Bib), Climax (Cli), Corsair (Cor), Eruption (Eru), Flashy Troutback (FT), Grand Rapids (GR), Green Forest (GF), Green Towers (GT), Ice Cube (IC), Iceberg (Ice), La Brillante (LB), Little Gem (LG), Lolla Rossa (LL), Merlot (Mer), Nansen (Nan), Pavane (Pav), Prizehead (Pri), Red Fox (RF), Red Leaf (RL), Redina (Red), RH08-0464 (RH), Salinas (Sal), SM09A (9A), SM09B (9B), SM13-L6 (L6), SM13-R1 (R1), Tom Thump (ToT), Triple Threat (TT), US96UC23 (US), Valmaine (Val), Winter Marvel (WM), and 56 F<sub>8</sub> recombinant inbred lines (RILs) randomly selected from the Salinas 88 (S88) × La Brillante population (Hayes et al., 2014; Simko et al., 2015b). Five plants per accession per treatment were used in all experiments. These plants were selected from a larger group of plants to minimize differences in plant size at the beginning of the experiment. In experiment 1 (described below), plants of a different size were selected purposely.

### Growing Conditions

Lettuce seeds were sown into square pots (68 mm × 68 mm, 95 mm depth) containing 1:1 mix of soil and sand. Seedlings were grown in a controlled environment growth chamber with 16 h photoperiod, 400  $\mu\text{mol m}^{-2} \text{s}^{-1}$  photosynthetic photon-flux density (PPFD), and constant temperature of 21°C (these conditions are called optimal throughout the text, OPT). Temperature stress was applied by decreasing temperature to 3°C (COLD), or increasing it to 39°C (HOT) for the duration described at individual experiments. Watering of plants at all treatments was performed as needed to keep approximately 70% substrate water content (SWC). SWC was determined by weighing the soil before and after drying (Granier et al., 2006). Salt stress (SALT) was imposed by adding NaCl to irrigation water to obtain salinity concentration of 100 mM.

### Tests on Seedlings

In experiments where cotyledons of lettuce seedlings were evaluated, the seeds were either germinated and grown on a wet filter paper in Petri dishes, or in plastic boxes used for holding

pipette tips. The boxes contained the same media as was used for growing plants in pots and the seedlings were kept in a grid using plastic that holds pipette tips (Badger et al., 2009). Seedlings were cultivated in OPT conditions until temperature treatments were applied.

## Visual Observations

Visual estimates of plant color intensity were performed on adaxial leaf surfaces. Green color intensity of lettuce leaves was rated light green, green, and dark green, while intensity of red color was rated as no red, light red, and red. These visual estimates of color intensity were compared to relative measurements of anthocyanin and chlorophyll content obtained from hyperspectral imaging.

## Hyperspectral and Chlorophyll Fluorescence Imaging

Hyperspectral imaging was performed with an A-Series VNIR Micro-Hyperspec Sensor (Headwall Photonics, Fitchburg, MA, USA) with the spectral range from 380 to 1,012 nm. The sensor was attached to a metal frame at the distance of 70 cm from scanned samples together with a broad-spectrum halogen lamp (ProLamp, Analytical Spectral Devices, Boulder, CO, USA). Reflectance calibration was performed using Spectralon SRT-MS-100 reflectance standard (Labsphere, North Sutton, NH, USA) that was placed next to the samples at each scan. Images were captured using XCAP v.3.7 software (EPIX, Buffalo Grove, IL, USA) and analyzed with ImageJ 1.49k software (National Institutes of Health, Bethesda, MD, USA).

Measurements of chlorophyll fluorescence were done with PlantScreen (in tray scanning format) or FluorCam 800MF (both from Photon Systems Instruments, Brno, Czech Republic). The protocol parameters were: camera distance 20 cm, TS 20 ms, shutter 1, sensitivity 52%, super 87.2%,  $F_0$  duration 2 s,  $F_0$  period 200 ms, and pulse duration 800 ms. All measurements were performed after 15 min of dark adaptations. Analyses of fluorescence data were carried out using FluorCam 7 (Photon Systems Instruments, Brno, Czech Republic) and ImageJ 1.49k software. Both hyperspectral imaging and chlorophyll fluorescence imaging scans of plants were taken from a top view only.

## Relative Content of Chlorophyll and Anthocyanin

Relative chlorophyll (RCC) and anthocyanin (RAC) content per leaf area were estimated by previously developed indices. These indices are based on measurements of tissue reflectance ( $R$ ) at the specific wavelengths (shown as subscript) obtained from hyperspectral imaging:  $\text{RCC} = (R_{728} - R_{720}) / (R_{728} + R_{720} - 2 \times R_{434})$  (Xue and Yang, 2009) and  $\text{RAC} = [R_{800} \times (1/R_{550} - 1/R_{700})]$  (Merzlyak et al., 2003). The indices showed very strong linear relationship ( $R^2 > 0.9$ ) with a relative content of total chlorophyll (chlorophyll  $a + b$ ) and anthocyanin, respectively (Merzlyak et al., 2003; Xue and Yang, 2009) when tested on a diverse set of samples.

## Extent of Photoinhibition

The extent of photoinhibition due to stress caused by abiotic factors was estimated through the analysis of maximum quantum efficiency of photosystem II (PSII) photochemistry. This parameter was calculated as  $F_v/F_m$ , where  $F_m$  is the maximum chlorophyll  $a$  fluorescence yield in the dark-adapted state, and  $F_v$  is the maximum variable fluorescence in the dark-adapted state (calculated as  $F_m - F_o$ ), and  $F_o$  is the minimum chlorophyll  $a$  fluorescence yield in the dark-adapted state (Maxwell and Johnson, 2000).

## Total Projected Leaf Area, Relative Growth Rate, and Fresh Weight

Total projected leaf area ( $A_{PT}$ ; Munns et al., 2010) in  $\text{mm}^2$  was determined from the images of chlorophyll fluorescence emission at the  $F_m$  level (Barbagallo et al., 2003). To compare  $A_{PT}$  values to the plant biomass that was produced above-ground, plants were cut at the soil level and their fresh weight (FW) was determined immediately. Relative growth rate (RGR) was calculated from  $A_{PT}$  as  $\text{RGR} = (\overline{\ln A_{PT2}} - \overline{\ln A_{PT1}}) / (t_2 - t_1)$ , where  $\overline{\ln A_{PT1}}$  and  $\overline{\ln A_{PT2}}$  are the means of natural logarithm transformed  $A_{PT}$  at the times  $t_1$  and  $t_2$ , respectively (Hoffmann and Poorter, 2002). The values of RGR were multiplied by 100 to obtain units in  $\text{mm}^2 \text{ per cm}^2 \text{ per day}$  ( $\text{mm}^2 \times \text{cm}^{-2} \times \text{d}^{-1}$ ).

## Description of Individual Experiments

### Experiment 1: Relationship between $A_{PT}$ and FW, and between Visual Observations of Leaf Color and RCC and RAC

Accessions: Ann, Bib, Cli, Eru, FT, GR, GF, GT, Ice, LB, LG, LL, Mer, Pav, RF, RL, RH, Sal, TT, and Val.

Growing conditions: 15 days in OPT.

Evaluations:  $A_{PT}$ , FW, visual observation of red and green leaf color, RCC and RAC; in addition, disks (1 cm diameter) were cut from selected leaves at the end of the experiment to make comparisons of RCC, RAC, and  $F_v/F_m$  at adaxial and abaxial surfaces of the leaves.

Note: Plants of highly different sizes were selected for the analysis of relationship between  $A_{PT}$  and FW.

### Experiment 2: Change in Growth and Photosynthesis in Suboptimal Temperature

Accessions: Eru, GR, GT, LB, LG, Pav, RL, Sal, and TT.

Growing conditions: 10 days in OPT, then 10 days in either OPT or COLD.

Evaluations:  $A_{PT}$  and  $F_v/F_m$ .

### Experiment 3: Change in Growth, Photosynthesis, RAC, and RCC in Suboptimal and Supraoptimal Temperatures

Accessions: 110, 112 (two RILs from the S88 × LB population), 9A, 9B, Eru, GR, GT, LB, LG, Pav, R1, RL, and TT. Five of the accessions (Eru, LB, LG, RL, and TT) were selected for hyperspectral analyses to determine changes in RCC and RAC after the temperature treatment.

Growing conditions: 10 days OPT, then 8 days in either OPT, COLD, or HOT, then 6 days in OPT (recovery).

Evaluations:  $A_{PT}$ ,  $F_v/F_m$ , RCC, and RAC.

#### **Experiment 4: Change in Growth and Photosynthesis under Increased Salinity**

Accessions: 9A, GT, LB, LG, Pav, RL, and R1.

Growing conditions: 10 days in OPT, then 8 days in either OPT or SALT.

Evaluations:  $A_{PT}$  and  $F_v/F_m$ .

#### **Experiment 5: Change in Photosynthesis in Suboptimal Temperature – Seedlings in Petri Dish**

Accessions: GR, LB, Sal, and TT.

Growing conditions: 4 days in OPT, then 1 day in either OPT or COLD.

Evaluations:  $F_v/F_m$ .

#### **Experiment 6: Change in Photosynthesis in Suboptimal Temperature – Seedlings in Plastic Box**

Accessions: 9A, 9B, Ann, BB, Bib, Cli, Cor, Eru, FT, GF, GR, GT, IC, Ice, L6, LB, LL, Mer, Nan, Pav, Pri, Red, RF, RH, RL, R1, Sal, ToT, TT, US, Val, and WM.

Growing conditions: 4 days in OPT, then 2 days in COLD.

Evaluations:  $F_v/F_m$ .

#### **Experiment 7: Detection of RCC and Mapping Locus – Seedlings in Plastic Box**

Accessions: 56 RILs from the S88 × LB population.

Growing conditions: 4 days in OPT.

Evaluations: RCC.

Note: Visual observations of leaf color were previously performed on the same RILs grown under field conditions (Hayes et al., 2014; Simko et al., 2015b). The evaluations were performed on adult plants at harvest maturity using the same scale (light green, green, and dark green) as in the experiment 1.

### **Statistical Analyses**

Differences between means of two groups were tested with *t*-test (or paired *t*-test), and among multiple groups with one-way analysis of variance (ANOVA). If ANOVA results were significant, the Tukey-Kramer HSD test was applied to compare all pairs of groups. The *F*-test was used to test if two variances are equal. The Pearson correlation coefficient was utilized to measure a linear correlation between two variables. All statistical analyses were calculated with JMP v. 11.1.1 (SAS Institute, Cary, NC, USA).

### **Quantitative Trail Locus (QTL) Mapping**

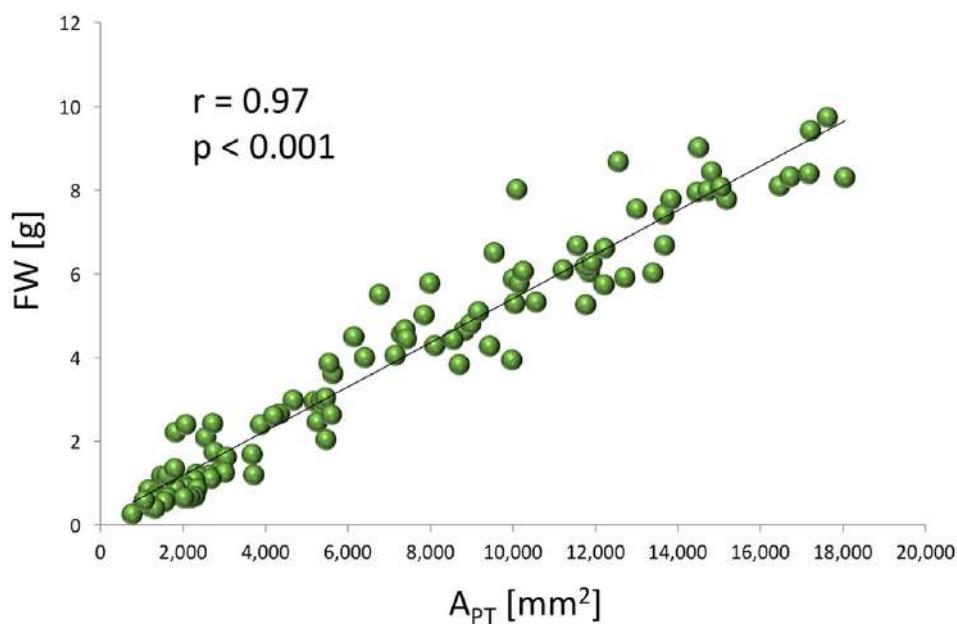
Quantitative Trail Locus for RCC and green leaf color visually observed on RILs were mapped with QGene v. 4.3.9 software (Joehanes and Nelson, 2008) using simple interval mapping. The significance threshold for QTL scores was determined empirically through permutations with 1,000 iterations (Churchill and Doerge, 1994). Molecular linkage map developed for this population was previously described in detail (Hayes et al., 2014; Simko et al., 2015b).

## **RESULTS**

A very strong, positive, linear correlation ( $r = 0.97, p < 0.0001$ ) was observed between  $A_{PT}$  and FW (Figure 1) in experiment 1, demonstrating that optical sensors can be used to accurately estimate plant above-ground biomass from top view imaging of leaf area in the early stages of lettuce development. Visual classification of leaf color into three green and three red groups was in good agreement with RCC and RAC (Figure 2) values calculated from hyperspectral imaging. Differences in RCC among three green groups were highly significant. When RAC values were compared, a relatively small difference was detected between 'no red' (mean of 2.61) and 'light red' (mean of 3.04) groups, but this difference was also significant at  $p < 0.01$ . These data indicate that visual classification of green and red leaf color can be used for initial estimates of total chlorophyll and anthocyanin concentration in lettuce leaves, and that the combination of chlorophyll and anthocyanin concentration has a substantial effect on a customer's visual perception of lettuce leaf color. When the measurements of RCC, RAC, and  $F_v/F_m$  were performed on the adaxial and abaxial surfaces of leaves, a strong, linear correlation was observed between surfaces for each parameter (RCC:  $r = 0.90, p < 0.0001$ ; RAC:  $r = 0.86, p < 0.0001$ ;  $F_v/F_m$ :  $r = 0.84, p = 0.001$ ; Figure 3). However, there were small, yet consistent differences between values measured on the two surfaces when compared across all tested leaves. The overall values of the parameters were higher on the adaxial surface (RCC: 0.15 vs. 0.09,  $p < 0.001$ ; RAC: 2.72 vs. 1.80,  $p = 0.005$ ; and  $F_v/F_m$ : 0.86 vs. 0.85,  $p = 0.037$ ) and the differences between the surfaces were generally more pronounced at greater values of each parameter (Figure 3).

RGR of all accessions significantly decreased when plants were cultivated at COLD conditions (experiment 2). The average RGR for nine tested accessions was 20.1 in OPT, while during the same period of time it was only 2.5 in COLD (Figure 4). Similarly,  $F_v/F_m$  significantly decreased for all accession when in COLD, and the overall mean dropped from 0.75 in OPT to 0.68 in COLD (Figure 4). No significant correlation was detected between RGR and  $F_v/F_m$  in OPT ( $r = 0.66, p = 0.055$ ), or in COLD ( $r = 0.21, p = 0.568$ ).

The growth of plants tested in the experiment 3 was substantially reduced in both low and high temperatures. The overall RGR was 34.7 in OPT, 1.9 in COLD, and 15.3 in HOT. Significant decrease in RGR was detected for all accessions in COLD or HOT when compared to OPT (Figure 5). However, while all accessions almost completely stopped growing in COLD, their growth was still noticeable in HOT. Similarly, as in the experiment 2, the  $F_v/F_m$  parameter significantly decreased for all accessions when cultivated in COLD (the overall average at OPT = 0.76 and in COLD = 0.66; Figure 5). Though the overall average of  $F_v/F_m$  remained the same in HOT (0.76) as in OPT, the change in this parameter varied widely across accessions. The  $F_v/F_m$  value significantly decreased in HOT compared to OPT in three accessions (GR, LG, and RL), did not change significantly in five accessions (112, Eru, LB, Pav, and TT), and significantly increased in five accessions (110, 9A, 9B, GT, and



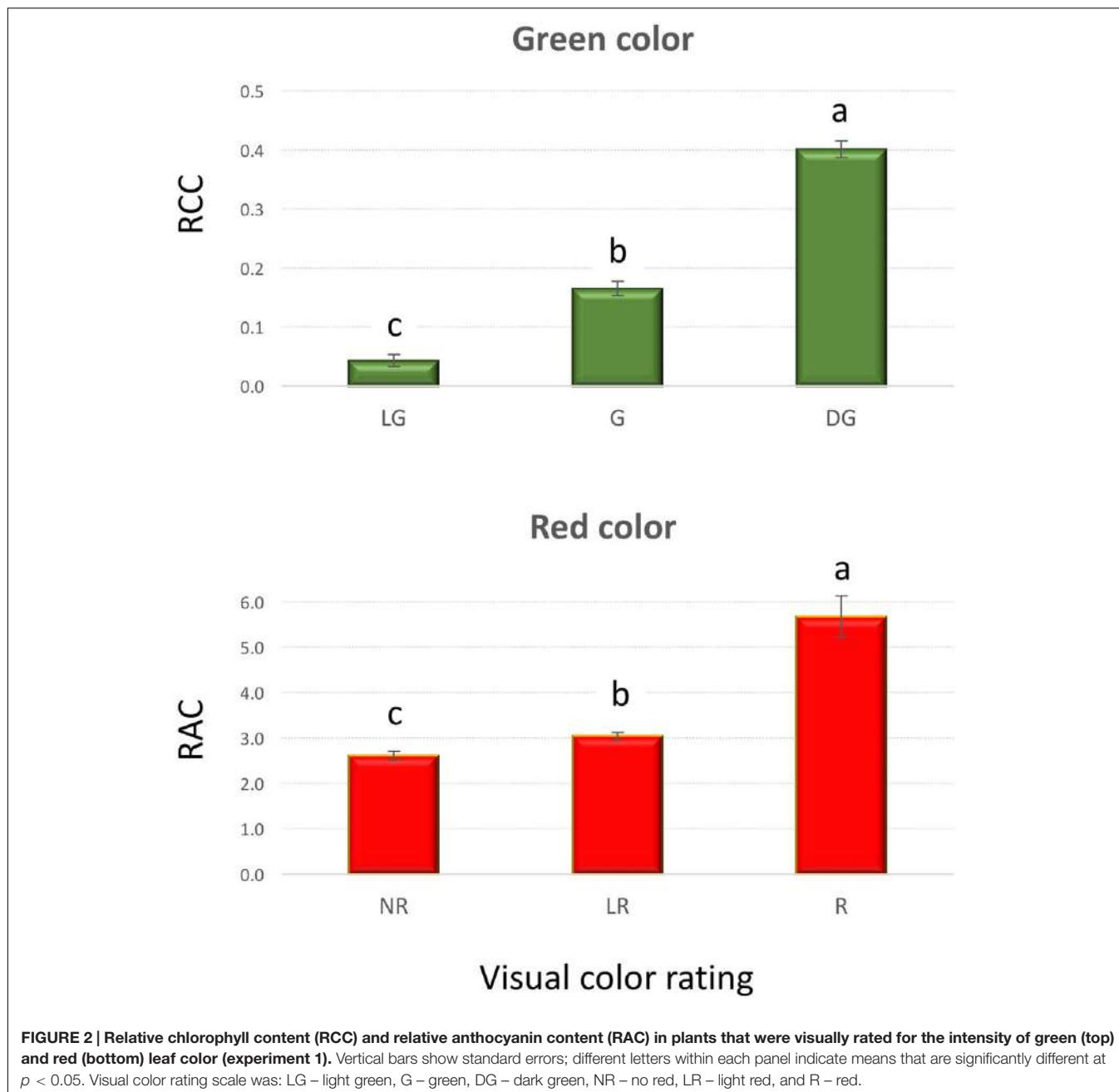
**FIGURE 1 |** Relationship between the total projected leaf area ( $A_{PT}$ ) determined from the images of chlorophyll fluorescence emission and the fresh weight (FW) of lettuce plants (experiment 1).

R1). No significant correlation was detected between  $F_v/F_m$  and RGR in any of the three growing conditions (OPT:  $r = 0.44$ ,  $p = 0.133$ ; COLD:  $r = 0.36$ ,  $p = 0.230$ ; and HOT:  $r = 0.52$ ,  $p = 0.068$ ). After returning plants cultivated in COLD to OPT conditions for 6 days their growth markedly improved (Figure 6). The overall RGR of plants constantly grown in OPT was 16.5, while for those previously grown in COLD it was 39.5. Significant increase in RGR (as compared to OPT) was detected for nine out of 13 accessions that were previously in COLD (9A, 9B, Eru, GR, GT, LB, LG, R1, and RL). The change in RGR was not so obvious for plants cultivated in HOT after they were returned to OPT. The overall RGR of these plants was 23.5, and only a single accession (RL) showed significantly higher RGR when compared to the plants constantly cultivated in OPT. The overall  $F_v/F_m$  for plants constantly in OPT was 0.76, for those moved from COLD 0.75, and for those previously in HOT 0.76 (Figure 6). These results show that the large drop in  $F_v/F_m$  in COLD (the overall value of 0.66) was not permanent and the plants recovered after moving to OPT. In cv. Eru, however, the  $F_v/F_m$  value for plants constantly in OPT was significantly higher (0.81) than for those moved in from HOT (0.78) or COLD (0.76) conditions. This difference does not seem to be caused by a damage to the light harvesting system but rather by a gradual increase in the  $F_v/F_m$  value for the plants in OPT. It has increased from 0.79 at the end of the temperature treatment period to 0.81 at the end of the recovery period. Again, no significant, linear correlation was detected between RGR and  $F_v/F_m$  after recovery period (OPT:  $r = 0.14$ ,  $p = 0.646$ ; COLD:  $r = 0.45$ ,  $p = 0.123$ ; and HOT:  $r = 0.03$ ,  $p = 0.920$ ). Five selected accession submitted to hyperspectral imaging showed significant changes in RCC and RAC when cultivated under

OPT, COLD, and HOT conditions. RCC gradually rose in all accessions with increasing temperature (Figure 7). The overall RCC values in COLD were 0.11, in OPT 0.17, and in HOT 0.33. In contrast, the overall RAC levels stayed almost the same in different conditions (COLD = 3.3, OPT = 3.4, and HOT = 3.2). Changes in RAC, however, varied across tested accessions (Figure 7). While RAC gradually decreased with the increasing temperature for the accession with the highest level of RAC in COLD (Eru), RAC increased in the accessions that had the lowest RAC in COLD (LG and LB). The changes in RCC and RAC appear to be reversible as readjustments in green and red coloring of foliage were already visually observable 1 day after moving plants from COLD and HOT to OPT conditions (Figure 8).

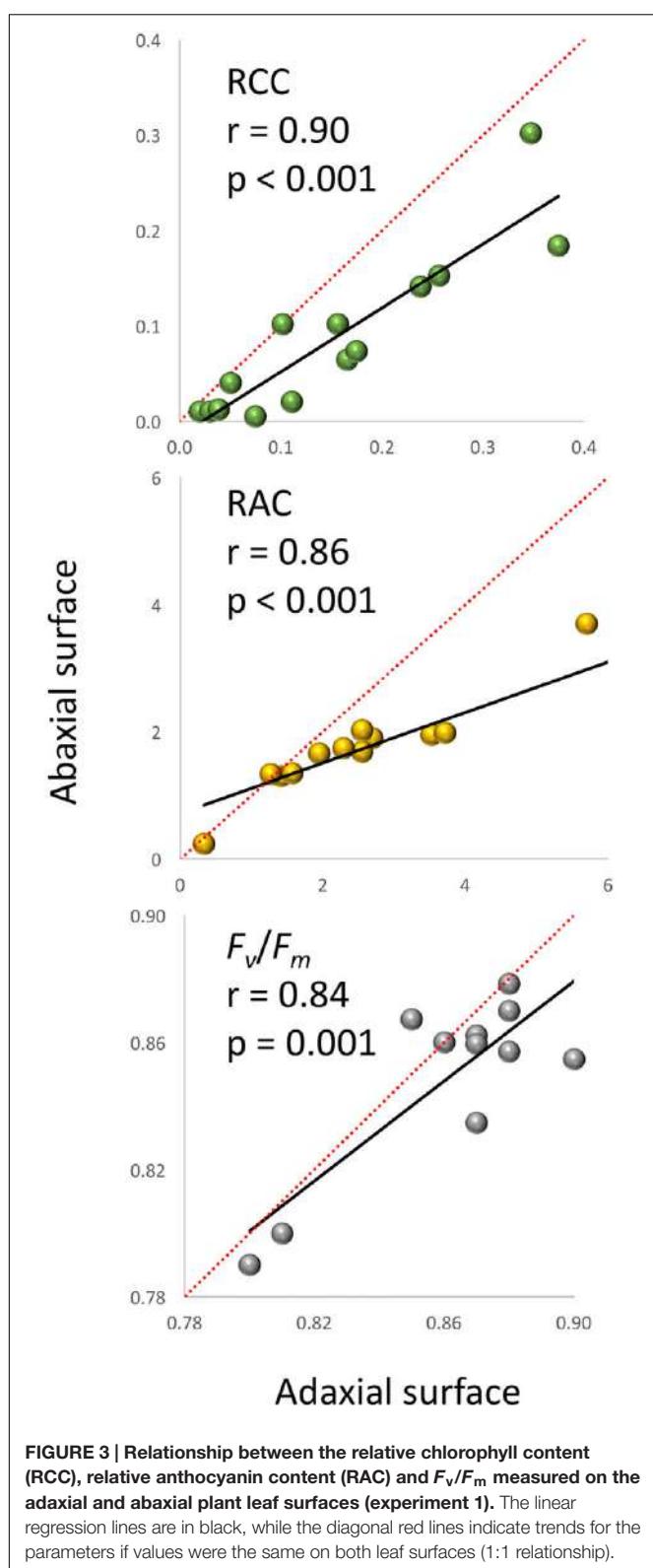
The addition of NaCl into irrigation water significantly affected both plant growth (RGR) and the efficiency of photosystem ( $F_v/F_m$ ). The overall RGR decreased from 30.5 in OPT to 17.9 in SALT (Figure 9). Though all accession had lower RGR in SALT, the difference was significant in only three (9A, LB, and RL) out of six tested accessions. Similarly, the  $F_v/F_m$  parameter decreased in SALT for all accessions, though the difference was not significant for R1 (Figure 9). The overall value of  $F_v/F_m$  dropped from 0.76 in OPT to 0.73 in SALT. No significant correlation was found between RGR and  $F_v/F_m$  in OPT ( $r = 0.49$ ,  $p = 0.320$ ), SALT ( $r = -0.36$ ,  $p = 0.478$ ), or the drop in the two parameters ( $r = 0.09$ ,  $p = 0.867$ ).

When young seedlings cultivated in Petri dishes were transferred to COLD conditions their  $F_v/F_m$  significantly decreased within a day as compared to the seedlings kept at OPT (Figure 10). While the overall value of  $F_v/F_m$  for cotyledons in OPT was 0.80, it was only 0.66 in COLD. Similar results were



observed on young seedlings of 33 accessions grown in a soil/sand mix in plastic boxes. After 2 days of temperature treatment, the overall  $F_v/F_m$  value in OPT was 0.85, while in COLD it was only 0.75. The  $F_v/F_m$  parameter significantly decreased in all accessions (drop ranged from 0.05 to 0.20) with the exception of cv. ToT, in which a drop of 0.03 occurred that, was not significant at  $p < 0.05$ . A relatively low decrease of  $F_v/F_m$  ( $\leq 0.07$ ) was also observed in accessions L6, Cli, Cor, US, BB, WM, Nan, and IC. In contrast, the largest decline in this parameter ( $\geq 0.12$ ) was observed in Eru, 9A, R1, RH, RL, Val, Pri, and 9B. There was a weak, but significant correlation ( $r = 0.47$ ,  $p = 0.005$ ) between  $F_v/F_m$  values in OPT and COLD (Figure 11).

Values of RCC measured on cotyledons of 56 randomly selected  $F_8$  RILs from the S88 × LB population were used to map locations of loci underlying this trait. A single, highly significant QTL (LOD = 5.5) was detected on linkage group 4 (LG 4), tightly linked to the marker Lsat\_1\_v3\_g\_0\_8627 (Figure 12; Table 1). Data from the field evaluations of lettuce leaf color on 90 RILs from the same population yielded also only a single QTL (LOD = 21.1) that was located at the same chromosomal region, indicating that the measurements of RCC on cotyledons and the visual assessment of green color on adult plants likely sense the same trait. The QTL for light-green color ( $qLG4$ ) explains 38 and 68% of the total phenotypic variation for RCC and the



**FIGURE 3 | Relationship between the relative chlorophyll content (RCC), relative anthocyanin content (RAC) and  $F_v/F_m$  measured on the adaxial and abaxial plant leaf surfaces (experiment 1).** The linear regression lines are in black, while the diagonal red lines indicate trends for the parameters if values were the same on both leaf surfaces (1:1 relationship).

visual color rating, respectively. The linear correlation between the phenotypic values of the two traits was highly significant ( $r = 0.71$ ,  $p < 0.0001$ ).

## DISCUSSION

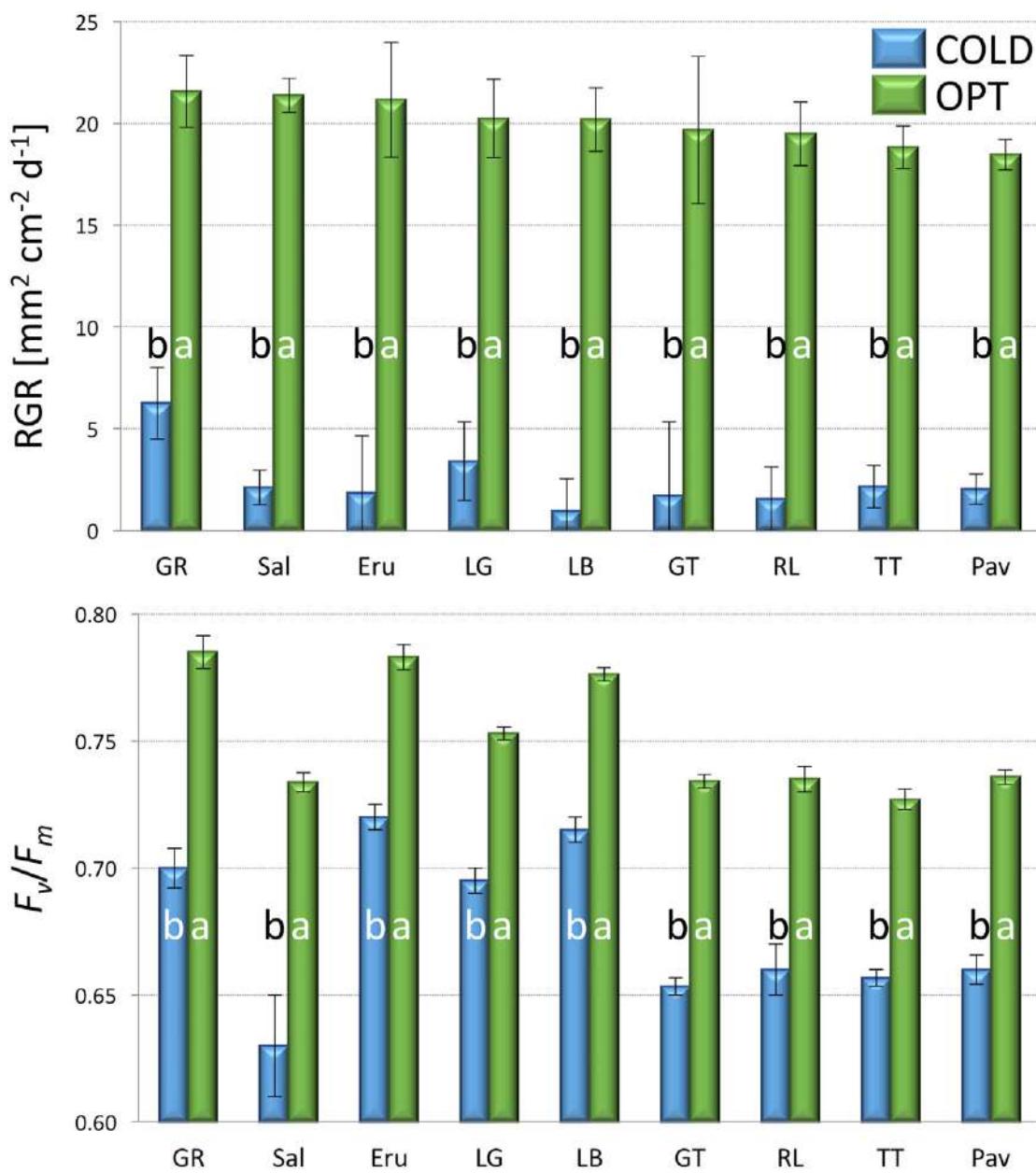
### Evaluation of Plant Size

Optical sensors that provide a top view image have commonly been used for the non-destructive estimates of leaf area for the plants with planophile growth habit (Barbagallo et al., 2003; Jansen et al., 2009; Flood et al., 2016). We have successfully used chlorophyll fluorescence imaging, rather than color or monochrome images, to estimate  $A_{PT}$  that in turn accurately predicts FW of young lettuce plants (Figure 1). This evaluation is possible because lettuce plants in the early stages of growth have almost planophile growth habit, with relatively flat, non-overlapping leaves. When the plants develop further, their growth habit become more erectophile, with leaves starting to overlap. Therefore, we do not recommend using the top view-based optical sensors alone for the estimates of lettuce size above 7 g FW without multiple view imaging, stereo imaging or other techniques to generate a height dimension.

### Evaluation of Leaf Color

Lettuce leaf color is critically evaluated by customers when making purchasing decisions. The amount and the distribution of leaf pigments contributes to the visual appeal of lettuce; thus it is important to analyze changes in the color under different growing conditions. We have determined that the visual rating of the green and red leaf color is strongly associated with the values of RCC and RAC obtained from hyperspectral imaging (Figure 2). Previously, a good correlation ( $r = 0.76$ ) was found between the ratings of lettuce red color intensity performed by human panelists and the direct measurements of anthocyanin levels (Gazula et al., 2007). Panelists, however, could not determine red color in a cultivar with a low anthocyanin level. Similarly, we have detected anthocyanin through hyperspectral analysis in the leaf samples with the 'no red' color rating, confirming that the threshold level of anthocyanin needed for the detection by visual observation is higher than is the actual level of anthocyanin in some lettuces (Gazula et al., 2007).

Several previous studies (e.g., Merzlyak et al., 2003; Xue and Yang, 2009) and our present analyses show that hyperspectral imaging can be successfully used for the quantification of chlorophyll and anthocyanin in leaves. There are certain aspects, however, that need to be considered when using optical sensors for such quantifications. The top view optical sensors quantify levels of pigments on the adaxial surface only, while extraction-based methods analyze samples that normally represent the cross section of the leaf. Though we detected very strong, linear correlations between RCC ( $r = 0.90$ ) and RAC ( $r = 0.86$ ) on the abaxial and adaxial leaf surfaces (Figure 3), the absolute difference between results obtained for the two surfaces increased as the level of pigments in leaves increased. This expanding difference is caused by a greater increase in pigments on the surfaces that have a direct contact with light. Therefore, scanning of both surfaces with optical sensors may be considered when such analysis is feasible; e.g., at the end of the experiment when plants can be removed from pots

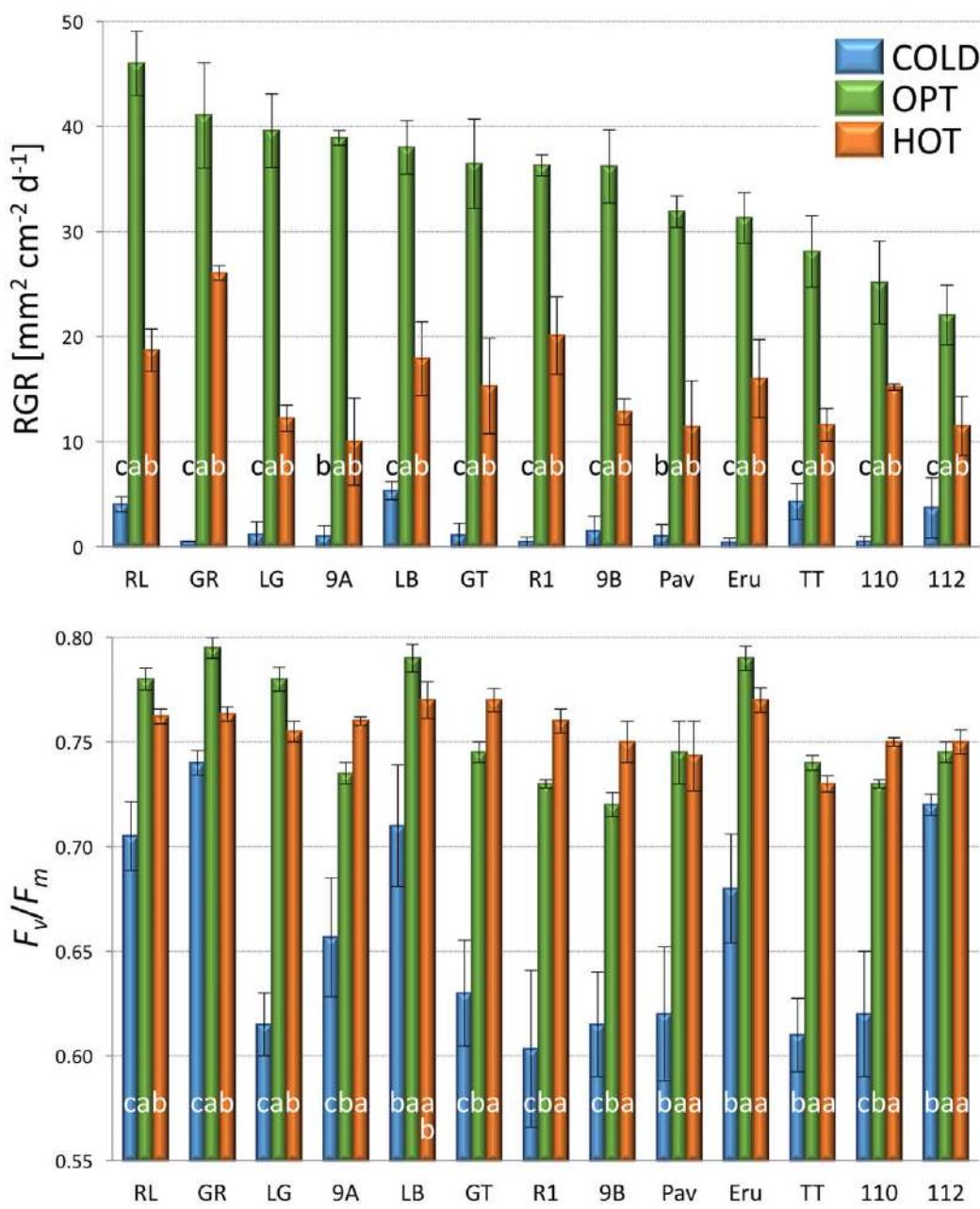


**FIGURE 4 |** Relative growth rate (RGR, top) and  $F_v/F_m$  (bottom) of plants cultivated in optimal (OPT, green color) or low (COLD, blue color) temperature (experiment 2). Vertical bars show standard errors ( $n = 5$ ); different letters within each accession indicate means that are significantly different at  $p < 0.05$ .

or leaves cut. Also, when plants age and their leaves begin overlapping, optical sensors cannot scan the hidden areas of leaves, thus detecting pigments only on the visible areas exposed to light. Nevertheless, the use of optical sensors has several major advantages compared to the quantifications performed on extracts from leaf tissue; it is much faster, it can analyze the whole plant surface (and even multiple plants) at once and promptly identify differences within a leaf, plant or between plants non-destructively, thus allowing analysis of the same leaf over time.

### Effect of Suboptimal Temperature

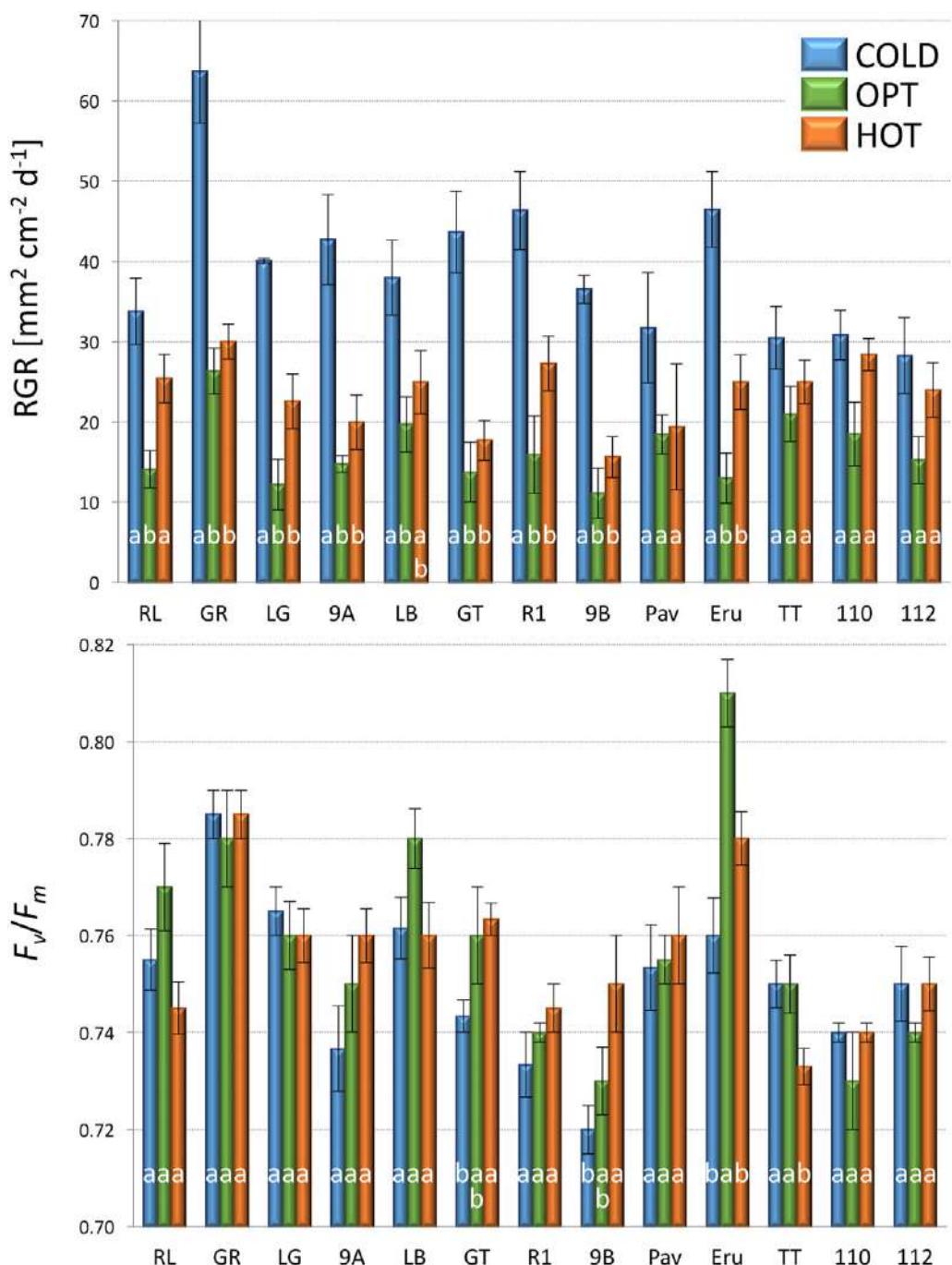
Our results show that the suboptimal temperatures have a major effect on the RGR, RCC and RAC levels (Figure 12), and the plant photochemistry as measured by the  $F_v/F_m$  parameter. When the plants were transferred from OPT to COLD conditions, their growth immediately decreased to almost zero and was followed by a drop in  $F_v/F_m$  and RCC. A similar decline of RGR and  $F_v/F_m$  was previously observed in *Arabidopsis thaliana* L. plants transferred from 22°C/18°C (day/night) to 5°C (Jansen et al., 2009), in  $F_v/F_m$  and chlorophyll content



**FIGURE 5 |** Relative growth rate (RGR, top) and  $F_v/F_m$  (bottom) of plants cultivated in optimal (OPT, green color), low (COLD, blue color), or high (HOT, orange color) temperature (experiment 3). Vertical bars show standard errors ( $n = 5$ ); different letters within each accession indicate means that are significantly different at  $p < 0.05$ .

of watermelon plants [*Citrullus lanatus* (Thunb.) Matsum. & Nakai] transferred to 12°C/10°C from 25°C/15°C (Hou et al., 2016), and in  $F_v/F_m$  of lettuce plants submitted to 4°C for 24 h (Oh et al., 2009). When compared to OPT conditions, RAC in COLD decreased in the accessions with low and intermediate RAC levels, but increased in the accession with high RAC (Figure 7). After returning plants to the OPT conditions, RCC, RAC (Figure 8), and  $F_v/F_m$  went back to the levels measured prior to the COLD treatment, indicating that the

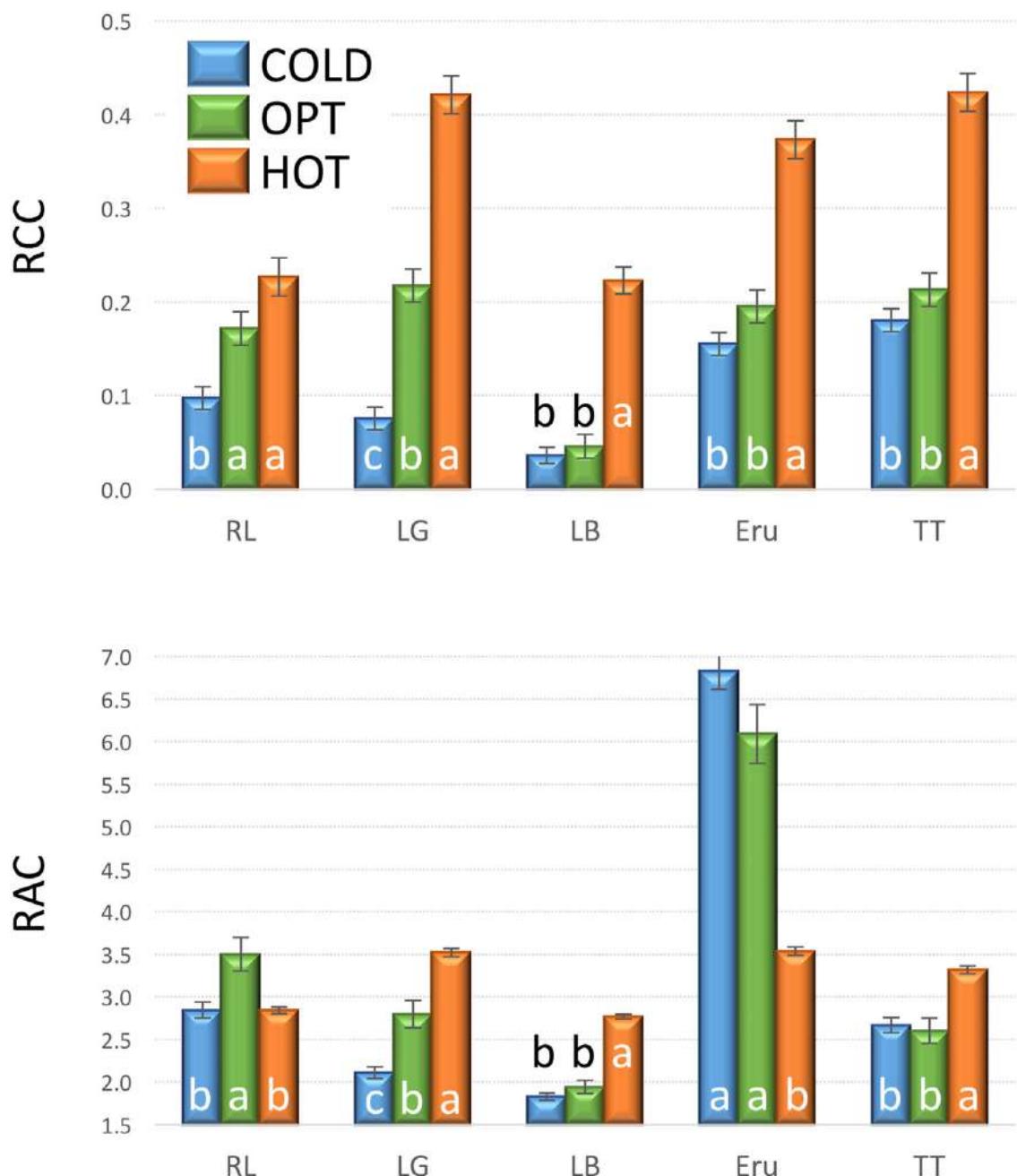
plants were not permanently damaged at 3°C. Remarkably, lettuce plants showed a very high resilience to low temperatures. In a single, unreplicated experiment, plants from each of the accessions tested in the experiment 2 were kept at COLD for 3 months (at 16 h photoperiod and 400  $\mu\text{mol m}^{-2} \text{s}^{-1}$  PPFD). After moving plants to OPT they immediately recovered without any obvious damage that could be visually observed (data not shown). Our results indicate that lettuce plants have a high adaptability to temperatures close to freezing, at least



**FIGURE 6 |** Relative growth rate (RGR, top) and  $F_v/F_m$  (bottom) of plants constantly cultivated in optimal (OPT, green color) temperature, or transferred to the optimal temperature after 8 days at low (COLD, blue color) or high (HOT, orange color) temperature (experiment 3). Vertical bars show standard errors ( $n = 5$ ); different letters within each accession indicate means that are significantly different at  $p < 0.05$ .

under the conditions tested in this study. We did not find any significant relationship between the decrease in RGR and  $F_v/F_m$  at COLD, or after moving plants from COLD to OPT. It was reported previously, that  $F_v/F_m$  decreases faster in the cold-sensitive *A. thaliana* plants than in the less sensitive plants when cultivated at 5°C (Jansen et al., 2009). These authors, however,

compared only two genotypes (wild type and transgenic); thus the results may not represent the general trend across many different genotypes. Still, it is possible, that  $F_v/F_m$  values in lettuce may show a relationship to RGR if tested under different environmental conditions (temperature, photoperiod, and/or PPFD).

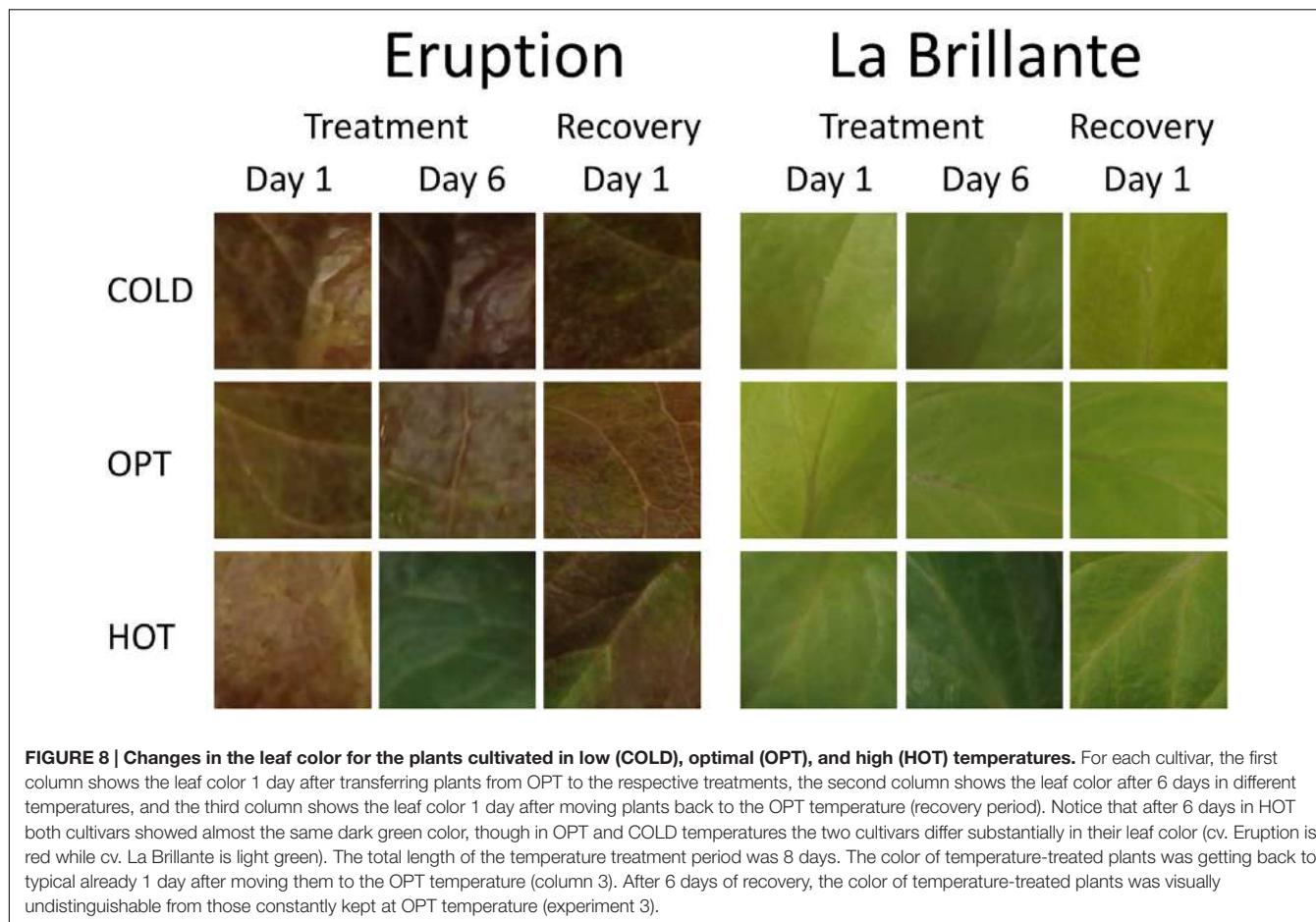


**FIGURE 7 |** Relative chlorophyll content (RCC, top) and relative anthocyanin content (RAC, bottom) in plants cultivated in optimal (OPT, green color), low (COLD, blue color), or high (HOT, orange color) temperature (experiment 3). Vertical bars show standard errors ( $n = 5$ ); different letters within each accession indicate means that are significantly different at  $p < 0.05$ .

### Effect of Supra-Optimal Temperature

The increase in temperature from 21°C (OPT) to 39°C (HOT) led to a significant decrease in RGR and increase in RCC in all accessions (Figures 5, 7, 8, and 13). RAC substantially decreased only in the accession with very high RCC at OPT, but increased in the majority of the accessions with the lower levels of RCC in OPT (Figures 7, 8, and 13). The change in the  $F_v/F_m$  parameter varied greatly among accessions (increasing,

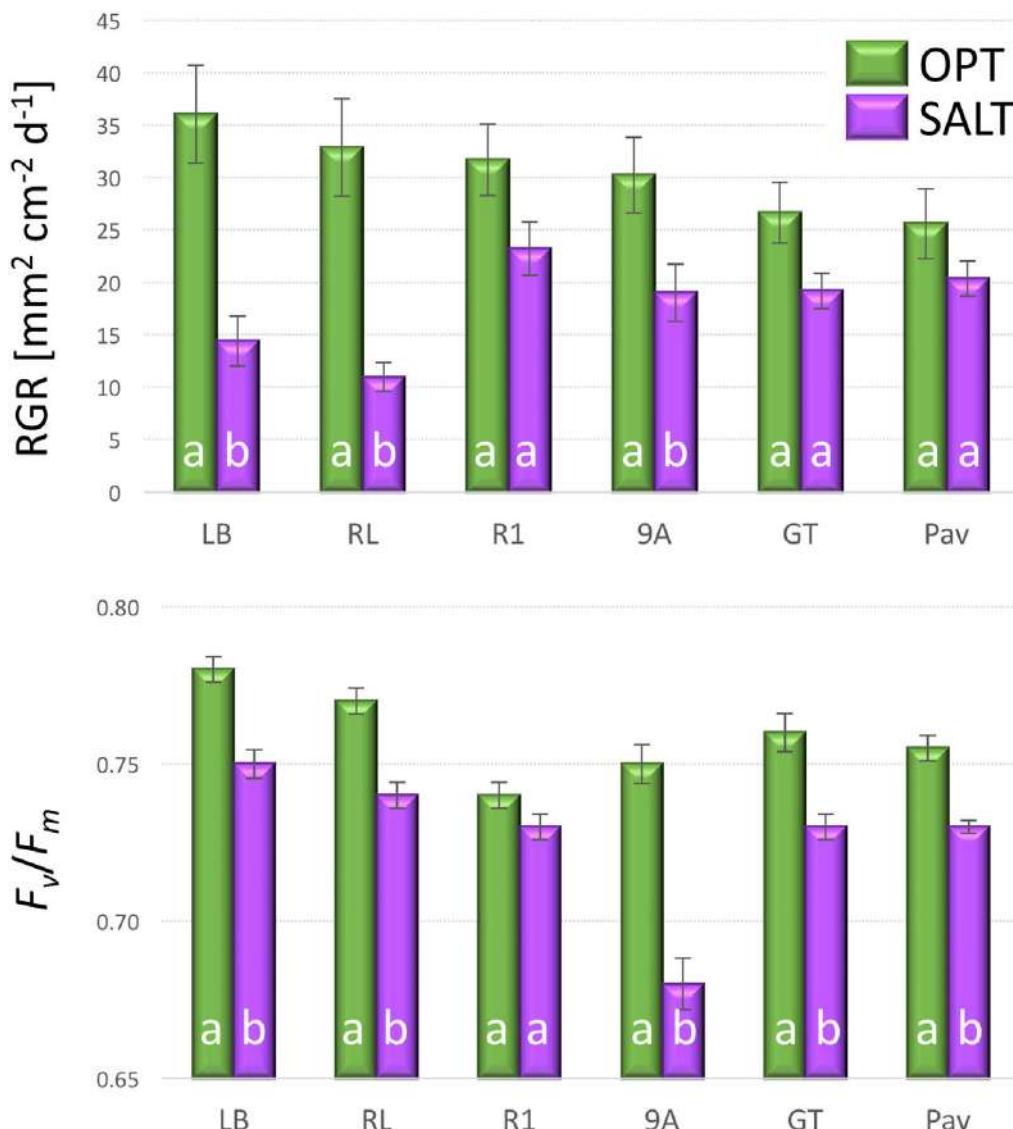
staying unchanged, or decreasing; Figure 5). A previous study on watermelon reported only a slight drop in  $F_v/F_m$  when temperature increased from 25°C/15°C to 42°C/40°C at the irradiance level of 250  $\mu\text{mol m}^{-2} \text{s}^{-2}$  (Hou et al., 2016). When lettuce plants of a single cultivar were exposed to 38°C for 3 h the  $F_v/F_m$  ratio somewhat decreased, but returned to almost the original values 21 h after the treatment (Oh et al., 2009).



Reduced concentrations of chlorophyll and anthocyanin have been observed in lettuce grown at supra-optimal temperatures (Gazula et al., 2005; Chon et al., 2012). These changes in the chlorophyll content are at odds with our observations that show overall increases in RCC with growing temperature (Figures 7, 8, and 13). Similar to our results, almost a 10-fold increase in chlorophyll content has been observed in plants of cultivar Grand Rapids when the average temperature was raised from 23 to 33°C (at 600  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ; Frantz et al., 2004). These large differences between studies may be caused by numerous factors, including accessions used in the studies, other environmental conditions interacting with temperature (photoperiod, PPFD, humidity, etc.), levels of nutrients in the growing substrate, watering regime, and the age of plants. The HOT temperature treatment used in our study (39°C) is probably close to the upper limit that cultivated lettuce could survive when continuously exposed to for several days. In the preliminary test (data not shown) we used the temperature of 42°C that led to severe stress and irreversible modifications in plants, such as chlorotic lesions, malformed leaves, dropping of leaves, and also plant death.

The overall mean of RAC stayed similar across treatments (COLD = 3.26, OPT = 3.39, and HOT = 3.20), while the variance among accessions radically decreased (significantly

smallest in HOT; Figure 14). Because the RAC variance within accessions did not substantially change under the different temperature treatments, the ANOVA *F*-value was over four times greater for RAC in COLD than in OPT or HOT. Hence, the breeders selecting for dark red color (high RAC) may consider temporarily subjecting lettuce plants to low temperatures where differences among genotypes are likely to be more pronounced (assuming that the pattern of changes in pigments is the same as in our study). These results are somewhat unexpected, because several previous studies reported that low temperatures lead to increased anthocyanin production in lettuce (Gazula et al., 2005, 2007; Boo et al., 2011; Chon et al., 2012; Becker et al., 2014a). However, when the regulation of anthocyanin biosynthesis [quantified as cyanidin-3-O-(6''-malonyl)-glucoside] was compared in three cultivars, a substantial difference in their response to varying temperatures was detected. While anthocyanin production in the red oak cultivar was negatively correlated with increasing temperature, the correlation was positive in the two Batavia cultivars (Marin et al., 2015). Nevertheless, it is problematic to compare results of temperature treatments from diverse studies, because several factors, including radiation (Tsormpatsidis et al., 2008, 2010; Marin et al., 2015), relative humidity (Marin et al., 2015), water availability (Rajabbeigi et al., 2013), light



**FIGURE 9 |** Relative growth rate (RGR, top) and  $F_v/F_m$  (bottom) of plants cultivated in optimal (OPT, green color) or elevated salinity (SALT, purple color) conditions (experiment 4). Vertical bars show standard errors ( $n = 5$ ); different letters within each accession indicate means that are significantly different at  $p < 0.05$ .

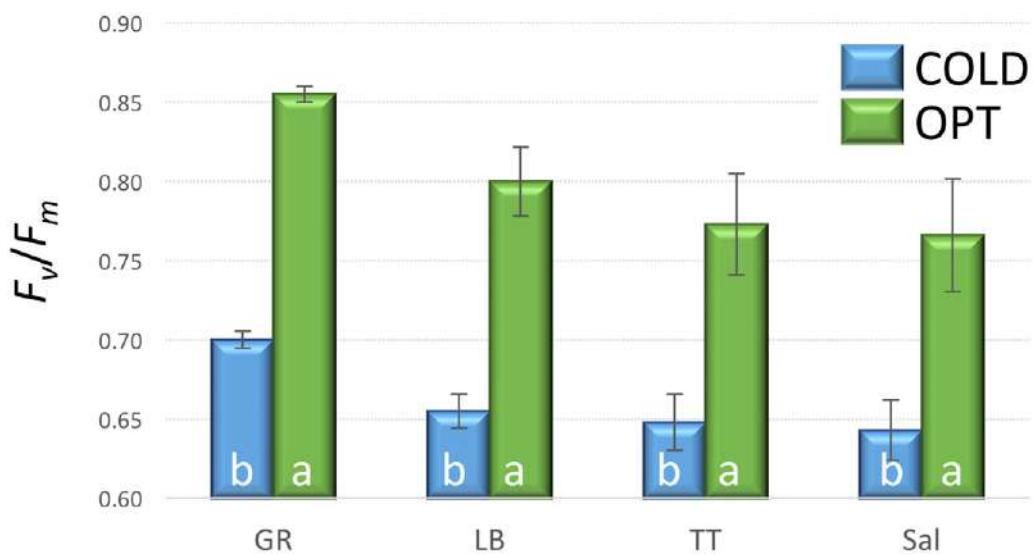
source (Park et al., 2012), CO<sub>2</sub> availability (Park et al., 2012), and plant growth stage (Becker et al., 2014b) affect biosynthesis of anthocyanin in lettuce either directly or in interaction. It is possible, that supra-optimal temperature (HOT) in our experiments caused a drought stress, despite regular watering. It was demonstrated that drought stress significantly increases anthocyanin content in lettuce (Rajabbeigi et al., 2013). Therefore, in future experiments, it may be useful to assess also differences in plant transpiration among tested accessions.

In contrast to RAC, the changes in temperature affected RCC in all accessions similarly. The overall mean of RCC grew from 0.11 in COLD, to 0.17 in OPT, to 0.33 in HOT, while the variance also gradually, but non-significantly, increased (Figure 14).

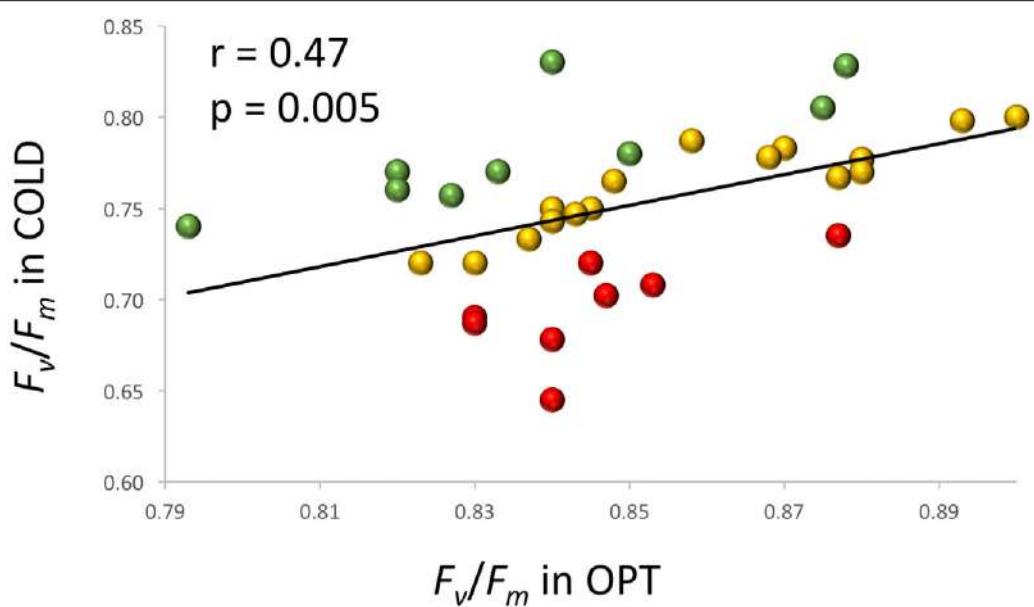
Differences in RCC among accessions suggest that the selection for higher RCC may be performed at somewhat higher than the optimal temperature, but the change in  $F$ -value was relatively minor (1.5 higher in HOT than in OPT) compared to that seen in RAC.

### Effect of Elevated Salinity

Elevated salinity inhibits growth of young leaves (the rapid, osmotic phase of the plant response to salt) and accelerates senescence of mature leaves (the slower, ionic phase of the plant response; Munns and Tester, 2008). Beside reducing plant growth, increased concentrations of salt lead to decreased water content in lettuce, lower concentrations of chlorophyll *a* and *b*, smaller intracellular spaces, increased elasticity of leaves, higher



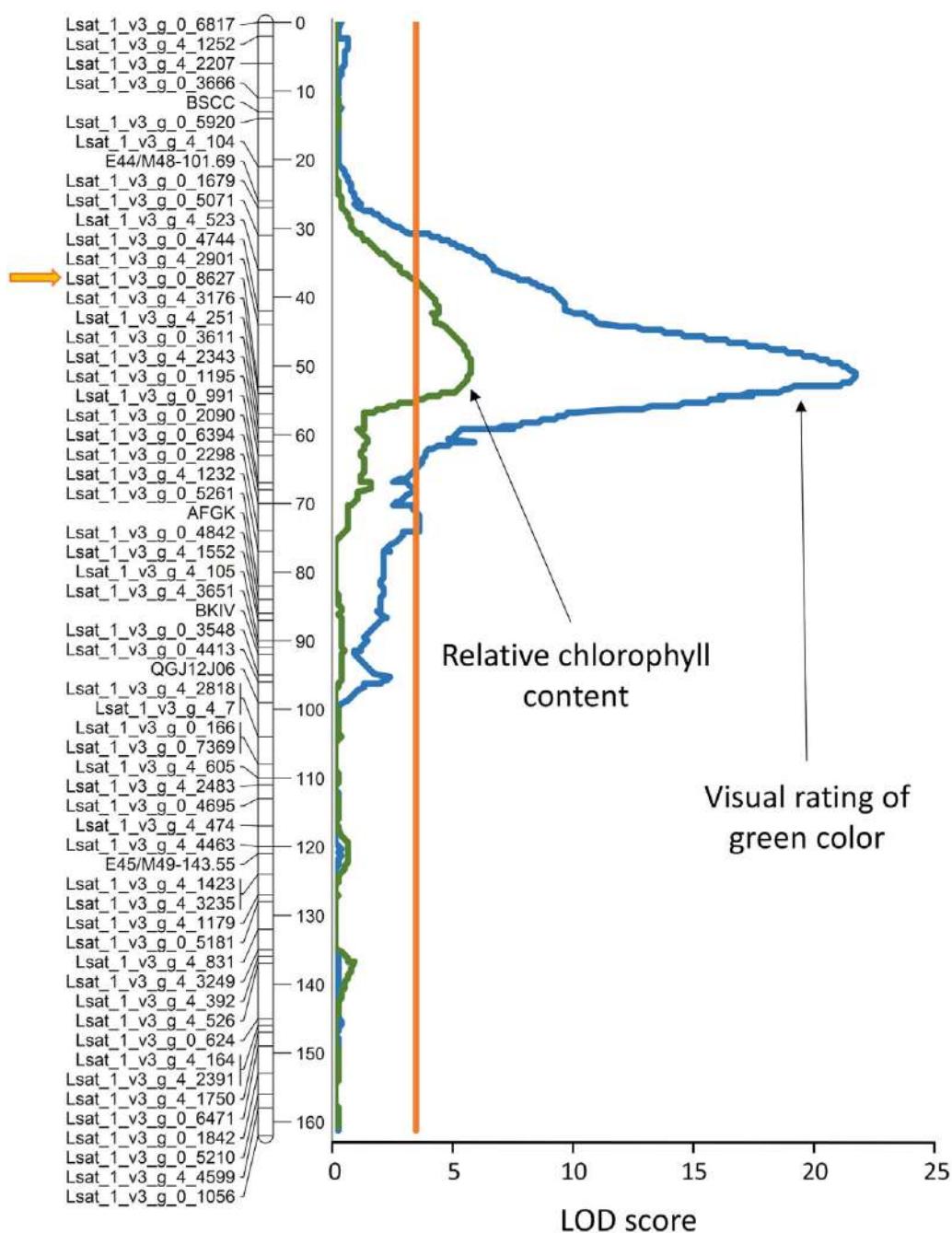
**FIGURE 10 |**  $F_v/F_m$  on cotyledons of seedlings cultivated in optimal (OPT, green color) or low (COLD, blue color) temperature in Petri dishes (experiment 5). Vertical bars show standard errors ( $n = 5$ ); different letters within each accession indicate means that are significantly different at  $p < 0.05$ .



**FIGURE 11 |** Relationship between  $F_v/F_m$  measured on cotyledons of seedlings cultivated in optimal (OPT, x-axis) or low (COLD, y-axis) temperature in plastic boxes (experiment 6). Red color indicates accessions with the largest decrease ( $\geq 0.12$ ) in the  $F_v/F_m$  values, green color indicates accessions with the lowest decrease ( $\leq 0.07$ ) in the  $F_v/F_m$  values, and yellow color indicates accessions with the intermediate decrease (0.08–0.11) in the  $F_v/F_m$  values. The  $F_v/F_m$  values in COLD were significantly ( $p \leq 0.05$ ) smaller for all but one accessions.

concentration of phenolic acids, and larger leaf areas occupied by palisade and spongy parenchyma (Garrido et al., 2014). Our study, focusing on detecting changes in RGR and photochemical efficiency ( $F_v/F_m$ ), determined that the growth of young plants was substantially reduced after exposing them to 100 mM NaCl for 8 days. The inhibition of plant growth did not significantly correlate with the reduction of  $F_v/F_m$  (Figure 9). Slower growth was previously observed for young lettuce plants (Xu and Mou,

2015) that were subjected to a mix of NaCl and CaCl<sub>2</sub>. Similar to our study, the reduction in FW was not correlated with the  $F_v/F_m$  values. These results suggest that  $F_v/F_m$  is not a robust indicator of the performance of lettuce under salt stress, possibly because growth may be more sensitive to the osmotic component of salt stress than photochemical efficiency (Munns and Tester, 2008). The impact of stress on  $F_v/F_m$  is strongly linked to the severity of the stress, as its value has been observed to gradually decrease



**FIGURE 12 | Genomic position of the quantitative trait locus (QTL) for light green color (qLG4) on linkage group 4.** Visual rating of the green color intensity was performed on adult plants in field, while the relative chlorophyll content (RCC) was determined from hyperspectral reflectance measured on cotyledons of seedlings cultivated in plastic boxes (experiment 7). The orange line parallel with the linkage map shows the significance threshold ( $\alpha = 0.05$ ). The allele for light green color and low RCC originates from cv. La Brillante. Detailed description of the linkage map for this population and its construction was published previously (Hayes et al., 2014; Simko et al., 2015b). Distance in cM is shown on the right site of the linkage map. LOD, logarithm of odds.

with increasing concentrations of NaCl applied (Bartha et al., 2010; Qin et al., 2013); potentially due to tissue damage from salt rather than an osmotic effect. Therefore, higher concentrations of NaCl, or a longer exposure to the salinity would likely yield different results.

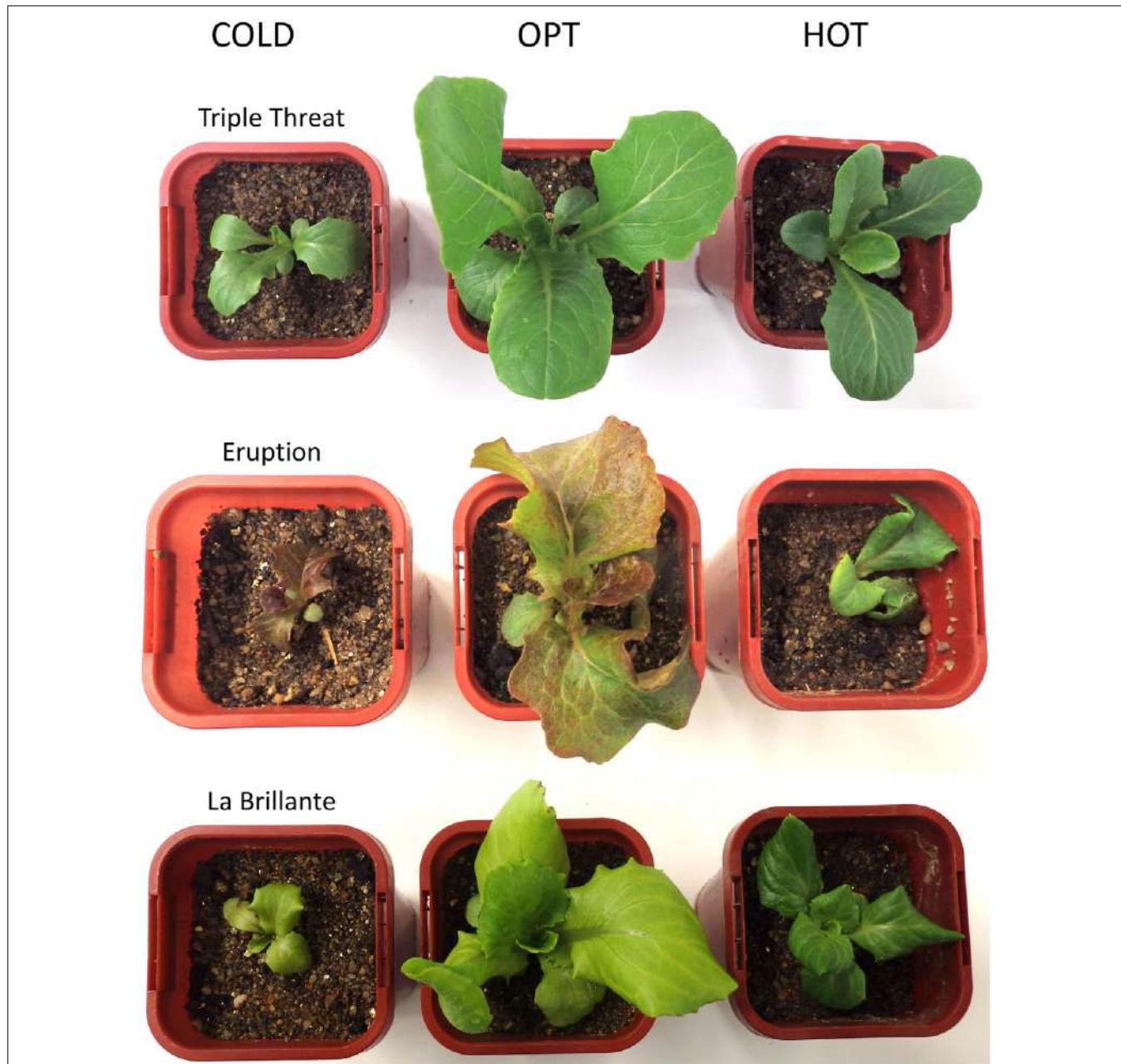
### Allele for Light Green Leaf Color

At least 10 genes related to the chlorophyll level in lettuce have been described previously (Robinson et al., 1983), including the *lg* gene for light green leaf color. Though the *lg* gene *per se* has not yet been mapped, the gene is loosely linked to the lettuce mosaic

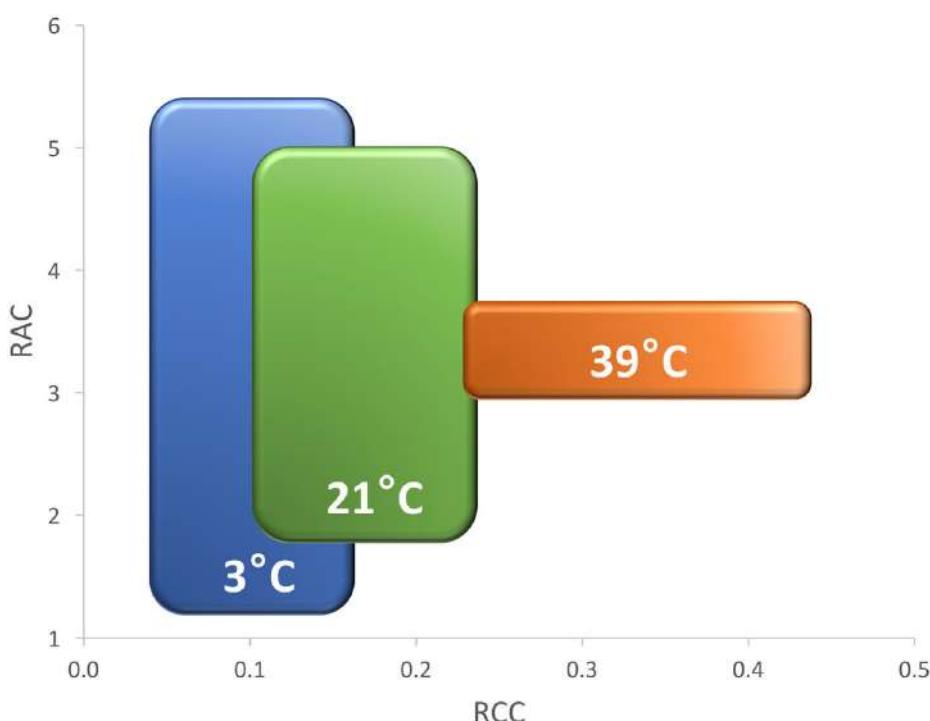
**TABLE 1 | Genomic location of the *qLG4* locus and its effect on the intensity of green color and the relative chlorophyll content.**

Trait	No. of RILs <sup>c</sup>	LG	Marker <sup>d</sup>	QTL location (cM) <sup>e</sup>	Support interval (cM) <sup>f</sup>	LOD	R <sup>2</sup> % <sup>g</sup>
Color <sup>a</sup>	90	4	Lsat_1_v3_g_0_8627	51.3	49.3–53.0	21.1	67.6
RCC <sup>b</sup>	56	4	Lsat_1_v3_g_0_8627	49.7	44.5–53.9	5.5	38.4

<sup>a</sup>Visual evaluation of green color on adult plants in field. <sup>b</sup>Relative chlorophyll content measured on cotyledons of seedlings grown in plastic boxes. <sup>c</sup>90 RILs from the S88 × LB population were visually evaluated for the leaf color, but only a subset of 56 RILs was evaluated for RCC. <sup>d</sup>Molecular marker nearest to the QTL. <sup>e</sup>Location of the QTL on the molecular linkage map. <sup>f</sup>The range of 1-LOD support interval. <sup>g</sup>Percent of the total phenotypic variation of the trait explained by the QTL. The allele for light green color and low RCC originates from cv. La Brillante.

**FIGURE 13 | Comparison of the size and the color of plants cultivated at optimal (OPT), low (COLD), and high (HOT) temperatures (experiment 3).**

Plants were initially grown at OPT for 10 days and then either continuously kept in OPT or transferred to COLD or HOT for 8 days. Sides of the square pots are 68 mm long.



**FIGURE 14 | Effect of increasing temperatures on relative chlorophyll content (RCC, x-axis) and relative anthocyanin content (RAC, y-axis) in five cultivars (Eruption, Little Gem, La Brillante, Red Leaf, and Triple Threat).** Plants were cultivated in low ( $3^{\circ}\text{C}$ , COLD, blue color), optimal ( $21^{\circ}\text{C}$ , OPT, green color), and high ( $39^{\circ}\text{C}$ , HOT, orange color) temperatures. Rectangles indicate the extend of standard deviations that were calculated from the mean values of all cultivars. The overall mean and the variance of RCC gradually grows with the increasing temperature. Opposite, the overall mean of RAC stays nearly stable while the variance significantly decreases (experiment 3). The large difference in RAC among cultivars in COLD ( $3^{\circ}\text{C}$ ) suggests that the selection for red color may be performed at somewhat lower than the optimal temperature.

virus resistance gene *mo-1* (Ryder, 1992) located on LG 4 (Nicaise et al., 2003; McHale et al., 2009). It is plausible then that *qLG4* (Figure 12) may be linked (or is allelic) to the *lg* gene. The *qLG4* locus detected in our study is different from QTLs for the total chlorophyll content that have previously been mapped to LGs 3, 7, and 9 (Damerum et al., 2015), or from those for chlorophyll *a* and chlorophyll *b* content located on LGs 1, 2, and 8 (Hayashi et al., 2012).

## CONCLUSION

An application of optical sensors for the analysis of plants is getting increased attention from plant scientists and growers, as the cost of sensors decreases and they become to be more widely available. Because sensors are amenable to automatization, high-throughput, automatic phenotyping is particularly attractive for the use in large-scale experiments (White et al., 2012; Fiorani and Schurr, 2013; Araus and Cairns, 2014), performed under field- or environmentally controlled conditions. Sensor-based phenotyping of lettuce is still, however, in only early stages of development. To our knowledge, automatic phenotyping with optical sensors is not yet commonly applied for analysis of lettuce plants in field, though optical sensor-based machines are already commercially used for precise thinning of lettuce crop

(Blue River Technology, Sunnyvale, CA, USA). More studies are needed to develop sensing and analytical tools and mathematical models that can be applied for the precise evaluation of lettuce plants in advanced stages of development when leaves from the same or nearby plants overlap and heads (a grouping of tightly packed, overlaying leaves) form. Such phenotyping tools would be valuable for the evaluations of crop development and its overall quality. Small-, and medium-size phenotyping studies performed in environmentally controlled areas (e.g., greenhouse, growth chamber, or laboratory) on young plants can be used by lettuce breeders to evaluate plant growth, architecture, and resistance and to select genotypes with desirable traits at early stages of development.

The present study was designed to test feasibility of using optical sensors for physiological evaluation of lettuce plants in early stages of their development with the long term aim of using these tools for breeding applications. Our results indicate that top view sensors can accurately determine plant size to approximately 7 g FW. Hyperspectral imaging analysis was able to detect changes in the total chlorophyll and anthocyanin levels, while chlorophyll fluorescence imaging revealed photoinhibition and reduction of plant growth caused by the extreme growing temperatures and salinity. Though no significant correlation was found between  $F_v/F_m$  and decrease in plant growth due to stress when comparisons were made across multiple accessions, it is

possible that this parameter may be used to determine the level of stress within an accession (e.g., gradual decrease in  $F_v/F_m$  values with falling temperatures) or be useful at higher levels of salt stress. It was demonstrated before that low temperatures, moderate heat, salt stress, and CO<sub>2</sub> limitation inhibit the repair of PSII by suppressing the synthesis of D1 protein that is required for the assembly of the active PSII complex (Murata et al., 2007). Therefore, more detailed studies are needed to investigate the genotype-specific effect of different stress factors on the decrease of  $F_v/F_m$  in lettuce. This study, however, serves as a proof of concept that optical sensors can be successfully used for non-destructive phenotyping of young lettuce plants. Moreover, we were able to identify the locus for light green leaf color (*qLG4*), and position this locus on the molecular linkage map of lettuce showing that these techniques have sufficient resolution to use in a genetic context in lettuce.

## AUTHOR CONTRIBUTIONS

IS designed and performed the experiments, statistically analyses and interpreted the data, and wrote the manuscript. RH developed the mapping population, contributed to data

interpretation, and revised the manuscript. RF developed approaches for the phenotypic analyses, provided crucial expertise for designing of the experiments and data interpretation, and revised the manuscript.

## FUNDING

IS acknowledges the receipt of a fellowship from the OECD Co-operative Research Programme: Biological Resource Management for Sustainable Agricultural Systems in 2013.

## ACKNOWLEDGMENTS

The authors wish to acknowledge the use of the Australian Plant Phenomics Facility in carrying out these experiments and the help of Helen Daily and the staff and students at HRPPC. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

## REFERENCES

- Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/j.tplants.2013.09.008
- Badger, M. R., Fallahi, H., Kaines, S., and Takahashi, S. (2009). Chlorophyll fluorescence screening of *Arabidopsis thaliana* for CO<sub>2</sub> sensitive photorespiration and photoinhibition mutants. *Funct. Plant Biol.* 36, 867–873. doi: 10.1071/FP09199
- Barbagallo, R. P., Oxborough, K., Pallett, K. E., and Baker, N. R. (2003). Rapid, noninvasive screening for perturbations of metabolism and plant growth using chlorophyll fluorescence imaging. *Plant Physiol.* 132, 485–493. doi: 10.1104/pp.102.018093
- Bartha, C., Fodorpataki, L., Székely, G., and Popescu, O. (2010). Physiological diversity of lettuce cultivars exposed to salinity stress. *Cont. Bot.* 45, 47–56.
- Becker, C., Klaering, H.-P., Kroh, L. W., and Krumbein, A. (2014a). Cool-cultivated red leaf lettuce accumulates cyanidin-3-O-(6'-O-malonyl)-glucoside and caffeoylmalic acid. *Food Chem.* 146, 404–411. doi: 10.1016/j.foodchem.2013.09.061
- Becker, C., Klaering, H.-P., Schreiner, M., Kroh, L. W., and Krumbein, A. (2014b). Unlike quercetin glycosides, cyanidin glycoside in red leaf lettuce responds more sensitively to increasing low radiation intensity before than after head formation has started. *J. Agric. Food Chem.* 62, 6911–6917. doi: 10.1021/jf404782n
- Boo, H.-O., Heo, B.-G., Gorinstein, S., and Chon, S.-U. (2011). Positive effects of temperature and growth conditions on enzymatic and antioxidant status in lettuce plants. *Plant Sci.* 181, 479–484. doi: 10.1016/j.plantsci.2011.07.013
- Chon, S.-U., Boo, H.-O., Heo, B.-G., and Gorinstein, S. (2012). Anthocyanin content and the activities of polyphenol oxidase, peroxidase and phenylalanine ammonia-lyase in lettuce cultivars. *Int. J. Food Sci. Nutr.* 63, 45–48. doi: 10.3109/09637486.2011.595704
- Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971.
- Damerum, A., Selmes, S. L., Biggi, G. F., Clarkson, G. J., Rothwell, S. D., Truco, M. J., et al. (2015). Elucidating the genetic basis of antioxidant status in lettuce (*Lactuca sativa*). *Hortic. Res.* 2:15055. doi: 10.1038/hortres.2015.55
- Fiorani, F., and Schurr, U. (2013). Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 64, 267–291. doi: 10.1146/annurev-arplant-050312-120137
- Flood, P. J., Kruijzer, W., Schnabel, S. K., Schoor, R., Jalink, H., Snel, J. F., et al. (2016). Phenomics for photosynthesis, growth and reflectance in *Arabidopsis thaliana* reveals circadian and long-term fluctuations in heritability. *Plant Methods* 12: 14. doi: 10.1186/s13007-016-0113-y
- Frantz, J. M., Ritchie, G., Cometti, N. N., Robinson, J., and Bugbee, B. (2004). Exploring the limits of crop productivity: beyond the limits of tipburn in lettuce. *J. Am. Soc. Hortic. Sci.* 129, 331–338.
- 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
- Garrido, Y., Tudela, J. A., Marín, A., Mestre, T., Martínez, V., and Gil, M. I. (2014). Physiological, phytochemical and structural changes of multi-leaf lettuce caused by salt stress. *J. Sci. Food Agric.* 94, 1592–1599. doi: 10.1002/jsfa.6462
- Gazula, A., Kleinhenz, M. D., Scheerens, J. C., and Ling, P. P. (2007). Anthocyanin levels in nine lettuce (*Lactuca sativa*) cultivars: Influence of planting date and relations among analytic, instrumented, and visual assessments of color. *HortScience* 42, 232–238.
- Gazula, A., Kleinhenz, M. D., Streeter, J. G., and Miller, A. R. (2005). Temperature and cultivar effects on anthocyanin and chlorophyll b concentrations in three related Lollo Rosso lettuce cultivars. *HortScience* 40, 1731–1733.
- Grahn, C. M., Benedict, C., Thornton, T., and Miles, C. (2015). Production of baby-leaf salad greens in the spring and fall seasons of northwest Washington. *HortScience* 50, 1467–1471.
- Granier, C., Aguirrezabal, L., Chenu, K., Cookson, S. J., Dauzat, M., Hamard, P., et al. (2006). PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytol.* 169, 623–635. doi: 10.1111/j.1469-8137.2005.01609.x
- Hayashi, E., You, Y., Lewis, R., Calderon, M. C., Wan, G., and Still, D. W. (2012). Mapping QTL, epistasis and genotype × environment interaction of antioxidant activity, chlorophyll content and head formation in domesticated lettuce (*Lactuca sativa*). *Theor. Appl. Genet.* 124, 1487–1502. doi: 10.1007/s00122-012-1803-0
- Hayes, R. J., Galeano, H. C., Luo, Y., Anotonise, R., and Simko, I. (2014). Inheritance of decay of fresh-cut lettuce in a recombinant inbred line population from 'Salinas 88' × 'La Brillante'. *J. Am. Soc. Hortic. Sci.* 139, 388–398.
- Hoffmann, W. A., and Poorter, H. (2002). Avoiding bias in calculations of relative growth rate. *Ann. Bot.* 90, 37–42. doi: 10.1093/aob/mcf140

- Hou, W., Sun, A., Chen, H., Yang, F., Pan, J., and Guan, M. (2016). Effects of chilling and high temperatures on photosynthesis and chlorophyll fluorescence in leaves of watermelon seedlings. *Biol. Plant.* 60, 148–154. doi: 10.1007/s10535-015-0575-1
- Jansen, M., Gilmer, F., Biskup, B., Nagel, K. A., Rascher, U., Fischbach, A., et al. (2009). Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in *Arabidopsis thaliana* and other rosette plants. *Funct. Plant Biol.* 36, 902–914. doi: 10.1071/FP09095
- Joehanes, R., and Nelson, J. C. (2008). QGene 4.0, an extensible Java QTL-analysis platform. *Bioinformatics* 24, 2788–2789. doi: 10.1093/bioinformatics/btn523
- Marin, A., Ferreres, F., Barberà, G. G., and Gil, M. I. (2015). Weather variability influences color and phenolic content of pigmented baby leaf lettuces throughout the season. *J. Agric. Food Chem.* 63, 1673–1681. doi: 10.1021/acs.jafc.5b00120
- Maxwell, K., and Johnson, G. N. (2000). Chlorophyll fluorescence — a practical guide. *J. Exp. Bot.* 51, 659–668. doi: 10.1093/jexbot/51.345.659
- McHale, L. K., Truco, M. J., Kozik, A., Wroblewski, T., Ochoa, O. E., Lahre, K. A., et al. (2009). The genomic architecture of disease resistance in lettuce. *Theor. Appl. Genet.* 118, 565–580. doi: 10.1007/s00122-008-0921-1
- Merzlyak, M. N., Solovchenko, A. E., and Gitelson, A. A. (2003). Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. *Postharvest Biol. Technol.* 27, 197–211. doi: 10.1016/S0925-5214(02)00066-2
- Munns, R., James, R. A., Sirault, X. R., Furbank, R. T., and Jones, H. G. (2010). New phenotyping methods for screening wheat and barley for beneficial responses to water deficit. *J. Exp. Bot.* 61, 3499–3507. doi: 10.1093/jxb/erq199
- 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
- Murata, N., Takahashi, S., Nishiyama, Y., and Allakhverdiev, S. I. (2007). Photoinhibition of photosystem II under environmental stress. *Biochim. Biophys. Acta* 1767, 414–421. doi: 10.1016/j.bbabi.2006.11.019
- Nicaise, V., German-Retana, S., Sanjuán, R., Dubrana, M.-P., Mazier, M., Maisonneuve, B., et al. (2003). The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the potyvirus *Lettuce mosaic virus*. *Plant Physiol.* 132, 1272–1282. doi: 10.1104/pp.102.017855
- Oh, M.-M., Trick, H. N., and Rajashekhar, C. (2009). Secondary metabolism and antioxidants are involved in environmental adaptation and stress tolerance in lettuce. *J. Plant Physiol.* 166, 180–191. doi: 10.1016/j.jplph.2008.04.015
- Park, Y. G., Park, J. E., Hwang, S. J., and Jeong, B. R. (2012). Light source and CO<sub>2</sub> concentration affect growth and anthocyanin content of lettuce under controlled environment. *Hortic. Environ. Biotechnol.* 53, 460–466. doi: 10.1007/s13580-012-0821-9
- Qin, L., Guo, S., Ai, W., Tang, Y., Cheng, Q., and Chen, G. (2013). Effect of salt stress on growth and physiology in amaranth and lettuce: implications for bioregenerative life support system. *Adv. Space Res.* 51, 476–482. doi: 10.1016/j.asr.2012.09.025
- Rajabbeigi, E., Eichholz, I., Beesk, N., Ulrichs, C., Kroh, L. W., Rohn, S., et al. (2013). Interaction of drought stress and UV-B radiation-impact on biomass production and flavonoid metabolism in lettuce (*Lactuca sativa* L.). *J. Appl. Bot. Food Qual.* 86, 190–197.
- Robinson, R., McCreight, J., and Ryder, E. (1983). The genes of lettuce and closely related species. *Plant Breed. Rev.* 1, 267–293. doi: 10.1007/978-1-4684-8896-8\_9
- Ryder, E. J. (1992). Lettuce genetics: inheritance, linkage, and epistasis. *J. Am. Soc. Hortic. Sci.* 117, 504–507.
- Seginer, I., Shina, G., Albright, L. D., and Marsh, L. S. (1991). Optimal temperature setpoints for greenhouse lettuce. *J. Agric. Eng. Res.* 49, 209–226. doi: 10.1016/0021-8634(91)80040-L
- Shannon, M., and Grieve, C. (1999). Tolerance of vegetable crops to salinity. *Sci. Hortic.* 78, 5–38. doi: 10.1016/S0304-4238(98)00189-7
- Shannon, M., McCreight, J., and Draper, J. (1983). Screening tests for salt tolerance in lettuce. *J. Am. Soc. Hortic. Sci.* 108, 225–230.
- Simko, I., Hayes, R. J., Mou, B., and McCreight, J. D. (2014). “Lettuce and Spinach. Pages 53–86,” in *Yield Gains in Major U.S. Field Crops*, eds S. Smith, B. Diers, J. Specht, and B. Carver (Madison, WI: American Society of Agronomy, Inc.).
- Simko, I., Jimenez-Berni, J. A., and Furbank, R. T. (2015a). Detection of decay in fresh-cut lettuce using hyperspectral imaging and chlorophyll fluorescence imaging. *Postharvest Biol. Technol.* 106, 44–52. doi: 10.1016/j.postharvbio.2015.04.007
- Simko, I., Jimenez-Berni, J. A., and Sirault, X. R. (2017). Phenomic approaches and tools for phytopathologists. *Phytopathology* 107, 6–17. doi: 10.1094/phyto-02-16-0082-rvw
- Simko, I., Ochoa, O. E., Pel, M. A., Tsuchida, C., Font I Forcada, C., Hayes, R. J., et al. (2015b). Resistance to downy mildew in lettuce ‘La Brillante’ is conferred by *Dm50* gene and multiple QTL. *Phytopathology* 105, 1220–1228. doi: 10.1094/PHYTO-02-15-0057-R
- Taubenhaus, J. J., Ezekiel, W. N., and Neblette, C. B. (1929). Airplane photography in the study of cotton root rot. *Phytopathology* 19, 1025–1029.
- Thompson, H. C., Langhans, R. W., Both, A.-J., and Albright, L. D. (1998). Shoot and root temperature effects on lettuce growth in a floating hydroponic system. *J. Am. Soc. Hortic. Sci.* 123, 361–364.
- Tsormpatsidis, E., Henbest, R. G. C., Battey, N. H., and Hadley, P. (2010). The influence of ultraviolet radiation on growth, photosynthesis and phenolic levels of green and red lettuce: potential for exploiting effects of ultraviolet radiation in a production system. *Ann. Appl. Biol.* 156, 357–366. doi: 10.1111/j.1744-7348.2010.00393.x
- Tsormpatsidis, E., Henbest, R. G. C., Davis, F. J., Battey, N. H., Hadley, P., and Wagstaffe, A. (2008). UV irradiance as a major influence on growth, development and secondary products of commercial importance in Lollo Rosso lettuce ‘Revolution’ grown under polyethylene films. *Environ. Exp. Bot.* 63, 232–239. doi: 10.1016/j.envexpbot.2007.12.002
- White, J. W., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., et al. (2012). Field-based phenomics for plant genetics research. *Field Crops Res.* 133, 101–112. doi: 10.1016/j.fcr.2012.04.003
- Xu, C., and Mou, B. (2015). Evaluation of lettuce genotypes for salinity tolerance. *HortScience* 50, 1441–1446.
- Xue, L., and Yang, L. (2009). Deriving leaf chlorophyll content of green-leafy vegetables from hyperspectral reflectance. *ISPRS J. Photogramm. Remote Sens.* 64, 97–106. doi: 10.1016/j.isprsjprs.2008.06.002

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

Copyright © 2016 Simko, Hayes and Furbank. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Accurate Digitization of the Chlorophyll Distribution of Individual Rice Leaves Using Hyperspectral Imaging and an Integrated Image Analysis Pipeline

Hui Feng<sup>1,2</sup>, Guoxing Chen<sup>1</sup>, Lizhong Xiong<sup>1</sup>, Qian Liu<sup>2\*</sup> and Wanneng Yang<sup>1\*</sup>

<sup>1</sup> National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research, Agricultural Bioinformatics Key Laboratory of Hubei Province, and College of Engineering, Huazhong Agricultural University, Wuhan, China, <sup>2</sup> Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, and Key Laboratory of Ministry of Education for Biomedical Photonics, Department of Biomedical Engineering, Huazhong University of Science and Technology, Wuhan, China

## OPEN ACCESS

### Edited by:

John Doonan,  
Aberystwyth University,  
United Kingdom

### Reviewed by:

Ming Chen,  
Zhejiang University, China  
Yuhui Chen,  
Noble Research Institute, LLC.,  
United States

Richard John Webster,  
Aberystwyth University,  
United Kingdom

### \*Correspondence:

Qian Liu  
qianliu@mail.hust.edu.cn  
Wanneng Yang  
ywn@mail.hzau.edu.cn

### Specialty section:

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

Received: 30 November 2016

Accepted: 30 June 2017

Published: 25 July 2017

### Citation:

Feng H, Chen G, Xiong L, Liu Q and Yang W (2017) Accurate Digitization of the Chlorophyll Distribution of Individual Rice Leaves Using Hyperspectral Imaging and an Integrated Image Analysis Pipeline. *Front. Plant Sci.* 8:1238.  
doi: 10.3389/fpls.2017.01238

Pigments absorb light, transform it into energy, and provide reaction sites for photosynthesis; thus, the quantification of pigment distribution is vital to plant research. Traditional methods for the quantification of pigments are time-consuming and not suitable for the high-throughput digitization of rice pigment distribution. In this study, using a hyperspectral imaging system, we developed an integrated image analysis pipeline for automatically processing enormous amounts of hyperspectral data. We also built models for accurately quantifying 4 pigments (chlorophyll a, chlorophyll b, total chlorophyll and carotenoid) from rice leaves and determined the important bands (700–760 nm) associated with these pigments. At the tillering stage, the  $R^2$  values and mean absolute percentage errors of the models were 0.827–0.928 and 6.94–12.84%, respectively. The hyperspectral data and these models can be combined for digitizing the distribution of the chlorophyll with high resolution (0.11 mm/pixel). In summary, the integrated hyperspectral image analysis pipeline and selected models can be used to quantify the chlorophyll distribution in rice leaves. The use of this technique will benefit rice functional genomics and rice breeding.

**Keywords:** chlorophyll, hyperspectral imaging, image analysis pipeline, rice, phenomics

## INTRODUCTION

Rice is a staple food for a majority of the world population (Zhang, 2007). To meet the increasing demand due to natural disasters, human factors and the increasing world population on rice growth and yield, it is important to breed new rice varieties. In breeding research, the plant phenotype is essential for the evaluation of breeding results and gene functional analysis (Yang et al., 2013; Jasinski et al., 2016; Montagnoli et al., 2016; Negi et al., 2016). Plants contain pigments such as chlorophylls and carotenoids, which absorb light and provide energy for photosynthesis (Blackburn, 1998b). Chlorophyll is the major nitrogenous substance in higher plants and can be used for measuring plant growth (Kochubey and Kazantsev, 2007; Xue and Yang, 2009). The amount of chlorophyll present also determines a plant's photosynthetic capability, productivity and

yield potential (Carter, 1998; Xue and Yang, 2009). Thus, quantification of these pigments is vital for rice phenomics and rice research.

Traditional methods for the quantification of plant pigments, including spectrophotometry (Ergun et al., 2004), paper chromatography (Sporer et al., 1954), thin-layer chromatography (Sievers and Hynninen, 1977), and high-performance liquid chromatography (Yuan et al., 1997), are time-consuming, destructive and not suitable for high-throughput phenotyping. Plant pigments have different absorption peaks under different wavelengths, which means that their spectral reflectance characteristics can be used for evaluating or distinguishing pigments (Benedict and Swidler, 1961; Gamon and Surfus, 1999). Using spectroscopy and a portable chlorophyll meter, several spectral indices have been identified, which can be used for predicting plant chlorophyll content non-destructively. Blackburn et al. reported that the amount of canopy chlorophyll a and b is related to the original reflectance at 676 and 810 nm (Blackburn, 1998a; Blackburn and Pitman, 1999). Because derivatization can reduce the noise caused by illumination, soil background, and atmosphere (Collins, 1978; Baret et al., 1992), derivative spectra have also been found to be more sensitive to the chlorophyll content and more effective than the original spectral index (Le Maire et al., 2004). Moreover, spectral indices calculated by the red edge can provide a more accurate estimation of pigment content (Miller et al., 1990; Zou et al., 2011). Researchers have also found that the ratio and normalized spectral indices are closely related to the pigment content (Moss and Rock, 1991; Chappelle et al., 1992). Yi et al. used partial least square regression and found that the reflectance at 515–550 nm, 715 and 750 nm regions had high sensitivity for detecting the carotenoid contents of cotton (Yi et al., 2014). A recent study has used hyperspectral imagery to estimate the spatial variability in the chlorophyll and nitrogen content of rice, with an  $R^2$  of 0.69–0.82 (Moharana and Dutta, 2016). Researchers also used canopy reflectance to estimate the durum wheat nitrogen status, with an RMSECV of 19.3–36.3% (Thorp et al., 2017). Portable chlorophyll meters, such as CCM-200 (Chlorophyll Content Meter) and SPAD-502 (Soil and Plant Analyzer Development), are widely used for measuring the chlorophyll content; however, manually operated portable chlorophyll meters are relatively subjective, and spectroscopy techniques cannot be used to digitize the chlorophyll distribution in rice leaves. Moreover, we summarized the recent studies on chlorophyll or nitrogen quantification that used spectral techniques (Supplementary Table 1). These studies showed that few efforts have been made to handle massive amounts of hyperspectral data and automatically digitalize the chlorophyll distribution in individual rice leaves with high-resolution.

In this study, we developed an integrated image analysis pipeline that can process extremely large amounts of hyperspectral data and built models to accurately measure 4 rice leaf pigments: chlorophyll a, chlorophyll b, total chlorophyll, and carotenoid. Moreover, by combining the hyperspectral data and the selected models, the distribution of these 4 pigments can be digitized with high resolution.

## MATERIALS AND METHODS

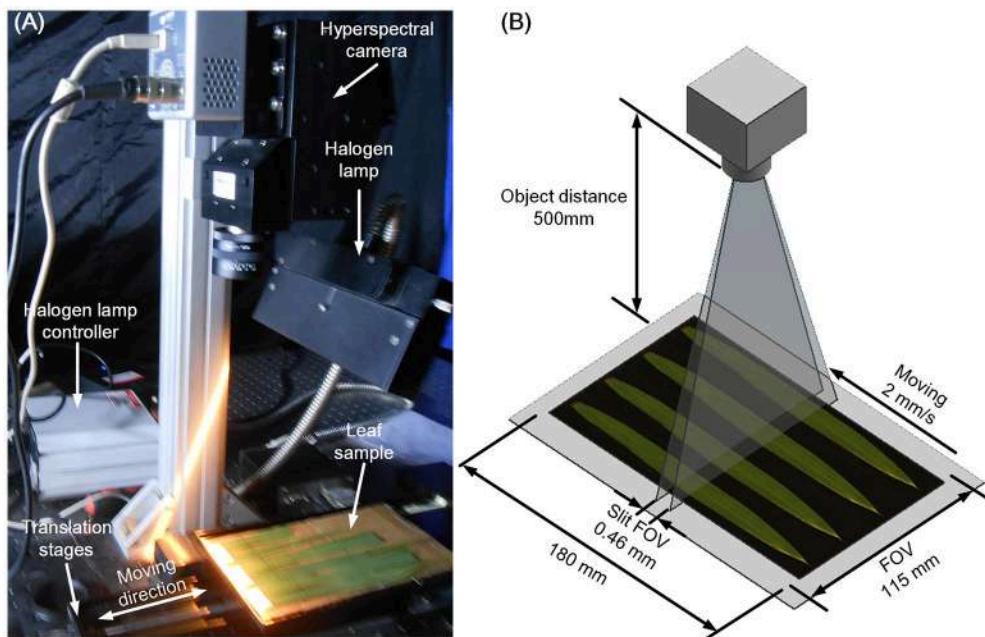
### Materials and Experimental Design

At the tillering stage, 10 rice accessions (*BLUE STICK*, *Chenwan3hao*, *PSBRC82*, *Manawthukha*, *Guantuibaihe*, *Xianggu*, *Wumanggaonuo*, *La110*, *Diantun502*, *TB154E-TB-2*, and *Ajaya*) were randomly selected from 533 rice core germplasm resources, and each accession was planted in 15 pots. The 15 pots were divided into 5 nitrogen application levels with 3 replicates: 0, 50, 100% (0.1 g *nitrogen per kg soil*), 150, and 200%. At the heading stage, 15 accessions (*RP2151-173-1-8*, *MR77* (*seberang*), *BASMATI 385*, *BLUE STICK*, *Chenwan3hao*, *PSBRC82*, *Manawthukha*, *Guantuibaihe*, *Xianggu*, *Wumanggaonuo*, *La110*, *Diantun502*, *TB154E-TB-2*, *Ajaya*, and *Bg90-2*) were randomly selected from 533 rice core germplasm resources, and 10 replicates of each accession were planted under the same nitrogen level (0.1 g of *nitrogen per kg of soil*). To test the relationship between the leaf nitrogen and hyperspectral indices, 90 accessions (seen in Supplementary Table 2) were randomly selected from 533 rice core germplasm resources and measured by an auto discrete analyzer (Smartchen 200, France), SPAD-502, and hyperspectral imaging. Detailed genetic information about these SNPs can be downloaded from the “RiceVarMap” database (<http://ricevarmap.ncgr.cn/>) (Narsai et al., 2013).

### Hyperspectral Imaging System and Hyperspectral Indices Extraction

Three leaves were selected from the main stem of each rice plant and scanned using the hyperspectral imaging system, which consisted of 4 major parts (Figure 1A): a halogen lamp, a translation stage, a hyperspectral camera (HyperspecTM VNIR, Headwall Photonics, USA), and a computer (OXPCO3, Dell, USA). To scan three leaves of one main stem simultaneously, the field of view was set at 115 × 180 mm. The major configurations of the hyperspectral imaging system are shown in Figure 1B, and the main parameters of the hyperspectral imaging system are shown in Table 1. The data were continuously stored as a binary data stream to acquire and store the original hyperspectral data as rapidly as possible. For each sample, the data size was 1.15 GBit.

After data acquisition, the binary data stream was reorganized to build 188 hyperspectral images under different wavelengths (Figures 2A–C). To process the massive number of images automatically, an integrated hyperspectral image analysis pipeline was developed (Figure 3). The detailed image analysis pipeline designed by LabVIEW is shown in Supplementary Figures 1–11, which included the following steps: (1) Open one binary data stream with the band interleaved by line format: The size of the hyperspectral data cube was 188 × 1,004 (*W*) × 1,637 (*H*). (2) The binary data stream was reorganized to build 188 hyperspectral images. (3) Image processing and ROI extracting: After image division, gray conversion, image binarization, horizontal open operation, removal of large areas, removal of noise, region growing, and extraction of the area of interest, a region of interest (ROI) was extracted for each leaf (Figures 2E–N). (4) ROI reflectance extracting: 188 original average reflectance indices



**FIGURE 1 |** Hyperspectral imaging system (A) and schematic diagram (B).

**TABLE 1 |** Main parameters of the hyperspectral imaging system.

Parameter	Value
Object distance	500 mm
Field of view (FOV)	115 × 180 mm
Slit FOV	0.46 mm
Scan speed	2 mm/s
Hyperspectral data of one plant	1.15 GBit
Frame number	1,637
Spectral resolution	3.2 nm
Spatial resolution	0.11 mm
Spectral range	400–1,000 nm
Focal length	35 mm
Band numbers	188
CCD resolution	1,004 × 1,002

(R) were obtained. (5) Derived indices extracting: These included 376 pseudo-absorption indices, 564 first derivative indices, 564 second derivative indices, 316,404 ratio indices, 316,404 normalized indices, 20 spectral indices, and 95 published indices. Finally, for each sample, 634,615 hyperspectral indices (in Table 2, among them, 20 spectral indices were shown in Supplementary Table 5, 95 published indices were shown in Supplementary Table 6) were saved. (6) Pearson's correlation coefficient was calculated, and the max correlation coefficient was obtained. (7) The binary data stream was closed.

## Manual Measurement

After hyperspectral imaging system acquisition, the ROI of each leaf was immersed in a 95% ethanol solution. When all of the

pigments had been dissolved, a spectrophotometer (L3, INESA, China) was used to measure the absorbance values of the solution at different wavelengths (470, 649, and 665 nm, Figure 2O). Finally, the contents of 4 pigments, chlorophyll a, chlorophyll b, carotenoid, and total chlorophyll, were calculated according to Equations (1)–(4) (Arnon, 1949).

$$C_a = 13.95A_{665} - 6.88A_{649} \quad (1)$$

$$C_b = 24.96A_{649} - 7.32A_{665} \quad (2)$$

$$C_{xc} = \frac{1000A_{470} - 2.05C_a - 114.8C_b}{245} \quad (3)$$

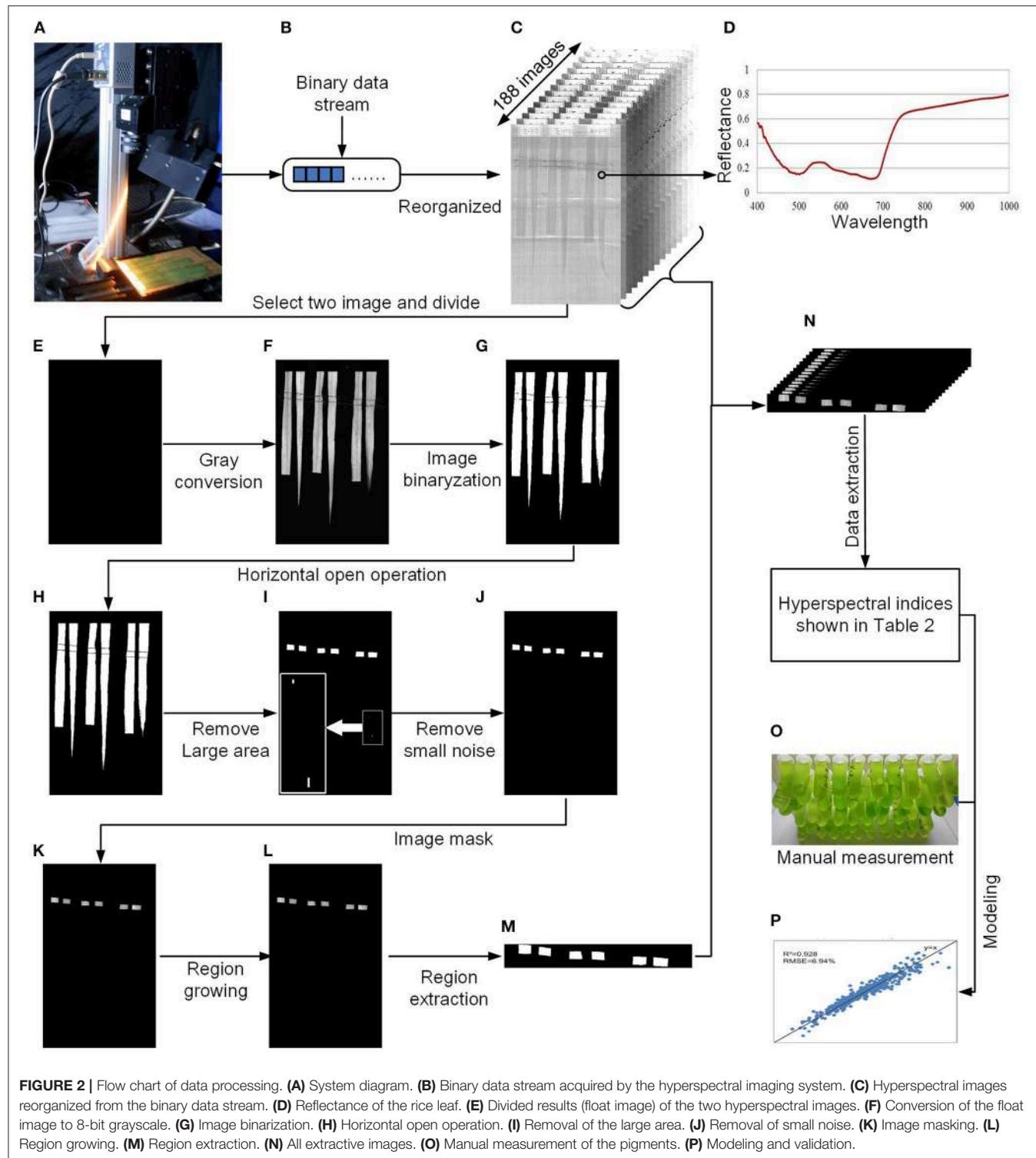
$$C = C_a + C_b \quad (4)$$

$C_a$  is the chlorophyll a content,  $C_b$  is the chlorophyll b content,  $C_{xc}$  is the carotenoid content, and  $C$  is the total chlorophyll content.  $A_{665}$ ,  $A_{649}$ , and  $A_{470}$  represent the absorbance at 665, 649, and 470 nm, respectively.

The distribution of the pigments at the two stages of plant growth is shown in Supplementary Table 3. For instance, at the tillering stage, the chlorophyll a content ranged from 61.24 to 573.63 mg/m<sup>2</sup>. The average value, the standard deviation, and the variable coefficient were 294.35 mg/m<sup>2</sup>, 92.19 mg/m<sup>2</sup>, and 31.32%, respectively. The correlation coefficients ( $r$ ) between the pigments for the two stages were all above 0.88 (Supplementary Table 4), demonstrating that the concentrations of the various pigments were highly correlated.

## Data Analysis and Modeling

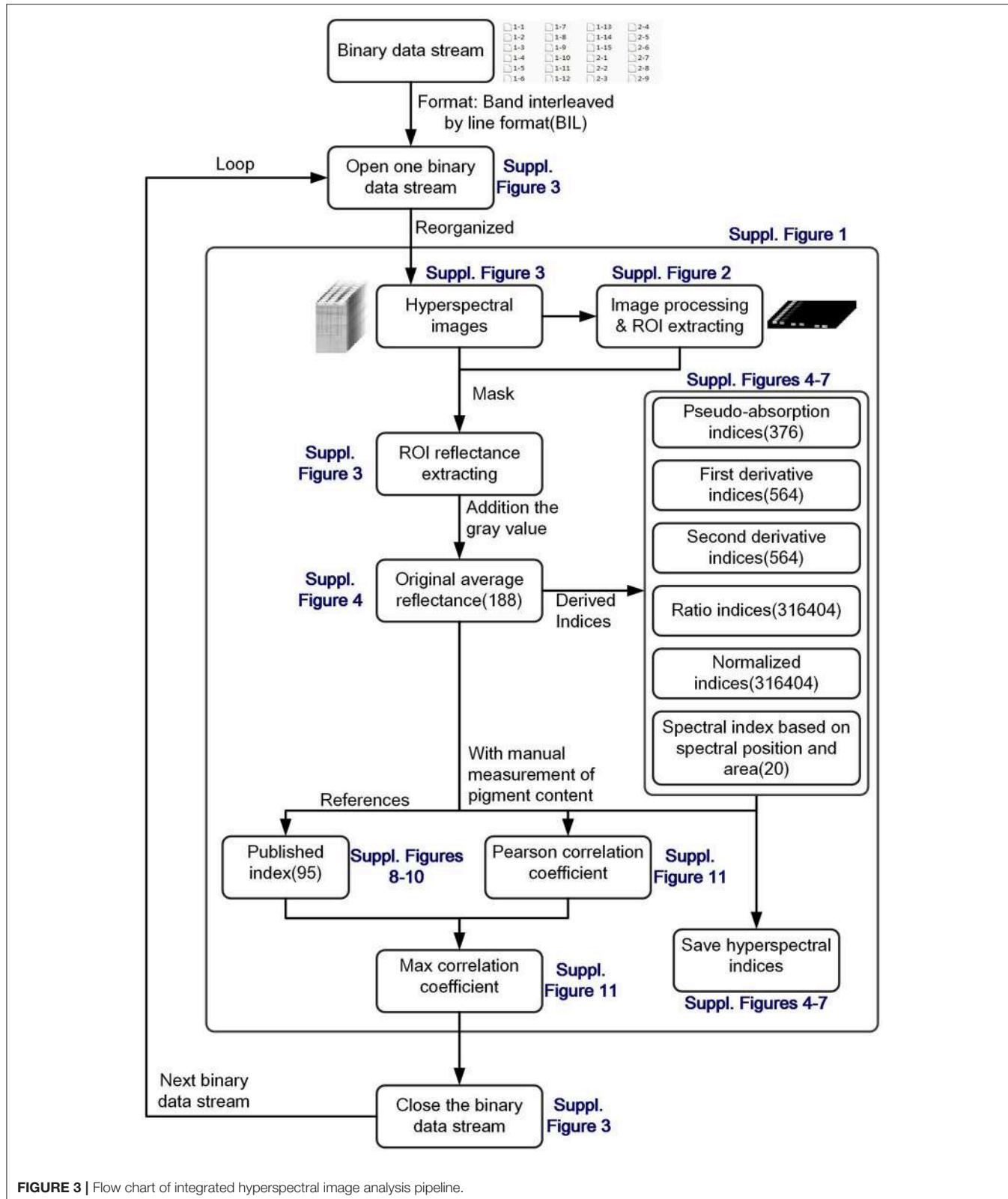
To determine the specific bands that are highly correlated with chlorophyll a, we calculated all of the correlation coefficients between 634,615 spectral indices and 4 pigments. The calculation



**FIGURE 2 |** Flow chart of data processing. **(A)** System diagram. **(B)** Binary data stream acquired by the hyperspectral imaging system. **(C)** Hyperspectral images reorganized from the binary data stream. **(D)** Reflectance of the rice leaf. **(E)** Divided results (float image) of the two hyperspectral images. **(F)** Conversion of the float image to 8-bit grayscale. **(G)** Image binarization. **(H)** Horizontal open operation. **(I)** Removal of the large area. **(J)** Removal of small noise. **(K)** Image masking. **(L)** Region growing. **(M)** Region extraction. **(N)** All extractive images. **(O)** Manual measurement of the pigments. **(P)** Modeling and validation.

of correlation coefficients was programmed using LabVIEW 8.6 (National Instruments, Inc., USA). The hot bands were found using the heat maps of the correlation coefficients, which were drawn using HemI software (Deng et al., 2014). After all of the indices were obtained, the best index with the highest

$r$  was identified and used to build 5 models (linear, power, exponential, logarithmic, and quadratic models). The statistical analyses of the 5 models (linear, power, exponential, logarithm, and quadratic model) for 4 pigments and cross-validation were implemented with LabVIEW 8.6 (National Instruments, Inc.,



**FIGURE 3 |** Flow chart of integrated hyperspectral image analysis pipeline.

USA). To evaluate the model performance with primary indices or multiple variables, stepwise regression analysis (SRA) was conducted using SPSS software (Statistical Product and Service

Solutions, Version 13.0, SPSS Inc., USA) (**Figure 2P**). Finally, the digitization of pigment distribution was performed using LabVIEW 8.6 (National Instruments, Inc., USA).

**TABLE 2 |** Table 2 Definition and calculation formulas of 634,615 hyperspectral indices.

Hyperspectral index	Number	Symbol*
Original average reflectance	188	$R_i$
Pseudo-absorption index	376	$\lg(R_i)$ , $\lg(1/R_i)$
First derivative index	564	$d(R_i)$ , $d(\lg(R_i))$ , $d(\lg(1/R_i))$
Second derivative index	564	$dd(R_i)$ , $dd(\lg(R_i))$ , $dd(\lg(1/R_i))$
Ratio index	316404	$\frac{R_i}{R_j}$ , $\lg(R_i)$ , $\lg(1/R_i)$ , $d(R_i)$ , $d(\lg(R_i))$ , $d(\lg(1/R_i))$ , $dd(R_i)$ , $dd(\lg(R_i))$ , $dd(\lg(1/R_i))$ , $R_i - R_j$ , $\lg(R_i) - \lg(R_j)$ , $\lg(1/R_i) - \lg(1/R_j)$ , $d(R_i) - d(R_j)$ , $d(\lg(R_i)) - d(\lg(R_j))$ , $d(\lg(1/R_i)) - d(\lg(1/R_j))$ , $R_i + R_j$ , $\lg(R_i) + \lg(R_j)$ , $\lg(1/R_i) + \lg(1/R_j)$ , $d(R_i) + d(R_j)$ , $d(\lg(R_i)) + d(\lg(R_j))$ , $d(\lg(1/R_i)) + d(\lg(1/R_j))$ , $dd(R_i) - dd(R_j)$ , $dd(\lg(R_i)) - dd(\lg(R_j))$ , $dd(\lg(1/R_i)) - dd(\lg(1/R_j))$ , $\frac{dd(\lg(R_i)) + dd(\lg(R_j))}{dd(\lg(R_i)) + dd(\lg(R_j))}$ , $\frac{dd(\lg(1/R_i)) + dd(\lg(1/R_j))}{dd(\lg(1/R_i)) + dd(\lg(1/R_j))}$
Normalized index	316404	
Spectral index based on spectral position and area	20	<b>Supplementary Table 5</b>
Published index	95	<b>Supplementary Table 6</b>

\* $0 \leq i \leq 187$ ,  $0 \leq j \leq 187$ .

## RESULTS AND DISCUSSION

### The Relationship between Chlorophyll a Concentration and Hyperspectral Indices

The number of total indices was too large to handle (634,615 indices for each sample); thus, to decrease the number of redundant indices, we first determined the relationship between the chlorophyll content and all the hyperspectral indices. Because the pigments were highly correlated with each other (**Supplementary Table 4**), we used chlorophyll a as an example to define the relationship between the pigments and the hyperspectral indices. In the 500–700 nm region (**Figure 4A**), the reflectance  $R$  was inversely correlated with the chlorophyll a content, indicating that the higher the reflectance was, the lower the chlorophyll a content was. This occurred because leaves with high chlorophyll content absorbed more light, causing the reflectance to decrease (**Figure 2D**). From **Figures 4A–F**, we found that compared with a logarithmic transformation, the use of derivative transformations such as  $dR$ ,  $ddR$ ,  $d(\lg(1/R))$ , and  $dd(\lg(1/R))$  could provide more abundant hyperspectral information.

**Figures 4G–I** show the correlation between the ratio index as defined in **Table 2** and chlorophyll a, and **Figures 4J–L** show the correlation between the normalized index (also defined in **Table 2**) and chlorophyll a. Each point on the heat map represents the correlation coefficient between a hyperspectral index and the chlorophyll a level. The correlation coefficients for other indices and the chlorophyll a level are shown in **Supplementary Figures 12, 13**. When  $R_i$  and  $R_j$  were both in the 500–750 nm region, the correlation coefficient was high, sometimes even close to 1. Thus, we can infer that useful information for estimating chlorophyll a can be obtained in the wavelength range 500–750 nm.

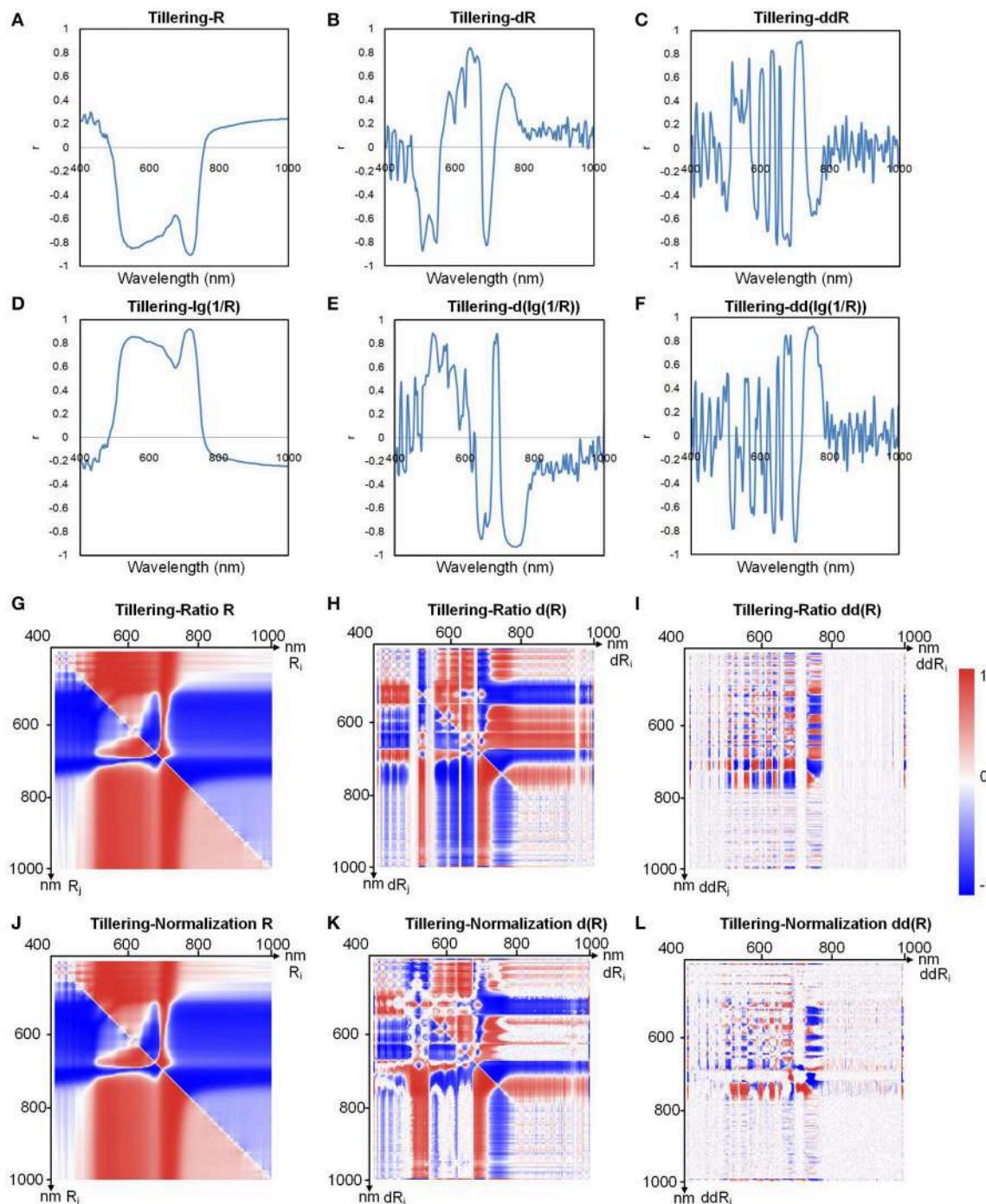
By comparing the data shown in **Figures 4G–I**, we found that for the ratio indices, the correlation between the derivative indices and chlorophyll a decreased, and the original hyperspectral index (average reflectance,  $R$ ) showed better correlation with chlorophyll a. As illustrated in **Figures 4J–L**, the same results could be obtained for the normalized indices. Thus, to decrease the redundant indices, primary indices, including the original average reflectance ( $R_i$ ), first derivative

index ( $d(R_i)$ ), second derivative index ( $dd(R_i)$ ), ratio index ( $R_i/R_j$ ), and normalized index ( $(R_i - R_j)/(R_i + R_j)$ ), were used for the subsequent modeling and prediction of chlorophyll levels. A combined heat map obtained by adding together all of the heat maps of ratio and normalization coefficients is shown in **Figure 5**. From this, we found that the region of the highest correlation was located between 700 and 760 nm. If only the primary indices in the 700–760 nm region were used, the number of indices would decrease from 634,615 to 483.

### Linear Modeling with a Single Variable

After all of the indices were calculated, the hyperspectral indices with the highest correlation coefficients ( $r$ ) of the pigments were selected for the modeling step, as shown in **Table 3**. The single-variable model for 4 pigments at the tillering and heading stages is shown in **Table 4**, which show that  $R^2$  ranged from 0.654 to 0.928, and the mean absolute percentage error (MAPE) was 6.94–12.84%. The scatter plots and the distribution of the relative error are shown in **Figure 6** and **Supplementary Figure 14**, respectively, which show the points to be evenly distributed around the line  $y = x$  and that most of the relative error within the range  $\pm 10\%$ . A 5-fold cross-validation of the single variable model for the 4 pigments at the two stages is shown in **Table 4**, which shows the ranges of  $R^2$  and MAPE as 0.671–0.930 and 7.49–13.02%, respectively.

To evaluate the model's robustness, we evaluated the relationship between  $\lg(R_{715})/\lg(R_{500})$  and the chlorophyll a level for different accessions grown under different nitrogen regimes at the tillering stage (**Figure 7**). The model was not sensitive to accession or the nitrogen application level. **Figure 7B** shows that the amount of chlorophyll a increased with increase in the nitrogen application level. Moreover, we also compared the best model for the 4 pigments in this study with the published indices, as shown in **Table 3** and **Supplementary Table 6**. The correlation between the pigments and the indices selected in this study (0.81–0.96) was higher than the correlation between the pigments and the published index with the highest  $r$  (0.67–0.92). On the other hand, all of the published indices with high  $r$  values were based on at least one wavelength in the range of 700–760 nm, implying that this range (700–760 nm) is important for the quantification of leaf chlorophyll.

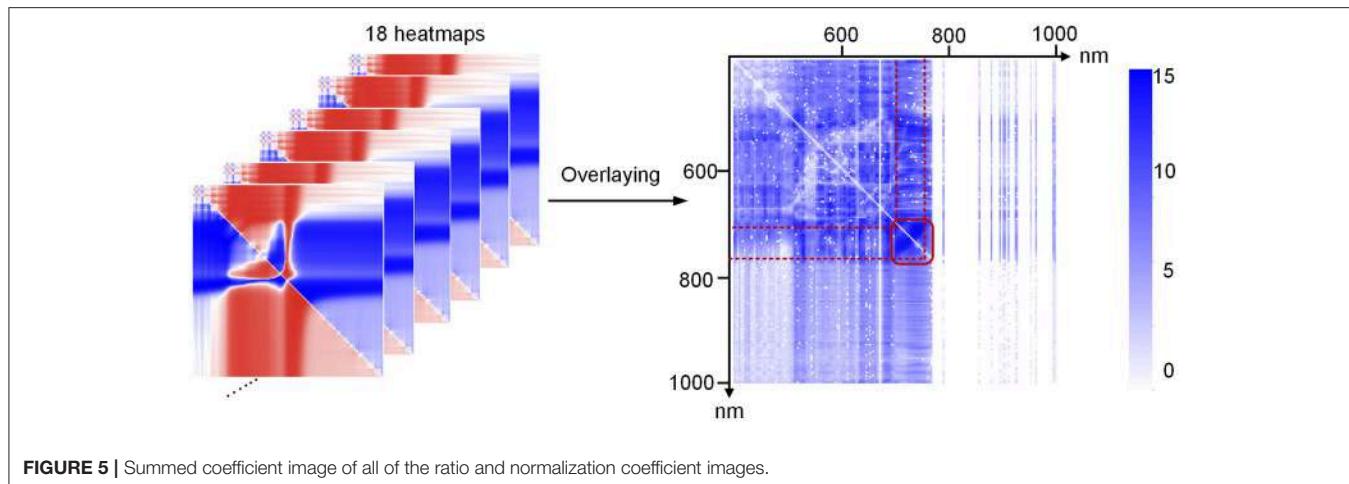


**FIGURE 4 |** Correlation coefficients between chlorophyll a and **(A)** R, **(B)** dR, **(C)** ddR, **(D)** lg(1/R), **(E)** d(lg(1/R)), **(F)** dd(lg(1/R)), **(G)** ratio R, **(H)** ratio dR, **(I)** ratio ddR, **(J)** normalization R, **(K)** normalization dR, and **(L)** normalization ddR at the tillering stage.

## Comparison of Linear and Non-linear Models

To determine the best model for determination of chlorophyll a levels, 5 models, including the linear, power, exponential, logarithmic, and quadratic models, were compared. The results are shown in **Table 5**. We found that the linear model had

the highest  $R^2$  (0.928) and lowest MAPE (6.94%). Based on the relative robustness of the models, the linear model was selected as the final model for the quantification of chlorophyll. The results also indicate that the best relationship between the chlorophyll content and the index value was linear in our study.



**FIGURE 5 |** Summed coefficient image of all of the ratio and normalization coefficient images.

**TABLE 3 |** Hyperspectral indices that displayed the highest  $r$  values selected from all indices or primary indices for the 4 pigments and comparison with published indices.

Stage	Pigment	Best index selected in all indices	$r$	Best index selected in primary indices	$r$	Published indices with the highest $r$	$r$
Tillering stage	Chlorophyll a	$\lg(R_{715})$ $\lg(R_{500})$	0.963	$R_{714}$	0.919	$D_{705}$ (Zarco et al., 2002) $D_{702}$	0.919
	Chlorophyll b	$\lg(R_{715})$ $\lg(R_{660})$	0.909	$R_{721}$	0.886	$R_{657}$ (Chappelle et al., 1992) $R_{700}$	0.876
	Total chlorophyll	$\lg(R_{715})$ $\lg(R_{500})$	0.957	$R_{718}$	0.920	$R_{657}$ (Chappelle et al., 1992) $R_{700}$	0.915
	Carotenoid	$\lg(R_{718})$ $\lg(R_{450})$	0.914	$dd(R_{724})$	0.848	$R_{657}$ (Chappelle et al., 1992) $R_{700}$	0.852
Heading stage	Chlorophyll a	$d(R_{997}) - d(R_{747})$ $d(R_{997}) + d(R_{747})$	0.873	$dd(R_{721})$	0.836	$R_{728} - R_{434}$ $R_{720} - R_{434}$	0.732
	Chlorophyll b	$d(R_{997}) - d(R_{728})$ $d(R_{997}) + d(R_{728})$	0.855	$R_{714}$	0.835	$R_{728} - R_{434}$ $R_{720} - R_{434}$	0.675
	Total chlorophyll	$d(R_{997}) - d(R_{747})$ $d(R_{997}) + d(R_{747})$	0.872	$R_{714}$	0.837	$R_{728} - R_{434}$ $R_{720} - R_{434}$	0.726
	Carotenoid	$d(\lg(1/R_{747})) - d(\lg(1/R_{792}))$ $d(\lg(1/R_{747})) + d(\lg(1/R_{792}))$	0.809	$dd(R_{721})$	0.782	$R_{728} - R_{434}$ $R_{720} - R_{434}$	0.668

**TABLE 4 |** Details of the single-variable models for the 4 pigments.

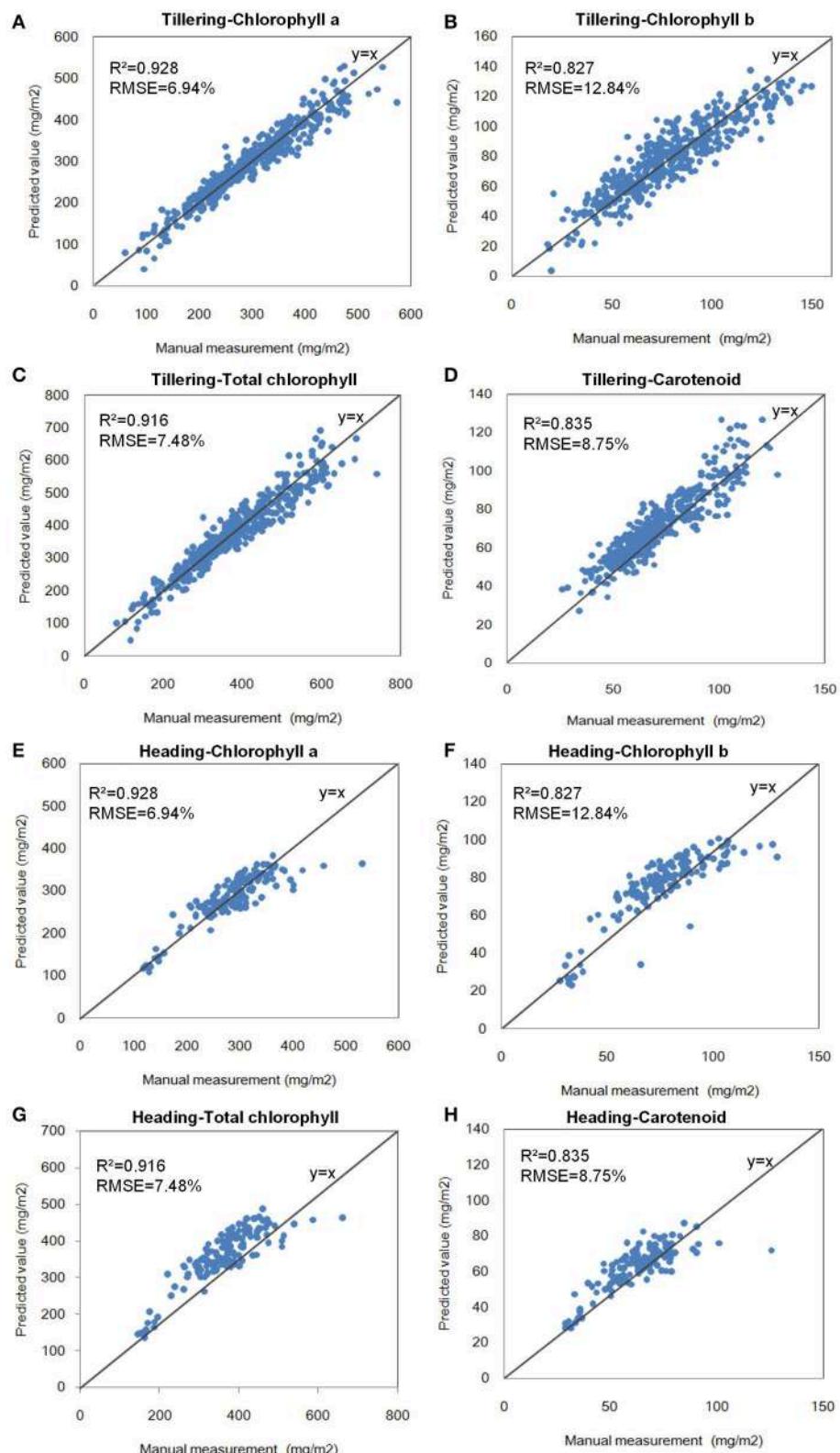
Stage	Chlorophyll	Single-variable model*	$R^2$	MAPE	RMSE (mg/m <sup>2</sup> )	5-fold cross validation			
						Modeling		Validation	
						$R^2$	MAPE	RMSE (mg/m <sup>2</sup> )	MAPE
Tillering stage	Chlorophyll a	$y = 1217.948x_1 - 301.306$	0.928	6.94%	24.73	0.930	6.93%	24.721	7.49%
	Chlorophyll b	$y = 557.723x_2 - 126.609$	0.827	12.84%	11.19	0.832	12.86%	11.184	13.02%
	Total chlorophyll	$y = 1596.104x_1 - 405.674$	0.916	7.48%	34.18	0.918	7.48%	34.179	7.53%
	Carotenoid	$y = 188.087x_3 - 22.582$	0.835	8.75%	7.76%	0.833	8.74%	7.760	8.91%
Heading stage	Chlorophyll a	$y = 7874.223x_4 - 8009.138$	0.761	8.25%	32.73	0.770	8.24%	32.703	8.46%
	Chlorophyll b	$y = 933.651x_5 - 1006.710$	0.731	10.93%	10.40	0.732	10.94%	10.384	11.24%
	Total chlorophyll	$y = 10159.684x_4 - 10340.279$	0.761	8.70%	42.32	0.768	8.68%	42.218	8.59%
	Carotenoid	$y = 175.113x_6 - 52.628$	0.654	9.14%	8.53	0.671	9.13%	8.528	9.29%

\* $x_1 = \log(R_{715}) / \log(R_{500})$ ,  $x_2 = \log(R_{715}) / \log(R_{660})$ ,  $x_3 = \log(R_{718}) / \log(R_{450})$ ,  $x_4 = (d(R_{997}) - d(R_{747})) / (d(R_{997}) + d(R_{747}))$ ,  $x_5 = (d(R_{997}) - d(R_{728})) / (d(R_{997}) + d(R_{728}))$ ,  $x_6 = (d(\lg(1/R_{747})) - d(\lg(1/R_{792}))) / (d(\lg(1/R_{747})) + d(\lg(1/R_{792})))$ .

## Comparison of Models with All Indices and Models with Primary Indices

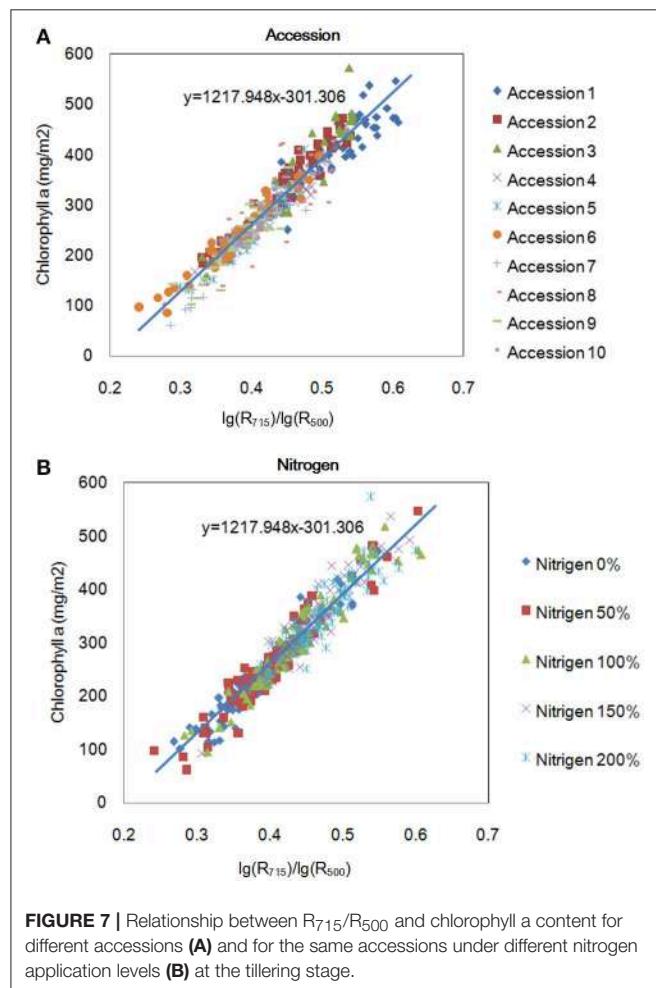
To compare the models that use all indices with those that use primary indices, we used 634,615 indices, and 483 primary indices for evaluating the model performance. The results

(Table 3) showed that the highest  $r$  of the models that used primary indices (0.782–0.920) was similar to the highest  $r$  of the models that used all indices (0.809–0.963), indicating that the models that use primary indices are sufficiently accurate for the quantification of the 4 pigments. If only the primary



**FIGURE 6 |** Scatter plots of the single-variable model of pigments at the two stages. **(A)** Chlorophyll a at the tillering stage. **(B)** Chlorophyll b at the tillering stage. **(C)** Total chlorophyll at the tillering stage. **(D)** Carotenoid at the tillering stage. **(E)** Chlorophyll a at the heading stage. **(F)** Chlorophyll b at the heading stage. **(G)** Total chlorophyll at the heading stage. **(H)** Carotenoid at the heading stage.

indices were extracted and analyzed, the volume of hyperspectral data decreased from hundreds of thousands to hundreds, which dramatically reduced the workload of data acquisition and data analysis. The results of this comparison are shown in **Table 3** and **Supplementary Table 7**.



**FIGURE 7 |** Relationship between  $R_{715}/R_{500}$  and chlorophyll a content for different accessions (A) and for the same accessions under different nitrogen application levels (B) at the tillering stage.

## Linear Modeling with Multi-Variables

We also evaluated the model performance using multi-variables. To facilitate the evaluation, only some primary indices, including  $R$ ,  $dR$ , and  $ddR$ , were used to build the model using a stepwise regression analysis. The results (**Supplementary Table 8**) showed that  $R^2$  and  $R_{adj}^2$  increased slightly and that MAPE and RMSE decreased slightly as the number of independent variables increased. The distribution of the relative error of the model using a stepwise regression analysis and multi-variables for chlorophyll a at the tillering stage is shown in **Supplementary Figure 15**, and 5-fold cross-validation of these models is shown in **Supplementary Table 8**.

## Digitization of Leaf Chlorophyll Distribution

After the best single-variable model was built, it was used to digitize the leaf chlorophyll distribution at a high resolution (0.11 mm/pixel), as shown in **Figure 8** (pseudo-color images). **Figures 8A–C** show the results obtained for one accession grown under different nitrogen application levels; with increasing nitrogen application, the chlorophyll a content increased dramatically. The chlorophyll a content of different accessions grown under the same nitrogen application level also varied (**Figures 8D–F**). **Figures 8A–F** show that for most samples, the chlorophyll concentration in the middle portion of the leaf was the highest, followed by the lower leaf and the upper leaf. Moreover, for the same leaf, the chlorophyll a content of the leaf vein was less than that of the leaf pulp, as shown in **Figure 8G**.

## Modeling Nitrogen with Hyperspectral Imaging

A recent study showed that  $R^2$  between the total chlorophyll content and leaf nitrogen content of Papaya plants (Castro et al., 2011) could reach 0.78, and hyperspectral reflectance measurements could reflect the canopy nitrogen content of winter wheat (Zhou et al., 2016). To test the correlation between the nitrogen and hyperspectral indices in rice, we measured 90 rice accessions, selected from 533 rice core germplasm resources, using an auto discrete analyzer (Smartchen 200, France), SPAD-502, and hyperspectral imaging. The correlation

**TABLE 5 |** Statistical summary of the 5 developed models for chlorophyll a estimation (sample size = 425)\*.

No.	Model classification	Model	Coefficients	$R^2$	MAPE	$SD_{APE}$
1	Linear	$y = a_0 + a_1 \times x$	$a_0 = -284.78$ $a_1 = 1351.04$	0.928	6.94%	7.86%
2	Power	$y = a_0 x^{a_1}$	$a_0 = 1469.65$ $a_1 = 1.92$	0.913	7.85%	10.16%
3	Exponential mode	$y = a_0 \times e^{a_1 \times x}$	$a_0 = 49.69$ $a_1 = 4.11$	0.887	9.65%	14.24%
4	Logarithmic	$y = a_0 \ln(a_1 x)$	$a_0 = 568.05$ $a_1 = 3.96$	0.911	8.23%	9.73%
5	Quadratic	$y = a_0 + a_1 \times x + a_2 \times x^2$	$a_0 = -305.10$ $a_1 = 1446.59$ $a_2 = -109.76$	0.922	7.54%	7.54%

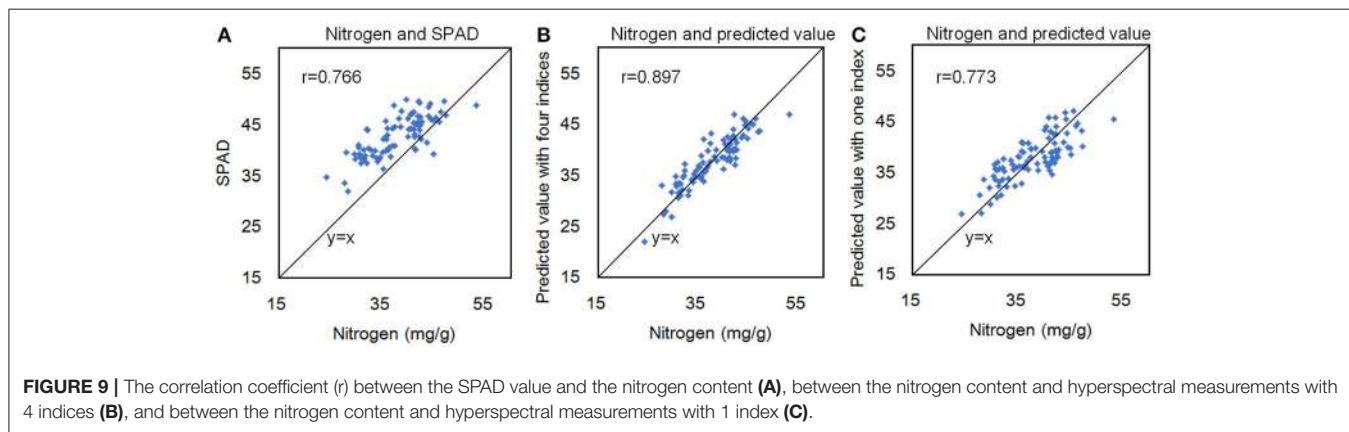
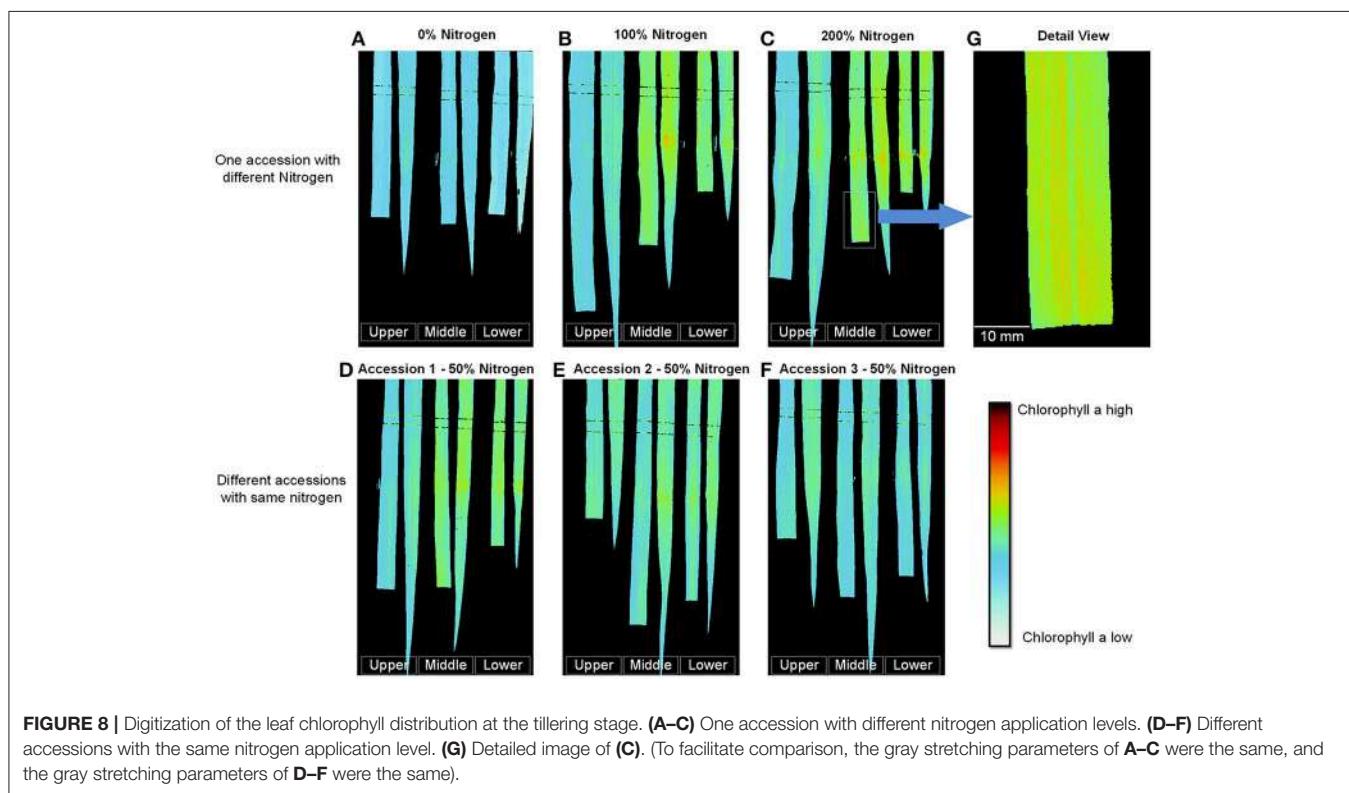
\*y is chlorophyll a, x is  $\frac{\lg(R_{715})}{\lg(R_{500})}$ .

coefficient ( $r$ ) between the SPAD value and the nitrogen content was 0.766 (Figure 9A), and  $r$  between the nitrogen content and hyperspectral measurements with 4 indices was 0.897 (Figure 9B). Moreover, only using one index, the  $r$  between the nitrogen content and hyperspectral measurements was 0.773 (Figure 9C). The results showed that nitrogen in rice plants could also be quantified using hyperspectral imaging.

## Comparison of Recent Related Studies for Quantifying Chlorophyll or Nitrogen Distribution

We compared the present research with recent related studies and found that several key wavelengths that reflect chlorophyll, such as cotton at 715 and 750 nm (Yi et al., 2014), winter

wheat at 705 nm and the red edge (Zhou et al., 2016), and grass at 690–750 (Tong and He, 2017), were co-determined. Moreover, the commonly adopted tools, such as ENVI and SAS, handled enormous amounts of hyperspectral data, particularly image analysis, with difficulty. To relieve the bottleneck, we developed an integrated image analysis pipeline in this study. With a single variable, the measuring accuracy of chlorophyll,  $R^2$ , ranged from 0.654 to 0.928. Moreover, due to using hyperspectral imaging in a higher resolution (0.11 mm/pixel), the distribution of leaf chlorophyll could be clearly visualized. The goal of this article was to quantify the chlorophylls in individual rice leaves, which should be tested and verified in the field in future. Combining the current field phenotyping tools, such as field phenotyping at the plot level (Andrade-Sanchez et al., 2014)



and movable imaging chambers in the field (Busemeyer et al., 2013), the integrated image analysis pipeline could be expanded to the field. Moreover, combined hyperspectral imaging with a novel sensor for structure imaging, such as a micro-CT (Mineyuki, 2014) and 3D laser scanning (Paulus et al., 2014), could also reconstruct the 3D distribution of chlorophyll in a high resolution.

## CONCLUSIONS

In this study, we used a hyperspectral imaging system to develop an integrated image analysis pipeline to handle extremely large amounts of hyperspectral data automatically. We also built models that could be used to accurately quantify 4 rice leaf pigments and identify the important spectral bands (700–760 nm) associated with these pigments. Moreover, by combining the hyperspectral data and these models, the distribution of chlorophyll could be digitized with high resolution (0.11 mm/pixel). In the future, the pipeline and selected models can potentially be applied to quantify the chlorophyll distribution in individual plants non-destructively. Evidence from related works shows that the image analysis pipeline combined with hyperspectral imaging could also be extended for co-determining wavelengths for quantifying chlorophyll in other crops.

## AUTHOR CONTRIBUTORS

HF and WY designed the research, performed the experiments, analyzed the data and wrote the manuscript. GC provided the rice samples and also performed experiments. LX and QL supervised the project, designed the research, and wrote the manuscript.

## ACKNOWLEDGMENTS

This work was supported by grants from the National Program on High Technology Development (2013AA102403), the National key research and development program (2016YFD0100101-18), the Scientific Conditions and Resources Research Program of Hubei Province of China (2015BCE044), the China Postdoctoral Science Foundation (2016M592345), and the Fundamental Research Funds for the Central Universities (2662017PY058).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.01238/full#supplementary-material>

**Supplementary Figure 1** | Flow chart of the program. The number represents the processing module of the following **Supplementary Figures 2–11**.

**Supplementary Figure 2** | Program for image processing and ROI extraction. **(A)** Reorganized two images from the binary data stream (1), **(B)** Segment the leaf part from the background with the two images (2), **(C)** Image processing for the ROI extraction from the leaf part (3), **(D)** ROI extraction and save (4).

**Supplementary Figure 3** | Program for ROI reflectance extraction. **(A)** The main program for ROI reflectance extraction of all samples (5), **(B)** Creating the Excel file (6), **(C)** Savin the Excel file (7), **(D)** The sub-program for ROI reflectance extraction

of single sample (8), **(E)** Applying the **Supplementary Figure 2** results to the current data processing (9).

**Supplementary Figure 4** | Program for calculation of the original average reflectance **(A)** (10) and the spectral index based on spectral position and area **(B)** (11).

**Supplementary Figure 5** | Program for calculation of the first and second derivatives (12).

**Supplementary Figure 6** | Program for calculation of the pseudo-absorption index. **(A)** The calculation of the IgR (13), **(B)** The calculation of the Ig(1/R) (14).

**Supplementary Figure 7** | Program for calculation of the ratio index. **(A)** The calculation of the 0–61 part of the ratio index (15), **(B)** The calculation of the 62–121 part of the ratio index (16), **(C)** The calculation of the 122–187 part of the ratio index (17). The programs for calculation of the normalized index are similar, except the  $R_i/R_j$  was changed into  $(R_i-R_j)/(R_i+R_j)$ .

**Supplementary Figure 8** | Program for calculation of the partial published index. **(A)** Published index 1–3 (18), **(B)** Published index 4–9 (19), **(C)** Published index 10–15 (20), **(D)** Published index 16–22 (21).

**Supplementary Figure 9** | Program for calculation of the partial published index. **(A)** Published index 23–31 (22), **(B)** Published index 32–35 (23), **(C)** Published index 36–43 (24), **(D)** Published index 44–56 (25).

**Supplementary Figure 10** | Program for calculation of the partial published index. **(A)** Published index 57–67 (26), **(B)** Published index 68–79 (27), **(C)** Published index 80–85 (28), **(D)** Published index 86–95 (29).

**Supplementary Figure 11** | Program for calculation of the correlation coefficient. **(A)** The program for calculation of the correlation coefficients between all the pigments and all the hyperspectral indices (30), **(B)** The program for combination the correlation coefficients of the ratio and normalized indices (31), **(C)** The program for building image with the correlation coefficients (32), **(D)** The program for finding the max correlation coefficient (33).

**Supplementary Figure 12** | Correlation coefficients between chlorophyll a and ratio Ig(R) **(A)**, normalization Ig(R) **(B)**, ratio d(lg(R)) **(C)**, normalization d(lg(R)) **(D)**, ratio dd(lg(R)) **(E)**, and normalization dd(lg(R)) **(F)** at the tillering stage.

**Supplementary Figure 13** | Correlation coefficients between chlorophyll a and ratio Ig(1/R) **(A)**, normalization Ig(1/R) **(B)**, ratio d(lg(1/R)) **(C)**, normalization d(lg(1/R)) **(D)**, ratio dd(lg(1/R)) **(E)**, and normalization dd(lg(1/R)) **(F)** at the tillering stage.

**Supplementary Figure 14** | Distribution of relative error of the single-variable models for the chlorophyll a **(A)**, chlorophyll b **(B)**, total chlorophyll **(C)**, and carotenoid **(D)** at the tillering stage. Distribution of relative error of the single-variable models for the chlorophyll a **(E)**, chlorophyll b **(F)**, total chlorophyll **(G)**, and carotenoid **(H)** at the heading stage.

**Supplementary Figure 15** | Distribution of relative error of the one independent variable **(A)**, two independent variables **(B)**, three independent variables **(C)**, four independent variables **(D)** models using stepwise regression analysis for chlorophyll a at the tillering stage.

**Supplementary Table 1** | The latest related papers of chlorophyll or nitrogen quantification with spectral methods.

**Supplementary Table 2** | Information about the 90 rice accessions and SPAD value.

**Supplementary Table 3** | Distribution of the pigments at the two stages.

**Supplementary Table 4** | Correlation coefficient ( $r$ ) between the pigments at the two stages.

**Supplementary Table 5** | Spectral index based on spectral position and area.

**Supplementary Table 6** | Published spectral indices\*.

**Supplementary Table 7** | Comparison of performance between models using all of the indices and models using original indices.

**Supplementary Table 8** | Details of the multiple-variable models for 4 pigments.

## REFERENCES

- Andrade-Sanchez, P., Gore, M. A., Heun, J. T., Thorp, K. R., Carmo-Silva, A. E., French, A. N., et al. (2014). Development and evaluation of a field-based high-throughput phenotyping platform. *Funct. Plant Biol.* 41, 68–79. doi: 10.1071/FP13126
- Arnon, D. I. (1949). Copper enzymes in isolated chloroplasts: polyphenoloxidase in *Beta vulgaris*. *Plant Physiol.* 24, 1–15. doi: 10.1104/pp.24.1.1
- Baret, F., Jacquemoud, S., Guyot, G., and Leprieur, C. (1992). Modeled analysis of the biophysical nature of spectral shifts and comparison with information content of broad bands. *Remote Sens. Environ.* 41, 133–142. doi: 10.1016/0034-4257(92)90073-S
- Benedict, H., and Swidler, R. (1961). Nondestructive method for estimating chlorophyll content of leaves. *Science* 133, 2015–2016. doi: 10.1126/science.133.3469.2015
- Blackburn, G. (1998a). Quantifying chlorophylls and carotenoids at leaf and canopy scales: an evaluation of some hyperspectral approaches. *Remote Sens. Environ.* 66, 273–285. doi: 10.1016/S0034-4257(98)00059-5
- Blackburn, G. (1998b). Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. *Int. J. Remote Sens.* 19, 657–675. doi: 10.1080/014311698215919
- Blackburn, G., and Pitman, J. (1999). Biophysical controls on the directional spectral reflectance properties of bracken (*Pteridium aquilinum*) canopies: results of a field experiment. *Int. J. Remote Sens.* 20, 2265–2282. doi: 10.1080/014311699212245
- Busemeyer, L., Ruckelshausen, A., Moeller, K., Melchinger, A. E., Alheit, K. V., Maurer, H. P., et al. (2013). Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. *Sci. Rep.* 3:2442. doi: 10.1038/srep02442
- Carter, G. A. (1998). Reflectance wavebands and indices for remote estimation of photosynthesis and stomatal conductance in pine canopies. *Remote Sens. Environ.* 63, 61–72. doi: 10.1016/S0034-4257(97)00110-7
- Castro, F. A., Campostrini, E., Netto, A. T., and Viana, L. H. (2011). Relationship between photochemical efficiency (JIP-Test Parameters) and portable chlorophyll meter readings in papaya plants. *Braz. J. Plant Physiol.* 23, 295–304. doi: 10.1590/S1677-04202011000400007
- Chappelle, E., Kim, M., and McMurtrey, J. (1992). Ratio analysis of reflectance spectra (RARS): an algorithm for the remote estimation of the concentrations of chlorophyll a, chlorophyll b, and carotenoids in soybean leaves. *Remote Sens. Environ.* 39, 239–247. doi: 10.1016/0034-4257(92)90089-3
- Collins, W. (1978). Remote sensing of crop type and maturity. *Engineering* 44, 43–55.
- Deng, W., Wang, Y., Liu, Z., Cheng, H., and Xue, Y. (2014). HemI: a toolkit for illustrating heatmaps. *PLoS ONE* 9:e111988. doi: 10.1371/journal.pone.0111988
- Ergun, E., Demirata, B., Gumus, G., and Apak, R. (2004). Simultaneous determination of chlorophyll a and chlorophyll b by derivative spectrophotometry. *Anal. Bioanal. Chem.* 379, 803–811. doi: 10.1007/s00216-004-2637-7
- Gamon, J., and Surfus, J. (1999). Assessing leaf pigment content and activity with a reflectometer. *New Phytol.* 143, 105–117. doi: 10.1046/j.1469-8137.1999.00424.x
- Jasinski, S., Lécureuil, A., Durandet, M., Bernard-Moulin, P., and Guerche, P. (2016). Arabidopsis seed content qtl mapping using high-throughput phenotyping: the assets of near infrared spectroscopy. *Front. Plant Sci.* 7:1682. doi: 10.3389/fpls.2016.01682
- Kochubey, S., and Kazantsev, T. (2007). Changes in the first derivatives of leaf reflectance spectra of various plants induced by variations of chlorophyll content. *J. Plant Physiol.* 164, 1648–1655. doi: 10.1016/j.jplph.2006.11.007
- Le Maire, G., Francois, C., and Dufrene, E. (2004). Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sens. Environ.* 89, 1–28. doi: 10.1016/j.rse.2003.09.004
- Miller, J., Hare, E., and Wu, J. (1990). Quantitative characterization of the vegetation red edge reflectance 1. An inverted-Gaussian reflectance model. *Remote Sens.* 11, 1755–1773. doi: 10.1080/01431169008955128
- Mineyuki, Y. (2014). 3D image analysis of plants using electron tomography and micro-CT. *Microscopy* 63(Suppl. 1), i8–i9. doi: 10.1093/jmicro/dfu036
- Moharana, S., and Dutta, S. (2016). Spatial variability of chlorophyll and nitrogen content of rice from hyperspectral imagery. *ISPRS J. Photogrammetry Remote Sens.* 122, 17–29. doi: 10.1016/j.isprsjprs.2016.09.002
- Montagnoli, A., Terzaghi, M., Fulgaro, N., Stoew, B., Wipenmyr, J., Ilver, D., et al. (2016). Non-destructive phenotypic analysis of early stage tree seedling growth using an automated stereovision imaging method. *Front. Plant Sci.* 7:1644. doi: 10.3389/fpls.2016.01644
- Moss, D. M., and Rock, B. N. (1991). “Analysis of red edge spectral characteristics and total chlorophyll values for red spruce (*Picea Rubens*) branch segments from Mt. Moosilauke, NH, USA,” in *Geoscience and Remote Sensing Symposium, (1991). IGARSS '91. Remote Sensing: Global Monitoring for Earth Management., International* (Durham, NC), 1529–1532.
- Narsai, R., Devenish, J., Castleden, I., Narsai, K., Xu, L., Shou, H., et al. (2013). Rice DB: an *Oryza* information portal linking annotation, subcellular location, function, expression, regulation, and evolutionary information for rice and *Arabidopsis*. *Plant J.* 76, 1057–1073. doi: 10.1111/tpj.12357
- Negi, M., Sanagala, R., Rai, V., and Jain, A. (2016). Deciphering phosphate deficiency-mediated temporal effects on different root traits in rice grown in a modified hydroponic system. *Front. Plant Sci.* 7:550. doi: 10.3389/fpls.2016.00550
- Paulus, S., Schumann, H., Kuhlmann, H., and Leon, J. (2014). High-precision laser scanning system for capturing 3D plant architecture and analysing growth of cereal plants. *Biosyst. Eng.* 121, 1–11. doi: 10.1016/j.biosystemseng.2014.01.010
- Sievers, G., and Hynninen, P. H. (1977). Thin-layer chromatography of chlorophylls and their derivatives on cellulose layers. *J. Chromatogr.* 134, 359–364. doi: 10.1016/S0021-9673(00)88534-9
- Sporer, A. H., Freed, S., and Sancier, K. M. (1954). Paper chromatography of chlorophylls. *Science* 119, 68–69. doi: 10.1126/science.119.3080.68
- Thorp, K. R., Wang, G., Bronson, K. F., Badaruddin, M., and Mon, J. (2017). Hyperspectral data mining to identify relevant canopy spectral features for estimating durum wheat growth, nitrogen status, and grain yield. *Comput. Electron. Agric.* 136, 1–12. doi: 10.1016/j.compag.2017.02.024
- Tong, A., and He, Y. (2017). Estimating and mapping chlorophyll content for a heterogeneous grassland: comparing prediction power of a suite of vegetation indices across scales between years. *ISPRS J. Photogrammetry Remote Sens.* 126, 146–167. doi: 10.1016/j.isprsjprs.2017.02.010
- Xue, L., and Yang, L. (2009). Deriving leaf chlorophyll content of green-leafy vegetables from hyperspectral reflectance. *ISPRS J. Photogram. Remote Sens.* 64, 97–106. doi: 10.1016/j.isprsjprs.2008.06.002
- Yang, W., Duan, L., Chen, G., Xiong, L., and Liu, Q. (2013). Plant phenomics and high-throughput phenotyping: accelerating rice functional genomics using multidisciplinary technologies. *Curr. Opin. Plant Biol.* 16, 180–187. doi: 10.1016/j.pbi.2013.03.005
- Yi, Q., Jiaapaer, G., Chen, J., Bao, A., and Wang, F. (2014). Different units of measurement of carotenoids estimation in cotton using hyperspectral indices and partial least square regression. *ISPRS J. Photogram. Remote Sens.* 91, 72–84. doi: 10.1016/j.isprsjprs.2014.01.004
- Yuan, J., Zhang, Y., Shi, X., Gong, X., and Chen, F. (1997). Simultaneous determination of carotenoids and chlorophylls in algae by high performance liquid chromatography. *Chin. J. Chromatogr.* 15, 133–135.

- Zarco, P., Miller, J., Mohammed, G., Noland, T., and Sampson, P. (2002). Vegetation stress detection through chlorophyll+ estimation and fluorescence effects on hyperspectral imagery. *J. Environ. Qual.* 31, 1433–1441. doi: 10.2134/jeq2002.1433
- Zhang, Q. F. (2007). Strategies for developing green super rice. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16402–16409. doi: 10.1073/pnas.0708013104
- Zhou, X., Huang, W., Kong, W., Ye, H., Luo, J., and Chen, P. (2016). Remote estimation of canopy nitrogen content in winter wheat using airborne hyperspectral reflectance measurements. *Adv. Space Res.* 58, 1627–1637. doi: 10.1016/j.asr.2016.06.034
- Zou, X., Shi, J., Hao, L., Zhao, J., Mao, H., Chen, Z., et al. (2011). *In vivo* noninvasive detection of chlorophyll distribution in cucumber (*Cucumis sativus*) leaves by indices based on hyperspectral imaging. *Anal. Chim. Acta* 706, 105–112. doi: 10.1016/j.aca.2011.08.026

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

The reviewer RJW and handling Editor declared their shared affiliation, and the handling Editor states that the process met the standards of a fair and objective review.

Copyright © 2017 Feng, Chen, Xiong, Liu and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# High-Throughput and Computational Study of Leaf Senescence through a Phenomic Approach

Jae IL Lyu<sup>1†</sup>, Seung Hee Baek<sup>2†</sup>, Sukjoon Jung<sup>2</sup>, Hyosub Chu<sup>1</sup>, Hong Gil Nam<sup>1,2</sup>, Jeongsik Kim<sup>1\*</sup> and Pyung Ok Lim<sup>2\*</sup>

<sup>1</sup> Center for Plant Aging Research, Institute for Basic Science, Daegu, South Korea, <sup>2</sup> Department of New Biology, Daegu Gyeongbuk Institute of Science and Technology, Daegu, South Korea

## OPEN ACCESS

### Edited by:

Marcos Egea-Cortines,  
Universidad Politécnica de Cartagena,  
Spain

### Reviewed by:

Biswapiya Biswas Misra,  
Texas Biomedical Research Institute,  
USA

Autar Krishen Mattoo,  
United States Department  
of Agriculture, USA

### \*Correspondence:

Jeongsik Kim  
yorus@postech.ac.kr  
Pyung Ok Lim  
polim@dgist.ac.kr

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

### Specialty section:

This article was submitted to  
Technical Advances in Plant Science,  
a section of the journal  
*Frontiers in Plant Science*

Received: 02 December 2016

Accepted: 09 February 2017

Published: 23 February 2017

### Citation:

Lyu JI, Baek SH, Jung S, Chu H, Nam HG, Kim J and Lim PO (2017) High-Throughput and Computational Study of Leaf Senescence through a Phenomic Approach. *Front. Plant Sci.* 8:250. doi: 10.3389/fpls.2017.00250

Leaf senescence is influenced by its life history, comprising a series of developmental and physiological experiences. Exploration of the biological principles underlying leaf lifespan and senescence requires a schema to trace leaf phenotypes, based on the interaction of genetic and environmental factors. We developed a new approach and concept that will facilitate systemic biological understanding of leaf lifespan and senescence, utilizing the phenome high-throughput investigator (PHI) with a single-leaf-basis phenotyping platform. Our pilot tests showed empirical evidence for the feasibility of PHI for quantitative measurement of leaf senescence responses and improved performance in order to dissect the progression of senescence triggered by different senescence-inducing factors as well as genetic mutations. Such an establishment enables new perspectives to be proposed, which will be challenged for enhancing our fundamental understanding on the complex process of leaf senescence. We further envision that integration of phenomic data with other multi-omics data obtained from transcriptomic, proteomic, and metabolic studies will enable us to address the underlying principles of senescence, passing through different layers of information from molecule to organism.

**Keywords:** time-series analysis, leaf senescence, lifespan, life history, high-throughput phenotyping, phenome, *Arabidopsis*

## INTRODUCTION

Leaf senescence, although a degenerative cellular process, is finely regulated and occurs by an intricate integration of multiple developmental and environmental signals. As a consequence, it is assumed that leaf senescence is a highly complex process involving the collective actions of thousands of genes and multiple pathways associated with aging, as well as their interplays, thereby complicating genetic and molecular analyses of senescence (Buchanan-Wollaston et al., 2005; Breeze et al., 2011; Schippers, 2015; Li et al., 2016; Liebsch and Keech, 2016;

**Abbreviations:** ABA, abscisic acid; BG1, blue/green index 1; Chlgreen, chlorophyll green index; DAE, days after emergence; Mac, Maccioni; MCARI, modified chlorophyll absorption in reflectance index; MES, methyl ester sulfonate; NDVI, normalized difference vegetation index; NORE, not oresara, ORE, oresara; PC, principal component; PCA, principal component analysis; PHI, phenome high-throughput investigator; PRI, photochemical reflectance index; PSNDc, pigment-specific normalized difference; QY\_max, maximum quantum yield of photosystem II; RGB, red-green-blue; SE, standard error; SIPI, structure intensive pigment index; SRWI2, simple ratio water index 2; SWIR, shortwave infrared; VNIR, visible and near-infrared.

Woo et al., 2016). Indeed, conventional molecular and genetic approaches in which one gene or mutant at a time is identified and characterized have been shown to be limited for revealing the global picture of molecular programs involved in leaf senescence (Guo, 2013; Kim et al., 2016). An additional pitfall experienced in previous studies is that leaf senescence has been achieved through a limited set of phenotypes in a narrow temporal window of senescence, mostly at an aged stage (Thomas, 2013).

The recent advances within omics technologies, including genomics, transcriptomics, proteomics, and metabolomics, have facilitated open innovation strategies toward systematic understanding of complex questions of plant growth, development, and responses to environments (Mochida and Shinozaki, 2011; Humplík et al., 2015; Rajasundaram and Selbig, 2016). However, the high-throughput phenotyping technologies for analyses of total physiological traits in plants lag behind our ability to investigate molecular omics, although measurement of physiological responses has been recognized as being essential to determine the implications of their reactions or responses (Furbank and Tester, 2011). One of the current technical challenges is therefore to advance the phenotyping system to allow numerous phenotypic analyses in an automated and high-throughput manner for a large set of plant populations under various conditions over time (Yang et al., 2013; Rahaman et al., 2015). These efforts are also being extended to address specific questions by establishing the phenotyping pipeline with specialized and sophisticated experimental designs and tools (Topp et al., 2013; Crowell et al., 2014; Slovak et al., 2014).

Toward this end, we are in the process of developing a cutting-edge plant phenotyping facility, the “phenome high-throughput investigator (PHI)”, which enables the evaluation of hundreds of traits through non-invasive approaches over time (Figure 1A). Our efforts further extend to establishing an operational pipeline for single-leaf-based quantitative phenotypic analyses that allow for the use of this efficient and powerful tool to study leaf senescence and its lifespan. Here, we present our current progress on the establishment of the PHI system and an evaluation of its performance. Moreover, we highlight potential strategies and tactics for phenome-level research toward understanding leaf senescence and lifespan in plants.

## **EXPERIMENTAL SCHEME: ESTABLISHMENT OF A SYSTEM FOR ASSESSMENT OF PHYSIOLOGICAL CHANGES IN *Arabidopsis* LEAVES DURING SENESCENCE**

Leaf senescence is the final stage of the life history of a leaf; thus, all previous experiences prior to the senescence stage can affect senescence and the lifespan process (Figure 1B-i). We assessed the morphological and physiological changes occurring during the entire leaf lifespan. In this regard, a quantitative phenotyping system on a single-leaf basis along with age information should be established. Measuring senescence parameters using a mixture of several leaves at a given age of a plant is not a valid analysis of leaf

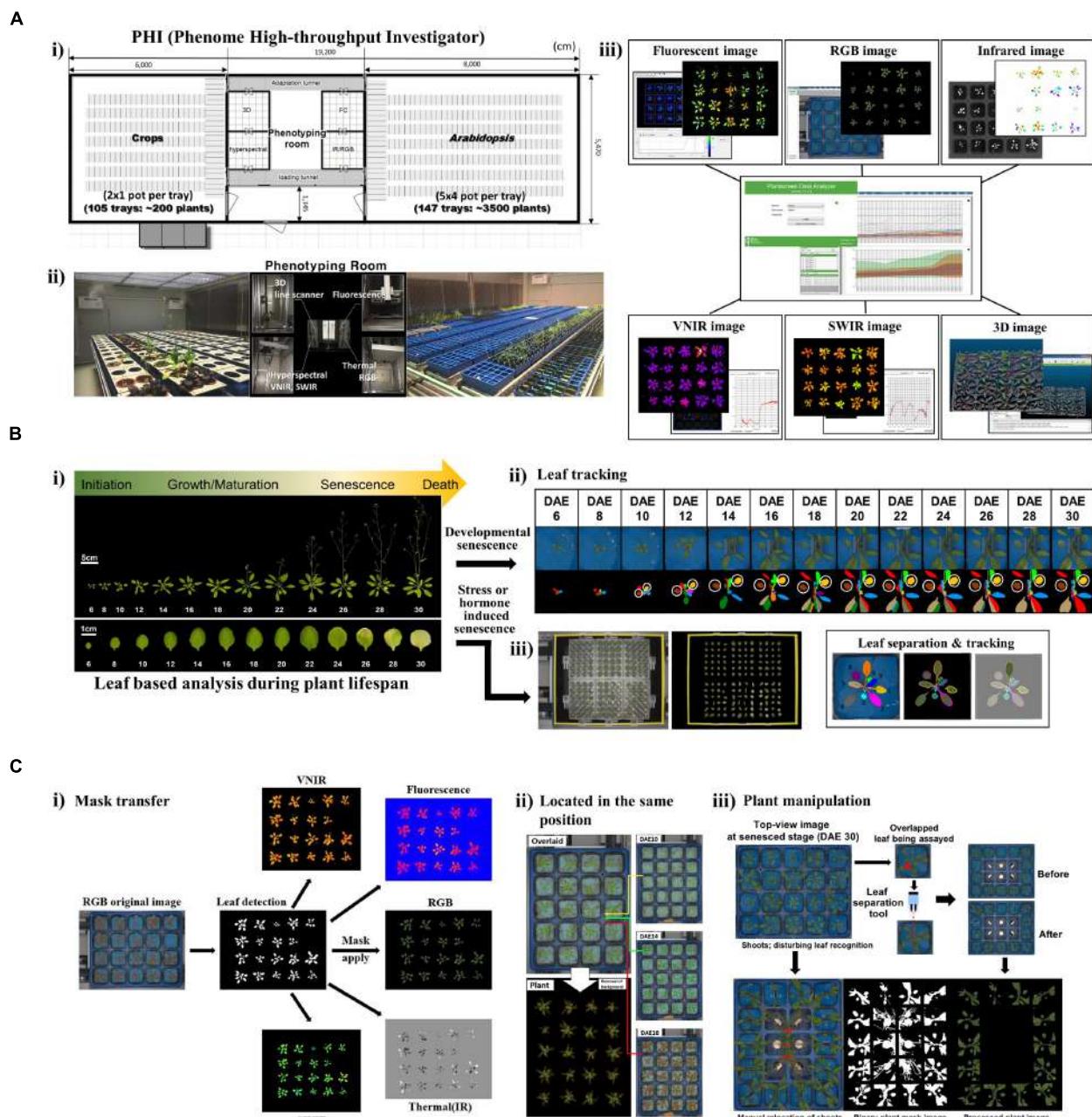
senescence and lifespan because individual leaves of a plant are of different ages (Zentgraf et al., 2004). Leaf developmental events such as senescence can also be modulated by external stresses or exogenous hormones; therefore, kinetic phenotyping analysis in leaves in response to these treatments is an additional valuable approach to dissect responses of leaf senescence (Lim et al., 2007; Schippers et al., 2015).

For the aforementioned purposes, we improved the PHI system to allow the assessment of the imaging-based phenotype through single-leaf-based analysis, either in intact plants or detached leaves in 24-well plates (Figures 1B-ii,iii). This leaf-based analysis requires a specialized experimental scheme and analytic modules beyond the configuration of a standardized phenotyping system, as detailed below. First, leaf segmentation and tracking in intact plants are necessary for chronological analyses in leaves. Second, a plant mask generated in a RGB image should be transferred and used for analyzing other images (Figure 1C-i). This is necessary when the plant signature is indistinguishable from the background soil or pot in a certain image (e.g., fluorescence images in fully senesced leaves). Third, plant trays should be located in the same position at each imaging unit. This could help to segment and track a single leaf of interest from the plants (Figure 1C-ii). Lastly, special manipulation is necessary to monitor phenotypes in leaves from the vegetative to senescing stages (Figure 1C-iii). Leaves are amenable to mature-stage phenomic analyses; however, the inaccessibility of old leaves covered by new leaves complicates the analyses of chronological events. Thus, leaf separation by placing blue clips on the petiole of the third and fourth leaves at DAE 14 is required for assays of later senescence. In addition, primary or axillary shoots should be directed to grow toward the central region of trays. On the basis of the aforementioned setup, high-throughput phenotypic traits occurring during leaf senescence in *Arabidopsis* would be assessed.

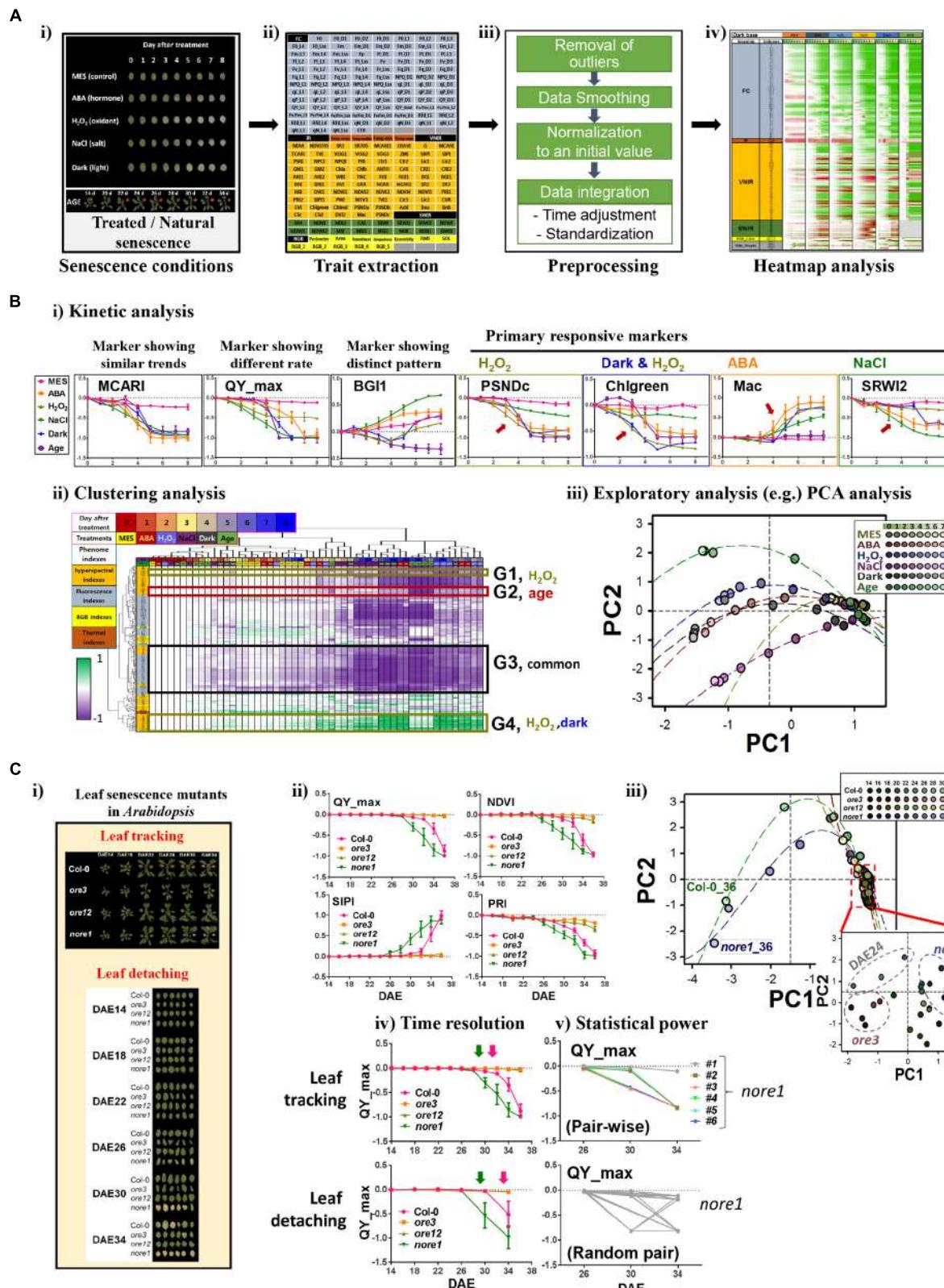
## **PROOF-OF-CONCEPT: PHENOMIC APPROACHES TO EVALUATE RESPONSES OF SENESCENCE IN *Arabidopsis* LEAVES**

Recent advances in non-invasive high-throughput imaging systems have allowed the monitoring of single to hundreds of plant traits to access plant physiological statuses from several thousands of plants in a kinetic manner (Chen et al., 2014; Rahaman et al., 2015; Cabrera-Bosquet et al., 2016). Leaf senescence occurs in an orderly and coordinated manner and involves changes in diverse metabolic processes, including catabolic processes of proteins, lipids, and carbohydrates, along with dismantlement of the photosynthetic apparatus (Lim et al., 2007; Watanabe et al., 2013). Thus, chronological analysis of various biological phenotypic traits is essential for understanding the processes of senescence.

Here, we explored the limited-scale feasibility of the PHI system for dissecting phenotypic responses during leaf senescence in *Arabidopsis*. System performance using PHI was



**FIGURE 1 | Scheme of the PHI and its application in leaf senescence research (A)** Overview of the PHI system. (i) The PHI is an automated high-throughput phenotyping system coupled with a controlled plant growth system. The image station is equipped with five non-invasive camera-based imaging units: fluorescent, RGB [top and side views (line scanner)], infrared, hyperspectral (VNIR; 400 to 1,000 nm, SWIR; 1,000 to 2,500 nm), and three-dimensional (3D) imaging. (ii) Two plant growth rooms supported the growth of different types of plants with an automatic and precise control of the environment. (iii) Screenshot of the PHI image analyzer software and image data. (B) Leaf-based analysis during plant lifespan. (i) Life cycles of leaf organs and entire plants in *Arabidopsis*. Shown are representative *Arabidopsis* plants and leaves when the age of the third rosette leaf is within DAE 6 to DAE 30. Responses of leaf senescence in *Arabidopsis* were assessed using developmental (ii) and stress- or hormone-induced (iii) senescence assays. (ii) Developmental senescence was monitored using leaf tracking and leaf separation. A representative plant with the third and fourth leaves of DAE 6 through DAE 30 in the pot where blue felt was placed on the top of the soil (Upper), and the pseudo-color image indicates correct recognition of individual leaves from plants (Lower). The third (brown) and fourth (yellow) leaves used for the senescence assay are marked with circles. (iii) Stress- or hormone-induced senescence was monitored using 24-well plates containing stress-inducing chemicals or hormones. (C) Leaf tracking for *in planta* senescence assay. (i) RGB image-based plant mask generation and its utilization in other image analyses. (ii) Location of a tray in the same position. Overlays of time-series images of a tray taken at DAE 10, 14, and 18 (Top, Middle, and Bottom of the Right panel). Overlaid original image (Left top) and plant image (Left bottom). (iii) Special manipulation of plants when assaying developmental senescence. Top-view RGB image at the senesced stage (DAE 30) encountered problems with the main or axillary shoots disturbing leaf recognition. Manual relocation of shoots to grow toward the central region of the part and leaf separation with blue clips was required.



**FIGURE 2 | Phenomic evaluation of leaf senescence responses triggered by various senescence-inducing factors as well as genetic mutations through PHI system. (A,B)** Phenomic evaluation of leaf senescence responses triggered by various senescence-inducing factors and its workflow. The workflow (Continued)

**FIGURE 2 | Continued**

involves trait extraction, preprocessing, heatmap analysis, and data mining. **(A)** Schematic workflow before data mining. **(i)** Picture of representative leaves incubated for indicated days under various senescence conditions such as 50  $\mu$ M ABA, 15 mM H<sub>2</sub>O<sub>2</sub>, 150 mM NaCl, darkness, or 3 mM MES as a control and the representative plants harboring the third rosette leaf at DAE 14 to 34. **(ii)** Trait extraction, after multimodal imaging, nearly 200 quantitative traits were extracted from the PHI data analyzer. **(iii)** Preprocessing, time-series numeric data were preprocessed by several steps as indicated. **(iv)** Heatmap analysis, temporal profiling of traits in senescence conditions was visualized using a heatmap for pattern comparison. **(B)** Data mining. Detailed trait and sample analyses were performed by several data mining techniques, including kinetic **(i)**, clustering **(ii)**, and exploratory PCA **(iii)** analyses. In kinetic analyses **(i)**, data represent mean  $\pm$  SE ( $n = 6$ ). Mean value for each genotype or for each age was used for clustering **(ii)** and PCA **(iii)** analysis. MCARI, modified chlorophyll absorption in reflectance index; QY\_max, maximum quantum yield of photosystem II; BG1, blue/green index 1; PSNDc, pigment-specific normalized difference; Chlgreen, chlorophyll green index; Mac, Macchioni; SRWI2, simple ratio water index 2. PC, Principal component. The percent variations explained by PC1 and PC2 were 67.2 and 10.1%, respectively. **(C)** PHI system-based phenomic evaluation of senescence responses in well-characterized leaf senescence mutants showing premature or delayed developmental senescence phenotypes. **(i)** Two senescence assays, leaf tracking and leaf detaching, are compared. Shown are the pictures representative of the third leaves of Col-0, ore3, ore12, and nore1 with different ages from DAE 14 to 34. **(ii)** Kinetic analysis of time-series data of QY\_max, NDVI, SIPI, and PRI in wild-type and senescence mutant leaves. NDVI, normalized difference vegetation index; SIPI, structure intensive pigment index; PRI, photochemical reflectance index. Data represent mean  $\pm$  SE ( $n = 6$ ). **(iii)** PCA-based analysis of senescence progression in wild-type and senescence mutant leaves with different ages. Mean value for each genotype or for each age was used for PCA analysis. The percent variations explained by PC1 and PC2 were 80.1 and 5.7%, respectively. **(iv,v)** Comparison of senescence assays in terms of time resolution **(iv)** and statistical power **(v)**. Data represent mean  $\pm$  SE ( $n = 6$ ).

first evaluated by monitoring the dynamics of phenotypic traits in *Arabidopsis* leaves treated with various senescence-inducing factors such as age, darkness, ABA, one of the stress-related phytohormones, as well as external stresses, including salinity (NaCl) and oxidative stress (H<sub>2</sub>O<sub>2</sub>; **Figure 2A-i**). These phenotypic traits (Supplementary Table S1 and **Figure 2A-ii**) include 208 indices that reflect multiple physiological statuses such as color and growth (12 indices; RGB), metabolic content, vital and vegetative status (77 indices; VNIR), water level or cellular components (16 indices; SWIR), chlorophyll-related photosynthetic performance (99 indices; fluorescence), and water evaporation-based guard cell activity (four indices; infrared). To analyze responses of senescence triggered by various senescence-inducing factors in a comprehensive manner, raw numeric trait datasets should be organized through a preprocessing pipeline that involves (1) removal of outliers, (2) smoothing of time-series data, (3) normalization to the initial value, and (4) data integration throughout time adjustment and data standardization (**Figure 2A-iii**, detailed in Supplementary Information). Such a data integration is necessary for comparative analysis of time-series data with different degrees of effectiveness. An organized and tabulated dataset can be displayed in a heatmap for visual summarization and intuitive comparison among different senescence conditions (**Figure 2A-iv**).

Using these datasets, further data mining, including kinetic, clustering, and exploratory analyses, was performed (**Figure 2B**). Kinetic analysis with individual phenotypic traits revealed informative traits for primary or acute responsiveness to each senescence-triggering factor (**Figure 2B-i**). Leaves at different senescent conditions show phenotypic similarity in most of the traits, as represented by a MCARI marker (detailed information of markers in Supplementary Table S1). In contrast, rapid changes of QY\_max, a conventional marker reflecting the photochemical quantum efficiency of photosystem II, were observed when treated with H<sub>2</sub>O<sub>2</sub> and NaCl, implying that QY\_max is the effective signature for monitoring responses of leaf senescence to these treatments. This finding also suggests that photosynthetic activity in chloroplasts might be affected as the primary target during senescence, which is consistent with the results of previous transcriptome and metabolome studies (Breeze et al.,

2011; Watanabe et al., 2013; Woo et al., 2016). In addition, a distinct temporal pattern in different senescence conditions was observed in some markers such as BG1. Other markers such as PSNDc, Chlgreen, Mac, and SRWI2 possess a feature of primary responsiveness for H<sub>2</sub>O<sub>2</sub>, both of darkness and H<sub>2</sub>O<sub>2</sub>, ABA, and NaCl, respectively. These traits can further assist in dissecting the temporal progression or coordination of the biological processes related to each condition. More comprehensive relationships among traits and samples can be dissected with further detailed analysis using clustering analysis of genome-wide data (**Figure 2B-ii**). Although many traits (e.g., belonging to G3) exhibited temporal changes by more than three factors, some groups of traits were associated with specific senescence-inducing factors, including H<sub>2</sub>O<sub>2</sub> (G1), age (G2), or both dark and H<sub>2</sub>O<sub>2</sub> (G4). Further detailed and comprehensive sample analysis to dissect their phenotypic relationship requires more sophisticated exploratory statistical techniques such as PCA (**Figure 2B-iii**). PCA indicated that the initial senescence responses, regardless of treatment, were similar among different senescence processes; however, as leaf senescence proceeded over time, the physiological status of leaf senescence caused by different senescence factors changed quite differentially, especially in the case of age and NaCl. It was also shown that darkness-induced senescence appeared to be more similar to that of ABA-induced senescence, although senescence responses induced by H<sub>2</sub>O<sub>2</sub> and darkness shared common markers in the clustering analysis. However, the possibility that different assay conditions among treatments or age interfere with certain reflected or fluorescent lights on the leaves cannot be excluded. Despite some limitations to this study, these results suggest that genome-wide analyses using a couple of hundreds traits enable us to dissect senescence responses triggered by various senescence inducers.

Next, we further validated the feasibility of this approach by interrogating genome-based senescence responses in the well-characterized leaf senescence mutants, *oresara 3* (*ore3*), *ore12*, and *not oresara 1* (*nore1*; **Figure 2C-i**). *ore3*, which is allelic to *ethylene insensitive 2*, is insensitive to ethylene signaling, whereas *ore12*, a dominant negative mutant of *ARABIDOPSIS HISTIDINE KINASE 3*, exhibits constitutive cytokinin responses, where both mutations delay leaf senescence

(Oh et al., 1997; Kim et al., 2006). In contrast, *nore1* accelerates leaf senescence with enhanced defense response (Lee et al., 2016). The chronological phenomic analyses using a leaf tracking approach were performed at the third and fourth leaves of the wild type (Col) and of these mutant plants from the maturation to senescence stages (from DAE 14 to 36 at 2 days intervals; **Figure 2C-ii**). As previously reported, *ore3* and *ore12* leaves showed delayed senescence phenotypes, whereas *nore1* showed early senescence phenotypes, based on the QY\_max value. Although QY\_max is widely used as a typical marker of senescence progression, it was found to be less sensitive than other vegetation indices such as NDVI, SIPI, or PRI in VNIR imaging. This finding indicates that reflectance changes due to loss of pigments occur earlier than loss of QY\_max during developmental senescence, and these appear to be more useful markers to detect early symptoms of developmental senescence. These kinetic analyses with a few valuable traits may also be evaluated for the progression or rate of senescence responses; *ore3* exhibited a slightly slower change in senescence progression relative to *ore12*. To further explore global changes of phenotypic responses in leaves of these mutants during senescence, PCA analysis was performed for all samples examined with all phenotypic traits (**Figure 2C-iii**). Plotting the individual samples against PC1 and PC2, which collectively explained 86.6% of the variation in samples, afforded a clear separation of Col, early, and delayed senescence mutants at a late senescence stage. Since a slight difference among samples could be masked due to drastic changes in old leaf samples, we further performed PCA analysis to investigate some differences among samples in the maturation to early senescence stages (DAE 14 to 24; **Figure 2C-iii**, embedded graph). From this test, we observed that *nore1* and *ore3* could be distinguished from Col and *ore12*, although visible differences between them were not detected. Interestingly, leaves of Col, *nore1*, and *ore3* from DAE 24 were also resolved from other young leaves, implying that physiological diversity might be explained by the interaction of genetic and developmental factors. Considered together, we conclude that quantitative measurement of phenotypic traits from leaves appears to be important for dissecting leaf senescence, and provides valuable information for phenotypic regulation by senescence-inducing factors or genetic components during senescence.

As the PHI system supports sequential leaf-based analysis from intact plants using leaf tracking, we further addressed advantages of the leaf tracking system by comparison with a conventional leaf detaching assay (**Figure 2C-i**). Practically, the non-invasive phenotyping system requires a much smaller number of plants. In addition, subtle differences among samples could be discerned; temporal analysis could be performed at a higher resolution using leaf tracking (**Figure 2C-iv**), and statistical powers could be increased with a larger number of samples and pair-wise analysis (**Figure 2C-v**). In addition, the performance of the association analysis between traits can be increased, based on the possibility of their one-to-one matching within one sample. Thus, not only is a PHI-based high-throughput system beneficial for performance but it also improves analytical capabilities.

## CONCLUSION

Here, we developed a specialized high-throughput phenotyping platform for analyzing senescence traits at a single-leaf basis, which will facilitate an alleviation of the phenotyping bottleneck in leaf senescence. As a proof of the concept, we dissected features of various senescence responses through kinetic and PCA analyses utilizing highly resolved and quantitative phenotyping data. In addition, we evaluated advantages of the leaf tracking system in a PHI high-throughput phenotyping system in terms of performance and analytic capabilities. Considered together, we demonstrated the pipeline of phenomics that allows the dissection of a system as complex as leaf lifespan and senescence.

## PERSPECTIVES

By virtue of great advances in omics technologies, big data generation has resulted in a major paradigm shift toward data-driven research in plant biology. Along with an increasing feasibility of molecule-based omics, the implementation of automated, high-throughput phenotyping at a similar level will offer new opportunities to understand the complex biological processes occurring in plants (Chen et al., 2014; Granier and Vile, 2014). Our establishment, including the experimental setup and phenotyping data analyses, will open up great opportunities to address concepts and premises that are critical to enhance our fundamental understanding of the as-yet incompletely understood complex process of leaf senescence.

First, our PHI system would allow dynamic, longitudinal, and multi-dimensional analyses that characterize the physiological and regulatory changes along the entire leaf lifespan at a system level. By taking advantage of the PHI system, systematic quantification analyses of all possible traits during the entire leaf development from a large population of plants, including many genetic resources, can be performed. This should result in more detailed insights into mechanisms governing developmental transitions during leaf life history, thereby elucidating important biological principles on how previous developmental programs contribute to the senescence process on a genetic basis. This would also contribute to infer the causal relationship between phenotypic traits at an earlier stage and responses of senescence, which might be valuable for screening during breeding programs. Our pipeline can be extended to the meta-analysis of multiplexed phenotyping data with large-scale quantitative phenotype collections, thereby allowing the depiction of the network relationship from gene- to senescence-related phenotypic traits along the leaf lifespan.

Second, leaf senescence was long believed to be an evolutionarily acquired beneficial process to maximize the fitness of plants. However, no clear evidence has yet emerged linking leaf senescence and fitness. A non-destructive senescence assay and its following fitness measurement such as seed yield will allow the elucidation of their relationship. Furthermore, high-throughput phenotypic analysis of various physiological and developmental traits from the large collection of genetic resources will allow the evaluation of the contribution of each trait to fitness factors,

which may thereby elucidate the importance of senescence for fitness, relative to other traits.

Third, a PHI-based high-throughput system supports controlled and precise environmental conditions. This facilitates the investigation of the direct relationship between environmental condition and senescence along with seed yield. It further infers how senescence may contribute to fitness under certain environmental conditions. In addition, high-throughput phenotyping with a large collection of natural accessions under different local simulated climates consisting of photoperiod, light spectrum, temperature, and relative humidity allow the identification of the relationship among senescence, environments, and adaptation to local environmental conditions (Li et al., 2010; Xu, 2016). Combined with genome-wide association analysis, these endeavors will eventually elucidate the mechanisms governing phenotypic plasticity and adaptive mechanisms (Todesco et al., 2010; Brachi et al., 2013; Yang et al., 2014).

Fourth, the main purpose of leaf senescence is the redistribution of nutrients from one part of the plant to another. Thus, senescence can be affected by the removal of sink or neighboring organs, which indicates the existence of inter-organ level coordination (Sekhon et al., 2012). Leaf-based analysis in a PHI system provides favorable tools to dissect inter-organ communication between individual leaves and leaves and other organs such as shoot or root.

Fifth, senescence is regarded as a typical irreversible phenomenon. However, depending on the leaf age and degree of treatments that induce leaf senescence, the primary response of senescence can be recovered. It is feasible to trace back phenotypic changes from leaves with different fates, which might provide some phenotypic clues on how the irreversible onset of senescence is determined.

Phenomic studies can contribute to validate their findings based on transcriptomic, genomic, proteomic, and metabolomic

data to senescence processes by providing the outer analytic layer to illustrate collective outputs of dynamic molecular changes such as genes, transcripts, proteins, and metabolites. Thus, combined with these multi-omics data, our phenotyping system is a very promising and valuable tool that allows the investigation of changes to morphology, physiology, and molecular behaviors in a comprehensive manner over leaf lifespan and senescence. This will facilitate an understanding of the mechanisms of life history and senescence over spatial and temporal scales.

## AUTHOR CONTRIBUTIONS

JL, JK, and PL conceived and designed the experiments. JL, SB, and HC performed the experiments. JL, SB, SJ, and JK analyzed the data. HN provided analysis tools. JK and PL wrote the paper. All authors carefully checked and approved this version of the manuscript.

## ACKNOWLEDGMENTS

This research was supported by the Institute for Basic Science (IBS-R013-D1) and the DGIST R&D Program (15-01-HRLA-01) from the Ministry of Science, ICT & Future Planning. We thank Martin Trtílek and his colleagues (Photon Systems Instruments) for the establishment of the PHI system and technical supports.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00250/full#supplementary-material>

## REFERENCES

- Brachi, B., Villoutreix, R., Faure, N., Haukeete, N., Piquot, Y., Pauwels, M., et al. (2013). Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*. *Mol. Ecol.* 22, 4222–4240. doi: 10.1111/mec.12396
- Breeze, E., Harrison, E., McHattie, S., Hughes, L., Hickman, R., Hill, C., et al. (2011). High-resolution temporal profiling of transcripts during *Arabidopsis* leaf senescence reveals a distinct chronology of processes and regulation. *Plant Cell* 23, 873–894. doi: 10.1105/tpc.111.083345
- Buchanan-Wollaston, V., Page, T., Harrison, E., Breeze, E., Lim, P. O., Nam, H. G., et al. (2005). Comparative transcriptome analysis reveals significant differences in gene expression and signalling pathways between developmental and dark/starvation-induced senescence in *Arabidopsis*. *Plant J.* 42, 567–585. doi: 10.1111/j.1365-313X.2005.02399.x
- Cabrera-Bosquet, L., Fournier, C., Brichet, N., Welcker, C., Suard, B., and Tardieu, F. (2016). High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. *New Phytol.* 212, 269–281. doi: 10.1111/nph.14027
- Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., et al. (2014). Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *Plant Cell* 26, 4636–4655. doi: 10.1105/tpc.114.129601
- Crowell, S., Falcao, A. X., Shah, A., Wilson, Z., Greenberg, A. J., and McCouch, S. R. (2014). High-resolution inflorescence phenotyping using a novel image-analysis pipeline, PANorama. *Plant Physiol.* 165, 479–495. doi: 10.1104/pp.114.238626
- 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
- Granier, C., and Vile, D. (2014). Phenotyping and beyond: modelling the relationships between traits. *Curr. Opin. Plant Biol.* 18, 96–102. doi: 10.1016/j.pbi.2014.02.009
- Guo, Y. (2013). Towards systems biological understanding of leaf senescence. *Plant Mol. Biol.* 82, 519–528. doi: 10.1007/s11103-012-9974-2
- Humplík, J. F., Lazár, D., Husičková, A., and Spíchal, L. (2015). Automated phenotyping of plant shoots using imaging methods for analysis of plant stress responses – a review. *Plant Methods* 11, 1–10. doi: 10.1186/s13007-015-0072-8
- Kim, H. J., Ryu, H., Hong, S. H., Woo, H. R., Lim, P. O., Lee, I. C., et al. (2006). Cytokinin-mediated control of leaf longevity by AHK3 through phosphorylation of ARR2 in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 814–819. doi: 10.1073/pnas.0505150103
- Kim, J., Woo, H. R., and Nam, H. G. (2016). Toward systems understanding of leaf senescence: an integrated multi-omics perspective on leaf senescence research. *Mol. Plant* 9, 813–825. doi: 10.1016/j.molp.2016.04.017
- Lee, I. H., Lee, I. C., Kim, J., Kim, J. H., Chung, E. H., Kim, H. J., et al. (2016). NORE1/SAUL1 integrates temperature dependent defense programs involving

- SGT1b and PAD4 pathways and leaf senescence in *Arabidopsis*. *Physiol. Plant.* 158, 180–199. doi: 10.1111/ppl.12434
- Li, L. H., Xing, Y. F., Chang, D., Fang, S. S., Cui, B. Y., Li, Q., et al. (2016). CaM/BAG5/Hsc70 signaling complex dynamically regulates leaf senescence. *Sci. Rep.* 6:31889. doi: 10.1038/srep31889
- Li, Y., Huang, Y., Bergelson, J., Nordborg, M., and Borevitz, J. O. (2010). Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21199–21204. doi: 10.1073/pnas.1007431107
- Liebsch, D., and Keech, O. (2016). Dark-induced leaf senescence: new insights into a complex light-dependent regulatory pathway. *New Phytol.* 212, 563–570. doi: 10.1111/nph.14217
- Lim, P. O., Kim, H. J., and Nam, H. G. (2007). Leaf senescence. *Annu. Rev. Plant Biol.* 58, 115–136. doi: 10.1146/annurev.arplant.57.032905.105316
- Mochida, K., and Shinozaki, K. (2011). Advances in omics and bioinformatics tools for systems analyses of plant functions. *Plant Cell Physiol.* 52, 2017–2038. doi: 10.1093/pcp/pcr153
- Oh, S. A., Park, J. H., Lee, G. I., Paek, K. H., Park, S. K., and Nam, H. G. (1997). Identification of three genetic loci controlling leaf senescence in *Arabidopsis thaliana*. *Plant J.* 12, 527–535.
- Rahaman, M. M., Chen, D., Gillani, Z., Klukas, C., and Chen, M. (2015). Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front. Plant Sci.* 6:619. doi: 10.3389/fpls.2015.00619
- Rajasundaram, D., and Selbig, J. (2016). More effort - more results: recent advances in integrative 'omics' data analysis. *Curr. Opin. Plant Biol.* 30, 57–61. doi: 10.1016/j.pbi.2015.12.010
- Schippers, J. H. (2015). Transcriptional networks in leaf senescence. *Curr. Opin. Plant Biol.* 27, 77–83. doi: 10.1016/j.pbi.2015.06.018
- Schippers, J. H., Schmidt, R., Wagstaff, C., and Jing, H. C. (2015). Living to die and dying to live: the survival strategy behind leaf senescence. *Plant Physiol.* 169, 914–930. doi: 10.1104/pp.15.00498
- Sekhon, R. S., Childs, K. L., Santoro, N., Foster, C. E., Buell, C. R., de Leon, N., et al. (2012). Transcriptional and metabolic analysis of senescence induced by preventing pollination in maize. *Plant Physiol.* 159, 1730–1744. doi: 10.1104/pp.112.199224
- Slovak, R., Göschl, C., Su, X., Shimotani, K., Shiina, T., and Busch, W. (2014). A scalable open-source pipeline for large-scale root phenotyping of *Arabidopsis*. *Plant Cell* 26, 2390–2403. doi: 10.1105/tpc.114.124032
- Thomas, H. (2013). Senescence, ageing and death of the whole plant. *New Phytol.* 197, 696–711. doi: 10.1111/nph.12047
- Todesco, M., Balasubramanian, S., Hu, T. T., Traw, M. B., Horton, M., Epple, P., et al. (2010). Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* 465, 632–636. doi: 10.1038/nature09083
- Topp, C. N., Iyer-Pascuzzi, A. S., Anderson, J. T., Lee, C. R., Zurek, P. R., Symonova, O., et al. (2013). 3D phenotyping and quantitative trait locus mapping identify core regions of the rice genome controlling root architecture. *Proc. Natl. Acad. Sci. U.S.A.* 110, E1695–E1704. doi: 10.1073/pnas.1304354110
- Watanabe, M., Balazadeh, S., Tohge, T., Erban, A., Giavalisco, P., Kopka, J., et al. (2013). Comprehensive dissection of spatiotemporal metabolic shifts in primary, secondary, and lipid metabolism during developmental senescence in *arabidopsis*. *Plant Physiol.* 162, 1290–1310. doi: 10.1104/pp.113.217380
- Woo, H. R., Koo, H. J., Kim, J., Jeong, H., Yang, J. O., Lee, I. H., et al. (2016). Programming of plant leaf senescence with temporal and inter-organellar coordination of transcriptome in *Arabidopsis*. *Plant Physiol.* 171, 452–467. doi: 10.1104/pp.15.01929
- Xu, Y. (2016). Envirotyping for deciphering environmental impacts on crop plants. *Theor. Appl. Genet.* 129, 653–673. doi: 10.1007/s00122-016-2691-5
- Yang, W., Duan, L., Chen, G., Xiong, L., and Liu, Q. (2013). Plant phenomics and high-throughput phenotyping: accelerating rice functional genomics using multidisciplinary technologies. *Curr. Opin. Plant Biol.* 16, 180–187. doi: 10.1016/j.pbi.2013.03.005
- Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5:5087. doi: 10.1038/ncomms6087
- Zentgraf, U., Jobst, J., Kolb, D., and Rentsch, D. (2004). Senescence-related gene expression profiles of rosette leaves of *Arabidopsis thaliana*: leaf age versus plant age. *Plant Biol.* 6, 178–183. doi: 10.1055/s-2004-815735

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

Copyright © 2017 Lyu, Baek, Jung, Chu, Nam, Kim and Lim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Arabidopsis Seed Content QTL Mapping Using High-Throughput Phenotyping: The Assets of Near Infrared Spectroscopy

**Sophie Jasinski<sup>1\*</sup>, Alain Lécureuil<sup>1</sup>, Monique Durandet<sup>1</sup>, Patrick Bernard-Moulin<sup>2</sup> and Philippe Guerche<sup>1</sup>**

<sup>1</sup> Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, Versailles, France, <sup>2</sup> ThermoFisher Scientific, Courtaboeuf, France

## OPEN ACCESS

### Edited by:

Marcos Egea-Cortines,  
Universidad Politécnica de Cartagena,  
Spain

### Reviewed by:

Veronique Storme,  
Ghent University, Belgium

Yuhui Chen,  
Samuel Roberts Noble Foundation,  
USA

### \*Correspondence:

Sophie Jasinski  
[sophie.jasinski@versailles.inra.fr](mailto:sophie.jasinski@versailles.inra.fr)

### Specialty section:

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

**Received:** 08 September 2016

**Accepted:** 25 October 2016

**Published:** 10 November 2016

### Citation:

Jasinski S, Lécureuil A, Durandet M, Bernard-Moulin P and Guerche P (2016) Arabidopsis Seed Content QTL Mapping Using High-Throughput Phenotyping: The Assets of Near Infrared Spectroscopy. *Front. Plant Sci.* 7:1682.  
doi: 10.3389/fpls.2016.01682

Seed storage compounds are of crucial importance for human diet, feed and industrial uses. In oleo-proteaginous species like rapeseed, seed oil and protein are the qualitative determinants that conferred economic value to the harvested seed. To date, although the biosynthesis pathways of oil and storage protein are rather well-known, the factors that determine how these types of reserves are partitioned in seeds have to be identified. With the aim of implementing a quantitative genetics approach, requiring phenotyping of 100s of plants, our first objective was to establish near-infrared reflectance spectroscopic (NIRS) predictive equations in order to estimate oil, protein, carbon, and nitrogen content in *Arabidopsis* seed with high-throughput level. Our results demonstrated that NIRS is a powerful non-destructive, high-throughput method to assess the content of these four major components studied in *Arabidopsis* seed. With this tool in hand, we analyzed *Arabidopsis* natural variation for these four components and illustrated that they all displayed a wide range of variation. Finally, NIRS was used in order to map QTL for these four traits using seeds from the *Arabidopsis thaliana* Ct-1 × Col-0 recombinant inbred line population. Some QTL co-localized with QTL previously identified, but others mapped to chromosomal regions never identified so far for such traits. This paper illustrates the usefulness of NIRS predictive equations to perform accurate high-throughput phenotyping of *Arabidopsis* seed content, opening new perspectives in gene identification following QTL mapping and genome wide association studies.

**Keywords:** *Arabidopsis thaliana*, seed storage contents, near infrared spectroscopy, plant, natural variation, quantitative trait loci

## BACKGROUND

Plant seeds constitute a key component of both human and livestock diets, as seed storage compounds are mainly composed of protein, oil and starch. Seed oil from oleaginous crops are composed mainly of triacylglycerols, which are structurally similar to long chain hydrocarbons derived from petroleum, and thus represent ecologically and economically competitive alternatives

**Abbreviations:** QTL, quantitative trait loci; RIL, recombinant inbred line.

to petroleum-based products for the production of molecules for green chemistry (e.g., detergents, paints, plastics, and lubricants) as well as for the production of biofuels (Durrett et al., 2008; Dyer et al., 2008). The increasing demand of plant-derived products for nutritional and industrial applications highlights the urgent need to develop new methodologies to increase the overall seed oil and protein content. Although, most of the biochemical steps involved in oil and protein biosynthesis are known and the key genes have been identified (Shewry et al., 1995; Beisson et al., 2003), the regulation of the processes that results in the final oil and protein content is not well-understood. Even more, the genetic factors that control the oil/protein ratio in the seeds have to be identified.

Seed oil and protein accumulation processes, like many important agronomical traits are quantitative and have a complex genetic basis. The method most commonly used for inferring the presence and position of such genes in the genome is based upon analysis such as QTL and more recently genome wide association studies (GWASs). Quantitative genetics has been used to search for genetic factors controlling oil and/or protein quantity in a variety of agronomically important species including rapeseed (Sun et al., 2012; Li et al., 2014), soybean (Eskandari et al., 2012; Hwang et al., 2014), maize (Zheng et al., 2008; Li et al., 2012; Yang et al., 2012), pea (Irzykowska and Wolko, 2004; Tar'an et al., 2004), rice (Ying et al., 2012), wheat (Plessis et al., 2013), oat (Kianian et al., 1999), sunflower (Mokrani et al., 2002), linseed (Kumar et al., 2015), cotton (Liu et al., 2015), and Jatropha (Liu et al., 2011). However, with the exception of one QTL affecting seed oil content in maize (Zheng et al., 2008), none of these studies have gone beyond the gene mapping stage. In the model species *Arabidopsis*, using natural variation resources, QTL involved in seed oil content and/or quality have also been detected (Hobbs et al., 2004; Jasinski et al., 2012; O'Neill et al., 2012; Sanyal and Linder, 2012) as well as “regions of interest” by GWAS (Branham et al., 2015), but only the study published by Jasinski et al. (2012) identified the gene involved. Since QTL cloning is often easier in model species for which substantial genetic resources exist, we implemented a QTL approach to study storage compound metabolisms in *Arabidopsis* seed (Jasinski et al., 2012; Chardon et al., 2014). Seed metabolism is very similar between *Arabidopsis* and *Brassica* species and the close relationship between them allows the use of comparative genetics to predict orthologous genes and alleles within the *Brassica* genome (Parkin et al., 2005). This will enable the translation of discoveries from *Arabidopsis* into Brassicaceae and other crops breeding programs.

Quantitative genetics relies on statistical links between phenotype and genotype, implying genotyping and phenotyping of 1000s of lines. In *Arabidopsis*, genotyping is not a limiting factor and many genetic and genomic resources are available including complete genome sequence of many accessions, data on gene structure, gene expression, DNA and seed stocks, genome maps, molecular markers. Seed oil and protein content are usually determined by standard analytical methods such as Soxhlet or gas chromatography (following fatty acid methyl ester extraction)

for oil content and combustion analysis or Kjeldahl for protein content. Although, these standard analytical techniques offer a high level of accuracy and precision, they also show some limitations, such as indirect determination (combustion analysis and Kjeldahl determine nitrogen content rather than actual protein content), high costs, time-consuming experiments and use of hazardous chemicals. For many of these reasons, they are not fully appropriate for high-throughput phenotyping required in genetics approaches. Near infrared spectroscopy (NIRS) is a vibrational spectroscopy technique, providing a spectrum representative of the “signature” of all components present in the analyzed sample. It possesses numerous advantages compared to classical analytical techniques. NIRS analyses show high degree of repeatability and are carried out with considerable saving of time (spectrum acquisition lasts only a few seconds), cost and without using hazardous chemicals. In addition, samples can be analyzed in their natural form without destruction neither any special sample preparation. However, a calibration has first to be established: regression modeling is used to relate NIRS spectra to chemical concentrations determined by a standard analytical method. After calibration, the developed regression equations allow accurate analysis of many other samples by prediction of data based on the spectra. Moreover, from only one spectrum, different components can be predicted using different predictive equations. In recent decades, NIRS has been widely used as a fast and reliable method for qualitative and quantitative analysis in many fields (Font et al., 2006) and International Standards Committees have formally accepted methods using NIRS for analysis of many compounds (Batten, 1998). Regarding *Brassica* seeds, many authors have reported NIRS models for different components, such as glucosinolates (Velasco and Becker, 1998; Font et al., 2004), fiber (Font et al., 2003), protein and oil contents (Tkachuk, 1981; Font et al., 2002a,b; Rossato et al., 2013).

Surprisingly, NIRS technique has not been applied to the analysis of *Arabidopsis* seed. Some people used nuclear magnetic resonance spectroscopy (NMR) as rapid technique to measure *Arabidopsis* seed oil content (O'Neill et al., 2003, 2012; Hobbs et al., 2004). However, NMR is not suitable for protein detection and was thus not suitable for our purpose. In this study, the potential of NIRS was evaluated for the simultaneous analysis of total oil and protein content of *Arabidopsis* seeds, as well as nitrogen and carbon contents, which allow studies of global metabolic fluxes. A calibration set of 90–112 seed samples was subjected to both NIRS and appropriate reference methods and predictive equations for seed (1) oil, (2) protein, (3) carbon, and (4) nitrogen content were developed.

These equations were further used to analyze *Arabidopsis* natural variation for these four major seed components. Finally, a search for genetic factors governing the accumulation of these four components in *Arabidopsis* seed was carried out by a QTL analysis (this work and Chardon et al., 2014), allowing the mapping of new QTL involved in seed oil and protein content.

## RESULTS

### Development of NIRS Predictive Equations for Seed Oil, Protein, Carbon, and Nitrogen Content

Four calibration models in order to predict oil, protein, carbon, and nitrogen content in *Arabidopsis* seeds were developed as indicated in Section "Methods."

The oil calibration set of 112 samples showed a wide range of variation for oil content from 18.70%, corresponding to the *wrl1* low-seed-oil T-DNA insertion mutant (Focks and Benning, 1998), to 46.90%, with a mean of 38.78% (Table 1) and a standard deviation (SD) of 3.88%. The predictive equation for seed oil content was developed with five partial least square (PLS) factors and first evaluated through cross-validation (leave-one-out method). Very high coefficient of determinations between Soxhlet and NIRS values were observed for both calibration and cross-validation ( $r^2_C = r^2_{CV} = 0.98$ , Table 1; Figure 1A). The standard error of cross-validation (SECV) was 0.606% (Table 1). The 36 additional seed samples were used to carry out an external validation to better assess the accuracy of this calibration model. This showed a coefficient of determination of 0.99 and a standard error of prediction (SEP) of 0.505% (Table 1; Figure 1A).

As for oil, the three other calibration sets showed a wide range of variation (Table 1). For each of the three components, a predictive equation was developed and the external validation sample set was further used to evaluate its performance. For the three developed models, a limited number of PLS factors ( $\leq 4$ ) was used and very high coefficients of determination between the reference method and NIRS values were observed for both calibration and cross-validation (Table 1; Figure 1).

The ratio of performance deviation (RPD), an indicator for the usefulness of the calibration model was calculated for each component (Table 1). According to the American Association of Cereal Chemists Method-39-00.01 (AACC International, 1999), a RPD  $\geq 2.5$  indicates that a calibration equation is useful for screening in breeding programs, a RPD  $\geq 5$  that a calibration is acceptable for quality control and a RPD  $\geq 8$  that a calibration is good for process control, development, and applied research. For all the models developed in this study, a RPD  $> 2.5$  was achieved, with a RPD of 7.68 (close to 8) for the oil content model and a RPD close to 5 for the nitrogen model. This indicates that the four models developed in this study are suitable for quantitative genetic approaches.

### NIRS Is a Suitable Tool for High-Throughput Phenotyping of *Arabidopsis* Seed

In order to fully demonstrate that the developed NIRS models were suitable to study seed composition in *Arabidopsis*, we analyzed mutants altered in seed filling. Pyruvate kinase (PK) catalyze the irreversible synthesis of pyruvate and ATP (Valentini et al., 2000), which are essential for fatty acid production in the plastids of maturing *Arabidopsis* embryos. Baud et al. have shown that the plastidial PK isoform PKp2 plays an important role in

seed oil synthesis, with *pkp2-1* mutant exhibiting a 50% reduction in seed oil content compared to wild-type (Baud et al., 2007). More recently, Chen et al. showed that seed filling in *Arabidopsis* requires sucrose transporters from the SWEET family (Chen et al., 2015). In particular, they showed that seed oil content was reduced by 34% in the *sweet11;12* double mutant (Chen et al., 2015). Seeds from *pkp2-1* and *sweet11;12* mutants were analyzed by NIRS. For *pkp2-1*, a 36% decrease in oil content compared to wild-type was observed (Figure 2A), which is comparable to the decrease described by Baud et al. (2007) on the same seed lot. *Sweet11;12* mutants displayed a 17% reduction in seed oil content compared to wild-type (Figure 2B), which is half the one described by Chen et al. (2015) on another seed lot. The *sweet11;12* seed lot measured by NIRS was then subjected to Gas Chromatography and resulted in a 15% decrease in seed oil (result not shown). This result suggested that the difference observed is probably due to environmental effect on seed filling more than to NIRS method.

Furthermore, since it is known that nitrogen nutrition impacts seed filling (Masclaux-Daubresse and Chardon, 2011), we analyzed seed content of wild-type plants (Col-0 and Ws) grown under low nitrogen (LowN; 2 mM nitrate) or under high nitrogen (HighN; 10 mM nitrate) nutrition conditions. As already published (Masclaux-Daubresse and Chardon, 2011), both accessions displayed higher seed nitrogen content under HighN compared to LowN (Figure 2C) and reversely a higher seed carbon content under LowN compared to HighN (Figure 2D).

These results demonstrate that NIRS is a powerful method to determine *Arabidopsis* seed composition.

### Natural Variation for Oil, Protein, Carbon, and Nitrogen Content in *Arabidopsis* Seed

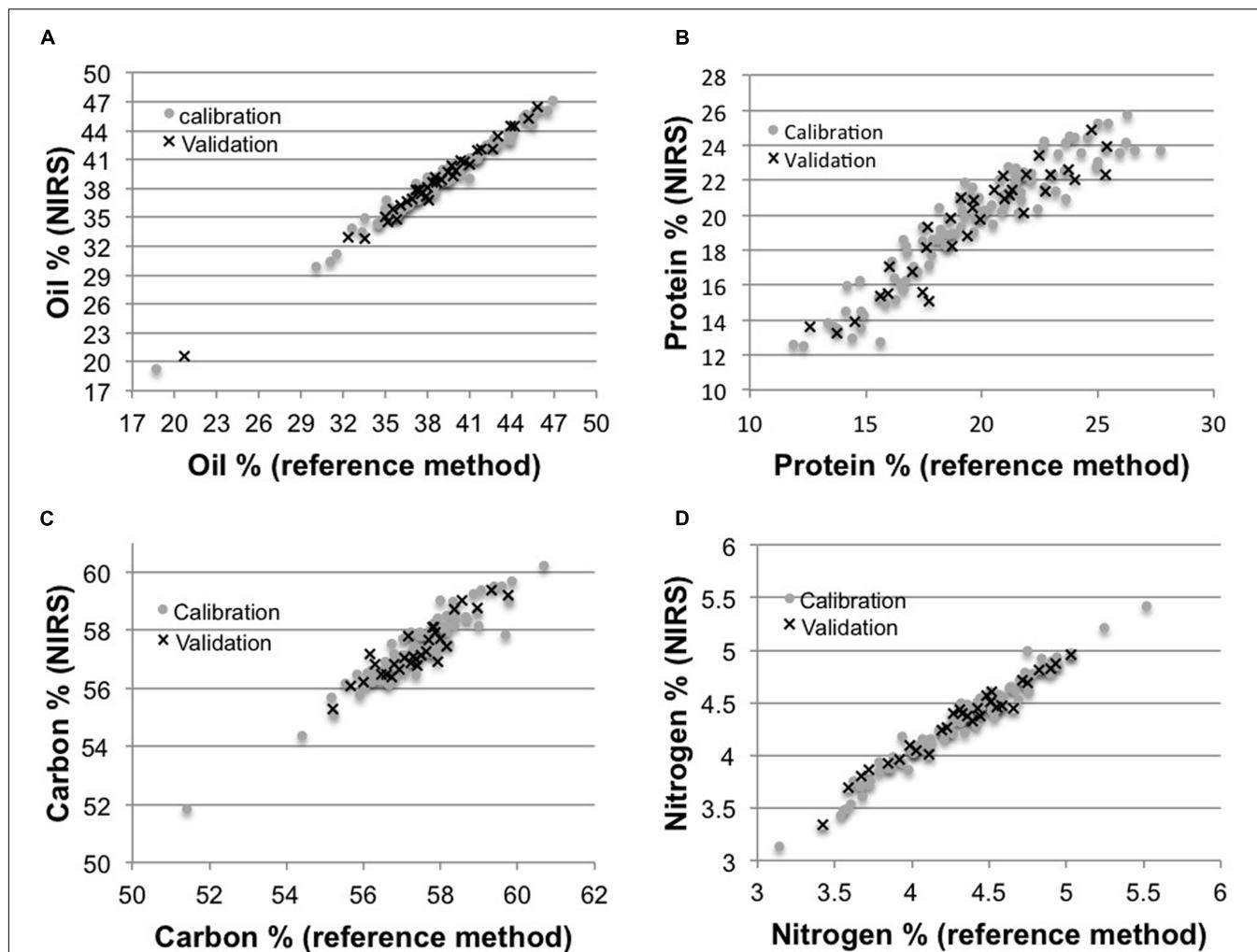
The development of NIRS predictive equations allowing high-throughput phenotyping opened the door to quantitative genetics study. First, we decided to explore *Arabidopsis* natural variability for oil, protein, carbon, and nitrogen content in seeds. For this purpose, we cultivated the Versailles BRC 48 core-collection of *Arabidopsis* in addition to the Col-0 accession and mini sets of 20 lines (maximizing genotypic variability, Simon et al., 2008) from eight populations (see Methods). Each genotype was cultivated in triplicate and three successive and independent cultures (C1, C2, and C3) were performed in growth chambers with similar global climate conditions. We estimated the natural variability of the four traits by NIRS phenotyping (Figure 3). The four traits displayed a wide range of variation in each culture, with C1 displaying the wider range going from 23.23 to 47.72% for oil, from 14.24 to 28.72% for protein, from 52.27 to 60.22% for carbon, and from 3.27 to 5.78% for nitrogen. The modal class is different in each culture, highlighting the environmental effect on these four traits.

In order to quantify the relative contribution of the genotype (G), the culture (C) and the G\*C interaction on the variation of these four traits, a global analysis of variance (ANOVA) was performed on the measures from the three cultures (Figure 4). The genotypic effects ranged from 53.0% for oil to 32.3%

**TABLE 1 |** Near-infrared reflectance spectroscopic (NIRS) calibration and cross validation statistics for seed oil, protein, carbon, and nitrogen contents (%).

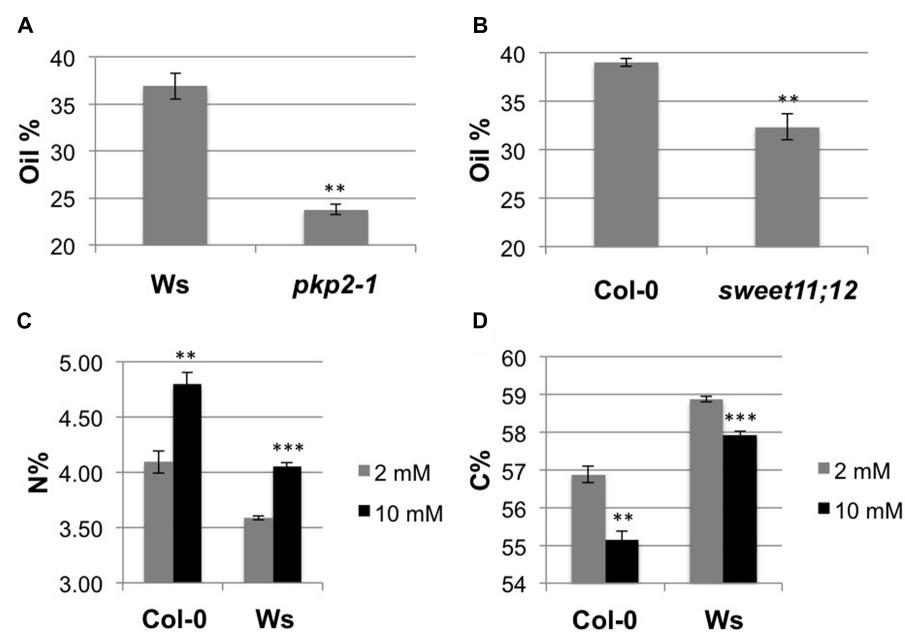
	n	Calibration				Cross-validation				External validation					
		Mean (%)	Range (%)	SD	PLS factors	SEC	$r^2_C$	SECV	$r^2_{CV}$	n	Mean (%)	SD	SEP	$r^2_V$	RPD
Oil	112	38.78	18.70–46.90	3.88	5	0.541	0.98	0.606	0.98	36	38.45	4.39	0.505	0.99	7.68
Protein	98	19.99	11.89–27.73	3.39	3	1.219	0.88	1.287	0.85	33	19.73	3.32	1.228	0.86	2.76
Carbon	91	57.39	51.41–60.69	1.25	3	0.424	0.88	0.460	0.86	30	57.39	1.04	0.432	0.82	2.90
Nitrogen	90	4.34	3.14–5.52	0.43	4	0.081	0.96	0.091	0.95	30	4.32	0.41	0.087	0.96	4.91

n, number of samples; SD, standard deviation; PLS, partial least square; SEC, standard error of calibration;  $r^2_C$ , coefficient of determination in calibration; SECV, standard error of cross-validation;  $r^2_{CV}$ , coefficient of determination in cross-validation;  $r^2_V$ , coefficient of determination in validation; SEP, standard error of prediction; RPD, ratio of performance deviation (SD/SEP).

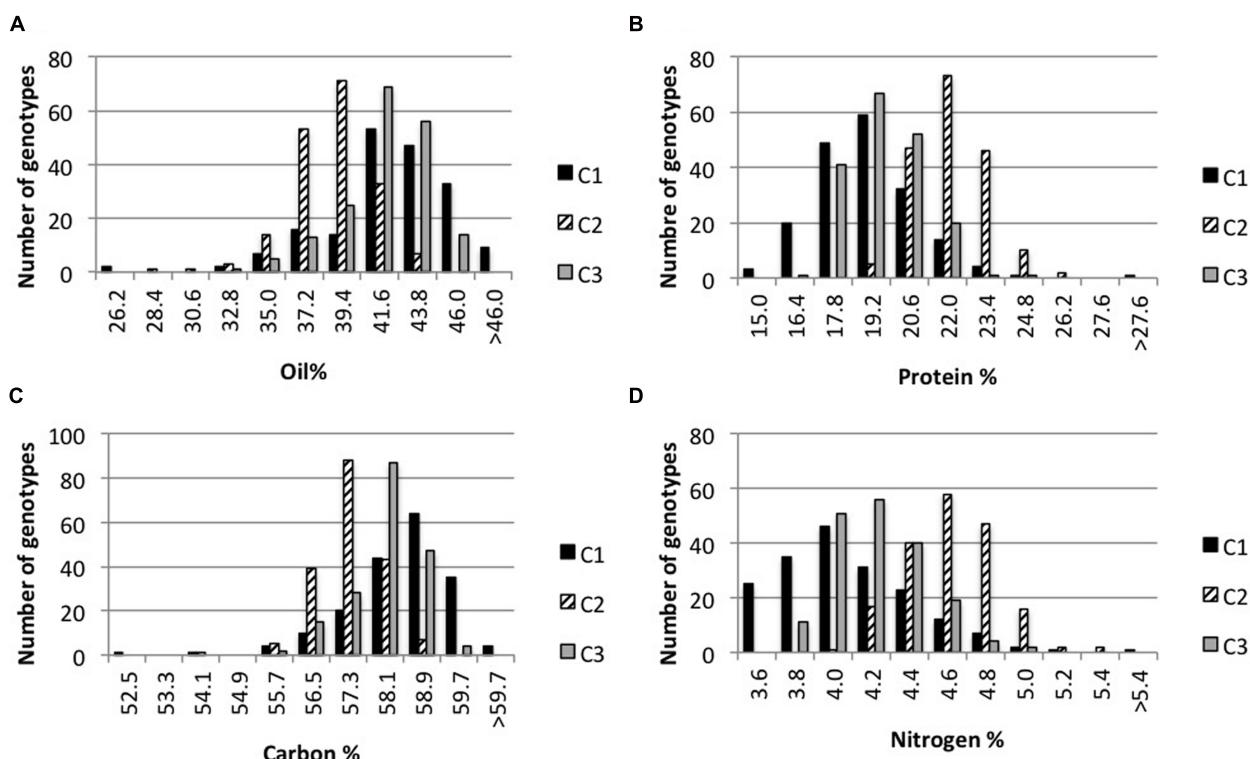
**FIGURE 1 |** Regression curves between near-infrared reflectance spectroscopic (NIRS) and reference methods for oil, protein, carbon, and nitrogen content. Scatter plots of NIRS versus reference method values for oil % (A), protein % (B), carbon (C), and nitrogen (D) in the calibration and validation sets.

for protein, explaining the most important part of the total phenotypic variation except for protein. However, culture effect explained an important part of the total phenotypic variation

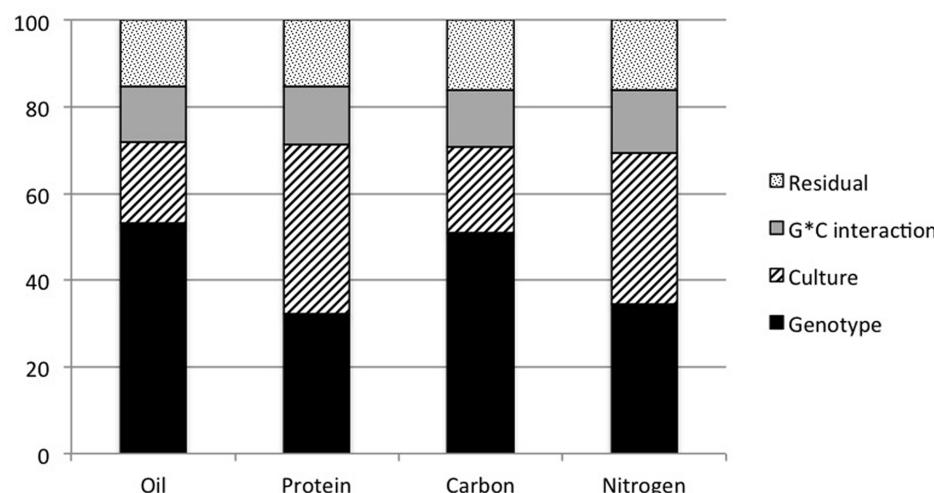
varying from 19 to 39% for oil and protein respectively. This result showed that nitrogen and protein contents were more influenced by the culture than oil and carbon.



**FIGURE 2 |** Near-infrared reflectance spectroscopic phenotyping allows detecting seed filling modifications. **(A,B)** Graphs showing seed oil content of two mutants: *pkp2-1* **(A)** and *sweet11;12* **(B)** compared to wild-type. **(C,D)** Graphs showing seed nitrogen **(C)** and carbon **(D)** content of two accessions (Col-0 and Ws) on two nitrogen nutrition conditions (2 and 10 mM). Bars represent SE ( $n = 3$  in **A,B**,  $n = 4$  in **C,D**). Significance in *t*-test, \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .



**FIGURE 3 |** Frequency distribution of seed oil, protein, carbon, and nitrogen content in three independent cultures. Frequency distributions of oil **(A)**, protein **(B)**, carbon **(C)**, and nitrogen **(D)** content mean values for 183 genotypes ( $n = 3$ ) corresponding to *Arabidopsis* accessions and RILs from eight populations. The same genotypes were cultivated in three independent cultures (C1, C2, C3).



**FIGURE 4 | Variance component analysis for seed oil, protein, carbon, and nitrogen content.** An ANOVA was performed on all genotypes from the three cultures for seed oil, protein, carbon, and nitrogen content. Histograms show the effects due to genotype, culture, interaction genotype\*culture and the residual as a percentage of the variation explained.

## Identification of QTL Involved in Seed Oil and Protein Content

The availability of NIRS predictive equations together with the large range of variation observed and the high contribution of genetic part to phenotypic variation for oil, protein, carbon, and nitrogen content opened the door to QTL study. From this previous study of natural variation, the Ct-1 × Col-0 RIL population was selected for QTL determination. A subset of 164 RILs, optimized for QTL mapping (Simon et al., 2008) was cultivated (see Methods) and seeds were phenotyped for oil, protein, carbon, and nitrogen content by NIRS. These RILs exhibited a wide range of values for these four traits as well as transgression beyond the parental line values (not shown), highlighting the potential of this subset to study the variation of these traits. QTL detection using standard procedures (see Methods) was carried out, allowing QTL detection for oil and protein content (this work) as well as for carbon, and nitrogen content (Chardon et al., 2014).

Five QTL for seed oil content were identified, explaining 44.5% of the total phenotypic variance observed (Table 2). The strongest QTL (Oil.4, explaining more than 15% of the phenotypic variance) is located between 31.6 and 46 cM on chromosome 4. Four QTL were detected for protein content, explaining 34% of the total phenotypic variance observed for this trait (Table 2). The strongest QTL (Prot.3, explaining more than 10% of the phenotypic variance) co-localized with Oil.4 on chromosome 4. Three out of five oil QTL, Oil.2, Oil.4, and Oil.5 overlapped with protein QTL Prot.1, Prot.3, and Prot4 respectively, but having an opposite effect on the corresponding traits, highlighting the strong negative correlation observed between oil and protein seed content, as already observed for carbon and nitrogen by Chardon et al. (2014). Interestingly, QTL specific to oil (Oil.1, i.e., without noticeable effect on protein) and protein (Prot.2, i.e., without noticeable effect on oil) were also identified in this

study. Altogether, these results illustrate that NIRS phenotyping of mature seeds allow identification of genetic factors involved in different pathways of oil and protein accumulation.

## DISCUSSION

Quantitative genetic relies on statistical links between phenotype and genotype of 100s of lines. In *Arabidopsis*, genotyping is no more a limiting factor, whereas high-throughput phenotyping can be an obstacle. Thus, our first objective was to establish near-infrared reflectance spectroscopic (NIRS) predictive equations in order to estimate oil, protein, carbon, and nitrogen content in *Arabidopsis* seed with high-throughput level.

**TABLE 2 | List of QTL detected for oil and protein in the Ct-1 × Col-0 RIL population.**

QTL name	Chr.	Position (cM)	LOD score	CI (cM)	Additive effect	R <sup>2</sup> (%)
Oil.1	2	66.9	3.47	47.7–70.8	0.8	4.93
Oil.2	3	6.0	4.57	0–15.8	-0.95	4.73
Oil.3	3	63.6	7.65	48.9–70.9	1.28	9.86
Oil.4	4	36.7	15.30	31.6–46	1.61	15.65
Oil.5	5	13.3	5.58	8.9–17.4	1.82	9.38
Prot.1	3	6.0	4.62	0–11.8	0.58	9.29
Prot.2	4	4.1	4.60	0–12.4	0.61	9.25
Prot.3	4	36.7	5.34	31.6–46	-0.65	10.23
Prot.4	5	13.3	3.84	4.5–17.4	-0.74	5.23

For each QTL, the chromosome (Chr.), the position of the nearest marker of the LOD score peak with the LOD score at the corresponding marker as well as the confidence interval (CI) are indicated. The additive effect represents the mean effect on trait of the replacement of both Col-0 alleles by Ct-1 alleles at the QTL. R<sup>2</sup> represents the proportion of phenotypic variance of the trait explained by the QTL. cM, centiMorgans.

Near-infrared reflectance spectroscopic calibration models for these four components were established on entire seeds using PLSs regression and leave-one-out cross-validation technique. To assess the accuracy of each model, an external validation with samples not included in the initial model was carried out.

The four developed models display good performances as evaluated by different parameters of the external validation set such as  $r_V^2$ , coefficient of determination; SEP and RPD (**Table 1**). The  $r_V^2$  values range from 0.82 for carbon content to 0.99 for oil content, indicating that the four models developed in this paper show good to excellent quantitative information (Font et al., 2006). As expected,  $\text{SECV} \geq \text{SEC}$  (and then  $R_{\text{CV}}^2 \leq R_{\text{C}}^2$ ) for the four models and  $\text{SEP} \geq \text{SEC}$  (and then  $R_{\text{V}}^2 \leq R_{\text{C}}^2$ ) for three out of the four models (**Table 1**).  $\text{SEP} < \text{SEC}$  (and  $R_{\text{V}}^2 > R_{\text{C}}^2$ ) for the oil model, which is unexpected and illustrates that the restricted validation set (36 samples, i.e., about one third of calibration sample number) fits better to the model. For each model, the validation set shows statistics very close to the calibration set, illustrating robustness and absence of overfitting of the models. Concerning oil content, the model described in this paper for *Arabidopsis* display better performance than the ones described for rapeseed (Tkachuk, 1981; Hom et al., 2007; Rossato et al., 2013). Interestingly, the models were developed on entire seeds without destruction neither any special sample preparation, which is a great advantage compared to calibration developed on powder or oil for example (Khamchum et al., 2013) as it's faster and allow the seeds to be used for other applications.

Seeds produced by a plant are heterogeneous (in size and composition) depending of their position on the mother plant and the environmental conditions during their development. This could induce huge phenotypic variation when phenotyping is performed on very little amount of seeds. The protocol described in this paper overcomes this problem since spectra are determined on a large number of seeds (160 mg, i.e., about 8000 seeds), allowing robust sampling. In favorable environmental condition, an *Arabidopsis* plant produces on average 1 g of seeds, highlighting that the quantity required for NIRS analysis is not a limiting factor. However, in stressful conditions, *Arabidopsis* may produce very few seeds. In this case, NIRS will not be suitable for seed content analysis.

Using *pkp2* and *sweet11;12* described mutants and two different nitrogen nutritions, we demonstrated that NIRS is a powerful method to determine *Arabidopsis* seed composition and that NIRS can probably replace labor intensive methods such as fatty acyl methyl ester extraction followed by Gas chromatography analysis for lipids or elemental analyzer measurements for nitrogen and carbon content.

With NIRS calibrations in hand, natural variation of *Arabidopsis* seed composition was explored. Three independent cultures (C1, C2, and C3) of a 48 core-collection, Col-0 and minimal sets of eight RIL populations were performed, allowing the estimation of environmental effect on the four seed traits analyzed (seed lipid, protein, carbon, and nitrogen content). As shown in **Figures 3** and **4**, the four traits display a wide range of variation and are strongly impacted by the environment. However, most of the genotypes (75, 97, and 91% from C1, C2, and C3 respectively) display oil content between 32.8 and

43.8%, as already observed by O'Neill et al. (2003) while studying 360 accessions. Similarly, in our three experiments, Cvi-0 was recorded with low oil content (36.21, 33.56, and 34.90%) while Ct-1 was recorded with high oil content (46.26, 42.40, and 45.3%) as in O'Neill et al. (2003). Even though seed composition is strongly impacted by environmental conditions, the four traits analyzed are also controlled by genetic factor as illustrated by **Figure 4**. Indeed nine QTL were identified for seed oil and protein content in the Ct-1 × Col-0 RIL population (**Table 2**). Most of the QTL for oil content co-localized with QTL for protein content but with opposite effect on each traits, highlighting the negative correlation between seed oil and protein content. Oil.1 and Oil.2 co-localized with seed oil content QTL previously identified by Hobbs et al. (2004) in the Ler × Cvi-0 RIL population and by O'Neill et al. (2012) in the Cvi-0 × Ag-0 RIL population for Oil.1. Interestingly, Oil.1 does not co-localize with seed protein QTL in the Ct-1 × Col-0 RIL population, suggesting that Oil.1 may regulate oil content without affecting protein content and thus represents a very good candidate to specifically modify oil content without affecting protein content in *Arabidopsis* seed. Conversely, Prot.2 may regulate protein content independently of oil content and could be used to solely modify seed protein content. Fine mapping is required to confirm Oil.1 and Prot.2 specificity as well as to identify the genes under the nine QTL identified.

## CONCLUSION

In summary, the results of the present work show that NIRS predictive equations developed in this study can be used to reliably predict oil, protein, nitrogen and carbon content of *Arabidopsis* seed samples without destruction neither any special sample preparation. This high-throughput method opens the way for quantitative genetic such as QTL cloning (up to gene identification and not only detection), as well as GWASs but also to mutant library screening. As a first attempt to identify genetic factor controlling seed oil and protein content, QTL for these traits have been mapped in the *Arabidopsis* Ct-1 × Col-0 RIL population. Some of the oil content QTL detected co-localized with QTL identified previously, thus validated our approach, but many novel QTL were also identified. In particular, to our knowledge, this is the first report of seed protein content QTL in *Arabidopsis*. The fine mapping of some of these QTL is underway and should give new insights on the regulatory pathway involved in *Arabidopsis* seed oil and protein accumulation.

## METHODS

### Plant Material

The 48 core-collection of *Arabidopsis* (McKhann et al., 2004) in addition to Col-0 accession, *wri1-3*, *wri1-4*, *tag1-2*, *pkp2-1*, and *sweet11;12* mutants, a minimal set (20 lines) of eight RIL populations (2RV, 3RV, 7RV, 8RV, 13RV, 17RV, 20RV, and 21RV) as well as the core-pop of 164 RILs of the Ct-1 × Col-0

population were used in this study (Simon et al., 2008). *wri1-3*, *wri1-4*, *pkp2-1*, and *tag1-2* seeds were provided by S. Baud and *sweet11;12* seeds were provided by R. Le Hir. The other seeds were obtained from the Versailles Biological Resource Centre for *Arabidopsis*<sup>1</sup>. Seeds were sown on damp Whatman filters, stratified for 3 days at 4°C and then transferred to a growth cabinet under long-day conditions at 21°C for 2 days. Three seedlings (with emerging radicle) per genotype were planted in soil in 7 cm pots and transferred to a non-heated and naturally lit greenhouse to be vernalized from November to February. After 8 weeks, one plantlet per pot was randomly retained without phenotype selection. After 12 weeks of vernalization, plants were transferred to a growth chamber under long-day conditions (16/8 h photoperiod at 150 mmol photons m<sup>-2</sup> s<sup>-1</sup>); 21°C day temperature and 18°C night temperature; relative humidity of 65%. From this time, three times a week the plant trays were moved around the growth chamber to reduce position effects. Bags were put over the plants to prevent seed dispersion as soon as the first silique had turned yellow. The plants were no longer watered once the youngest silique had turned yellow. Plants were kept in the growth chamber until dry and then harvested.

Three cultures (C1, C2, and C3) including the 48 core-collection, Col-0, the minimal sets of the 8 RIL populations and *wri1-3*, *wri1-4*, and *tag1-2* mutants for C3, as well as one culture (C4) of the core-pop of 164 RILs of the Ct-1 × Col-0 population were performed following this protocol.

## Near-Infrared Spectroscopy

### NIR Spectra Acquisition

Seed samples were placed in a 9 mm diameter clear glass bottle (Agilent, 5182-0714) on 4 mm height for NIRS spectra acquisition and were analyzed as intact (without any treatment). This corresponds to about 300 µl of *Arabidopsis* seeds (about 160 mg or 8 000 seeds).

Spectra acquisition was performed with a Fourier transform near-infrared (FTNIR) analyzer (Antaris II spectrometer; ThermoFisher Scientific, France). Spectra were collected in reflectance mode with an 8 cm<sup>-1</sup> optical resolution and were obtained as an average of 16 scans. Spectra were collected over the range 4000 to 10000<sup>-1</sup> and calibrations done using four spectral ranges: from 4100 to 4940 cm<sup>-1</sup>; from 5390 to 6690 cm<sup>-1</sup>; from 6900 to 7130 cm<sup>-1</sup>, and from 7185 to 9000 cm<sup>-1</sup>. These spectral regions provide useful information about the organic signature of the *Arabidopsis* samples and exclude the water spectral regions. They have been selected by looking at the regression vector from the PLS (see Development of NIRS Calibration Models) and using a Thermo proprietary pure component algorithm.

### Selecting the Samples for NIRS Calibration

The robustness and accuracy of a NIRS model are strongly dependent on the accuracy of the reference method but also of the samples chosen for calibration development. Indeed, the calibration samples have to be representative of the

spectral variability and must cover the range of the component concentration of the samples that will be further monitored.

As the NIR spectral variability of *Arabidopsis* seeds was not known, NIR spectra of 650 samples (one spectrum per sample) from two independent cultures (C1 and C2) were collected. Spectra were treated with a multiplicative signal correction (MSC) to correct multiplicative effects due to light scattering in spectral data and a Principal component analysis (PCA) was performed in order to select samples maximizing spectral variability. PC1 and PC2 explained 84.5 and 11.9% of the spectral variation respectively. Their graphic representations were similar to a seed spectrum and suggested that they reflect variations due to differences in spectra baseline or particle size for example. PC3 explained 1.9% of the spectral variation and its graphic representation displayed peaks at wavelengths specific to seed storage compounds. Thus the PC1/PC3 graph was used as a criterion for selecting 100 samples in the population as being more variable on the basis of spectra features (Shenk and Westerhaus, 1991).

Seed oil content of these 100 samples was determined by the Soxhlet reference method and a preliminary calibration model was set up. Using this model, seed oil content of samples available at this time (1788 samples from three independent cultures and including *wri1* and *tag1* low-seed-oil insertion mutants) was predicted. This prediction allowed the selection of 48 additional samples with extreme values (maximal and minimal) in order to extend the range of concentration of the final calibration set. Seed oil content was measured with the Soxhlet method on these 48 additional samples. The same procedure was applied to choose samples for seed protein, carbon, and nitrogen content calibration models.

### Development of NIRS Calibration Models

Calibration models were developed using TQ Analyst software (ThermoFisher Scientific, France) using PLSs regression and leave-one-out cross-validation technique. Prior to the PLS regression, all spectra were pre-treated with the scatter correction MSC and by applying a first derivative transformation and a Norris derivative filter (segment length 5, gap size: 5). The use of derivative spectra instead of the raw optical data to perform calibration is a way of solving problems associated with offsets and overlapping peaks.

Near-infrared reflectance spectroscopic calibration models were established for oil, protein, carbon, and nitrogen content by using a number of PLS factors optimal for each component (i.e., only the primary, most important factors were used, the “noise” being encapsulated in the less important factors). The optimal number of PLS factors was determined as the minimum of the PRESS (predicted residual error sum of squares) curve when doing a leave-one-out cross validation method.

The quality of each calibration model was then evaluated by several parameters: the determination coefficients between concentrations predicted from NIRS and from reference analysis,  $r_c^2$  and  $r_{CV}^2$ , calculated for calibration and cross-validation (leave-one-out) data processing respectively, and their respective standard errors [calibration (SEC) and cross-validation (SECV)].

<sup>1</sup><http://publiclines.versailles.inra.fr/>

To assess the accuracy of each newly developed calibration model, an external validation with samples not included in the initial model was carried out. The total number of samples was divided into calibration and external validation sets in a rate 3:1. For that purpose, the samples were ranked according to their reference values and then about one sample every four was assigned to the external validation set. In addition, to account for environmental variation in seed composition, the seed samples chosen for the calibration and external validation sets were derived from the three cultures (C1, C2, and C3). The prediction quality of NIRS analyses was then quantified by the SEP and the determination coefficient ( $r^2_V$ ) between concentrations obtained from NIRS and from reference analysis for the validation set. The RPD was calculated as the ratio between the SD of the reference values and the SEP. RPD is indicative of the usefulness of the NIRS calibrations.

## Analysis of Seed Oil Content (Reference Method)

Oil was extracted following the standard NF V03-908 protocol (extraction by hot solvent with a “Soxhlet” extractor). About 1 g of seed was dried (103°C during 20 h) and ground in hexane with a grinder. Oil was then extracted with hexane by the Soxhlet method. The total seed oil content was expressed as percentage of the dried seed weight.

## Analysis of Seed Protein Content (Reference Method)

Phenol extraction of seed protein was adapted from Meyer et al. (1988). Ten mg of seeds were homogenized in a 2 ml tube containing a ceramic bead and 1 ml of an emulsion of 50% (v/v) phenol (previously equilibrated in 1 M Tris HCl pH8) in 0.1 M Tris HCl pH8 1% SDS using a Fastprep-24 Instrument (MP-Biomedical, maximal intensity, twice 1 min). After centrifugation (13 000 g, 20 min), 200 µl of the phenolic phase was accurately delipidated twice with 500 µl of hexane. One hundred µl of the phenol phase was taken after centrifugation (13 000 g, 10 min) and the proteins were precipitated with five volumes of methanol containing 0.1 M ammonium acetate at -20°C overnight. The precipitate was collected by centrifugation and washed four times with methanol (-20°C) containing 0.1 M ammonium acetate, and twice with 80% acetone (in water). The resulting pellets were dried under reduced pressure and then resuspended in 1 ml of 0.1 M Tris-HCl pH8 1% SDS. After overnight agitation, the fully dissolved solution was then cleared by centrifugation (13 000 g, 10 min) and the protein concentration was determined by spectrometry at 280 nm, assuming that 1 OD corresponds to 1mg/ml protein solution.

## Seed Nitrogen and Carbon Content (Reference Method)

Five mg of seeds, dried overnight at 100°C, were weighed on a lab balance model M2P (Sartorius, Göttingen, Germany) with a readability of 0.001 mg, then analyzed for nitrogen and carbon concentration by the Dumas combustion method (Anonymous,

1990) with an automated CN analyzer (Heraeus CN-Rapid, Hanau, Germany).

## QTL Detection

For each RIL, the mean value from three plants was taken for each measured trait for QTL analysis.

Quantitative trait loci analyses were performed using R/qtl library in the R environment (Broman et al., 2003; Arends et al., 2010) with standard methods for interval mapping (IM) and multiple QTL mapping (MQM) (Arends et al., 2010). First, IM was carried out to determine putative QTL involved in the variation of the trait, and then MQM model was performed on the same data: the closest marker to each local logarithm-of-odds (LOD) score peak (putative QTL) was used as a cofactor to control the genetic background while testing at another genomic position. The significance threshold ( $p < 0.05$ ) of LOD was determined by permutation test ( $n = 1000$ ) for each trait (Churchill and Doerge, 1994). The estimated additive effect (representing the mean effect of the replacement of the *Col-0* alleles by *Ct-1* alleles at the locus) and the percentage of variance explained by each QTL ( $R^2$ ) affecting a trait were obtained for the final MQM model.

## AUTHOR CONTRIBUTIONS

PG, PB-M, and SJ established NIRS models. PG and SJ performed the statistical analysis and wrote the manuscript. AL and MD carried out the plant cultures, seed protein content (reference method) measurements and NIR spectrum acquisitions and predictions. PG and AL performed seed carbon and nitrogen content analysis (reference method). SJ performed the QTL detection experiments. All authors read and approved the final manuscript.

## FUNDING

This work was supported by the INRA “Biology and Plant Breeding” Department. The IJPB benefits from the support of the Labex Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS).

## ACKNOWLEDGMENTS

Soxhlet oil extractions were performed by M. Krouti (Terres Inovia, Ardon, France). We thank F. Chardon and M. Reymond for critical reading of the manuscript. We thank S. Baud and R. Le Hir for *pkp2-1* and *sweet11;12* seeds respectively, as well as A. Marmagne for *Col-0* and *Ws* seeds from plants grown on 2 and 10 mM nitrate. We also thank Lilian Dahuron, Philippe Marechal, and Sébastien Bénard for plant care.

## REFERENCES

- AACC International (1999). *Approved Methods of Analysis, Method 39-00.01. Near-Infrared Methods—Guidelines for Model Development and Maintenance*. Approved November 3, 1999, 11th Edn. St. Paul, MN: AACC International. doi: 10.1094/AACCIIntMethod-39-00.01
- Anonymous (1990). "Crude protein in animal feed combustion method," in *Official Methods of Analysis*, 15th Edn, ed. K. Helrich (Rockville, MD: Association of Official Analytical Chemists, Inc.).
- Arends, D., Prins, P., Jansen, R. C., and Broman, K. W. (2010). R/qtL: high-throughput multiple QTL mapping. *Bioinformatics* 26, 2990–2992. doi: 10.1093/bioinformatics/btq565
- Batten, G. D. (1998). Plant analysis using near infrared reflectance spectroscopy: the potential and the limitations. *Aust. J. Exp. Agric.* 38, 697–706. doi: 10.1071/ea97146
- Baud, S., Wuilleme, S., Dubreucq, B., de Almeida, A., Vuagnat, C., Lepiniec, L., et al. (2007). Function of plastidial pyruvate kinases in seeds of *Arabidopsis thaliana*. *Plant J.* 52(3), 405–419. doi: 10.1111/j.1365-313X.2007.03232.x
- Beisson, F., Koo, A. J., Ruuska, S., Schwender, J., Pollard, M., Thelen, J. J., et al. (2003). *Arabidopsis* genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. *Plant Physiol.* 132, 681–697. doi: 10.1104/pp.103.022988
- 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
- Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtL: QTL mapping in experimental crosses. *Bioinformatics* 19, 889–890. doi: 10.1093/bioinformatics/btg112
- Chardon, F., Jasinski, S., Durandet, M., Lecureuil, A., Soulay, F., Bedu, M., et al. (2014). QTL meta-analysis in *Arabidopsis* reveals an interaction between leaf senescence and resource allocation to seeds. *J. Exp. Bot.* 65, 3949–3962. doi: 10.1093/jxb/eru125
- Chen, L. Q., Lin, I. W., Qu, X. Q., Sosso, D., McFarlane, H. E., Londono, A., et al. (2015). A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the *Arabidopsis* embryo. *Plant Cell* 27, 607–619. doi: 10.1105/tpc.114.134585
- Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971.
- Durrett, T. P., Benning, C., and Ohlrogge, J. (2008). Plant triacylglycerols as feedstocks for the production of biofuels. *Plant J.* 54, 593–607. doi: 10.1111/j.1365-313X.2008.03442.x
- Dyer, J. M., Stymne, S., Green, A. G., and Carlsson, A. S. (2008). High-value oils from plants. *Plant J.* 54, 640–655. doi: 10.1111/j.1365-313X.2008.03430.x
- Eskandari, M., Cober, E. R., and Rajcan, I. (2012). Genetic control of soybean seed oil: I. QTL and genes associated with seed oil concentration in RIL populations derived from crossing moderately high-oil parents. *Theor. Appl. Genet.* 126, 483–495. doi: 10.1007/s00122-012-1995-3
- Focks, N., and Benning, C. (1998). wrinkled1: a novel, low-seed-oil mutant of *Arabidopsis* with deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant Physiol.* 118, 91–101. doi: 10.1104/pp.118.1.91
- Font, R., del Rio, M., Fernandez, J. M., and de Haro, A. (2003). Acid detergent fiber analysis in oilseed *Brassicas* by near-infrared spectroscopy. *J. Agric. Food Chem.* 51, 2917–2922. doi: 10.1021/jf021011
- Font, R., del Rio, M., Fernandez-Martinez, J. M., and de Haro-Bailon, A. (2004). Use of near-infrared spectroscopy for screening the individual and total glucosinolate contents in Indian mustard seed (*Brassica juncea* L. Czern. & Coss.). *J. Agric. Food Chem.* 52, 3563–3569. doi: 10.1021/jf0307649
- Font, R., del Rio-Celestino, M., and de Haro-Bailon, A. (2006). The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. *Ind. Crops Prod.* 24, 307–313. doi: 10.1016/j.indcrop.2006.06.012
- Font, R., Rio, M. D., Fernandez, J., Arthur, E., Bancroft, I., Chinoy, C., et al. (2002a). Seed oil content analysis of Ethiopian mustard (*Brassica carinata* A. Braun) by near infrared spectroscopy. *Cruciferae Newsletter* 24, 5–6.
- Font, R., Rio, M. D., Sillero, A., Arthur, E., Bancroft, I., Chinoy, C., et al. (2002b). Using near infrared spectroscopy for determining protein content in Ethiopian mustard (*Brassica carinata* A. Braun). *Cruciferae Newsletter* 24, 7–8.
- Hobbs, D. H., Flintham, J. E., and Hills, M. J. (2004). Genetic control of storage oil synthesis in seeds of *Arabidopsis*. *Plant Physiol.* 136, 3341–3349. doi: 10.1104/pp.104.049486
- Hom, N. H., Becker, H. C., and Mollers, C. (2007). Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds. *Euphytica* 153, 27–34. doi: 10.1007/s10681-006-9195-3
- Hwang, E. Y., Song, Q., Jia, G., Specht, J. E., Hyten, D. L., Costa, J., et al. (2014). A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics* 15:1. doi: 10.1186/1471-2164-15-1
- Irzykowska, L., and Wolko, B. (2004). Interval mapping of QTLs controlling yield-related traits and seed protein content in *Pisum sativum*. *J. Appl. Genet.* 45, 297–306.
- Jasinski, S., Lecureuil, A., Miquel, M., Loudet, O., Raffaele, S., Froissard, M., et al. (2012). Natural variation in seed very long chain fatty acid content is controlled by a new isoform of KCS18 in *Arabidopsis thaliana*. *PLoS ONE* 7:e49261. doi: 10.1371/journal.pone.0049261
- Khamchum, C., Punsuvon, V., Kasemsumran, S., and Suttiwijitpukdee, N. (2013). A feasibility study of oil content and fatty acid composition of seed powder and seed oil of *Pongamia pinnata* by near infrared spectroscopy. *Scienceasia* 39, 384–391. doi: 10.2306/scienceasia1513-1874.2013.39.384
- Kianian, S. F., Egli, M. A., Phillips, R. L., Rines, H. W., Somers, D. A., Gengenbach, B. G., et al. (1999). Association of a major grain oil content QTL and an acetyl-CoA carboxylase gene in oat. *Theor. Appl. Genet.* 98, 884–894. doi: 10.1007/s001220051147
- Kumar, S., You, F. M., Duguid, S., Booker, H., Rowland, G., and Cloutier, S. (2015). QTL for fatty acid composition and yield in linseed (*Linum usitatissimum* L.). *Theor. Appl. Genet.* 128, 965–984. doi: 10.1007/s00122-015-015-3
- Li, F., Chen, B., Xu, K., Wu, J., Song, W., Bancroft, I., et al. (2014). Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). *DNA Res.* 21, 355–367. doi: 10.1093/dnaresearch/dsu002
- Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., et al. (2012). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* 45, 43–50. doi: 10.1038/ng.2484
- Liu, G. Z., Mei, H. X., Wang, S., Li, X. H., Zhu, X. F., and Zhang, T. Z. (2015). Association mapping of seed oil and protein contents in upland cotton. *Euphytica* 205, 637–645. doi: 10.1007/s10681-015-1450-z
- Liu, P., Wang, C. M., Li, L., Sun, F., and Yue, G. H. (2011). Mapping QTLs for oil traits and eQTLs for oleosin genes in jatropha. *BMC Plant Biol.* 11:132. doi: 10.1186/1471-2229-11-132
- Masclaux-Daubresse, C., and Chardon, F. (2011). Exploring nitrogen remobilization for seed filling using natural variation in *Arabidopsis thaliana*. *J. Exp. Bot.* 62, 2131–2142. doi: 10.1093/jxb/erq405
- McKhann, H. I., Camilleri, C., Berard, A., Bataillon, T., David, J. L., Reboud, X., et al. (2004). Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *Plant J.* 38, 193–202. doi: 10.1111/j.1365-313X.2004.02034.x
- Meyer, Y., Grosset, J., Chartier, Y., and Cleyetmarel, J. C. (1988). Preparation by two-dimensional electrophoresis of proteins for antibody-production - antibodies against proteins whose synthesis is reduced by auxin in tobacco mesophyll protoplasts. *Electrophoresis* 9, 704–712. doi: 10.1002/elps.1150091105
- Mokrani, L., Gentzbittel, L., Azanza, F., Fitamant, L., Al-Chaarani, G., and Sarrafi, A. (2002). Mapping and analysis of quantitative trait loci for grain oil content and agronomic traits using AFLP and SSR in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 106, 149–156.
- O'Neill, C. M., Gill, S., Hobbs, D., Morgan, C., and Bancroft, I. (2003). Natural variation for seed oil composition in *Arabidopsis thaliana*. *Phytochemistry* 64, 1077–1090. doi: 10.1016/S0031-9422(03)00351-0
- O'Neill, C. M., Morgan, C., Hattori, C., Brennan, M., Rosas, U., Tschoep, H., et al. (2012). Towards the genetic architecture of seed lipid biosynthesis and accumulation in *Arabidopsis thaliana*. *Heredity* 108, 115–123. doi: 10.1038/hdy.2011.54
- Parkin, I. A., Gulden, S. M., Sharpe, A. G., Lukens, L., Trick, M., Osborn, T. C., et al. (2005). Segmental structure of the *Brassica napus* genome based on

- comparative analysis with *Arabidopsis thaliana*. *Genetics* 171(2), 765–781. doi: 10.1534/genetics.105.042093
- Plessis, A., Ravel, C., Bordes, J., Balfourier, F., and Martre, P. (2013). Association study of wheat grain protein composition reveals that gliadin and glutenin composition are trans-regulated by different chromosome regions. *J. Exp. Bot.* 64, 3627–3644. doi: 10.1093/jxb/ert188
- Rossato, R., Prete, C. E. C., de Castro, C., Tomm, G. O., Leite, R. S., Mandarino, J. M. G., et al. (2013). Predicting rapeseed oil content with near-infrared spectroscopy. *Pesqui. Agropecuaria Bras.* 48, 1601–1604. doi: 10.1590/s0100-204x2013001200010
- Sanyal, A., and Linder, C. R. (2012). Quantitative trait loci involved in regulating seed oil composition in *Arabidopsis thaliana* and their evolutionary implications. *Theor. Appl. Genet.* 124, 723–738. doi: 10.1007/s00122-011-1742-1
- Shenk, J. S., and Westerhaus, M. O. (1991). Population structuring of near-infrared spectra and modified partial least-squares regression. *Crop Sci.* 31, 1548–1555. doi: 10.2135/cropsci1991.0011183X003100060034x
- Shewry, P. R., Napier, J. A., and Tatham, A. S. (1995). Seed storage proteins: structures and biosynthesis. *Plant Cell* 7, 945–956. doi: 10.1105/tpc.7.7.945
- Simon, M., Loudet, O., Durand, S., Berard, A., Brunel, D., Sennesal, F. X., et al. (2008). Quantitative trait loci mapping in five new large recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. *Genetics* 178, 2253–2264. doi: 10.1534/genetics.107.083899
- Sun, M., Hua, W., Liu, J., Huang, S., Wang, X., Liu, G., et al. (2012). Design of new genome- and gene-sourced primers and identification of QTL for seed oil content in a specially high-oil *Brassica napus* cultivar. *PLoS ONE* 7:e47037. doi: 10.1371/journal.pone.0047037
- Tar'an, B., Warkentin, T., Somers, D. J., Miranda, D., Vandenberg, A., Blade, S., et al. (2004). Identification of quantitative trait loci for grain yield, seed protein concentration and maturity in field pea (*Pisum sativum* L.). *Euphytica* 136, 297–306. doi: 10.1023/B:EUPH.0000032721.03075.a0
- Tkachuk, R. (1981). Oil and protein-analysis of whole rapeseed kernels by near-infrared reflectance spectroscopy. *J. Am. Oil Chem. Soc.* 58, 819–822. doi: 10.1007/bf02665588
- Valentini, G., Chiarelli, L., Fortin, R., Speranza, M. L., Galizzi, A., and Mattevi, A. (2000). The allosteric regulation of pyruvate kinase. *J. Biol. Chem.* 275, 18145–18152. doi: 10.1074/jbc.M001870200
- Velasco, L., and Becker, H. C. (1998). Analysis of total glucosinolate content and individual glucosinolates in *Brassica* spp. by near-infrared reflectance spectroscopy. *Plant Breed.* 117, 97–102. doi: 10.1111/j.1439-0523.1998.tb01459.x
- Yang, X., Ma, H., Zhang, P., Yan, J., Guo, Y., Song, T., et al. (2012). Characterization of QTL for oil content in maize kernel. *Theor. Appl. Genet.* 125, 1169–1179. doi: 10.1007/s00122-012-1903-x
- Ying, J. Z., Shan, J. X., Gao, J. P., Zhu, M. Z., Shi, M., and Lin, H. X. (2012). Identification of quantitative trait loci for lipid metabolism in rice seeds. *Mol. Plant* 5, 865–875. doi: 10.1093/mp/ssr100
- Zheng, P., Allen, W. B., Roesler, K., Williams, M. E., Zhang, S., Li, J., et al. (2008). A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Nat. Genet.* 40, 367–372. doi: 10.1038/ng.85

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

Copyright © 2016 Jasinski, Lécureuil, Durandet, Bernard-Moulin and Guerche. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Spectral Knowledge (SK-UTALCA): Software for Exploratory Analysis of High-Resolution Spectral Reflectance Data on Plant Breeding

Gustavo A. Lobos<sup>1\*</sup> and Carlos Poblete-Echeverría<sup>2,3\*</sup>

<sup>1</sup> Plant Breeding and Phenomic Center, Facultad de Ciencias Agrarias, PIEI Adaptación de la Agricultura al Cambio Climático, Universidad de Talca, Talca, Chile, <sup>2</sup> Escuela de Agronomía, Pontificia Universidad Católica de Valparaíso, Quillota, Chile,

<sup>3</sup> Department of Viticulture and Oenology, Faculty of AgriSciences, Stellenbosch University, Matieland, South Africa

## OPEN ACCESS

### Edited by:

Marcos Egea-Cortines,  
Universidad Politécnica de Cartagena,  
Spain

### Reviewed by:

Ankush Prashar,  
Newcastle University, UK  
Alan Gay,  
Aberystwyth University, UK

### \*Correspondence:

Carlos Poblete-Echeverría  
cpe@sun.ac.za  
Gustavo A. Lobos  
globosp@utalca.cl

### Specialty section:

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

Received: 17 July 2016

Accepted: 16 December 2016

Published: 09 January 2017

### Citation:

Lobos GA and Poblete-Echeverría C (2017) Spectral Knowledge (SK-UTALCA): Software for Exploratory Analysis of High-Resolution Spectral Reflectance Data on Plant Breeding. *Front. Plant Sci.* 7:1996.  
doi: 10.3389/fpls.2016.01996

This article describes public, free software that provides efficient exploratory analysis of high-resolution spectral reflectance data. Spectral reflectance data can suffer from problems such as poor signal to noise ratios in various wavebands or invalid measurements due to changes in incoming solar radiation or operator fatigue leading to poor orientation of sensors. Thus, exploratory data analysis is essential to identify appropriate data for further analyses. This software overcomes the problem that analysis tools such as Excel are cumbersome to use for the high number of wavelengths and samples typically acquired in these studies. The software, Spectral Knowledge (SK-UTALCA), was initially developed for plant breeding, but it is also suitable for other studies such as precision agriculture, crop protection, ecophysiology plant nutrition, and soil fertility. Various spectral reflectance indices (SRIs) are often used to relate crop characteristics to spectral data and the software is loaded with 255 SRIs which can be applied quickly to the data. This article describes the architecture and functions of SK-UTALCA and the features of the data that led to the development of each of its modules.

**Keywords:** phenotyping, phenomic, scan, wavelength, noise, outlier, spectral reflectance index (SRI), collinearity

## INTRODUCTION

The responses of any living organism are ultimately controlled by genes (G), but the expression of these are modulated in several ways, partly because of the action of other genes, and the complex interaction between them, but mostly in response to the environment (E) where the plant grows and develops (GxE interaction). Gene sequencing is becoming more routine, economical, and fast, but for proper analysis and interpretation of the information an adequate phenotypic characterization is essential, even though it poses one of the greatest difficulties (Lörz and Wenzel, 2005; Finkel, 2009; Lobos et al., 2014; Estrada et al., 2015).

Progress in science and technology have made it possible to study different processes involved in multiple areas of knowledge. In agronomy and biological sciences, sensors, and instrumentation have been developed to characterize the behavior of a particular organism, or a group of them, under a specific environmental condition or situation.

Currently, equipment, techniques, and analyses are available that have proved helpful in characterizing the phenotype (phenotyping), and in the case of remote sensing, quick and high predictive power (Lobos and Hancock, 2015; Camargo and Lobos, 2016).

Among the available remote sensing tools, spectrometers, or spectroradiometers mainly exploit the principle of quantifying the proportion of reflected radiation by an object relative to the incident radiation (Borengasser et al., 2008). Reflectance (graphically represented by the spectral signature) is related to the absorption and transmission of each wavelength, thus representing plant status under ambient or experimental conditions (Garriga et al., 2014). For example, compared to a senescing plant, a healthy one should absorb more in the visible (blue and red light) and reflect more in the near infrared range.

Nowadays plant reflectance can be measured from space by satellites (with certain limitations on interpretation due to pixel resolution) or from the troposphere by manned and unmanned aerial vehicles (problems related to the number and resolution of the spectrum bands; Araus and Cairns, 2014). Equipment used on the ground covers a wider range of the spectrum, with a better resolution. The most modern devices not only measure into the near infrared region (700–1300 nm; NIR), but also from the ultraviolet (~200 nm) up to the short wavelength infrared (~2500 nm) (Cabrera-Bosquet et al., 2012). This high-resolution technology allows examination of plants beyond the 1000 nm region of the spectrum, with great potential for phenotype prediction (Garbulsky et al., 2011; White et al., 2012; Araus and Cairns, 2014; Lobos et al., 2014).

Several critical issues for making good reflectance measurements in the field have been reported in the literature (e.g., Curtiss and Goetz, 1994; Milton et al., 1995; Salisbury, 1998; Schaeppman, 1998; Curtiss and Goetz, 2001). Independent of the equipment used on the ground, a correct measurement of the reflectance in the field is mandatory. Ideally, measurements should be restricted to clear sky conditions, performing a radiometric calibration every 10–15 min to limit variations in reflectance induced by changes in the angle of the sun, and taking in account basic but important considerations such as maintaining the same orientation, angle, and distance to the canopy on each assessed plot or ensuring dark colored clothing for the operators. Although instrument settings vary among brands and models, a number of steps should be followed to optimized data capture. The equipment should be turned on in advance to allow the device to equilibrate with the ambient temperature, the integration time for a single scan or sample needs to be defined (maximizing sensitivity, but avoiding saturation), the number of scans per sample or samples per plot and the convenience of averaging them before data processing should be determined, and the exact sequence for checking darks and standards recommended by the manufacturer needs to be ascertained.

In general, due to its simplicity and ability to forecast several phenotypic characteristics, reflectance is used to calculate “Spectral Reflectance Indices” (SRIs) (Lobos and Hancock, 2015). SRIs are based on relationships between wavelengths or spectrum bands, usually designed to be relatively immune to changes in solar radiation between measurements, relating them quantitatively to changes in plant phenotype (Mullan, 2012). Today there are hundreds of SRIs proposed to estimate different traits (e.g., leaf area index, yield, gas exchange, fluorescence, pigment content, plant water status, carbon isotopic

**TABLE 1 | Nomenclature related to spectrometer data collection.**

In this article	In other articles or manuals	Definition
Plot		Land area where a single genotype is growing; in other studies it could be considered as a replication.
Scan	Data collection, spectrum or spectra collection, scanning	Action oriented to collect the spectrum or spectra by one scan or shoot (informal terminology).
Samples	Sample spectra, samples of scan, scanned samples, scanned data, artifacts, features	Some spectrometers are able to register several samples within the same scan; number of spectral signatures captured per scan.
Integrations per sample	Spectrum average or averaging	Integration of spectra within the same sample.

discrimination, etc.) but because of the lack of tools capable of assessing several SRIs at the same time, most of the published works focus on a small percentage of them (e.g., SR, NDVI, WI, NDWI, PRI, SAVI, etc.).

In breeding programs, there is a need to regularly evaluate hundreds or thousands of genotypes in a short time. Therefore, due to the time and cost involved breeders have not been able to perform a thorough phenotypic characterization of the material, limiting their evaluations to the yield, its components and some others traits that are relatively easy to assess (Kipp et al., 2014; Lobos and Hancock, 2015; Camargo and Lobos, 2016). With the emergence of phenomics, which is the acquisition of high-dimensional phenotypic data (high-throughput phenotyping) for characterization of the phenotype of organisms in a multidimensional manner (Houle et al., 2010; Kipp et al., 2014), measurements that used to take weeks or months can now be performed in a few hours (White et al., 2012; Lobos and Hancock, 2015). The implementation of phenomics in plant breeding programs is relatively new, and is an area where more development is likely to be needed (Lobos and Hancock, 2015).

For a correct interpretation of spectral reflectance data, it is essential to have reliable and representative information, especially when it comes from field measurements. The use of reflectance data in breeding programs has several advantages but probably the major problem is the amount of data originated by the numbers of wavelengths and genotypes assessed. If the reflectance data is analyzed in a conventional way (e.g., Excel files), the detection of measurement errors, the study of the spectral noise (originating from absorption by environmental compounds such as water or CO<sub>2</sub>) or the relationship between a specific wavelength and a response variable become difficult or subjective. Nevertheless, as far as we are aware there is currently no free software available that allows detailed exploratory analysis of high-resolution spectral reflectance data. Therefore, the aim of this article is to present an overview of the architecture and functions of Spectral Knowledge (SK-UTALCA), software that has been specially developed for exploratory analysis of

high-resolution spectral reflectance data, with applications in plant breeding research and also in many other fields.

Due to the broad nomenclature related to spectral measurements, some definitions are given in order to facilitate the understanding of this article and software (**Table 1**).

## MAIN SK-UTALCA ARCHITECTURE AND FUNCTIONALITIES

Spectral Knowledge (SK-UTALCA) is a software package developed in Matlab® and is available compiled for use in a Windows 64-bit environment from a download link or as source

code in supplementary material. This program allows, in an efficient and versatile manner, two types of actions: (i) cleaning of the data matrix by studying the spectral noise, and detecting within- and between- measurement errors; and (ii) application of a preliminary analysis of wavelength collinearity, and the detection of wavelengths or SRIs related to a response variable (**Table 2**).

After X and Y are loaded, the user can perform any main command, without a specific order. At the same time, each analysis can be run considering the previous exploration (*Run from current data*) or from the original data (loaded as X) (*Run from original data*). On each section of the cleaning data matrix

**TABLE 2 | Main SK-UTALCA functionalities according to the program menu.**

Main objective	Main commands	Secondary commands	Description
Input and output of information	Import X and Y data	Spectral data (X)	Import spectral data: first column or row (depends on the equipment) must include the assessed wavelengths.
		Samples per plot	Indicate samples per scan (definitions in <b>Table 1</b> ).
		Transpose data	Software works only with wavelengths as columns; the user will be able to transpose their data.
		Response variable (Y)	Import response variables data (on columns) where the three first columns must be codes (free criteria).
	Export data	Average	It is possible to export the average of the samples per scan or each sample individually.
		Empty data	Data can be exported including or excluding cells deleted during the cleaning of the data matrix.
Cleaning data matrix	Noise analysis	Wavelength segments	Ten different segments to analyze in relation to the percentage change among a determined neighbor size. Noise elimination can be applied equally to all data (Group) or for each sample (Individual). Additionally, negative values can be also deleted.
		Maximum variation coefficient	Criteria to select samples within a same scan where the variation coefficient, at any wavelength, is lower than the established threshold (Scans without problems) and those that exceeded it (Scans with problems).
		Samples to delete	If there are inconsistencies in one or more samples within the same scan, it is possible to select and delete them.
	Outlier analysis		Through a graphical analysis of the cloud of data points (response variable vs. SRI), it is possible to detect those out of range, identify the source of the problem and delete them in the case of clear evidence of a mistake.
	Preliminary analysis	Collinearity analysis	For a given response variable, through linear or artificial neural network (ANN) analysis, it is possible to identify wavelengths without collinearity.
		Individual wavelength analysis	Through different regression models and statistical parameters, it is possible to identify wavelengths better associated with a given response variable.
		SRI analysis	Through different regression models and a coefficient of determination threshold, it is possible to identify SRIs that are better associated with a given response variable. The software will be launched with a database of 255 SRIs (Supplementary Table 1).
		Detailed index report	For subsequent graphical representation it is possible to export, for each genotype or measurement, individual values of SRIs and response variables.

or the preliminary analysis, the user will be able to export the analyzed information (csv format).

## Main Menu (Data Set)

In the main menu, users are required to load the spectral reflectance data [*Spectral data (x)*]. The first step in the use of SK-UTALCA software is to load the data set in Microsoft Excel format (xlsx format). To read the spectral data it is necessary to indicate the number of samples taken in each plot (*Samples per plot*). Depending on the equipment used, reflectance data is organized in columns or rows. In the software, the spectral bands need to be located in columns and the spectral data in rows (it is possible to use the “*Transpose data*” option to relocate the data set as needed). The first wavelength measured must be in the second column (the first column is for codification purposes and depends only on the user).

The second file needed (xlsx format) contains the values for the independent variables [*Response variables (y)*] for each plot. In this case, the spreadsheet must consider three codification columns and the first variable should be allocated to the fourth column.

The software has no limitation on the number of spectral data points (wavelengths or measurements) or response variables.

## Noise Analysis

This first filter removes the spectral noise originated by the natural presence of certain elements in the atmosphere, such as water and carbon dioxide, which absorb specific wavelengths (Salisbury, 1998; Curtiss and Goetz, 2001; Psomas et al., 2005; Ma and Chen, 2006; Clevers et al., 2008). Researchers who screen hundreds or thousands of genotypes under field conditions usually consider at least three or four spectral samples per plot, generating a matrix of data that makes it difficult to objectively select the noise segment(s) for deletion. Furthermore, spreadsheet graphical options are usually restricted to a maximum number of data series per chart (e.g., ~255 in Excel for Windows or Mac), so there is no easy way to take a decision based on this tool. For this reason, for breeding purposes, conventional visual noise elimination is not a real alternative, restricting the criteria to the assumption of arbitrary limits, usually following thresholds from a third person or related articles.

With this module, it will be possible to analyze the spectral noise by considering up to ten independent segments. This will allow the user to set up different criteria in each segment, being more or less strict depending on the wavelengths analyzed, the data collected or previous knowledge. To apply the filter on each spectral signature, it is necessary to indicate the lower and upper limit for each segment (*Wavelength segments*), the maximum accepted percentage of variations (%) between two neighboring wavelengths, and the number of neighbors (*N size*) where the previous condition is found consecutively. The graphic window will show red crosses where the first criterion is satisfied and black ones where both have been met, this last condition determining where the software will perform the cleaning.

However, an objective selection is not the only important aspect of spectral noise. During the day there are environmental

changes (e.g., relative humidity) that not only affect the magnitude of each problematic wavelength, but also the number of wavelengths involved. For instance, measurements performed under conditions of higher relative humidity (usually before midday) produce wider noise segments beyond 1000 nm; if the determination of the number of wavelengths to eliminate considers measurements across the whole day, the noise edges will be established by genotypes evaluated early in the day (broader noise segments), risking the loss of important spectral information from those assessed under lower relative humidity (usually after midday) and therefore possessing narrower noise segments.

Because of this, after the noise selection criteria (% and *N Size*) are established, the user has an opportunity to filter by considering all the measurements as one group (*Group*) or as individual scans (*Individual*). When the *group* filter is selected in a specific segment, the program analyzes each sample where the selection criteria are met, identifying the minor and major wavelengths that have problems in the spectral data file, and uses these two wavelengths to eliminate the noise from each sample uniformly. This is very similar to what is done visually, but with an objective approach. For the *individual* option, each sample will be filtered independently from the others, rescuing important information for modeling, or the use of SRIs.

## Scan Analysis

In this module, the user will be able to analyze, identify and correct inconsistencies between spectral signatures from the same scan or plot, a problem that is often unnoticed. In general, for simplicity or to dilute any errors generated while collecting the data, there is a tendency to average samples within the same scan, which most of the time is done without any deeper analysis. As mentioned before, this should not be a complication when the data analysis considers a few measurements, but in breeding programs this search would be time consuming.

There are several aspects influencing the homogeneity between samples within the same scan, especially if the measurements were performed under field conditions. Unnoticed modification of the measurement angle during plot screening is probably the main source of variability. In practical terms it is difficult to maintain the exact angle of measurement, even for a few seconds (hand steadiness of the operator, distractions, or fatigue); each sample is derived from several integrations, usually more than 10, so the chance of making a mistake is not uncommon. When a plot is screened, it can be performed by keeping the fiber aimed at a single point (lower variability and representation) or across several plants (higher representation but greater variability); when the second option is taken, the chances of integrating other materials into a single scan or sample (e.g., soil, weeds, or air) are increased, and also enhanced by changes in measurement angles. Other considerations such as the effect of the wind speed or turbulence on the measured surface would be detected.

The user needs to set up the *Maximum variation coefficient* accepted for the samples belonging to the same scan. The software will find the scans where the limit is exceeded, at any wavelength, and this will be reported in the *Scans with problems*

section. The samples that need to be checked can be individually analyzed on the graphical window, where it is possible to visualize all the samples in a single graph, identifying (zooming in and out) and deleting those spectral signatures with problems.

It is important to mention that the samples selected with problems within a same scan, do not necessarily need to be modified. This decision will depend on the magnitude of the differences between samples and the number of wavelengths involved. In cases where the user decides to intervene in a scan, it is possible to select and delete one or more samples from the *Samples to delete* section.

## Outlier Analysis

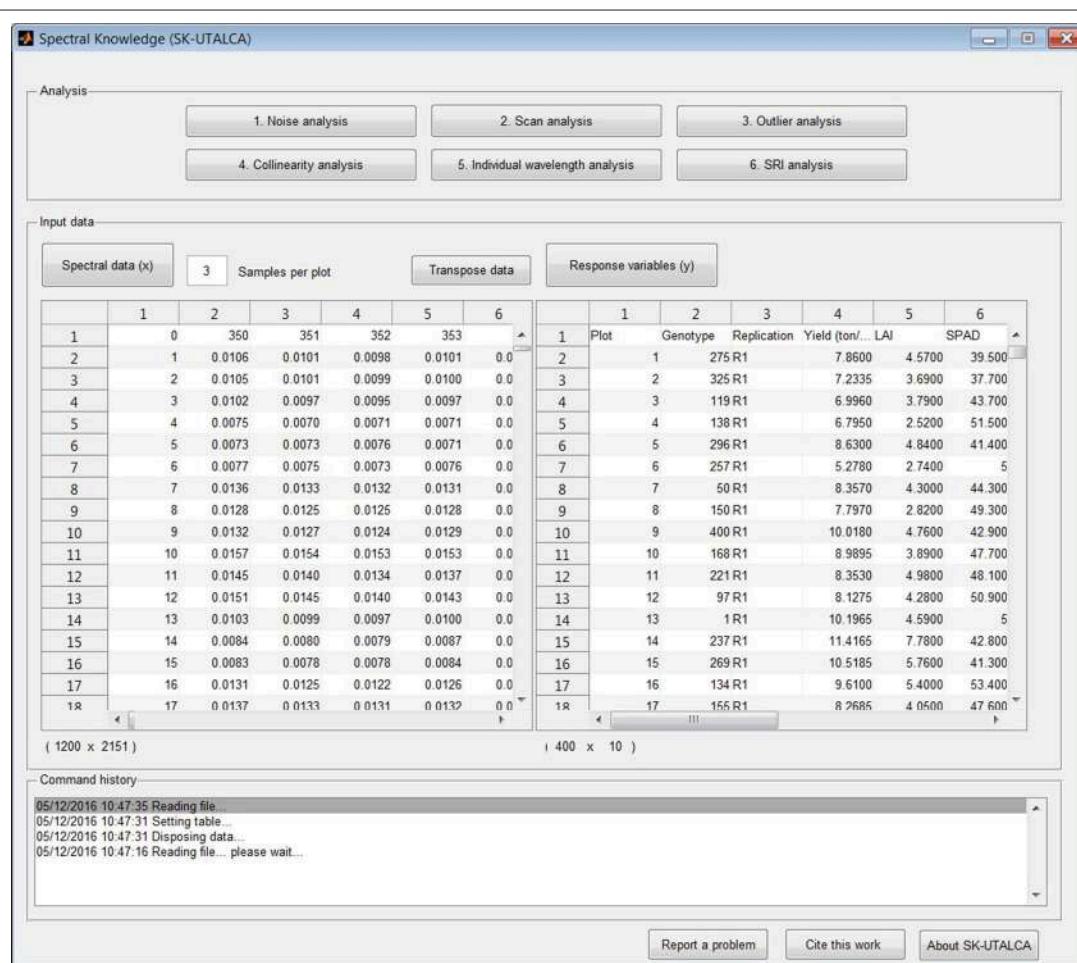
This third filter is designed for rapid identification of problems associated with inconsistencies within spectral data. When outlier data is found, it will be necessary to evaluate the permanence of these in the data matrix.

Because of the high number of genotypes and samples per scan, it is difficult to identify data points that do not follow the general trends. Field experience has proven that is common to find small clouds of data whose main source of error comes

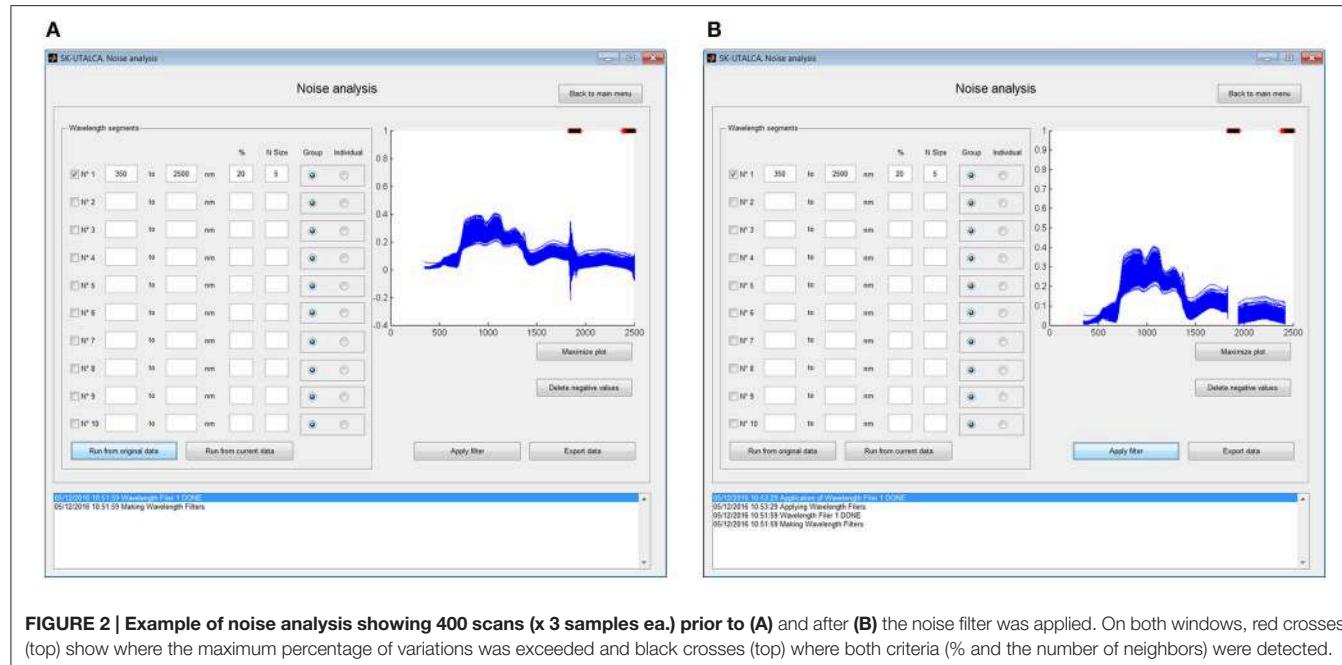
from the calibration process. For example, the sun's movement throughout the day requires calibrations to be performed every 10–15 min. Due to distractions or tiredness during long working hours, the calibration can be forgotten, generating differences in the sun's incidence angle and therefore variations in the reflectance readings. Another form of user error, although less common and related to specific devices, may occur if the user has left the mouse cursor on one of the calibration icons (optimization, dark current, or white reference), performing an unconscious and incomplete calibration with a random click and thus generating undetectable reading errors.

In this module, it is possible to integrate a visual analysis of the reflectance and the response variable data at the same time. The user has four graphs to explore outlier information, evaluating different SRIs, and traits. In this section, it is also possible to *Edit* each graph, selecting data that need to be removed from the data matrix.

For these actions, the software will average the samples per scan to generate each SRI. This is important because the user should check the *Noise Analysis* and *Scan Analysis* modules first.



**FIGURE 1 | Main screen divided horizontally into three sections: analysis, input data, and command history.** Screen shows loaded databases (spectral and response variable data files); the transpose data option is also available for the spectral matrix.



**FIGURE 2 | Example of noise analysis showing 400 scans (x 3 samples ea.) prior to (A) and after (B) the noise filter was applied. On both windows, red crosses (top) show where the maximum percentage of variations was exceeded and black crosses (top) where both criteria (% and the number of neighbors) were detected.**

## Collinearity Analysis

Collinearity or multicollinearity is a problem in regression analysis where the predictor variables “X” are themselves highly correlated (Draper and Smith, 2003). With the use of high-resolution spectral reflectance data, the collinearity problem is inherent to the data collection method employed because several wavelengths are highly correlated. If the goal is to understand how several predictor variables impact on a specific response variable “Y,” the collinearity is a big issue. Therefore, depending on the modeler’s interest, it may be necessary to implement a collinearity analysis before construction of complex models (e.g., multilinear regression model).

In this module, the user can identify wavelengths that deliver the same predictive information for a given response variable, keeping only those that best explain it. This analysis can be performed (*collinearity test setting*) by *linear regression*, indicating the threshold coefficient of determination (*R square cutoff*), or through Artificial Neural Networks (ANN), considering a training process by Levenberg-Marquardt (trainlm), and Mean Squared Error (MSE) as a performance indicator. Depending on the data matrix and computer performance, the non-linear approach (ANN) could take several minutes or hours.

## Individual Wavelength Analysis

For the construction of new SRIs and regression models, it would be desirable to know the degree of dependency between individual wavelengths and the response variable. In this module, the researcher can study the behavior of each wavelength relative to each variable under study, considering one, or more of the following models:

- (1) Polynomial 1:  $y = p_1 \cdot x + p_2$
- (2) Polynomial 2:  $y = p_1 \cdot x^2 + p_2 \cdot x + p_3$
- (3) Weibull:  $y = p_1 \cdot p_2 \cdot x^{(p_2-1)} \cdot e^{(-p_1 \cdot x^{p_2})}$
- (4) Exponential:  $y = p_1 \cdot e^{(p_2 \cdot x)}$
- (5) Power:  $y = p_1 + p_2 \cdot x^{(p_3)}$
- (6) Logarithmic:  $y = p_1 \cdot \ln(x) + p_2$

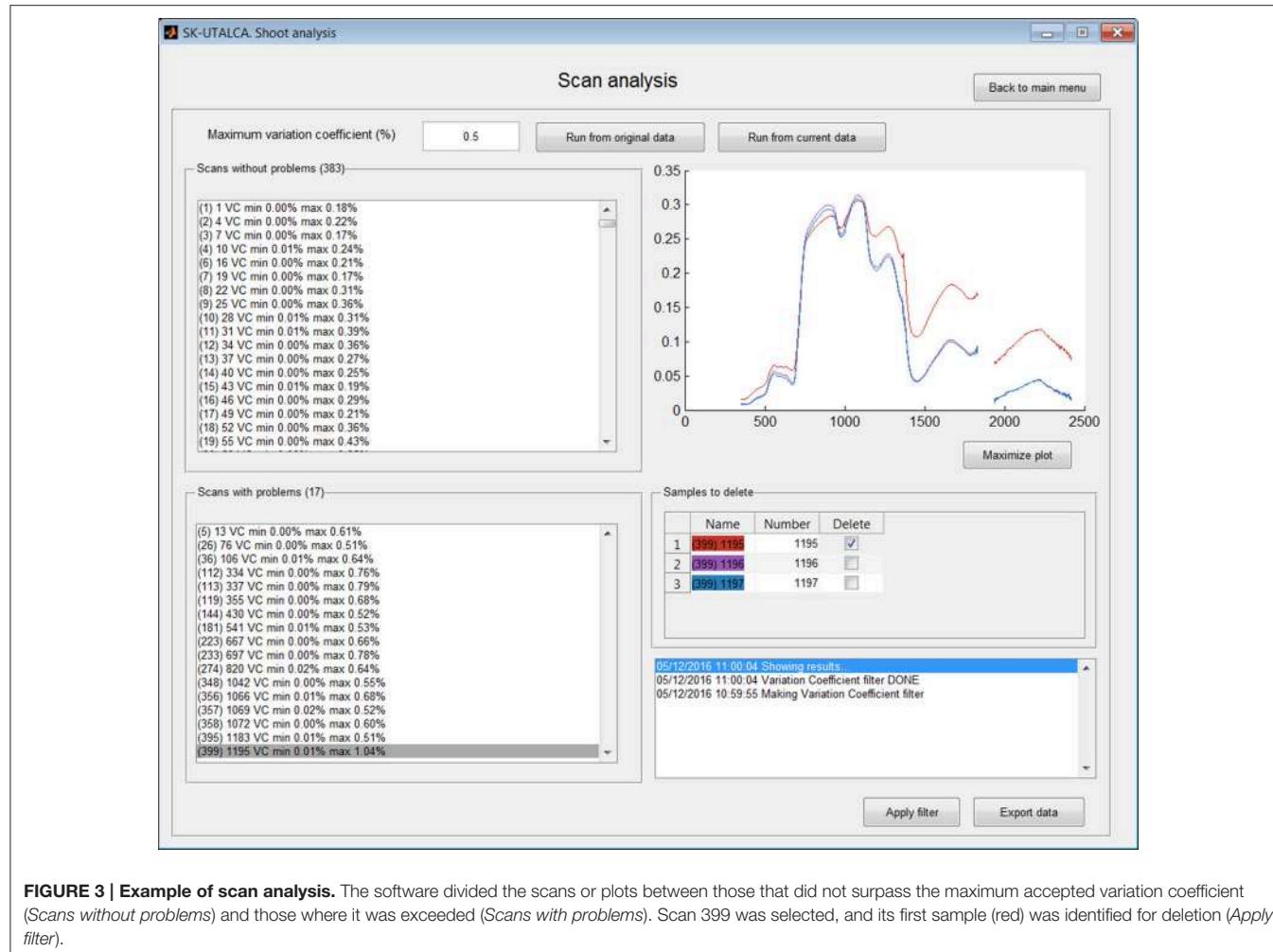
For this analysis, the user can select different statistics to sort the results (adjusted and non-adjusted determination coefficient, root mean squared error, sum of squares due to errors, and degree of freedom). It is also necessary to set up a minimum or maximum value for the selected statistics in order to export just those results (*Values above or below*). The exported file will show, for each wavelength, the statistics values for the selected model(s) where those minimum or maximum values were met.

This module and the following one (SRI analysis) work with sample averages, forcing the user to perform a deep preliminary analysis, thus avoiding any error in the data matrix.

## Spectral Reflectance Index (SRI) Analysis

The implementation of concatenate formulas in spreadsheets is helpful for automating time-consuming procedures. However, due to the number of scans, samples per scans, measured wavelengths, evaluated response variables, and tested SRIs, the physical size of the resulting spreadsheets (several MB) implies the need for high performance computers.

By evaluating the same regression models reviewed with the previous function, the user will be able to identify the SRIs (initially 255: Jordan, 1969; Rouse et al., 1973; Rouse, 1974; Tucker, 1979; Hardisky et al., 1983; Guyot and Baret, 1988;

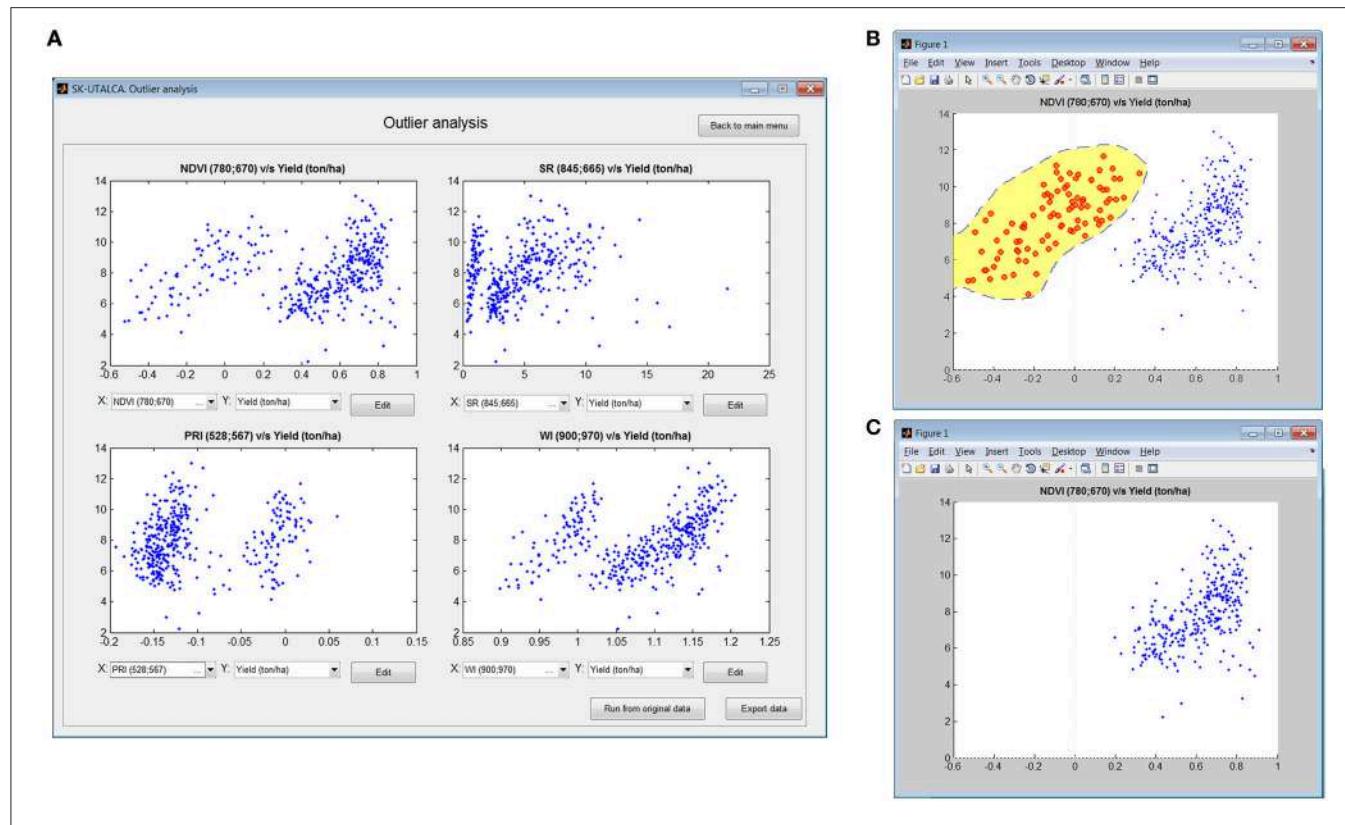


**FIGURE 3 | Example of scan analysis.** The software divided the scans or plots between those that did not surpass the maximum accepted variation coefficient (Scans without problems) and those where it was exceeded (Scans with problems). Scan 399 was selected, and its first sample (red) was identified for deletion (Apply filter).

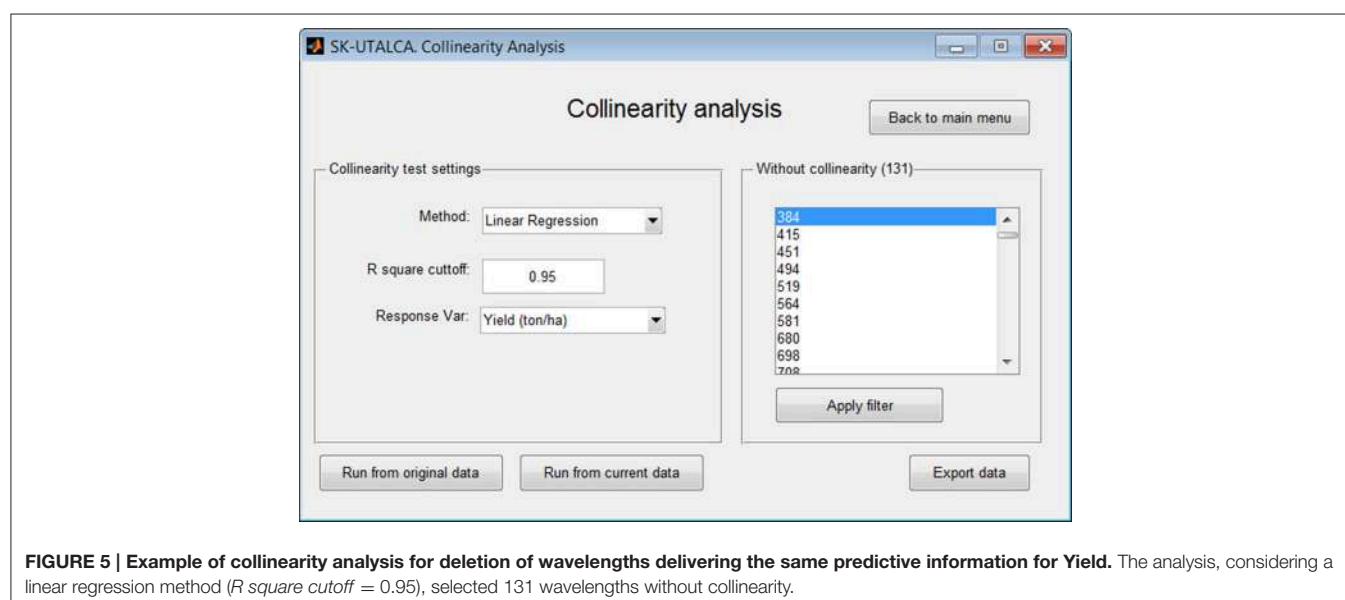
Guyot et al., 1988; Huete, 1988; Baret et al., 1989; Clevers, 1989; Curran, 1989; Hunt and Rock, 1989; Major et al., 1990; Barnes et al., 1992, 2000; Chappelle et al., 1992; Gamon et al., 1992; Peñuelas et al., 1993a,b, 1994, 1995, 1997; Vogelmann et al., 1993; Carter, 1994; Gitelson and Merzlyak, 1994, 1997; McMurtrey et al., 1994; Qi et al., 1994; Roujeau and Breon, 1995; Smith et al., 1995; Chen, 1996; Chen and Cihlar, 1996; Filella et al., 1996; Fourt et al., 1996; Gao, 1996; Ma et al., 1996; Rondeaux et al., 1996; Huete et al., 1997; van Deventer et al., 1997; Blackburn, 1998, 1999; Datt, 1998, 1999; Merton, 1998; Peñuelas and Filella, 1998; Gamon and Surfus, 1999; Gitelson et al., 1999, 2001, 2003, 2005, 2006; Merzlyak et al., 1999; Peñuelas and Inoue, 1999; Daughtry et al., 2000; Marshak et al., 2000; Thenkabail et al., 2000; Broge and Leblanc, 2001; Raun et al., 2001; Zarco-Tejada et al., 2001, 2003a,b, 2005; Broge and Mortensen, 2002; Haboudane et al., 2002, 2004; Read et al., 2002; Serrano et al., 2002; Sims and Gamon, 2002; Gupta et al., 2003; Hansen and Schjoerring, 2003; Steddom et al., 2003; Viña, 2003; Dash and Curran, 2004; Gandia et al., 2004; Le Maire et al., 2004, 2008; Schlemmer et al., 2005; Zhao et al., 2005; Vincini et al., 2006; Babar et al., 2006a,b; Mirik et al., 2006a,b; Inoue et al., 2007,

2008; Prasad et al., 2007; Rodríguez-Pérez et al., 2007; Zhu et al., 2007; Rama Rao et al., 2008; White et al., 2008; Wu et al., 2008a,b; Richter et al., 2009; Serbin et al., 2009; Stroppiana et al., 2009; Yañez et al., 2009; Dzikiti et al., 2010; Herrmann et al., 2010; Mistele and Schmidhalter, 2010; Yao et al., 2010, 2011; Garrity et al., 2011; Hernández-Clemente et al., 2011; Main et al., 2011; Pimstein et al., 2011; Tian et al., 2011, 2014; Winterhalter et al., 2011; Wang et al., 2011a,b) having the higher adjusted coefficients of determination (*Adj. RSquare values above*) in relation to a response variable. Internally, the software will select all combinations (regression model, SRI, and response variable) where the adjusted coefficient of determination was reached. The *Export data* option will generate a report that includes all the statistics analyzed in the previous function for the best-evaluated regression model and for each one (in the case that more than two were tested).

For publication purposes this module also includes an exportable *Detailed index report*, where it is possible to select specific SRIs and response variables. The report will include the SRI and variable values for each of the measurements, allowing the user to create XY graphs.



**FIGURE 4 | Example of outlier analysis showing four scatterplot graphs (NDVI, SR, PRI and WI vs. Yield) (A). Using the *Edit* function, NDVI vs. Yield was used to select scans with NDVI values below 0.31 (B) for deletion (C).**



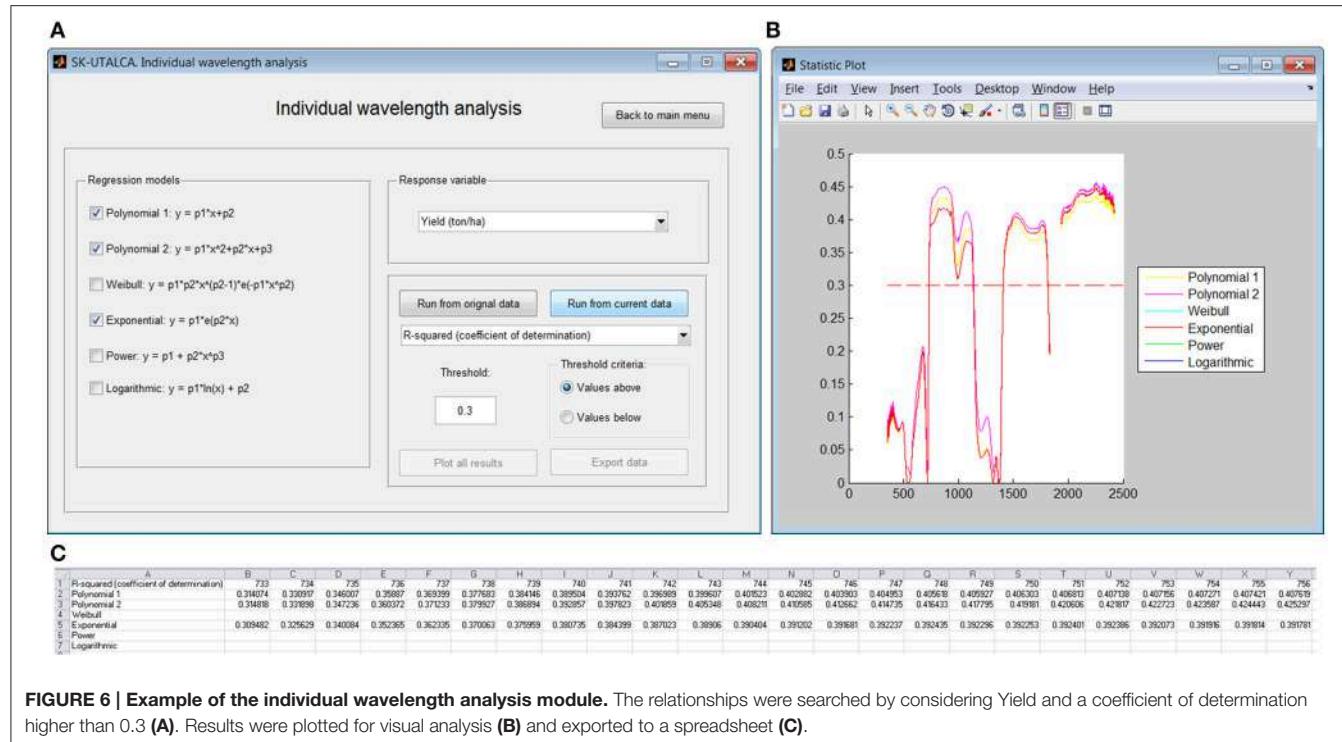
**FIGURE 5 | Example of collinearity analysis for deletion of wavelengths delivering the same predictive information for Yield.** The analysis, considering a linear regression method ( $R^2$  square cutoff = 0.95), selected 131 wavelengths without collinearity.

## OPERATIONAL EXAMPLES OF SK-UTALCA

### Testing Data Sets

During the 2011/12 growing season, 386 genotypes of wheat (*Triticum spp.* L.) from different breeding programs (INIA-Chile,

INIA-Uruguay and CIMMYT) were assessed under three water regimens (fully irrigated, mild water deficit and severe water deficit). This trial was established at Santa Rosa Experimental Station ( $36^{\circ} 32' S$ ,  $71^{\circ} 55' W$ ; 217 m.a.s.l.), Regional Research Center INIA Quilamapu (Chillán, VIII Region, Chile),



considering an alpha-lattice design (386 genotypes + 2 cvs. replicated seven times to assess field variability) and two replications.

Reflectance measurements were performed using a portable spectroradiometer (FieldSpec® 3 Jr, ASD Inc., Boulder, CO, USA) (350–2500 nm), between 12:00 and 16:00 h, on clear days (solar radiation higher than 800 W m<sup>-2</sup>). Prior to the first measurement and every 15 min, the equipment was calibrated using a field reference panel (Spectralon, ASD Inc., Boulder, CO, USA). The equipment was configured to read three samples per scan. Each plot (genotype) was scanned once.

A detailed methodology can be found in Lobos et al. (2014). For purposes of this article, only one environment (fully irrigated), one phenological stage (grain filling) and one replicate will be considered.

## Data Analysis

In this section, we highlight some of the key results of the analysis performed using the SK-UTALCA software.

### Setting Up

Prior to analysis the user needs to: (i) load the spectral data file (denoted as "x"); (ii) load the response variable(s) file (denoted as "y"); and (iii) define the number of samples per scan (in this case three). Wavelengths need to be placed in columns and samples in rows; the *transpose data* function is available.

The file format for the spectra (*Genotype*, *Wavelength<sub>1</sub>*, *Wavelength<sub>2</sub>*, *Wavelength<sub>3</sub>*, ... *Wavelength<sub>n</sub>*) and the response variables (*Plot*, *Genotype*, *Replication*, *Variable<sub>1</sub>*, *Variable<sub>2</sub>*, ..., *Variable<sub>n</sub>*) are presented in Figure 1. If for any reason the user

realizes that there are missing plots (no spectral information) before the spectral data is uploaded, keeping in mind the sample number per scan, those rows can be left empty. If calibration data is among the spectral data output from the spectrometer, it should be removed prior to uploading the reflectance data (x).

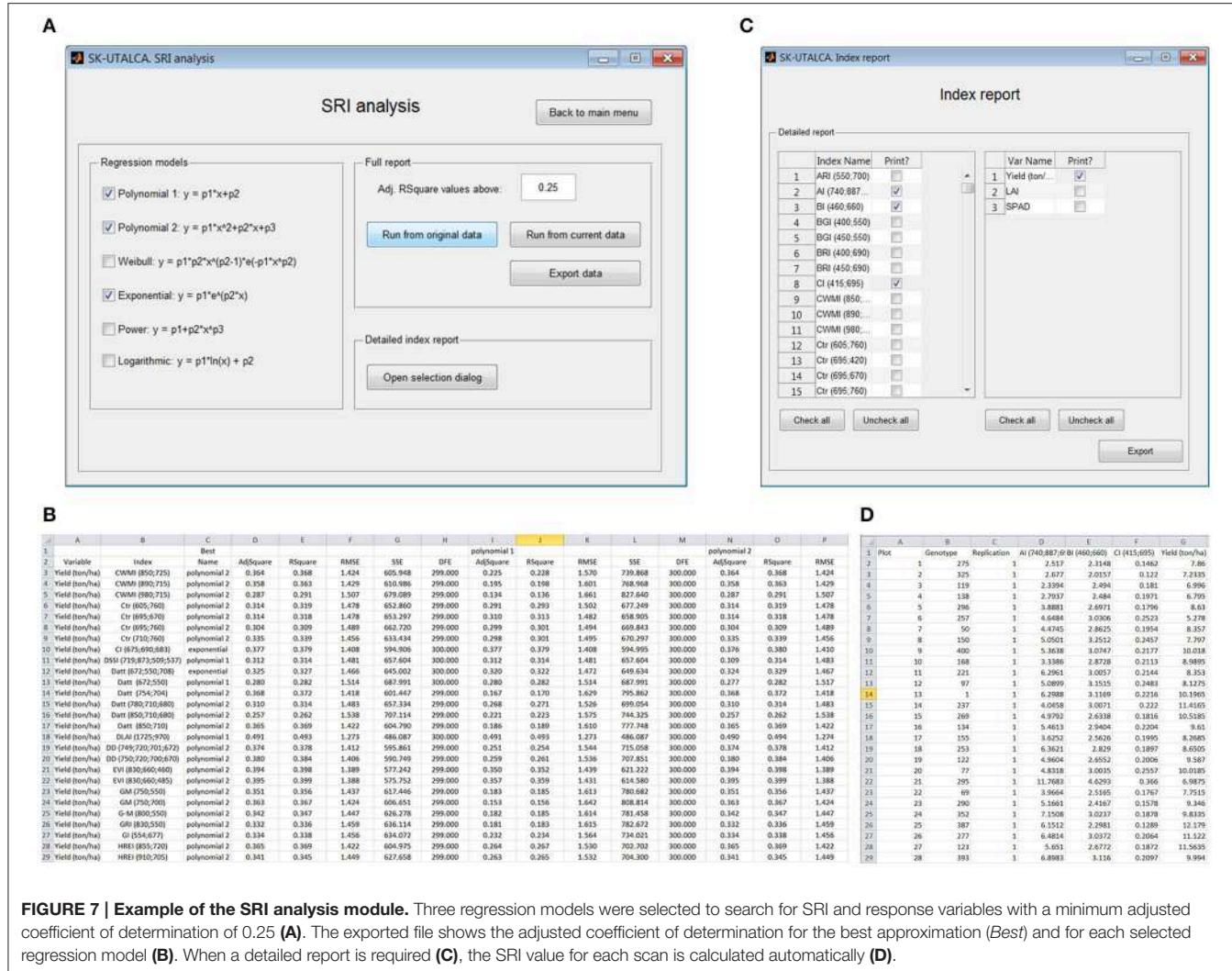
Once the data has been loaded into the software and the wavelengths are arranged into columns, it is possible to start the analysis.

### Noise Analysis

To apply this filter it is necessary to indicate the wavelength segment for analysis, the cutting criteria (*Group* or *Individual*), the maximum percentage of variations accepted (%), and the number of neighbors (*N size*). The selection of each wavelength segment and the criteria for each one (% and *N Size*) will depend on the user experience and the environmental conditions where the measurements were taken; for example, noise at 1800–1950 nm and 2350–2500 nm is usually wider and stronger than at 1300–1400 nm, so the criteria should consider higher values of % and *N Size* for the first two segments. In this operational example, the filter was applied to the whole spectral range (350–2500 nm) considering a group filter, with five wavelengths as *N size* and a maximum accepted variation among them of 20%. Figure 2 shows the results prior to (A) and after (B) the filter was applied. In this case, the filter was able to detect two main noise zones from 1833 to 1935 nm and from 2422 to 2500 nm.

### Scan Analysis

The Scan analysis module allows detection of abnormal variations among samples within the same scan. In this operational example, the Scan analysis was applied using the



function *Run from the current data*, that is to say, considering the results obtained using previous filter (without spectral noise). The *Maximum variation coefficient* was set at 0.5%. The software was able to select 383 scans or plots without problems and 17 where the threshold was exceeded (5, 26, 36, 112–113, 119, 144, 181, 223, 233, 274, 348, 356–358, 395, and 399). In the Figure 3, scan or plot 399 is graphed, and the first sample (1195 on red) was selected for deletion. This result could be an indicator of a measurement problem associated with the operator (modification of the measurement angle) or external conditions (e.g., wind speed) during the first sample integrations. In case of all samples from a specific scan need to be deleted, the software will maintain this scan as empty rows, avoiding problems in further analyses.

### Outlier Analysis

This module is a simple and exploratory analysis to identify outlier scans, allowing the user to detect field measurement problems (e.g., calibration). Four scatterplot graphs will show the relationship between any SRI available on the software database

and the loaded response variables. If a problem is detected, it is possible to use the *Edit* option to manually remove the samples. In this operational example, the relationship between NDVI and Yield was used to inspect the possible outlier samples. Figure 4A shows how different SRIs (NDVI, SR, PRI, and WI) can generate different data distributions, helping the user in cases where problems are not so evident; on the top left graph (NDVI vs. Yield) two clouds of data points can be identified, divided at the NDVI value of 0.31.

Once the information from the smaller data cloud was analyzed ( $\text{NDVI} < 0.31$ ), it was evident that the data set corresponded to 96 contiguous scans or plots (104–200), suggesting that there were problems associated with the measurement. When information from the spectrometer was checked, it was concluded that the operator had skipped one calibration. It is always important to check the pertinence of negative SRI values because they are probably related to measurement errors.

After identification of the origin of a particular problem, any graph can be selected for editing. In this example (NDVI vs.

Yield), the scan with the problem can be selected (**Figure 4B**) and deleted (**Figure 4C**).

### Collinearity Analysis

Using linear regression or ANN, the collinearity analysis module identifies wavelengths that delivering the same predictive information for a given response variable, keeping only those that best explain it. In this operational example, collinearity analysis was applied by considering the results obtained from the scan analysis (*Run from current data*), with Yield being the response variable in the linear regression (*R square cutoff* = 0.95). Results of this analysis found 131 wavelengths without collinearity (**Figure 5**).

### Individual Wavelength Analysis

In this module it is possible to assess the relationship of individual wavelengths and a given response variable. Three regression models were selected (*Polynomial 1* and *2*, and *Exponential*) to search for wavelengths with determination coefficients above 0.3 in relation to Yield (**Figure 6A**). If the user selects *Plot all results*, a graph will show the wavelengths below and above the determination coefficient cutoff (**Figure 6B**). These results can be exported to a spreadsheet; for each selected regression model, only wavelengths where the chosen statistic surpassed the cutoff will be shown (**Figure 6C**). In this operational example, there were three groups of wavelengths with determination coefficients above the threshold: 733–1139, 1409–1815, and 1936–2421 nm.

### Spectral Reflectance Index (SRI) Analysis

As in the previous module, when different regression models were considered, SRI analysis has the option to evaluate the relationship between all loaded response variables and all SRIs available in the software database. For this example, three regression models were selected (*Polynomial 1* and *2*, and *Exponential*) to search the SRIs and response variables with an adjusted determination coefficient higher than 0.25 (**Figure 7A**). When *Export data* is selected, all relationships with adjusted determination coefficients higher than 0.25 will be reported (**Figure 7B**); the results, which are organized according to the loaded variables (column A) and SRIs (column B), show which regression model had the highest determination coefficient (*Best*) for each SRI, as well as its statistics [adjusted and non-adjusted determination coefficient, root mean squared error (RMSE), sum of squares due to errors (SSE) and degree of freedom (DFE)] (columns C–H). The results for each evaluated regression model are also described (*Polynomial 1*: columns I–M; *Polynomial 2*: columns N–R, and so on). In this screen example, the adjusted  $R^2$  varied between 0.257 (Datt 850;710;680) and 0.406 (DLAI 1725;970), with these SRIs having the highest and lowest RMSEs, respectively (**Figure 7B**).

The selection of *Open selection dialog (Detailed index report)* enables the user to select specific SRIs and response variables for figure elaboration (*Index report*, **Figure 7C**). In this operational example, three SRIs (AI, BI, and CI) and one response variable (Yield) were selected. The SRI value for each scan or plot is given

(**Figure 7D**) so the user can generate XY scatter plots for each tested SRI (X) and response variable (Y).

## CONCLUSIONS

Spectral Knowledge (SK-UTALCA) is a software package that allows an easy and fast exploratory analysis of high-resolution spectral reflectance data, providing the user with tools to detect measurement problems and the generation of key information for later modeling. SK-UTALCA is especially useful for plant breeding or any other research area where the number of measurements (big data files) involves long working hours that increase the risk of making involuntarily mistakes. This freely-available software is the result of several years of measurements and analysis of spectral data oriented toward the prediction of traits in plant breeding.

## AUTHOR CONTRIBUTIONS

GL, CP-E contributed equally toward the intellectual input into the final version of this paper, including development of the software, data analysis, interpreting, and discussion of the results, and writing and editing the manuscript.

## ACKNOWLEDGMENTS

This work was supported and financed by the research program “Adaptation of Agriculture to Climate Change (A2C2),” “Nucleo Científico Multidisciplinario” and the “Vicerrectoría de Innovación, Desarrollo y Transferencia Tecnológica (VIDTT)” from Universidad de Talca. We also received significant funds from the National Commission for Scientific and Technological Research CONICYT (FONDEF IDEA 14I10106 and FONDECYT N°11130601). We would like to express our gratitude to Sebastian Romero, Félix Estrada, Miguel Garriga, and especially to Alejandro Escobar for continued technical assistance in field experiments and laboratory analysis. We thank also Rodrigo Aguilar and Felipe Ojeda for their outstanding programming work, Genberries Ltda. for equipment support, and Ivan Matus (INIA- Chile) for genetic material and trial maintenance. Finally, our special gratitude goes to Greg Matyjewicz (ASD Inc., Boulder, CO, USA) for technical definitions and valuable discussion.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01996/full#supplementary-material>

## Software Availability

The compiled version of the software will be available for free downloading at <http://www.fenomica.utalca.cl/> and source code is available as Supplementary Material.

## REFERENCES

- Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/j.tplants.2013.09.008
- Babar, M. A., Reynolds, M. P., van Ginkel, M., Klatt, A. R., Raun, W. R., and Stone, M. L. (2006a). Spectral reflectance indices as a potential indirect selection criteria for wheat yield under irrigation. *Crop Sci.* 46, 578–588. doi: 10.2135/cropsci2005.0059
- Babar, M. A., Reynolds, M. P., Van Ginkel, M., Klatt, A. R., Raun, W. R., and Stone, M. L. (2006b). Spectral reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll, and canopy temperature in wheat. *Crop Sci.* 46, 1046–1057. doi: 10.2135/cropsci2005.0211
- Baret, F., Guyot, G., and Major, D. J. (1989). “TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation,” in *Proceeding of the 1989 International Geoscience and Remote Sensing Symposium (IGARSS '89) and the 12th Canadian Symposium on Remote Sensing* (Vancouver, BC), 1355–1358.
- Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., et al. (2000). “Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data,” in *Proceedings of the Fifth International Conference on Precision Agriculture*, eds P. C. Robert, R. H. Rust, and W. E. Larson (Madison, WI: ASA/CSSA/SSSA).
- Barnes, J. D., Balaguer, L., Manrique, E., Elvira, S., and Davison, A. W. (1992). A reappraisal of the use of DMSO for the extraction and determination of chlorophylls a and b in lichens and higher plants. *Environ. Exp. Bot.* 32, 85–100. doi: 10.1016/0098-8472(92)90034-Y
- Blackburn, G. A. (1998). Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. *Int. J. Remote Sens.* 19, 657–675. doi: 10.1080/014311698215919
- Blackburn, G. A. (1999). Relationships between spectral reflectance and pigment concentrations in stacks of deciduous broadleaves. *Remote Sens. Environ.* 70, 224–237. doi: 10.1016/S0034-4257(99)00048-6
- Borengasser, M., Hungate, W., and Watkins, R. (2008). *Hyperspectral Remote Sensing. Principles and Applications*. Boca Raton, FL: CRC Press.
- Broge, N. H., and Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sens. Environ.* 76, 156–172. doi: 10.1016/S0034-4257(00)00197-8
- Broge, N. H., and Mortensen, J. V. (2002). Deriving green crop area index and canopy chlorophyll density of winter wheat from spectral reflectance data. *Remote Sens. Environ.* 81, 45–57. doi: 10.1016/S0034-4257(01)00332-7
- Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., Serret, M. D., and Araus, J. L. (2012). High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. *J. Integr. Plant Biol.* 54, 312–320. doi: 10.1111/j.1744-7909.2012.01116.x
- Camargo, A., and Lobos, G. A. (2016). Latin America: a development pole for phenomics. *Front. Plant Sci.* 7:1729. doi: 10.3389/fpls.2016.01729
- Carter, G. A. (1994). Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *Int. J. Remote Sens.* 15, 697–703. doi: 10.1080/01431169408954109
- Chappelle, E., Kim, M., and McMurtrey, J. (1992). Ratio analysis of reflectance spectra (RARS): an algorithm for the remote estimation of the concentrations of chlorophyll A, chlorophyll B, and carotenoids in soybean leaves. *Remote Sens. Environ.* 39, 239–247. doi: 10.1016/0034-4257(92)90089-3
- Chen, J. M. (1996). Optically-based methods for measuring seasonal variation of leaf area index in boreal conifer stands. *Agr. Forest Meteorol.* 80, 135–163. doi: 10.1016/0168-1923(95)02291-0
- Chen, J. M., and Cihlar, J. (1996). Retrieving leaf area index of boreal conifer forests using landsat TM images. *Remote Sens. Environ.* 55, 153–162. doi: 10.1016/0034-4257(95)00195-6
- Clevers, J. G. P. W. (1989). Application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture. *Remote Sens. Environ.* 29, 25–37. doi: 10.1016/0034-4257(89)90076-X
- Clevers, J. G. P. W., Kooistra, L., and Schaeppman, M. E. (2008). Using spectral information from the NIR water absorption features for the retrieval of canopy water content. *Int. J. Appl. Earth Obs.* 10, 388–397. doi: 10.1016/j.jag.2008.03.003
- Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sens. Environ.* 30, 271–278. doi: 10.1016/0034-4257(89)90069-2
- Curtiss, B., and Goetz, A. F. H. (1994). “Field spectrometry: techniques and instrumentation,” in *Proceedings of an International Symposium on Spectral Sensing Research* (San Diego, CA), 10–15.
- Curtiss, B., and Goetz, A. F. H. (2001). “Field spectrometry: techniques and instrumentation,” in *Technical Guide*, 4th Edn, ed D. C. Hatchell (Boulder, CO: Analytical Spectral Devices Inc), Section 12–1 to 12–10.
- Dash, J., and Curran, P. (2004). The MERIS terrestrial chlorophyll index. *Int. J. Remote Sens.* 25, 5403–5413. doi: 10.1080/0143116042000274015
- Datt, B. (1998). Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+b, and total carotenoid content in eucalyptus leaves. *Remote Sens. Environ.* 66, 111–121. doi: 10.1016/S0034-4257(98)00046-7
- Datt, B. (1999). Visible/near infrared reflectance and chlorophyll content in Eucalyptus leaves. *Int. J. Remote Sens.* 20, 2741–2759. doi: 10.1080/014311699211778
- Daughtry, C., Walther, C., Kim, M., de Colstoun, E., and McMurtrey, J. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens. Environ.* 74, 229–239. doi: 10.1016/S0034-4257(00)00113-9
- Draper, N. R., and Smith, H. (2003). *Applied Regression Analysis*, 3rd Edn. New York, NY: Wiley.
- Dzikiti, S., Verreyne, J. S., Stuckens, J., Strever, A., Verstraeten, W. W., Swennen, R., et al. (2010). Determining the water status of Satsuma mandarin trees [*Citrus Unshiu Marcovitch*] using spectral indices and by combining hyperspectral and physiological data. *Agr. Forest Meteorol.* 150, 369–379. doi: 10.1016/j.agrformet.2009.12.005
- Estrada, F., Escobar, A., Romero-Bravo, S., González-Talice, J., Poblete-Echeverría, C., Caligari, P. D. S., et al. (2015). Fluorescence phenotyping in blueberry breeding for genotype selection under drought conditions, with or without heat stress. *Sci. Hort.* 181, 147–161. doi: 10.1016/j.scienta.2014.11.004
- Filella, I., Serrano, L., Serra, J., and Peñuelas, J. (1996). Evaluating wheat nitrogen status with canopy reflectance indices and discriminant analysis. *Crop Sci.* 35, 1400–1405. doi: 10.2135/cropsci1995.0011183X003500050023x
- Finkel, E. (2009). With ‘Phenomics’ plant scientists hope to shift breeding into overdrive. *Science* 325, 380–381. doi: 10.1126/science.325\_380
- Fourty, T., Baret, F., Jacquemoud, S., Schmuck, G., and Verdebout, J. (1996). Leaf optical properties with explicit description of its biochemical composition: direct and inverse problems. *Remote Sens. Environ.* 56, 104–117. doi: 10.1016/0034-4257(95)00234-0
- Gamon, J., Peñuelas, J., and Field, C. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens. Environ.* 41, 35–44. doi: 10.1016/0034-4257(92)90059-S
- Gamon, J., and Surfus, J. S. (1999). Assessing leaf pigment content and activity with a reflectometer. *New Phytol.* 143, 105–117. doi: 10.1046/j.1469-8137.1999.00424.x
- Gandia, S., Fernández, G., García, J. C., and Moreno, J. (2004). “Retrieval of vegetation biophysical variables from CHRIS/PROBA data in the SPARC campaign,” in *Proceedings of the 2nd CHIRS/Proba Workshop, ESA/ESRIN* (Frascati), 40–48.
- Gao, B. C. (1996). NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58, 257–266. doi: 10.1016/S0034-4257(96)00067-3
- Garbulsky, M. F., Peñuelas, J., Gamon, J., Inoue, Y., and Filella, I. (2011). The photochemical reflectance index (PRI) and the remote sensing of leaf, canopy and ecosystem radiation use efficiencies: a review and meta-analysis. *Remote Sens. Environ.* 115, 281–297. doi: 10.1016/j.rse.2010.08.023
- Garriga, M., Retamales, J. B., Romero, S., Caligari, P. D. S., and Lobos, G. A. (2014). Chlorophyll, anthocyanin and gas exchangechanges assessed by spectroradiometry in *Fragaria chiloensis* under salt stress. *J. Integr. Plant Biol.* 56, 505–515. doi: 10.1111/jipb.12193
- Garrison, S. R., Eitel, J. U., and Vierling, L. A. (2011). Disentangling the relationships between plant pigments and the photochemical reflectance index reveals a new approach for remote estimation of carotenoid content. *Remote Sens. Environ.* 115, 628–635. doi: 10.1016/j.rse.2010.10.007
- Gitelson, A. A., Buschmann, C., and Lichtenthaler, H. K. (1999). The Chlorophyll fluorescence ratio F735/F700 as an accurate measure of

- the chlorophyll content in plants. *Remote Sens. Environ.* 69, 296–302. doi: 10.1016/S0034-4257(99)00023-1
- Gitelson, A. A., Keydan, G. P., and Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophys. Res. Lett.* 33, L11402. doi: 10.1029/2006GL026457
- Gitelson, A. A., and Merzlyak, M. N. (1994). Quantitative estimation of chlorophyll-a using reflectance spectra: experiments with autumn chestnut and maple leaves. *J. Photoch. Photobiol. B* 22, 247–252. doi: 10.1016/1011-1344(93)06963-4
- Gitelson, A. A., and Merzlyak, M. N. (1997). Remote estimation of chlorophyll content in higher plant leaves. *Int. J. Remote Sens.* 18, 2691–2697. doi: 10.1080/014311697217558
- Gitelson, A. A., Merzlyak, M. N., and Chivkunova, O. B. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochem. Photobiol.* 74, 38–45. doi: 10.1562/0031-8655(2001)074<0038:OPANE>2.0.CO;2
- Gitelson, A. A., Vi-a, A., Ciganda, V., Rundquist, D. C., and Arkebauer, T. J. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* 32, 1–4. doi: 10.1029/2005GL022688
- Gitelson, A., Gritz, Y., and Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* 160, 271–282. doi: 10.1078/0176-1617-00887
- Gupta, R. K., Vijayan, D., and Prasad, T. (2003). Comparative analysis of red-edge hyperspectral indices. *Adv. Space Res.* 32, 2217–2222. doi: 10.1016/S0273-1177(03)90545-X
- Guyot, G., and Baret, F. (1988). “Utilisation de la haute resolution spectrale pour suivre l'état des couverts végétaux. Spectral signatures of objects in remote sensing,” in *Proceedings of the Conference Held in Aussois (Modane)*, 279–286.
- Guyot, G., Baret, F., and Major, D. (1988). High spectral resolution: determination of spectral shifts between the red and near infrared. *Int. Arch. Phot. Remote Sens.* 8, 1307–1317.
- Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., and Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* 90, 337–352. doi: 10.1016/j.rse.2003.12.013
- Haboudane, D., Miller, J., Tremblay, N., Zarco-Tejada, P., and Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ.* 81, 416–426. doi: 10.1016/S0034-4257(02)00018-4
- Hansen, P. M., and Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens. Environ.* 86, 542–553. doi: 10.1016/S0034-4257(03)00131-7
- Hardisky, M. A., Klemas, V., and Smart, R. M. (1983). The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of *Spartina alterniflora* canopies. *Photogramm. Eng. Remote Sens.* 49, 77–83.
- Hernández-Clemente, R., Navarro-Cerrillo, R. M., Suárez, L., Morales, F., and Zarco-Tejada, P. J. (2011). Assessing structural effects on PRI for stress detection in conifer forests. *Remote Sens. Environ.* 115, 2360–2375. doi: 10.1016/j.rse.2011.04.036
- Herrmann, I., Karniel, A., Bonfil, D. J., Cohen, Y., and Alchanatis, V. (2010). SWIR-based spectral indices for assessing nitrogen content in potato fields. *Int. J. Remote Sens.* 31, 5127–5143. doi: 10.1080/01431160903283892
- Houle, D., Govindaraju, D. R., and Omholt, S. (2010). Phenomics: the next challenge. *Nat. Rev. Genet.* 11, 855–866. doi: 10.1038/nrg2897
- Huete, A. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 25, 295–309. doi: 10.1016/0034-4257(88)90106-X
- Huete, A. R., Liu, H. Q., Batchily, K., and van Leeuwen, W. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sens. Environ.* 59, 440–451. doi: 10.1016/S0034-4257(96)00112-5
- Hunt, E. R. Jr., and Rock, B. N. (1989). Detection of changes in leaf water content using Near- and Middle-Infrared reflectances. *Remote Sens. Environ.* 30, 43–54. doi: 10.1016/0034-4257(89)90046-1
- Inoue, Y., Peñuelas, J., Miyata, A., and Mano, M. (2008). Normalized difference spectral indices for estimating photosynthetic efficiency and capacity at a canopy scale derived from hyperspectral and CO<sub>2</sub> flux measurements in rice. *Remote Sens. Environ.* 112, 156–172. doi: 10.1016/j.rse.2007.04.011
- Inoue, Y., Qi, J., Olioso, A., Kiyono, Y., Horie, T., Asai, H., et al. (2007). Traceability of slash-and-burn land-use history using optical satellite sensor imagery: a basis for chronosequential assessment of ecosystem carbon stock in Laos. *Int. J. Remote Sens.* 28, 5641–5647. doi: 10.1080/01431160701656323
- Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology* 50, 663–666. doi: 10.2307/1936256
- Kipp, S., Mistele, B., Baresel, P., and Schmidhalter, U. (2014). High-throughput phenotyping early plant vigour of winter wheat. *Eur. J. Agron.* 52(Part B), 271–278. doi: 10.1016/j.eja.2013.08.009
- Le Maire, G., Francois, C., and Dufrene, E. (2004). Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sens. Environ.* 89, 1–28. doi: 10.1016/j.rse.2003.09.004
- Le Maire, G., Francois, C., Soudani, K., Berveiller, D., Pontailler, J. Y., Bréda, N., et al. (2008). Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sens. Environ.* 112, 3846–3864. doi: 10.1016/j.rse.2008.06.005
- Lobos, G. A., and Hancock, J. (2015). Breeding blueberries for a changing global environment: a review. *Front. Plant Sci.* 6:782. doi: 10.3389/fpls.2015.00782
- Lobos, G. A., Matus, I., Rodriguez, A., Romero-Bravo, S., Araus, J. L., and del Pozo, A. (2014). Wheat genotypic variability in grain yield and carbon isotope discrimination under Mediterranean conditions assessed by spectral reflectance. *J. Integr. Plant Biol.* 56, 470–479. doi: 10.1111/jipb.12114
- Lörz, H., and Wenzel, G. (2005). *Molecular Marker Systems in Plant Breeding and Crop Improvement*. New York, NY: Springer.
- Ma, B. L., Morrison, M. J., and Dwyer, L. M. (1996). Canopy light reflectance and field greenness to assess nitrogen fertilization and yield of maize. *Agron. J.* 88, 915–920. doi: 10.2134/agronj1996.00021962003600060011x
- Ma, J., and Chen, X. (2006). Adjacency effect estimation by ground spectra measurement and satellite optical sensor synchronous observation data. *Chin. Opt. Lett.* 4, 546–549. Retrieved from: <http://col.osa.org/abstract.cfm?URI=col-4-9-546>
- Main, R., Cho, M. A., Mathieu, R., O'Kennedy, M. M., Ramoelo, A., and Koch, S. (2011). An investigation into robust spectral indices for leaf chlorophyll estimation. *ISPRS J. Photogramm.* 66, 751–761. doi: 10.1016/j.isprsjprs.2011.08.001
- Major, D. J., Baret, F., and Guyot, G. (1990). A ratio vegetation index adjusted for soil brightness. *Int. J. Remote Sens.* 11, 727–740. doi: 10.1080/01431169008955053
- Marshak, A., Knyazikhin, Y., Davis, A., Wiscombe, W., and Pilewskie, P. (2000). Cloud – vegetation interaction: use of normalized difference cloud index for estimation of cloud optical thickness. *Geophys. Res. Lett.* 27, 1695–1698. doi: 10.1029/1999GL010993
- McMurtrey, J. E., Chappelle, E. W., Kim, M. S., Meisinger, J. J., and Corp, L. A. (1994). Distinguishing nitrogen fertilization levels in field corn (*Zea mays* L.) with actively induced fluorescence and passive reflectance measurements. *Remote Sens. Environ.* 47, 36–44. doi: 10.1016/0034-4257(94)90125-2
- Merton, R. (1998). “Monitoring community hysteresis using espectral shift analysis and the red-edge vegetation estress index,” in *Proceedings of the Seventh Annual JPL Airborne Earth Science Workshop. Jet Propulsion Laboratory* (Pasadena, CA).
- Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., and Rakitin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol. Plantarum* 106, 135–141. doi: 10.1034/j.1399-3054.1999.106119.x
- Milton, E. J., Rollin, E. M., and Emery, D. R. (1995). “Advances in field spectroscopy,” in *Advances in Environmental Remote Sensing*, eds. F. M. Danson and S. E. Plummer (Chichester, IL: John Wiley & Sons Ltda.), 9–32.
- Mirik, M., Michels, G. J., Kassymhanova-Mirik, S., Elliott, N. C., Catana, V., Jones, D. B., et al. (2006b). Using digital image analysis and spectral reflectance data to quantify damage by greenbug (Hemiptera: Aphididae) in winter wheat. *Comput. Electron. Agric.* 51, 86–98. doi: 10.1016/j.compag.2005.11.004
- Mirik, M., Michels, Jr., G. J., Mirik, S. K., Elliott, N. C., and Catana, V. V. (2006a). Spectral sensing of aphid (Hemiptera: Aphididae) density using field

- spectrometry and radiometry. *Turk. J. Agric. For.* 30, 421–428. Retrieved from: <http://dergipark.gov.tr/tbtkagriculture/issue/11618/138426>
- Mistele, B., and Schmidhalter, U. (2010). Tractor-based quadrilateral spectral reflectance measurements to detect biomass and total aerial nitrogen in winter wheat. *Agron. J.* 102, 499–506. doi: 10.2134/agronj2009.0282
- Mullan, D. (2012). “Spectral radiometry,” in *Physiological Breeding I: Interdisciplinary Approaches to Improve Crop Adaptation*, eds M. P. Reynolds, A. J. D. Pask, and D. M. Mullan (Mexico: CIMMYT). 69–80.
- Peñuelas, J., Baret, F., and Filella, I. (1995). Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica* 31, 221–230.
- Peñuelas, J., and Filella, I. (1998). Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends Plant Sci.* 3, 151–156. doi: 10.1016/S1360-1385(98)01213-8
- Peñuelas, J., Filella, I., Biel, C., Serrano, L., and Savé, R. (1993a). The reflectance at the 950–970 nm region as an indicator of plant water status. *Int. J. Remote Sens.* 14, 1887–1905. doi: 10.1080/01431169308954010
- Peñuelas, J., Filella, I., Gamon, J. A., and Field, C. (1997). Assessing photosynthetic radiation-use efficiency of emergent aquatic vegetation from spectral reflectance. *Aquat. Bot.* 58, 307–315. doi: 10.1016/S0304-3770(97)00042-9
- Peñuelas, J., Gamon, J. A., Fredeen, A. L., Merino, J., and Field, C. B. (1994). Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. *Remote Sens. Environ.* 48, 135–146. doi: 10.1016/0034-4257(94)90136-8
- Peñuelas, J., Gamon, J. A., Griffin, K. L., and Field, C. B. (1993b). Assessing community type, plant biomass, pigment composition, and photosynthetic efficiency of aquatic vegetation from spectral reflectance. *Remote Sens. Environ.* 46, 110–118. doi: 10.1016/0034-4257(93)90088-F
- Peñuelas, J., and Inoue, Y. (1999). Reflectance indices indicative of changes in water and pigment contents of peanut and wheat leaves. *Photosynthetica* 36, 355–360. doi: 10.1023/A:1007033503276
- Pimstein, A., Karnieli, A., Bansal, S. K., and Bonfil, D. J. (2011). Exploring remotely sensed technologies for monitoring wheat potassium and phosphorus using field spectroscopy. *Field Crop. Res.* 121, 125–135. doi: 10.1016/j.fcr.2010.12.001
- Prasad, B., Carver, B. F., Stone, M. L., Babar, M. A., Raun, W. R., and Klatt, A. R. (2007). Potential use of spectral reflectance indices as a selection tool for grain yield in winter wheat under great plains conditions. *Crop Sci.* 47, 1426–1440. doi: 10.2135/cropsci2006.07.0492
- Psomas, A., Kneubuehler, M., Huber, S., Itten, K. I., and Zimmermann, N. E. (2005). “Seasonal variability in spectral reflectance for discriminating grasslands along a dry-mesic gradient in Switzerland,” in *Proceedings of the 4th EARSEL Workshop on Imaging Spectroscopy* (Warsaw), 709–722.
- Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., and Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sens. Environ.* 48, 119–126. doi: 10.1016/0034-4257(94)90134-1
- Rama Rao, N., Garg, P. K., Ghosh, S. K., and Dadhwal, V. K. (2008). Estimation of leaf total chlorophyll and nitrogen concentrations using hyperspectral satellite imagery. *J. Agric. Sci.* 146, 65–75. doi: 10.1017/S0021859607007514
- Raun, W. R., Solie, J. B., Johnson, G. V., Stone, M. L., Lukina, E. V., Thomason, W. E., et al. (2001). In-season prediction of potential grain yield in winter wheat using canopy reflectance. *Agron. J.* 93, 131–138. doi: 10.2134/agronj2001.931131x
- Read, J. J., Tarpley, L., McKinion, J. M., and Reddy, K. R. (2002). Narrow-waveband reflectance ratios for remote estimation of nitrogen status in cotton. *J. Environ. Qual.* 31, 1442–1452. doi: 10.2134/jeq2002.1442
- Richter, K., Atzberger, C., Vuolo, F., Weihs, P., and D'Urso, G. (2009). Experimental assessment of the Sentinel-2 band setting for RTM-based LAI retrieval of sugar beet and maize. *Can. J. Remote Sen.* 35, 230–247. doi: 10.5589/m09-010
- Rodríguez-Pérez, J. R., Ria-o, D., Carlisle, E., Ustin, S., and Smart, D. R. (2007). Evaluation of hyperspectral reflectance indexes to detect grapevine water status in vineyards. *Am. J. Enol. Viticul.* 58, 302–317. Retrieved from: <http://www.ajeonline.org/content/58/3/302>
- Rondeaux, G., Steven, M., and Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* 55, 95–107. doi: 10.1016/0034-4257(95)00186-7
- Roujean, J. -L., and Breon, F. -M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.* 51, 375–384. doi: 10.1016/0034-4257(94)00114-3
- Rouse, J. W. Jr. (1974). *Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation*. NASA/GSFC Type III Final Report. 371.
- Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1973). “Monitoring vegetation systems in the Great Plains with ETRS,” in *Third ETRS Symposium*, NASA SP353, Vol. 1 (Washington, DC), 309–317.
- Salisbury, J. W. (1998). *Spectral Measurement Field Guide*. Earth Satellite Corporation, April 23 1998, Published by the Defence Technology Information Centre as Report No. ADA362372.
- Schaepman, M. E. (1998). *Calibration of a Field Spectroradiometer*. Calibration and Characterisation of a non-imaging field spectroradiometer supporting imaging spectrometer validation and hyperspectral sensor modelling, Remote Sensing Laboratories, Department of Geography, University of Zurich, 1998.
- Schlemmer, M. R., Francis, D. D., Shanahan, J. F., and Schepers, J. S. (2005). Remotely measuring chlorophyll content in corn leaves with differing nitrogen levels and relative water content. *Agron. J.* 97, 106–112. doi: 10.2134/agronj2005.0106
- Serbin, G., Hunt, E. R. Jr., Daughtry, C. S. T., McCarty, G. W., and Doraiswamy, P. C. (2009). An improved ASTER index for remote sensing of crop residue. *Remote Sens.* 1, 971–991. doi: 10.3390/rs1040971
- Serrano, L., Peñuelas, J., and Ustin, S. L. (2002). Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: decomposing biochemical from structural signals. *Remote Sens. Environ.* 81, 355–364. doi: 10.1016/S0034-4257(02)00011-1
- Sims, D. A., and Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.* 81, 337–354. doi: 10.1016/S0034-4257(02)00010-X
- Smith, R. C. G., Adams, J., Stephens, D. J., and Hick, P. T. (1995). Forecasting wheat yield in a Mediterranean-type environment from the NOAA satellite. *Aust. J. Agric. Res.* 46, 113–125. doi: 10.1071/AR9950113
- Steddom, K., Heidel, G., Jones, D., and Rush, C. M. (2003). Remote detection of rhizomania in sugar beets. *Phytopathology* 93, 720–726. doi: 10.1094/PHYTO.2003.93.6.720
- Stroppiana, D., Boschetto, M., Brivio, P. A., and Bocchi, S. (2009). Plant nitrogen concentration in paddy rice from field canopy hyperspectral radiometry. *Field Crop. Res.* 111, 119–129. doi: 10.1016/j.fcr.2008.11.004
- Thenkabail, P. S., Smith, R. B., and De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens. Environ.* 71, 158–182. doi: 10.1016/S0034-4257(99)00067-X
- Tian, Y. C., Gu, K. J., Chu, X., Yao, X., Cao, W. X., and Zhu, Y. (2014). Comparison of different hyperspectral vegetation indices for canopy leaf nitrogen concentration estimation in rice. *Plant Soil* 376, 193–209. doi: 10.1007/s11104-013-1937-0
- Tian, Y. C., Yao, X., Yang, J., Cao, W. X., Hannaway, D. B., and Zhu, Y. (2011). Assessing newly developed and published vegetation indices for estimating rice leaf nitrogen concentration with ground- and space-based hyperspectral reflectance. *Field Crop. Res.* 120, 299–310. doi: 10.1016/j.fcr.2010.11.002
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8, 127–150. doi: 10.1016/0034-4257(79)90013-0
- van Deventer, A. P., Ward, A. D., Gowda, P. M., and Lyon, J. G. (1997). Using thematic mapper data to identify contrasting soil plains and tillage practices. *Photogramm Eng. Rem.* 63, 87–93.
- Viña, A. (2003). *Remote Detection of Biophysical Properties of Plant Canopies*. Available online at: [http://calamps.unl.edu/snrscoq/SNRS\\_Colloquium\\_2002\\_Andres\\_Vina.ppt](http://calamps.unl.edu/snrscoq/SNRS_Colloquium_2002_Andres_Vina.ppt)
- Vincini, M., Frazzi, E., and D'Alessio, P. (2006). “Angular dependence of maize and sugar beet VIs from directional CHRIS/Proba data” in *Proceeding 4th ESA CHRIS PROBA Workshop* (Frascati: ESRIN), 19–21.
- Vogelmann, J. E., Rock, B. N., and Moss, D. M. (1993). Red edge spectral measurements from sugar maple leaves. *Int. J. Remote Sens.* 14, 1563–1575. doi: 10.1080/01431169308953986

- Wang, L., Hunt, J. E. R., Qu, J. J., Hao, X., and Daughtry, C. S. T. (2011a). Towards estimation of canopy foliar biomass with spectral reflectance measurements. *Remote Sens. Environ.* 115, 836–840. doi: 10.1016/j.rse.2010.11.011
- Wang, T. C., Ma, B. L., Xiong, Y. C., Saleem, M. F., and Li, F. M. (2011b). Optical sensing estimation of leaf nitrogen concentration in maize across a range of water-stress levels. *Crop Pasture Sci.* 62, 474–480. doi: 10.1071/CP10374
- White, D. C., Williams, M., and Barr, S. L. (2008). “Detecting sub-surface soil disturbance using hyperspectral first derivative band ratios of associated vegetation stress,” in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Beijing: ISPRS), 243–248.
- White, J. W., Andrade-Sánchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., et al. (2012). Field-based phenomics for plant genetics research. *Field Crops Res.* 133, 101–112. doi: 10.1016/j.fcr.2012.04.003
- Winterhalter, L., Mistele, B., Jampatong, S., and Schmidhalter, U. (2011). High throughput phenotyping of canopy water mass and canopy temperature in well-watered and drought stressed tropical maize hybrids in the vegetative stage. *Eur. J. Agron.* 35, 22–32. doi: 10.1016/j.eja.2011.03.004
- Wu, A., Xiong, X., Cao, C., and Angal, A. (2008a). “Monitoring MODIS calibration stability of visible and near-IR bands from observed top-of-atmosphere BRDF-normalized reflectances over Libyan Desert and Antarctic surfaces,” in *Proceedings of SPIE Conference on Earth Observing Systems XIII*, eds J. J. Butler and J. Xiong (Bellingham, WA: SPIE).
- Wu, C., Niu, Z., Tang, Q., and Huang, W. (2008b). Estimating chlorophyll content from hyperspectral vegetation indices: modeling and validation. *Agric. Forest Meteorol.* 148, 1230–1241. doi: 10.1016/j.agrformet.2008.03.005
- Yañez, P., Retamales, J. B., Lobos, G. A., and del Pozo, A. (2009). Light environment within mature rabbit-eared blueberry canopies influences flower bud formation. *Acta Hort. (ISHS)* 810, 471–473. doi: 10.17660/ActaHortic.2009.810.61
- Yao, H., Tang, L., Tian, L., Brown, R. L., Bhatnagar, D., and Cleveland, T. E. (2011). “Using hyperspectral data in precision farming applications,” in *Hyperspectral Remote Sensing of Vegetation*, eds P. S. Thenkabail, J. G. Lyon, and A. Huete (Boca Raton, FL: CRC Press), 591–608.
- Yao, X., Zhu, Y., Tian, Y., Feng, W., and Cao, W. (2010). Exploring hyperspectral bands and estimation indices for leaf nitrogen accumulation in wheat. *Int. J. Appl. Earth Obs. Geoinf.* 12, 89–100. doi: 10.1016/j.jag.2009.11.008
- Zarco-Tejada, P. J., Berjón, A., López-Lozano, R., Miller, J. R., Martín, P., Cachorro, V., et al. (2005). Assessing vineyard condition with hyperspectral indices: leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sens. Environ.* 99, 271–287. doi: 10.1016/j.rse.2005.09.002
- Zarco-Tejada, P. J., Miller, J. R., Noland, T. L., Mohammed, G. H., and Sampson, P. H. (2001). Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. *IEEE T. Geosci. Remote* 39, 1491–1507. doi: 10.1109/36.934080
- Zarco-Tejada, P. J., Pushnik, J. C., Dobrowski, S., and Ustin, S. L. (2003a). Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. *Remote Sens. Environ.* 84, 283–294. doi: 10.1016/S0034-4257(02)00113-X
- Zarco-Tejada, P. J., Rueda, C. A., and Ustin, S. L. (2003b). Water content estimation in vegetation with MODIS reflectance data and model inversion methods. *Remote Sens. Environ.* 85, 109–124. doi: 10.1016/S0034-4257(02)00197-9
- Zhao, D., Reddy, K. R., Kakani, V. G., Read, J. J., and Koti, S. (2005). Selection of optimum reflectance ratios for estimating leaf nitrogen and chlorophyll concentrations of field-grown cotton. *Agron. J.* 97, 89–98. doi: 10.2134/agronj2005.0089
- Zhu, Y., Tian, Y., Yao, X., Liu, X., and Cao, W. (2007). Analysis of common canopy reflectance spectra for indicating leaf nitrogen concentrations in wheat and rice. *Plant Prod. Sci.* 10, 400–411. doi: 10.1626/pps.10.400

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

Copyright © 2017 Lobos and Poblete-Echeverría. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Comprehensive Phenotypic Investigation of the “Pod-Shattering Syndrome” in Common Bean

Maria L. Murgia<sup>1</sup>, Giovanna Attene<sup>1</sup>, Monica Rodriguez<sup>1</sup>, Elena Bitocchi<sup>2</sup>, Elisa Bellucci<sup>2</sup>, Davide Fois<sup>1</sup>, Laura Nanni<sup>2</sup>, Tania Gioia<sup>3</sup>, Diego M. Albani<sup>4</sup>, Roberto Papa<sup>2\*</sup> and Domenico Rau<sup>1\*</sup>

<sup>1</sup> Dipartimento di Agraria, Sezione di Agronomia, Colture Erbacee e Genetica, Università degli Studi di Sassari, Sassari, Italy

<sup>2</sup> Dipartimento di Scienze Agrarie, Alimentari ed Ambientali, Università Politecnica delle Marche, Ancona, Italy, <sup>3</sup> Scuola di

Scienze Agrarie, Alimentari, Forestali ed Ambientali, Università degli Studi della Basilicata, Potenza, Italy, <sup>4</sup> Dipartimento di

Agraria, Sezione di Economia e Sistemi Arborei e Forestali, Università degli Studi di Sassari, Sassari, Italy

## OPEN ACCESS

### Edited by:

John Doonan,  
Aberystwyth University, UK

### Reviewed by:

M. Teresa Sanchez-Ballesta,  
Institute of Food Science, Technology  
and Nutrition (CSIC), Spain  
Yuhui Chen,  
Samuel Roberts Noble Foundation,  
USA

### \*Correspondence:

Roberto Papa  
r.papa@univpm.it  
Domenico Rau  
dmrau@uniss.it

### Specialty section:

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

Received: 30 September 2016

Accepted: 09 February 2017

Published: 03 March 2017

### Citation:

Murgia ML, Attene G, Rodriguez M, Bitocchi E, Bellucci E, Fois D, Nanni L, Gioia T, Albani DM, Papa R and Rau D (2017) A Comprehensive Phenotypic Investigation of the “Pod-Shattering Syndrome” in Common Bean.

Front. Plant Sci. 8:251.

doi: 10.3389/fpls.2017.00251

Seed shattering in crops is a key domestication trait due to its relevance for seed dispersal, yield, and fundamental questions in evolution (e.g., convergent evolution). Here, we focused on pod shattering in common bean (*Phaseolus vulgaris* L.), the most important legume crop for human consumption in the world. With this main aim, we developed a methodological pipeline that comprises a thorough characterization under field conditions, including also the chemical composition and histological analysis of the pod valves. The pipeline was developed based on the assumption that the shattering trait itself can be treated in principle as a “syndrome” (i.e., a set of correlated different traits) at the pod level. We characterized a population of 267 introgression lines that were developed *ad-hoc* to study shattering in common bean. Three main objectives were sought: (1) to dissect the shattering trait into its “components,” of *level* (percentage of shattering pods per plant) and *mode* (percentage of pods with twisting or non-twisting valves); (2) to test whether shattering is associated to the chemical composition and/or the histological characteristics of the pod valves; and (3) to test the associations between shattering and other plant traits. We can conclude the following: Very high shattering levels can be achieved in different modes; shattering resistance is mainly a qualitative trait; and high shattering levels are correlated with high carbon and lignin contents of the pod valves and with specific histological characteristics of the ventral sheath and the inner fibrous layer of the pod wall. Our data also suggest that shattering comes with a “cost,” as it is associated with low pod size, low seed weight per pod, high pod weight, and low seed to pod-valves ratio; indeed, it can be more exhaustively described as a syndrome at the pod level. Our work suggests that the valve chemical composition (i.e., carbon and lignin content) can be used for a high throughput phenotyping procedures for shattering phenotyping. Finally, we believe that the application of our pipeline will greatly facilitate comparative studies among legume crops, and gene tagging.

**Keywords:** domestication, domestication syndrome, shattering, common bean, phenotypic analysis, element composition analysis, cell wall analysis

## INTRODUCTION

The loss of seed shattering occurred independently in several crops and in different areas of the world during the domestication of many food crops, as this loss was crucial for adaptation of the plants to the agro-ecosystem, to provide ancient farmers with easier and more abundant harvests (Tang et al., 2013). Non-shattering/indehiscent types emerged in maize, barley, and rice (see Li and Olsen, 2016, for a review). Maize was domesticated in the New World, in Mexico, while barley and rice were domesticated in the Fertile Crescent of the Old World and in south-east Asia, respectively. Similarly, among the legume crops, indehiscent phenotypes emerged in soybean and common bean, which were domesticated in the Old World and the New World, respectively (Hymowitz, 1970; Harlan, 1992; Bitocchi et al., 2012; Schmutz et al., 2014). However, fully indehiscent phenotype emerged in common bean only after domestication with the development of snap varieties that are used for the production of green beans due to the absence of fiber strings along the pod valves. In other domesticated commercial classes (e.g., dry beans) shattering traits it is only reduced from that observable in wild populations.

Thus, deciphering the genetic basis of pod shattering is important for evolutionary studies, particularly to unravel the mechanisms of parallel evolution (Lin et al., 2012; Dong and Wang, 2015), and also because this will provide breeders with key information to manipulate this trait to reduce yield loss (Singh, 2001; Santalla et al., 2004). For the same reason, the genomic information would be a great tool to facilitate the exploitation of exotic germplasm in common bean breeding. The potential of these studies is well-represented by those that have been conducted in cereals (Lin et al., 2012). However, in legumes, “studies of the identification of pod-shattering genes lag far behind those of the cereal crops” (Li and Olsen, 2016).

The shattering system of legume crops is distinct from that of cereals (Li and Olsen, 2016). In legumes, dehiscence is subsequent to the “hygroscopic movement” of the pod valves following dehydration. The release of the accumulated elastic tension during dehydration results in the splitting of the valves along their suture lines (Elbaum and Abraham, 2014). The ability to undergo this movement has often been attributed to specific patterns of lignification of the pod-valve tissues.

Among legumes, The most relevant studies on pod dehiscence have been conducted in soybean. Histological analysis has shown that shattering wild genotypes differ from non-shattering varieties in terms of the degree of lignification of the cells along the suture lines of the pod valves (Dong et al., 2014). Among the cultivated germplasm, differential lignification of the lignin-rich inner sclerenchyma of the pod walls also influences the level of shattering (Funatsu et al., 2014). Single major genes underlying these histological differences have also been cloned (Dong et al., 2014; Funatsu et al., 2014). The loss of pod dehiscence has been studied to some extent in lupin, chickpea, pigeonpea, pea yardlong bean, and wild cowpea (Ladizinsky, 1979; Muehlbauer et al., 1998; Boersma et al., 2007, 2009; Weeden, 2007; Abbo et al., 2009; Suanum et al., 2016).

In common bean, there are few such data available. The pioneering studies date to almost a century ago (Lamprecht, 1932; Prakken, 1934). These attempted to index pod-shattering resistance not only based on the occurrence of valve splitting (presence/absence), but also depending on the *mode* of shattering; i.e., based on the degree of torsion (twisting/spiral coiling) of the pod valves after dehiscence (Lamprecht, 1932), and on suggested histological differences between shattering and non-shattering types, mainly in the lignification patterns of the valves tissues (Prakken, 1934). Oligogenic (Lamprecht, 1932) and monogenic (Prakken, 1934) bases for the genetic control of this trait were also proposed. Several decades later, in the pioneer study of Koinange et al. (1996) the pod strings locus (St) was mapped on chromosome 2, and it was proposed to control the differences in shattering between the wax snapbean Midas, an Andean commercial cultivar, and the wild Mesoamerican accession G12873 (Koinange et al., 1996). This locus did not cosegregate with two candidate genes *PvSHP1* and *PvIND*, even if *PvIND* is linked to the St locus (Nanni et al., 2011; Gioia et al., 2012).

The aim of this study was to conduct a comprehensive phenotypic investigation of pod shattering in common bean. With this aim, we also set up a phenotyping pipeline that comprises characterization under field conditions, including the chemical composition, and histological analysis of the pod valves. Following this pipeline, we characterized a population of 267 introgression lines (ILs) that were developed *ad-hoc* to study pod shattering in common bean. In more detail, we pursued the following three goals:

- 1) To dissect the shattering trait into its “components,” as *level* (percentage of shattering pods per plant) and *mode* (percentage of twisting and non-twisting pods per plant). This will answer the question of how the *level* and the *mode* of shattering predict the resistance to (manual) shattering.
- 2) To test whether the *occurrence*, *level*, and *mode* of shattering depend on the chemical composition and histological characteristics of the pod valves. This will help to determine the mechanism of shattering, and it will also allow identification of traits that are useful to surrogate or complement the phenotypic characterization. This might be useful for the development of screening methods that are implementable for high-throughput phenotyping platforms (Fiorani and Schurr, 2013).
- 3) To test the relationships between pod shattering and other plant traits. This will allow the question to be answered in terms of whether pod shattering itself can be treated as a “syndrome” (i.e., a set of correlated phenotypic characteristics), particularly at the pod level, instead of as a single individual trait.

## MATERIALS AND METHODS

### Plant Materials

A population of 267 introgression lines was phenotyped, which was representative of a larger set of about 1200 introgression lines developed in the Papa laboratory (Università Politecnica

delle Marche, Ancona, Italy) in collaboration with the Attene laboratory (Università degli Studi di Sassari, Sassari, Italy). The population was developed starting from a backcross between the line MG38 belonging to the recombinant inbred line (RIL) population used by Koinange et al. (1996) and the recurrent parent MIDAS (**Figure 1**). The MG38 line is a RIL obtained from a cross between the wild Mesoamerican genotype, G12873, and the Andean snap bean variety MIDAS. The MG38 genotype was selected for some wild pod traits (small size, curved shape, pigmented valves, pod shattering), and seed characteristics (very small size). However, for other traits (e.g., determinacy, seed dormancy, photoperiod sensitivity), MG38 was selected for domesticated phenotypes to facilitate the population development and increase. Based on amplified fragment-length polymorphism analysis, MG38 has 55% of the genome attributable to the wild Mesoamerica parent G12873 (Papa, unpublished data). MIDAS is characterized by large and relatively straight, yellow and snap bean non-shattering pods, with relatively large seeds.

To obtain the introgression lines, MG38 was back-crossed with MIDAS as a recurrent parent, and different cycles of back-crossing and selfing were carried out together with selection for the wild characteristics of the pods and seeds. Among the 267 lines analyzed in this study, 70 belonged to BC<sub>3</sub>/F<sub>4</sub>:F<sub>5</sub> families, and 217 to BC<sub>3</sub>/F<sub>6</sub>:F<sub>7</sub> families. Overall, in the field 130 families were represented. Among these families, 101 families were represented by at least two ILs. In some case, there were three ILs per family (i.e., 29 families were represented by one individual). Precisely, there were 19 BC<sub>3</sub>/F<sub>4</sub>:F<sub>5</sub> families and 82 BC<sub>3</sub>/F<sub>6</sub>:F<sub>7</sub> families represented with at least two ILs summing up to 232 ILs.

## Phenotypic Characterization

The phenotyping data presented here were obtained in 2014, between May and October (sowing date, May 19). The experimental layout comprised eight rows, each with 35–38 holes; the distance between rows was 1.5 m; the distance between holes (within the rows) was 0.8 m. For each line, a single plant was grown in each single hole. The two parents, MIDAS and MG38, were replicated three times. The positions of the lines were completely randomized. A plastic sheet was used along each row to facilitate weed control (Supplementary Figure 1A). Standard agronomic practices were adopted, in terms of irrigation, fertilization, and pest control. The meteorological conditions were hot and dry with many days with maximum temperature over 30° (Supplementary Figure 1B). Under these conditions, ILs had the opportunity to fully express their shattering phenotype.

## Measuring Pod Shattering in the Field

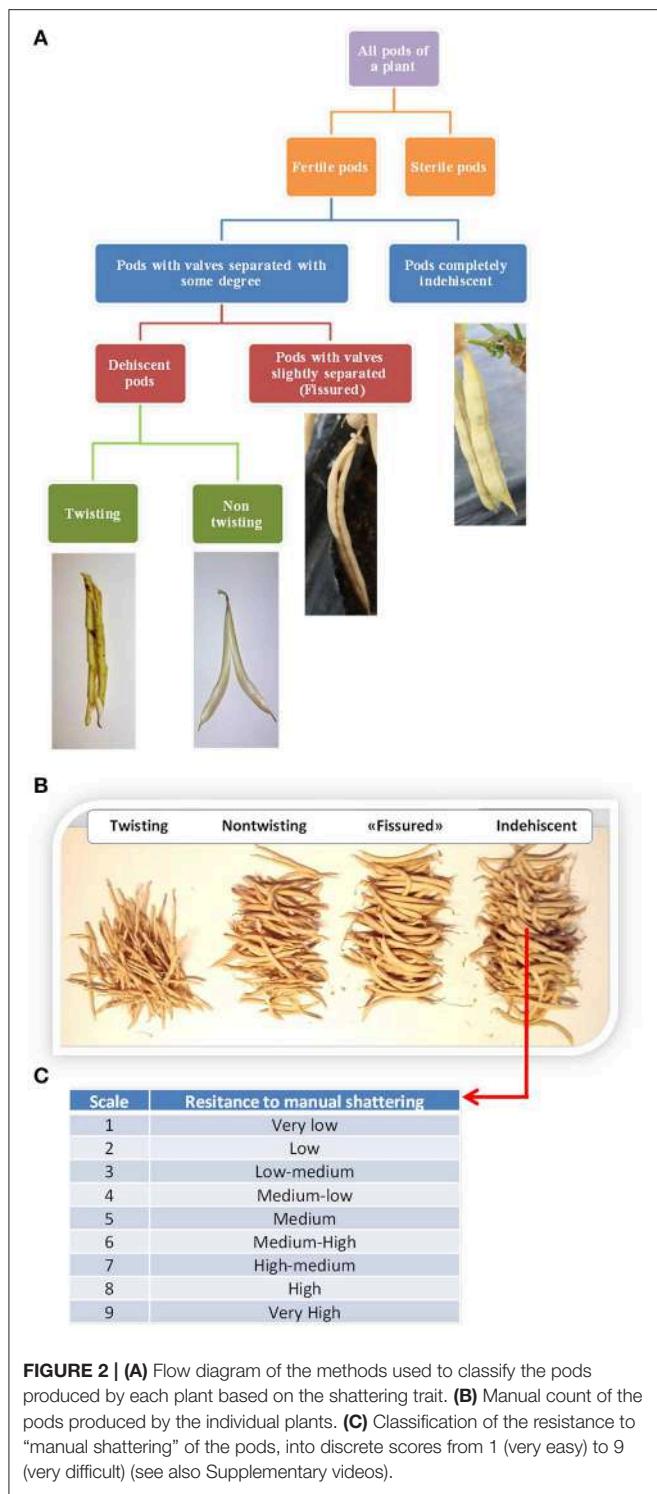
We evaluated shattering after each plant reached full maturity. For each plant, we first distinguished between fertile and sterile pods. The numbers of “naturally” shattering and non-shattering pods were then counted. Fertile pods were further classified into different types, as exemplified in **Figure 2A**. Four pod categories were recognized: Indehiscent; “fissured,” with valves that were not perfectly closed along the ventral suture; dehiscent



**FIGURE 1 | Representative photographs of the differences between the MG38 (Left) and MIDAS (Right) common beans for pod and seed traits (photographs: M.L. Murgia, D. Rau).**

with non-twisting valves; and dehiscent with twisting valves. It was sometimes difficult to distinguish between these last two categories because of the presence of intermediate cases. Nonetheless, on the basis of this classification, the number of pods falling into each category was counted for each plant (**Figure 2B**). For the statistical analysis, the number of pods was expressed as the percentage of the fertile pods produced by each plant.

Furthermore, for each line separately, the shattering of indehiscent pods was promoted by hand, for the evaluation of the “resistance to manual shattering,” based on a scale from 1 (i.e., very low resistance to shattering, where valves abruptly shattered under very light pressure on the distal part of the pod) to 9 (i.e., very high resistance to shattering, where valves did not separate and it was necessary to “break” them) (**Figure 2C**; see also Supplementary Information and videos). To avoid bias, the determination of the resistance to manual shattering was



**FIGURE 2 | (A)** Flow diagram of the methods used to classify the pods produced by each plant based on the shattering trait. **(B)** Manual count of the pods produced by the individual plants. **(C)** Classification of the resistance to “manual shattering” of the pods, into discrete scores from 1 (very easy) to 9 (very difficult) (see also Supplementary videos).

performed independently (i.e., at a different time) from the pod classification.

### Chemical Characterization of the Pod Valves

The chemical composition of the pod valves was investigated to determine whether the pod shattering was correlated to these

characteristics. This element composition analysis looked at carbon, hydrogen, and nitrogen. Here, for each introgression line, 2 g dried pod valves was pulverized in a grinder (18,000 rpm, 1 min). The pulverized tissue was transferred into plastic 50-mL tubes and stored for a few days at room temperature in a cool, dry place. The analyses were performed using 0.080 g pulverized tissue from each line. The samples were combusted at 1,000°C in an excess of oxygen using an element analyser (LECO CHN 628; Leco Corporation, St. Joseph, MI, USA), to determine the carbon, hydrogen, and nitrogen contents. The instrument was calibrated using the “oat meal 502276” forage standard with 46.43% carbon and 2.64% nitrogen. For each run, three independent samples of the standard were included.

The analysis was first performed considering the two parental lines, MIDAS and MG38, with each as three biological replicates (i.e., three plants were grown for each parent). For each biological replicate, there were three technical replicates (i.e., three independent analyses). As there were highly significant differences between the two parents (see Results), the analysis was extended to all of the introgression lines. Three technical replicates were performed for each introgression line.

### Cell-wall Analysis

For each individual plant, 6 g dried valves were pulvressed in a mill (Retsch SM 100) for 10 min. The procedure of Van Soest and Wine (1967) was then followed. First, the neutral detergent fiber was quantified, which represents the total content in the cell wall of the analyzed sample. Thus, the acid detergent fiber was determined, which mainly represents an intermediate step that is necessary to extract the acid detergent lignin, which correlates with the lignin content of the sample analyzed. We also calculated the differences for the neutral detergent fiber minus the acid detergent fiber, and the acid detergent fiber minus the acid detergent lignin, which provided rough estimations of the hemicellulose and cellulose contents, respectively (Van Soest and Wine, 1967). All of these chemical fractions are expressed as percentages of the dried organic matter after subtracting the weight of the ashes (see Supplementary Information for further details).

This analysis was initially performed for MIDAS and MG38, for which three biological replicates were available. For each biological replicate, three technical replicates were included. As the analysis of variance (ANOVA) showed clear-cut differences between the parent lines, the cell-wall analysis was extended to 12 indehiscent introgression lines, and 12 high-shattering introgression lines (>65% shattering, as seen for MG38, the wild-like parent).

### Anatomical and Histological Study of the Pod Valves

This study had the specific aim to look for differential patterns between the shattering and non-shattering lines, particularly for lignin deposition. The analysis was conducted considering 5- and 20-days-old pods, and pods at the maturation stage. The pods were kept in a solution of 95% ethanol and glacial acetic acid (5:2, v/v) for 3 days, and then stored at 4°C in 70% ethanol. Sections of the ventral and dorsal suture sheath were obtained manually. The sections were treated with Javelle water (an aqueous solution

containing sodium hypochlorite and some sodium chloride, used as a bleach and disinfectant) for ~10 min. After this washing, the sections were immersed in 50% acetic acid for a few minutes.

The pod valves were also embedded in paraffin, and 10- $\mu\text{m}$  sections were obtained using a sliding microtome (Reicher) (see Supplementary Information for further details). The manually obtained sections were stained according to two different methods: Toluidine blue O (TBO), and carmine-iodine green; whole microtome sections were stained only with toluidine blue O. The toluidine blue O was used to differentially stain polysaccharides and lignin, whereby cells with thick lignified walls are sky blue, and cellulose and hemicellulose are dark blue (Mitra and Loqué, 2014). With carmine-iodine, lignin is green, and cellulose is pink (Deysson, 1954).

### Phenotyping of the Other Plant Characteristics

To allow the study of the relationships between shattering and the other plant traits, a total of 27 traits were recorded (7 qualitative, 20 quantitative). These were: Number of cotyledonary leaves (two, three); angle of the cotyledonary leaves ( $60^\circ$ ,  $120^\circ$ ,  $180^\circ$ ); lobature of the cotyledonary leaves; stem color (green, red); growth type (non-climbing, intermediate, climbing); flower color (white, light purple, dark purple); pod color (yellow, striped, with from 1 to 3 stripes); plant height (cm); plant vigor (height per width;  $\text{cm}^2$ ); flowering time and pod setting (days from May 19); pod weight per plant (g); valve weight per plant (g); seed weight per plant (g); number of pods per plant; number of seeds per plant; mean pod weight (g); mean valve weight (g); 100-seed weight (g); weight of seeds per pod (g); number of seeds per pod; Harvest Index at pod level. To avoid loss of seeds at the maturation stage in the shattering plants, mature pods were (almost) enveloped in plastic nets (Supplementary Figure 2). Moreover, at the end of the ripening stage (i.e., before shattering occurred), 10 pods per introgression line were randomly sampled. These were scanned, and the acquired images (600 dpi) were processed with the Tomato Analyzer software (Rodríguez et al., 2010), to determine the following pod traits: Perimeters; area; curved height; maximum height; maximum width (Supplementary Figure 3). All of these measures were in pixels. We also calculated the ratio of the curved length to maximum height, where a ratio of 1.0 indicates a perfectly straight pod, while ratios <1 indicate a more or less marked "C" shapes of the pods. All of these variables must be referred to the projection area of the pod on the scanner glass. The procedure was first set up for the parental lines, MG38 and MIDAS. The analysis was then extended to all of the other introgression lines. For statistical analysis, 10 pods per introgression line were considered, and the means were calculated.

### Statistical Analysis

For each variable used to describe shattering, the frequency distribution was first determined. Associations between variables were quantified using Pearson " $r$ " coefficient (quantitative traits) or contingency analysis (qualitative traits). Differences among groups of lines for the various phenotypic and chemical traits were tested using one-way analysis of variance (ANOVA), considering each line as a "replicate" of the group.

Resistance to manual shattering was modeled based on the other six indicators of pod shattering: Indehiscent (%); valves separated to some degree (%); fissured (%); shattering (%); non-twisting (%); twisting (%) (see Figure 2). With this aim, the method of recursive partitioning was adopted, which is also known as decision-trees analysis. This is particularly indicated to investigate relationships among variables without having an *a priori* model, and it is particularly powerful as it considers a very high number of possible partitions, and takes into consideration only the best one (see JMP version 7, User Manual; SAS Institute Inc., Cary, NC, USA). In this case, the categorical X variable was the degree of "resistance to manual shattering" (scored as 1 to 9), while all of the other six indicators of shattering were considered as possible explanatory Y variables. Thus, it was possible to obtain a hierachal system of (dichotomic) criteria that allowed the prediction of the manual shattering *resistance* from the observed *level* and *mode* of shattering. Statistical analysis of the phenotypic data was all performed using JMP version 7 (SAS Institute Inc., Cary, NC, USA).

## RESULTS

### Shattering Level and Mode

As expected, the two parental lines showed highly contrasting phenotypes for pod shattering: MIDAS was completely indehiscent, while MG38 was highly dehiscent, with a mean of 65% shattering pods per plant. Moreover, 98% of the variance for shattering occurrence was located among-families indicating a very limited role of environmental factors influencing this trait in the population of ILs grown under our field conditions (see Supplementary Information and Supplementary Table 1).

**Table 1** gives the descriptive statistics for the six variables measured to characterize the introgression lines for the pod-shattering trait.

The introgression lines were highly variable for the pod-shattering trait, as the indehiscent pods per plant ranged from 3.9 to 100%, with a mean of 50.4%. Twenty-nine introgression lines (~10% of the total) were completely indehiscent. The pods per plant with valves separated to some degree ranged from 0 to 96.1%, with a mean of 49.6%. The distribution of these two variables tended to bimodality (Supplementary Figure 4). The fissured pods per plant ranged from 0 to 71.7%, with a mean of 18.0%.

The *levels* of shattering were highly variable, as the shattering pods per plant ranged from 0 to 82.6%, with a mean of 31.6%. The *modes* of shattering were also highly variable, as the non-twisting and twisting pods per plant both ranged from 0 to ~60%. Non-twisting pods were more frequent than twisting pods, with means of 11.1 and 20.1%, respectively.

The distribution of the trait "resistance to manual shattering" is illustrated in Figure 3. MIDAS had a score of 8 (i.e., high resistance), while MG38 had a score of 2 (i.e., low resistance). The mean for this trait was 4.12 (i.e., medium-low resistance;  $\sigma = 1.96$ ; S.E. = 0.12), and the distribution appeared to be bimodal. About 15% of the introgression lines showed scores of 1 and 2 (i.e.,  $\leq$ MG38), while about 10% showed scores of 8 and 9 (i.e.,  $\geq$ MIDAS).

**TABLE 1 | Descriptive statistics for the six variables used to measure the pod shattering trait in the population of bean introgression lines.**

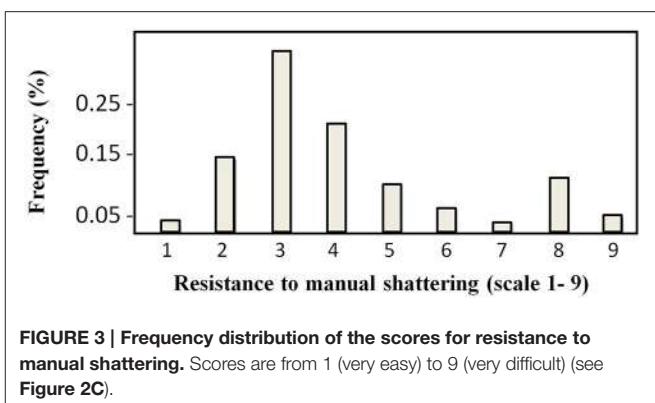
Shattering trait	Statistics			
	Mean (%)	Standard deviation (%)	Standard error of the mean	95% confidence interval
Completely non-dehiscent	50.4	26.7	1.6	47.2–53.6
Valves separated to some degree	49.6	26.7	1.6	46.4–52.8
– Fissured	18.0	13.4	0.8	16.4–19.7
– Dehiscent	31.6	21.3	1.3	29.1–34.1
– Twisting	11.5	10.9	0.7	10.2–12.8
– Non-twisting	20.1	14.3	0.9	18.4–21.8

### Relationships among the Measures of Shattering

**Figure 4A** shows the relationships between the *levels* and *modes* of shattering. Introgression lines with the same or very similar levels of shattering (percentage shattering pods per plant) showed a very different ratio between the twisting and non-twisting types. For example, among the introgression lines with very high levels of shattering (>65%), the ratio of non-twisting to twisting pods per plant varied from ~25:50 (1:2) to about 60:15 (4:1). Moreover, these data also suggested that transgressive variation probably occurred for ~10% of the lines, which showed higher shattering than MG38 (>65%), the highly shattering parental line (**Figure 4A**).

The level of shattering was more strongly correlated with the frequency of non-twisting pods ( $R^2 = 0.71, P < 10^{-4}$ ) than with the frequency of twisting pods ( $R^2 = 0.57; P < 10^{-4}$ ) (**Figure 4B**). In particular, while a low number of twisting pods corresponded to different levels of shattering, a low number of non-twisting pods was more indicative of low levels of shattering.

The resistance to manual shattering was modeled considering the six variables measured to dissect out the shattering trait (**Figure 2**, **Table 1**), with recursive partition analysis applied (**Figure 5**). The variable that best predicted resistance to manual shattering was the shattering pods per plant; i.e., the *level* of shattering. Indeed, the threshold of 10% shattering pods defined two groups of plants with different mean shattering resistance scores of 3.3 and 7.1; this partition captured 65% of the total variance for shattering resistance ( $P < 0.0001$ ; **Figure 5**). A second partition suggested a role for the *mode* of shattering. Indeed, within the group of introgression lines showing <10% shattering pods per plant, a threshold of 9% twisting pods defined two subgroups of plants that had mean shattering resistance scores of 2.9 and 4.1. This partition captured a small portion, 6%, of the total variance for resistance to manual shattering ( $P < 0.0001$ ; **Figure 5**). A third partition was found within the group of introgression lines with shattering pods  $\geq 10\%$ . Here, the threshold of 4.2% of non-sigaroid pods defined two subgroups of plants with mean shattering resistance scores of 5.3 and 7.5, and these explained an additional 4% of the total variance for shattering resistance (**Figure 5**). Thus, cumulatively these three partitions explained 75% of the total variance. The

**FIGURE 3 | Frequency distribution of the scores for resistance to manual shattering.** Scores are from 1 (very easy) to 9 (very difficult) (see **Figure 2C**).

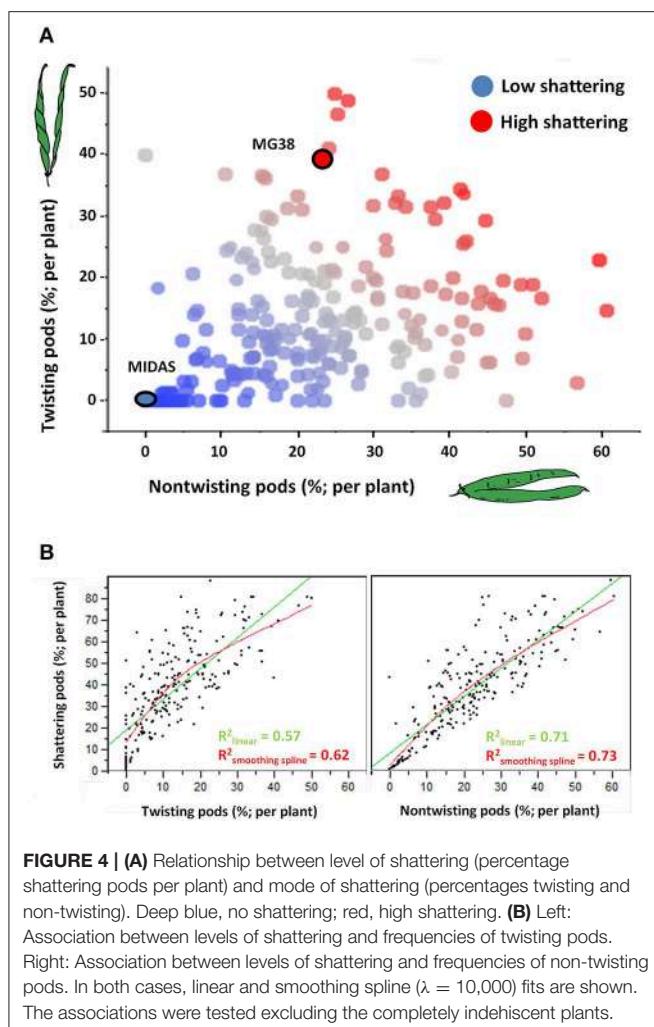
fourth partition (not shown) explained 0.8% of the total variance for shattering resistance, which indicated that dealing with a more complex model was not necessary.

### Chemical Analysis

The pod valves of the two parental lines, MG38 and MIDAS, had significantly different carbon contents (ANOVA,  $P < 0.0001$ ; **Table 2**). The highly dehiscent MG38 had a carbon content of 43.8% dry weight, which gave a 6.8% increase in the carbon content of the indehiscent MIDAS, from 41.0% dry weight (**Table 3**). ANOVA also revealed a marginally significant difference for the hydrogen contents ( $P < 0.047$ ; **Table 2**), again in favor of MG38 (6.7% dry weight) compared to MIDAS (6.5% dry weight; **Table 3**). The difference in the nitrogen contents was not significant ( $P = 0.502$ ) (**Tables 2, 3**).

The comparison of the indehiscent vs. dehiscent introgression lines was highly significant for the carbon contents (**Table 2**). The dehiscent introgression lines showed a 6.9% increase in the carbon content of the indehiscent introgression lines, according to dry weight (**Table 3**). The indehiscent introgression lines had the same carbon content as MIDAS, while the dehiscent introgression lines had the same carbon content as MG38. The difference between these dehiscent and indehiscent introgression lines was small, but significant for the hydrogen content although not for the nitrogen content (**Tables 2, 3**). The frequency distribution for the carbon contents tended to be bimodal, while this was less evident for the hydrogen and nitrogen contents (Supplementary Figure 5).

The relationship between the carbon contents and the shattering pods per plant (**Figure 6**) showed an abrupt transition in the carbon content that occurred between 5 and 10% shattering pods per plant (**Figure 6A**). Partition analysis showed that this transition occurred at 7.14% shattering pods per plant (not shown). The definition of the introgression lines into two groups based on this transition, with the first with <7.14% and the second with  $\geq 7.14\%$  shattering pods per plant, captured 47% of the total variance for the carbon contents (**Table 2**). When the introgression lines with <7.14% shattering pods were excluded from the analysis, there was a weak, but significant, negative correlation ( $r = -0.296; P < 0.0001$ ) between shattering level and carbon content (**Figure 6B**).



## Cell-wall Analysis

The pod valves of MIDAS and MG38 had significantly different total fiber contents (ANOVA,  $P < 0.0001$ ; **Table 4**). The fiber content of the highly dehiscent MG38 (62.0%) showed a ~48% increase of that of the completely indehiscent MIDAS (42.0%; **Figure 7A**). The contents of the three cell-wall components of lignin, hemicellulose, and cellulose were always higher for MG38 than MIDAS, with the greatest difference seen for lignin, followed by hemicellulose and cellulose (**Figures 7B–D**). Statistical analysis was carried out to compare the two groups of introgression lines, the first comprising the completely indehiscent plants (i.e., non-shattering), and the second including the plants with  $>65\%$  shattering pods per plant (i.e., high shattering) (**Table 4**, **Figures 7A–D**). These two groups strongly differed in their lignin contents, with the difference for the latter representing a 180% increase of the former, which was much greater than the increase from the MIDAS to MG38 parental lines (80%) (**Figure 7B**). The high shattering group also showed increases of the non-shattering group for hemicellulose (33.9%) and cellulose (7.6%) contents, which were here less than for the parental lines (79.5%, 30.1%, respectively) (**Figures 7C,D**). These highly dehiscent (i.e., high shattering) introgression lines showed lower total fiber content than MG38 (**Figure 7A**), which was mainly due to reduction in the cellulose content, and to a slight, although not statistically significant, reduction in the hemicellulose content (**Figure 7D**). In contrast, these lines had higher lignin content compared to MG38 (**Figure 7B**). This suggested that the achievement of the very high pod shattering ability (here even higher than the wild-like MG38) is associated with an increase in the proportion of lignin in the cell wall.

respectively) (**Figures 7C,D**). These highly dehiscent (i.e., high shattering) introgression lines showed lower total fiber content than MG38 (**Figure 7A**), which was mainly due to reduction in the cellulose content, and to a slight, although not statistically significant, reduction in the hemicellulose content (**Figure 7D**). In contrast, these lines had higher lignin content compared to MG38 (**Figure 7B**). This suggested that the achievement of the very high pod shattering ability (here even higher than the wild-like MG38) is associated with an increase in the proportion of lignin in the cell wall.

## Correlations between the Element Compositions and the Cell-wall Analysis

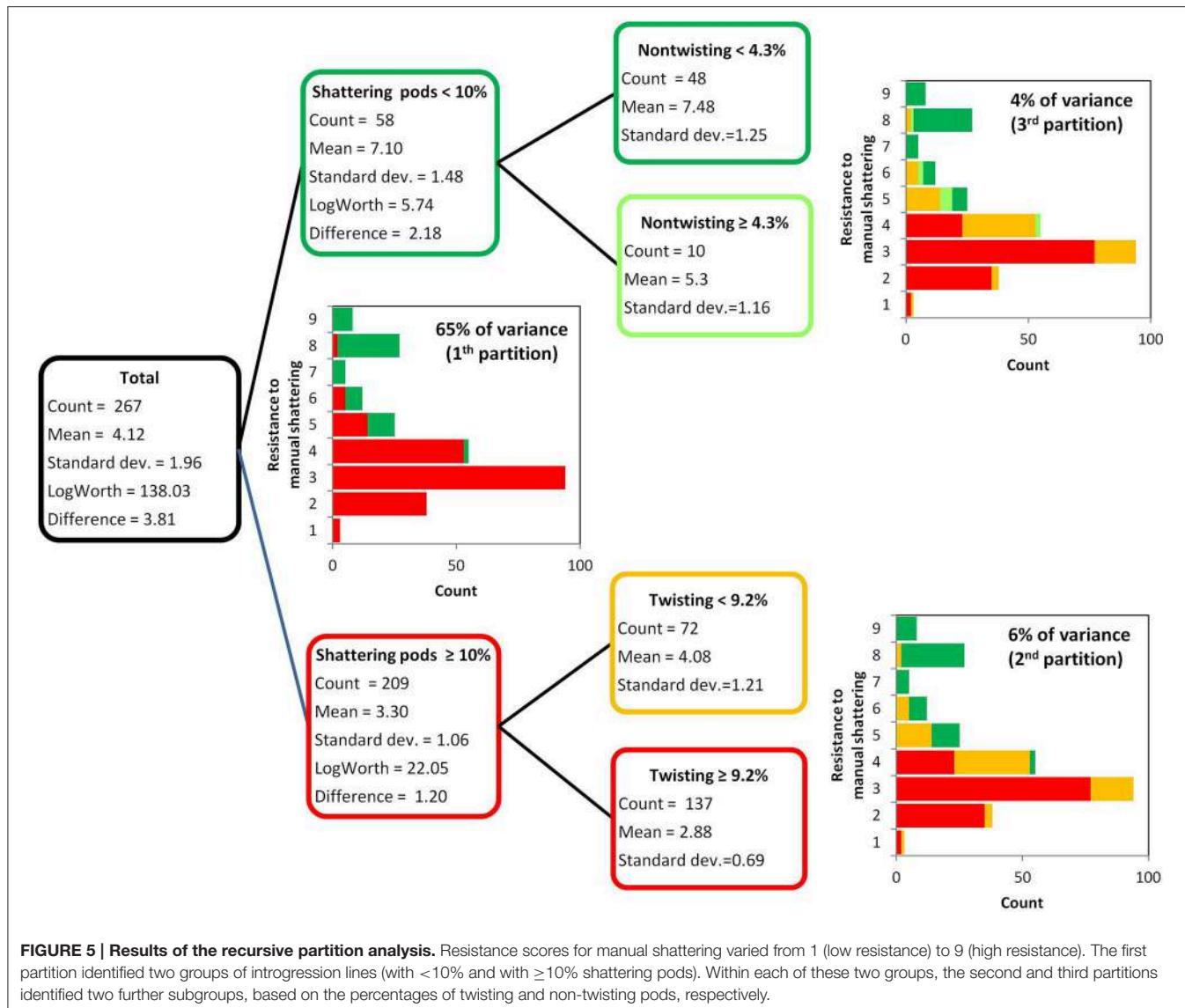
The carbon content was strongly correlated with the total fiber content of the pod valves ( $r = 0.685$ ); moreover, among the three cell-wall components (i.e., lignin, hemicellulose, cellulose), the carbon content showed the best correlation with the lignin content ( $r = 0.672$ ; **Table 5**). Stepwise multiple regression analysis was performed with the carbon content as the dependent variable and lignin, hemicellulose, and cellulose as the independent variables (Supplementary Table 2). Here, the only variable that entered into the model was the lignin content ( $P = 1.85 \times 10^{-5}$ ). This thus indicates that the hemicellulose and cellulose correlation to the carbon content was mainly due to their correlation with the lignin content.

## Anatomical and Histological Analysis of the Pod Valves

The analysis conducted with 5-day-old pods showed no obvious differences between the shattering and non-shattering genotypes (Supplementary Figure 6). The ventral sheath showed only a few cells with very low levels of lignification. A similar situation was observed for the dorsal sheath (not shown). There was no lignin deposition in correspondence with the inner parenchyma cells of the pod walls.

Analyses of 20-day-old pods showed evident lignin deposition in the ventral sheath of the pod valves, and a clear-cut difference between the shattering (i.e., MG38) and non-shattering (i.e., MIDAS) genotypes (**Figures 8A,B**). Indeed, the proportion of cells with thick secondary cell-wall formation (i.e., sclerenchymatic cells) was clearly greater for MG38 (highly dehiscent), compared to MIDAS (indehiscent). The absence of cells with thick secondary cell-wall formation for MG38 was limited only to the external layer of the cells of the sheath and to the dehiscence zone, while for MIDAS this involved all of the sheath (**Figure 8B**). Moreover, for MG38, the cell-wall thickness tended to reduce when moving from the sheath to the dehiscence zone (**Figure 8A**), where there was the tendency to easily “fracture” (**Figure 8A**). A similar pattern was observed for the dorsal sheath (Supplementary Figure 7). A clear-cut difference was also seen between the parental MG38 and MIDAS for the degree of lignification in the inner cells of the pod walls, with very strong lignification (i.e., sclerenchyma) for MG38, and complete absence of lignin deposition for MIDAS (**Figure 9**).

As not all anatomical or histological differences between the wild-like parent (MG38) and the cultivated varieties



**FIGURE 5 | Results of the recursive partition analysis.** Resistance scores for manual shattering varied from 1 (low resistance) to 9 (high resistance). The first partition identified two groups of introgression lines (with <10% and with  $\geq 10\%$  shattering pods). Within each of these two groups, the second and third partitions identified two further subgroups, based on the percentages of twisting and non-twisting pods, respectively.

**TABLE 2 | Results for the ANOVA performed for the chemical element analysis of the pod valves.**

Element	MIDAS vs. MG38				Indehiscent vs. dehiscent				<7.14 vs. $\geq 7.14\%$ shattering			
	M.S.	$F_{1,4}$	P	$R^2_{adj}$	M.S.	$F_{1, 227}$	P	$R^2_{adj}$	M.S.	$F_{1, 227}$	P	$R^2_{adj}$
Carbon	12.30	256.0	< 0.0001	0.92	191.93	105.2	< 0.0001	0.31	286.04	202.92	< 0.0001	0.47
Hydrogen	0.04	8.1	0.047	0.15	0.72	6.7	0.010	0.02	0.29	2.63	0.110	0.01
Nitrogen	0.00	0.5	0.502	0.00	0.10	2.3	0.130	0.01	0.05	1.34	0.250	0.00

ANOVA was applied to the following comparisons: Parental lines MG38 (dehiscent) vs. MIDAS (indehiscent); indehiscent vs. dehiscent introgression lines; and introgression lines with <7.14 vs.  $\geq 7.14\%$  shattering pods. M.S., Mean Square;  $F_{x,y}$ , F ratio with x and y degrees of freedom for the numerator and denominator, respectively.  $R^2_{adj}$ , adjusted  $R^2$ ; P, significance level.

(MIDAS) are necessarily correlated with the shattering traits, two introgression lines (one with shattering >MG38, the other without shattering) were also compared (Supplementary Figures 8A–C). Encouragingly, here the patterns were similar to those observed between MG38 and MIDAS, which suggests

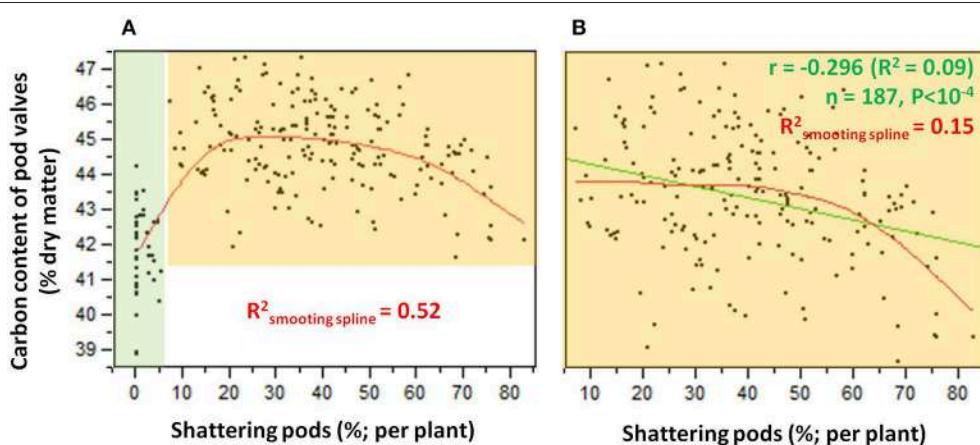
that the histological differences seen do indeed underlie the shattering/non-shattering phenotypes.

It was difficult to obtain good sections of the pod valves at the maturation stage because of the fragility of the tissue. However, it can be noted that at this stage, the ventral sheath of MIDAS had

**TABLE 3 | Mean contents of carbon, hydrogen, and nitrogen of the pod valves as applied to the following comparisons: two parental lines MG38 (dehiscent) vs. MIDAS (inindehiscent); indehiscent vs. dehiscent introgression lines; and introgression lines with <7.14 vs. ≥7.14% shattering pods.**

Element	Comparison	N	Statistic (% dry weight)		Comparison	N	Statistic (% dry weight)		Comparison	N	Statistic (% dry weight)	
			Mean	95% confidence interval			Mean	95% confidence interval			Mean	95% confidence interval
Carbon	Midas	3	41.0	40.7–41.4	Indehiscent	26	41.7A	41.2–42.2	<7.14%	41A	41.8	41.5–42.2
	MG38	3	43.9	43.5–44.2	Dehiscent	203	44.6B	44.4–44.8	≥7.14%	188B	44.8	44.6–44.9
Hydrogen	Midas	3	6.5	6.4–6.6	Indehiscent	26	6.3A	6.2–6.5	<7.14%	41	6.4	6.3–6.5
	MG38	3	6.7	6.6–6.8	Dehiscent	203	6.5B	6.5–6.6	≥7.14%	188	6.5	6.5–6.6
Nitrogen	Midas	3	0.6	0.5–0.8	Indehiscent	26	0.5	0.5–0.6	<7.14%	41	0.5	0.5–0.6
	MG38	3	0.6	0.5–0.7	Dehiscent	203	0.5	0.5–0.5	≥7.14%	188	0.5	0.5–0.5

Groups of introgression lines with different letters have significantly different means ( $P < 0.05$ ; Tukey-Kramer multiple comparison tests).



**FIGURE 6 | (A)** Relationship between carbon contents and levels of shattering. Green shading, individuals for which shattering was <7.14%; orange shading, individuals for which shattering was ≥7.14%. **(B)** Relationship between carbon contents and shattering levels excluding individuals with low or no shattering.  $R^2$  given for smoothing spline ( $\lambda = 10,000$ ) (red) and linear fit (green).

**TABLE 4 | Results for the ANOVA performed for the cell-wall analysis.**

Cell-wall component	Midas vs. MG38				Indehiscent vs. highly dehiscent			
	M.S.	F <sub>1, 4</sub>	P	R <sup>2</sup> <sub>adj</sub>	M.S.	F <sub>1, 22</sub>	P	R <sup>2</sup> <sub>adj</sub>
Total fiber (NDF)	600.2	98.7	<10 <sup>-3</sup>	0.95	825.15	273.04	<.0001	0.92
Lignin (ADL)	9.5	163.0	<10 <sup>-3</sup>	0.97	115.74	305.35	<.0001	0.93
Hemicellulose (NDF-ADF)	129.3	14.0	0.02	0.72	151.43	140.95	<.0001	0.86
Cellulose (ADF-ADL)	100.9	134.1	<10 <sup>-3</sup>	0.96	32.05	14.64	0.0009	0.37

ANOVA was applied to the following comparisons: Parental lines MG38 (dehiscent) vs. MIDAS (inindehiscent); and two groups of 12 introgression lines, one completely indehiscent vs. one highly dehiscent (i.e., with percentage shattering pods >65% of MG38). M.S., Mean Square; F<sub>x/y</sub>, F ratio with x and y degrees of freedom for the numerator and denominator, respectively. R<sup>2</sup><sub>adj</sub>, adjusted R<sup>2</sup>; P, significance level.

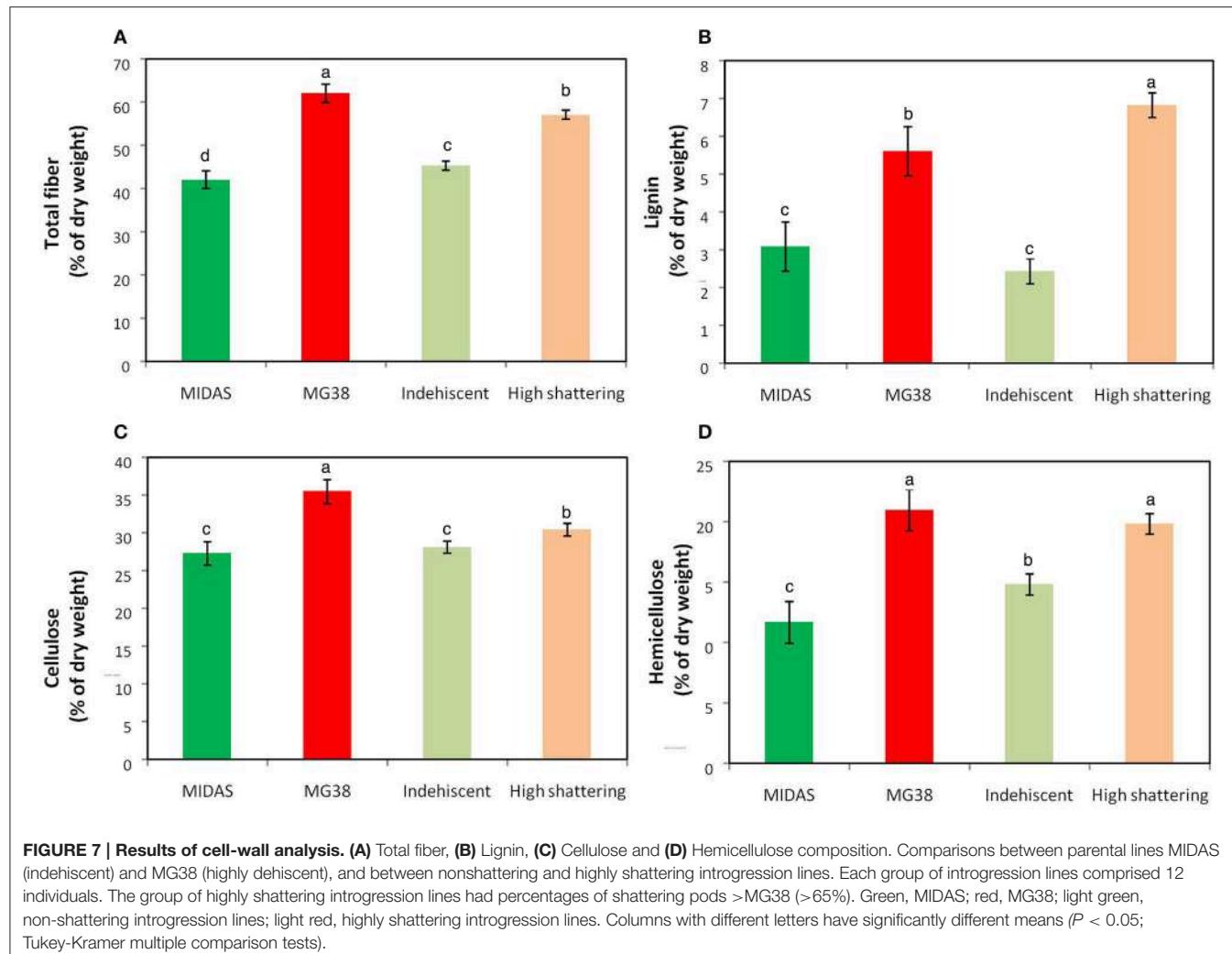
NDF, neutral detergent fiber; ADL, acid detergent lignin; ADF, acid detergent fiber.

more mechanical resistance than that of MG38, which appeared to be very fragile instead (Supplementary Figure 9).

## Relationships between Pod Shattering and the Other Plant Characteristics

Table 6 gives the associations between the levels of shattering (percentage shattering pods per plant) and the other 28

phenotypic traits, of which seven are qualitative and 21 are quantitative. These included morphological, phenological, and productive traits. Overall, shattering was very poorly correlated with all of these plant traits considered. Significant weak associations were detected for three qualitative traits, four quantitative-productive traits, and six precision phenotyping traits that describe pod size and shape (Table 6).



For the associations with the qualitative traits, it was observed that: Plants with red stems showed higher shattering ( $59.6 \pm 12.66\%$ ) than those with green stems ( $31.25 \pm 1.30\%$ ); plants with white flowers showed higher shattering ( $36.52 \pm 1.44\%$ ) than those with purple ( $22.85 \pm 2.67\%$ ) and light purple ( $13.66 \pm 3.96\%$ ) flowers; plants with yellow pods showed higher shattering ( $35.89 \pm 1.44\%$ ) than those with striped/yellow pods ( $17.58 \pm 2.80\%$ ), with an intermediate position seen for those with striped pods ( $30.55 \pm 6.33\%$ ) and yellow/striped pods ( $24.92 \pm 5.77\%$ ).

For the correlations with the productive traits, valve weight per plant and mean valve weight increased when the shattering level increased, with the opposite for 100-seed weight and Harvest Index at pod level, which decreased when the shattering level increased (Table 6). In more detail, the one-way ANOVA between the dehiscent vs. indehiscent introgression lines showed that the former had significantly higher valve weight per plant and mean valve weight than the latter, with increases in the indehiscent introgression lines of  $35.4\%$  ( $P = 0.0085$ ;  $t$ -test) and  $26.2\%$  ( $P = 0.0004$ ). In contrast, the opposite was seen for the 100-seed weight and the Harvest Index at pod level, where the indehiscent introgression lines showed an increase of  $15.9\%$  ( $P$

=  $0.0002$ ) and  $8.9\%$  ( $P < 0.0001$ ), respectively, to the dehiscent introgression lines.

The correlations between the levels of shattering with the six precision phenotyping variables that describe the pod morphology were significant (from  $P < 0.0001$  to  $= 0.021$ ) and all negative (Table 6).

## DISCUSSION

### Field Phenotyping of the Shattering Trait in Common Bean

High variations for both *levels* and *modes* of pod shattering were recorded. All of the shattering types were distinguishable, which varied from completely indehiscent to “twisting,” passing through the two defined “intermediate” states of “fissured” and “shattering but non-twisting” (Lamprecht, 1932). Each introgression line was characterized by counting and classifying the pods into these four categories, with the degree of resistance to manual shattering also independently measured.

As shown by the partition analysis, the best predictor of resistance to manual shattering was the *level* of shattering

**TABLE 5 | Correlations between the element compositions and cell-wall fiber contents for the pod valves of the 24 introgression lines, 12 non-shattering, and 12 very high shattering.**

Cell-wall component	Element composition		
	%Carbon	%Hydrogen	%Nitrogen
Total fiber (NDF)	0.685***	0.038	0.218
Lignin (ADL)	0.672***	0.081	0.247
Hemicellulose (NDF-ADF)	0.623**	0.140	0.195
Cellulose (ADF-ADL)	0.527*	-0.183	0.134

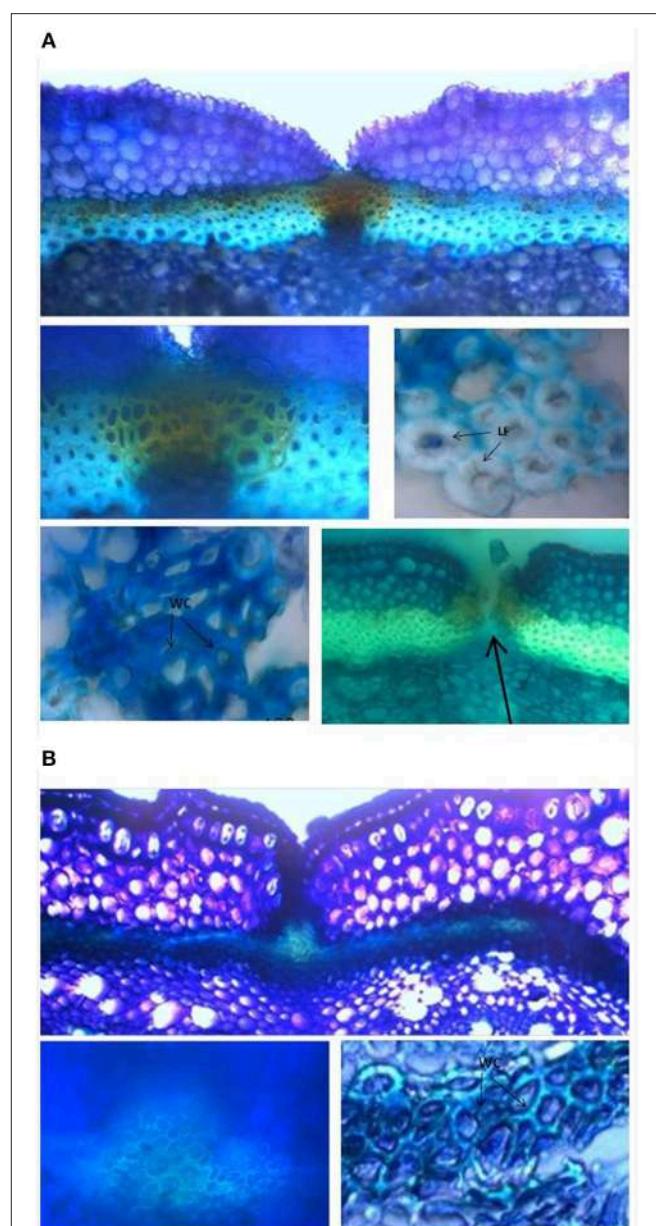
NDF, neutral detergent fiber; ADL, acid detergent lignin; ADF, acid detergent fiber; \*P < 0.05; \*\*P < 0.01; \*\*\*P < 0.001.

(percentage of shattering pods per plant), while the *mode* of shattering (twisting/non-twisting) was less relevant. Moreover, a low threshold of shattering pods per plant (10%) was sufficient to distinguish between the low and medium-high resistant introgression lines. All this suggests that shattering might be controlled by the “switching” of the mechanism of control that determines the abrupt change in the possibility of splitting the pod valves. These data also indicate that when considering both natural or artificial plant populations, genetic studies aimed at deciphering the genetic architecture of the pod shattering trait would benefit from a step-wise approach that comprises the following: (1) comparing indehiscent vs. dehiscent introgression lines (regardless of the degree of shattering); (2) considering only dehiscent introgression lines (regardless of the mode); and (3) considering separately among the dehiscent introgression lines those with twisting and non-twisting pods. Indeed, this approach would allow the genetic basis of the *occurrence* of shattering (yes/no) and also its *tuning* (low/high) and *mode* (twisting/non-twisting) to be described. It should also be noted that the variable of “fissured pods” did not prove useful to predict resistance to manual shattering; this suggests that this trait would be better investigated separately from the others.

## Element Composition and Cell-wall Analysis of the Pod Valves

These shattering and non-shattering genotypes clearly differed in their carbon contents. The contents of carbon, hydrogen, and nitrogen are expected to be stoichiometrically correlated to the amount of organic matter in the tissues (Chiariello et al., 2000), and thus to the cumulative content of carbohydrate, protein, lipid, and all other organic compounds. However, in plants, differences in the carbon content have frequently been correlated to differences in lignin content (Loader et al., 2003). This was also the case for the valves of common bean; indeed, the cell-wall analysis here confirmed that the differences in the carbon contents between the shattering and non-shattering types were mainly correlated with the differences in the lignin contents, in comparison with the other cell-wall components (i.e., hemicellulose, cellulose).

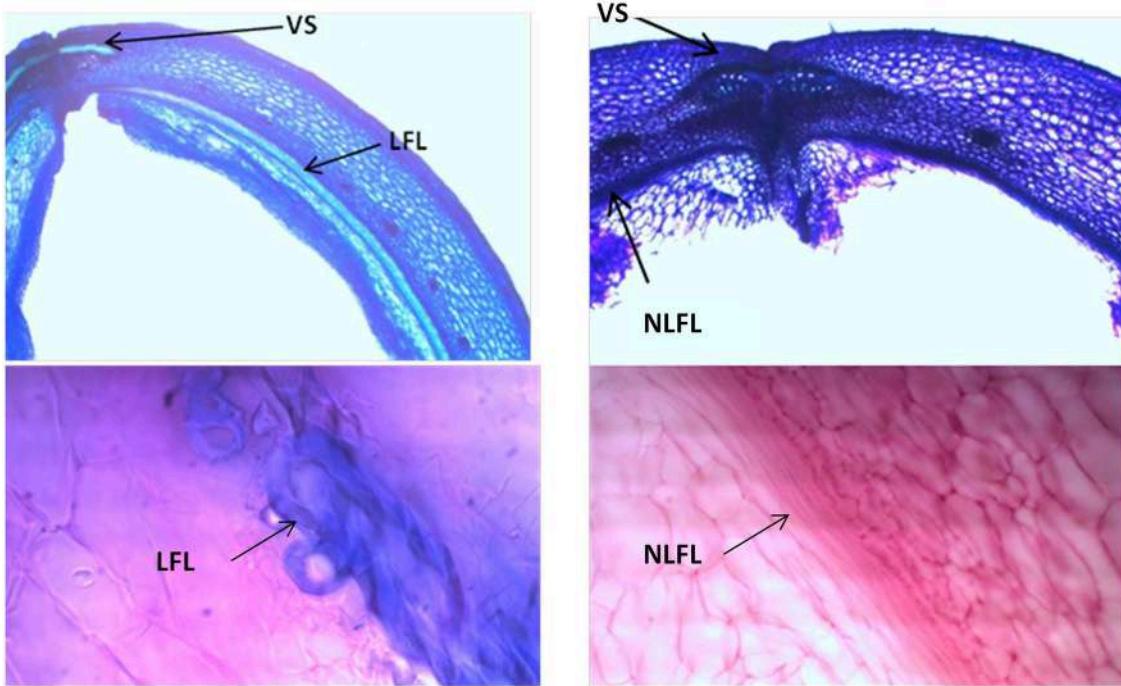
There was an abrupt increase in the carbon content at a level of shattering of ~7.14%. This value was similar to the threshold of 10% of the shattering pods per plants that explained the largest proportion (65%) of the variance for



**FIGURE 8 | Representative images from the histological study of the ventral sheath of the pod valves. (A)** toluidine blue O (TBO) staining, showing ventral sheath of pod valves from MG38 (highly dehiscent), and details of dehiscent zone, thick lignified fibers (sclerenchyma), wood cells, dehiscent zone after cracking (arrow). **(B)** TBO staining, showing ventral sheath of pod valves from MIDAS (indehiscent), with two details of the indehiscence zone. LF, lignified fibers; WC, wood cells.

resistance to manual shattering. These observations suggest that environmental effects might act on the level of shattering, and that the complementation of the whole-plant characterization and chemical element composition analysis can lead to more precise and alternative or complementary phenotyping option.

The comparison of the data in the present study with those from the literature reveal differences between common bean and



**FIGURE 9 | Representative images with toluidine blue O (TBO) (Upper panels) and carmine-iodine green staining (Lower panels), illustrating differences between pod valves from MG38 (shattering; left) and MIDAS (non-shattering; right) in terms of the degree of lignification of the inner layer of the sclerenchymatic cells of the pod wall. VS, ventral sheath; LFL, lignified fiber layer; NLFL, non-lignified fiber layer.**

soybean. Indeed, in soybean, the high-shattering cultivars were shown to have similar lignin contents (not higher, as observed here for common bean) to the low-shattering cultivars (see Table 1 of Romkaew et al., 2008). Moreover, in F2 and back-crossed populations between yardlong and wild cowpea, among these three fiber components, the contents of hemicellulose showed the highest correlation with pod shattering (Suanum et al., 2016). All this suggests that there are histological differences between other legumes species and common bean, which appear to be due to differences in the patterns of cell-wall lignification of the pod tissues, or differences in the prevalent fiber type.

## Histological Characterization of the Pod Valves

The data from the histological characterization of pod valves in the present study parallel the observations at the chemical level. Indeed, overall, cell-wall lignification is much more pronounced in the shattering type than the non-shattering type for common bean. Specifically, the ventral sheath of the wild-like genotype (MG38) was characterized by very strong sclerenchymatization of the cells, while the opposite was seen for the domesticated cultivar (MIDAS). This observation is consistent with Prakken (1934), who indicated this anatomical difference at the basis of the presence/absence of pod strings, and on the basis of the shattering/non-shattering phenotypes. Moreover, the presence/absence of pod strings has been shown to be under the control of the *St* gene (Koinange et al., 1996).

The histological differences between the shattering and non-shattering genotypes for common bean in the present study appear to be more pronounced than those for soybean. Indeed, for soybean, the differences were limited to the dehiscent zone, where excessive lignification of the fiber cap cells was seen in the cultivated non-shattering genotypes, as compared to the wild shattering genotypes (Dong et al., 2014). Furthermore, a major gene, known as *SHA1-5*, was identified as being responsible for lignin deposition in the fiber cap cells of soybean (2014). The present study did not show any clear histological differences in the common bean dehiscence zone. Albeit it cannot be completely excluded that there were some undetected histological differences here in the dehiscence zone in common bean, these data suggest that the histological basis of pod shattering in bean and soybean are different, at least partially.

Furthermore, the shattering genotypes here had a fibrous and strongly lignified cell layer between the inner and outer parenchyma of the pod wall, while this was not seen for the non-shattering genotypes. This difference was also noted for common bean by Prakken (1934), in their comparison of the “stringy” and “stringless” types. Funatsuki et al. (2014) noted that in cultivated soybean, the differential lignification in the lignin-rich inner sclerenchyma of the pod walls influenced valve twisting and pod shattering. Furthermore, they showed that lignin deposition in this layer was under the control of a major gene, known as *PDH1*. We note here that in soybean, the difference between the shattering and non-shattering types appears to be in the *degree* of lignification of this inner sclerenchyma layer (Funatsuki et al.,

**TABLE 6 | Correlations between level of shattering (percentage shattering pods per plant) and the 28 phenotypic traits.**

Trait	Association			
	$R^2_{adj}$	M.S.	F	P
Qualitative				
Number of cotyledonary leaves	0.000	106.86	0.24	0.628
Angle of cotyledonary leaves	0.000	99.21	0.21	0.810
Lobature of cotyledonary leaves	0.010	672.75	1.48	0.229
Stem color	<b>0.020</b>	<b>2446.21</b>	<b>5.50</b>	<b>0.020</b>
Growth type	0.000	15.59	0.04	0.950
Flower color	<b>0.140</b>	<b>8396.23</b>	<b>21.45</b>	<b>&lt;10<sup>-4</sup></b>
Pod color	<b>0.120</b>	<b>4708.32</b>	<b>11.75</b>	<b>&lt;10<sup>-4</sup></b>
Quantitative	$R^2$	R	N	P
Plant height	0.010	0.102	264	0.099
Flowering time	0.000	0.000	266	0.998
Pod setting	0.001	0.034	264	0.058
Pods weight per plant	0.003	0.056	267	0.366
Valve weight per plant	<b>0.029</b>	<b>0.169</b>	<b>264</b>	<b>0.006</b>
Seed weight per plant	0.000	-0.003	265	0.949
Number of pods per plant	0.014	0.117	265	0.055
Number of seeds per plant	0.010	0.100	266	0.103
Mean pod weight	0.001	-0.031	267	0.611
Mean valve weight	<b>0.071</b>	<b>0.267</b>	<b>264</b>	<b>&lt;10<sup>-4</sup></b>
100-seed weight	<b>0.045</b>	<b>-0.212</b>	<b>265</b>	<b>0.001</b>
Seeds per pod	0.014	-0.119	265	0.052
Number of seeds per pod	0.006	0.770	266	0.211
Harvest Index at the pod level	<b>0.106</b>	<b>-0.327</b>	<b>265</b>	<b>&lt;10<sup>-4</sup></b>
Pod perimeter	<b>0.089</b>	<b>-0.298</b>	<b>259</b>	<b>&lt;10<sup>-4</sup></b>
Pod area	<b>0.022</b>	<b>-0.148</b>	<b>259</b>	<b>0.017</b>
Pod maximum width	<b>0.181</b>	<b>-0.425</b>	<b>259</b>	<b>&lt;10<sup>-4</sup></b>
Pod maximum height	<b>0.020</b>	<b>-0.143</b>	<b>259</b>	<b>0.021</b>
Pod curved height	<b>0.071</b>	<b>-0.266</b>	<b>259</b>	<b>&lt;10<sup>-4</sup></b>
Pod maximum/ curved height	<b>0.120</b>	<b>-0.347</b>	<b>259</b>	<b>&lt;10<sup>-4</sup></b>

*Bold*, significant associations. M.S., Mean Square; F, F ratio;  $R^2_{adj}$ , adjusted  $R^2$ ; r, Pearson correlation coefficient; P, significance level.

2014), while the present study indicates that in common bean this difference is much stronger, with the presence/absence of the lignified layer seen. This suggests that the role of this lignified layer in the shattering might be more relevant (or at least different) in common bean compared to soybean. This might mark another difference between these two closely related crops, of common bean and soybean. Prakken (1934) suggested that in common bean, the control of the traits of “stringlessness” (which depends on the characteristics of the ventral sheaths) and “parchment” (which depends on the layer between the inner and outer parenchyma of the pod wall) was independent, and in both cases was under simple monogenic control. Another study suggested oligogenic control for the stringless trait, with the contribution of either environmental effects or epistatic interactions (Dong and Wang, 2015). Thus, in common bean, the artificial selection might have targeted multiple genes to minimize the seed loss during domestication. This evokes a

scenario that arises from the joint consideration of the data obtained in soybean by the independent studies of Dong et al. (2014) and Funatsuki et al. (2014).

Based on the data in the present study, two further conclusive considerations can be made that might be useful to support the identification of shattering genes in common bean. First, it is likely that the genes underlying the shattering trait in common bean are involved in the regulation of the secondary cell-wall deposition or fiber-cell differentiation. This is well-supported by the data from the chemical analysis and the anatomical-histological investigations here; indeed, fibers are mainly composed of sclerenchymatic cells, that have well-developed secondary cell walls. This possibility is also suggested by the data of Suanum et al. (2016), who reported co-localization of QTLs for pod fiber content and pod shattering in back-cross populations between yearlong bean and wild cowpea. Secondly, as the comparison with the literature indicates some chemical and histological differences between soybean and common bean, it might be useful to consider as candidate genes not only those involved in the shattering of soybean, but also those from other phylogenetically more distant species (Dong and Wang, 2015; Li and Olsen, 2016).

## Relationships between Shattering and the Other Plant Traits

The shattering levels were very poorly correlated with the other morpho-phenological traits and productive characteristics of the plants investigated here. However, an interesting consideration arises from the observation that shattering is significantly (albeit poorly) associated with low 100-seed weight, small pod size, and low Harvest Index at pod level. This suggests that pod shattering might have an “energy cost” for the plant (McGinley and Charnov, 1988; Chiariello et al., 2000); i.e., the synthesis of the biomolecules and the creation of the tissues needed for shattering might reduce the resources available for seed and pod development. In agreement with this energy cost hypothesis, the carbon contents of the pod valves were strongly and positively correlated with the levels of shattering. This all suggests that shattering can be better viewed as a syndrome at the pod level. However, as the same data can be explained by pleiotropic effects or linkage drag, more data will need to be collected also in other species to confirm this hypothesis.

## CONCLUSIONS

Pod shattering in common bean was investigated in the present study. With this objective, we set up and adopted a pipeline for phenotypic characterization of this trait. Four main results were achieved: (1) very high shattering levels can be obtained with a high percentage of either twisting or non-twisting pods, or with a balanced combination between these two; i.e., in common bean, the modes of shattering do not have any great impact on the levels of shattering; (2) shattering appears to be controlled by a “switching” mechanism that determines an abrupt change in the ability to split the pod valves; (3) high shattering

levels is correlated with high carbon and lignin contents of the pod valves, and with specific histological characteristics of the ventral sheath and the inner sclerenchymatic layer of the pod wall; and (4) shattering appears to have a “cost,” and it might be more exhaustively described as a “syndrome” at the pod level.

Overall, our pipeline will help with the deciphering of the genetic architecture of shattering in different crops, thus facilitating comparative studies in legumes.

## AUTHOR CONTRIBUTIONS

Designed the project: DR and RP. Managed the project: DR, GA, and RP. Wrote the article: MM and DR. Contributed to the drafting and the critical revision of the article: DR, MM, MR, GA, EBi, EBe, DA, DF, LN, TG, and RP. Contributed plant materials: EBi, EBe, LN, GA, and RP. Performed phenotypic characterization under field conditions MM, DF, TG, and DR. Histological analysis: DA, MM, and DR. Collect data from element composition and cell-wall analyses: MM. Analyzed and interpreted data: DR, MM, MR, RP, and GA. Edited the article: DR, MM, MR, EBi, EBe, GA, and RP. All of the author approve the final version of the manuscript.

## REFERENCES

- Abbo, S., Saranga, Y., Peleg, Z., Kerem, Z., Lev-Yadun, S., and Gopher, A. (2009). Reconsidering domestication of legumes versus cereals in the ancient Near East. *Q. Rev. Biol.* 84, 29–50. doi: 10.1086/596462
- Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., Giardini, A., Zeuli, P. S., et al. (2012). Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl. Acad. Sci. U.S.A.* 109, E788–E796. doi: 10.1073/pnas.1108973109
- Boersma, J. G., Buirchell, B. J., Sivasithamparam, K., and Yang, H. (2007). Development of two sequence-specific PCR markers linked to the le gene that reduces pod shattering in narrow-leaved Lupin (*Lupinus angustifolius* L.). *Genet. Mol. Biol.* 30, 623–629. doi: 10.1590/S1415-47572007000400020
- Boersma, J. G., Nelson, M. N., Sivasithamparam, K., and Yang, H. (2009). Development of sequence-specific PCR markers linked to the Tardus gene that reduces pod shattering in narrow-leaved lupin (*Lupinus angustifolius* L.). *Mol. Breed.* 23, 259–267. doi: 10.1007/s11032-008-9230-2
- Chiariello, N. R., Mooney, H. A., and Williams, K. (2000). “Growth, carbon allocation and cost of plant tissues,” in *Plant Physiological Ecology* eds R. W. Pearcy, J. R. Ehleringer, H. Mooney, and P. W. Rundel (Dordrecht: Springer Netherlands), 327–365.
- Deysson, G. (1954). *Éléments D'anatomie Des Plantes Vasculaires*. Paris: Société d'édition d'enseignement supérieur.
- Dong, Y., and Wang, Y.-Z. (2015). Seed shattering: from models to crops. *Front. Plant Sci.* 6:476. doi: 10.3389/fpls.2015.00476
- 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
- Elbaum, R., and Abraham, Y. (2014). Insights into the microstructures of hygroscopic movement in plant seed dispersal. *Plant Sci.* 223, 124–133. doi: 10.1016/j.plantsci.2014.03.014
- Fiorani, F., and Schurr, U. (2013). Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 64, 267–291. doi: 10.1146/annurev-arplant-050312-20137
- Funatsuki, H., Suzuki, M., Hirose, A., Inaba, H., Yamada, T., Hajika, M., et al. (2014). Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17797–17802. doi: 10.1073/pnas.1417282111
- Gioia, T., Logozzo, G., Kami, J., Zeuli, P. S., and Gepts, P. (2012). Identification and characterization of a homologue to the *Arabidopsis INDEHISCENT* gene in common bean. *J. Hered.* 104, 273–286. doi: 10.1093/jhered/ess102
- Harlan, J. (1992). Origins and processes of domestication. *Grass Evol. Domest.* 159:175.
- Hymowitz, T. (1970). On the domestication of the soybean. *Econ. Bot.* 24, 408–421. doi: 10.1007/BF02860745
- Koinange, E. M., Singh, S. P., and Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Sci.* 36, 1037–1045. doi: 10.2135/cropsci1996.0011183X003600040037x
- Ladizinsky, G. (1979). Species relationships in the genus *Lens* as indicated by seed-protein electrophoresis. *Bot. Gaz.* 140, 449–451. doi: 10.1086/337112
- Lamprecht, H. (1932). Beiträge zur genetik von *Phaseolus vulgaris*. *Hereditas* 16, 169–211. doi: 10.1111/j.1601-5223.1932.tb02567.x
- Li, L.-F., and Olsen, K. (2016). Chapter three-to have and to hold: selection for seed and fruit retention during crop domestication. *Curr. Top. Dev. Biol.* 119, 63–109. doi: 10.1016/bs.ctdb.2016.02.002
- 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
- Loader, N., Robertson, I., and McCarroll, D. (2003). Comparison of stable carbon isotope ratios in the whole wood, cellulose and lignin of oak tree-rings. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 196, 395–407. doi: 10.1016/S0031-0182(03)00466-8
- McGinley, M. A., and Charnov, E. L. (1988). Multiple resources and the optimal balance between size and number of offspring. *Evol. Ecol.* 2, 77–84. doi: 10.1007/BF02071590
- Mitra, P. P., and Loqué, D. (2014). Histochemical staining of *Arabidopsis thaliana* secondary cell wall elements. *J. Vis. Exp.* 87:e51381. doi: 10.3791/51381
- Muehlbauer, F. J., Summerfield, R., and Kaiser, W. (1998). Principles and Practice of Lentil Production. US Dep. Agric; USDA Agric. Res. Serv; ARS 141, 26.
- Nanni, L., Bitocchi, E., Bellucci, E., Rossi, M., Rau, D., Attene, G., et al. (2011). Nucleotide diversity of a genomic sequence similar to SHATTERPROOF (PvSHP1) in domesticated and wild common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 123, 1341–1357. doi: 10.1007/s00122-011-1671-z

## ACKNOWLEDGMENTS

Thanks for technical assistance at University of Sassari to: A. Ara, B. Scalas, and M. Pinna for phenotyping under field conditions; M. Deroma for CHN analysis; R. Rubattu for cell-wall analysis; G. Becca for histological analysis. This research represents part of the PhD project carried out by MM at the doctoral school of Science and Biotechnology of Agricultural and Forestry Science and Food Production, curriculum Crop Productivity of the University of Sassari (Supervisor: DR; Tutor: GA). MM gratefully acknowledge Sardinia Regional Government that partially funded her PhD scholarship (P.O.R. Sardegna F.S.E. 2007–2013—Obiettivo competitività regionale e occupazione, Asse IV Capitale umano, Linea di Attività 1.3.1. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007–2013—Axis IV Human Resources, Objective 1.3, Line of Activity 1.3.1.).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00251/full#supplementary-material>

- Prakken, R. (1934). Inheritance of colours and pod characters in *Phaseolus vulgaris*. *L. Genetica* 16, 177–296. doi: 10.1007/BF02071498
- Rodríguez, G. R., Moyseenko, J. B., Robbins, M. D., Morejón, N. H., Francis, D. M., and van der Knaap, E. (2010). Tomato Analyzer: a useful software application to collect accurate and detailed morphological and colorimetric data from two-dimensional objects. *Vis. Exp.* 37:e1856. doi: 10.3791/1856
- Romkaew, J., Nagaya, Y., Goto, M., Suzuki, K., and Umezaki, T. (2008). Pod dehiscence in relation to chemical components of pod shell in soybean. *Plant Prod. Sci.* 11, 278–282. doi: 10.1626/pps.11.278
- Santalla, M., Menéndez-Sevillano, M. C., Monteagudo, A. B., and De Ron, A. M. (2004). Genetic diversity of Argentinean common bean and its evolution during domestication. *Euphytica* 135, 75–87. doi: 10.1023/B:EUPH.0000009543.46471.72
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi: 10.1038/ng.3008
- Singh, S. P. (2001). Broadening the genetic base of common bean cultivars. *Crop Sci.* 41, 1659–1675. doi: 10.2135/cropsci2001.1659
- Suanum, W., Somta, P., Kongjaimun, A., Yimram, T., Kaga, A., Tomooka, N., et al. (2016). Co-localization of QTLs for pod fiber content and pod shattering in F2 and backcross populations between yardlong bean and wild cowpea. *Mol. Breed.* 36, 1–11. doi: 10.1007/s11032-016-0505-8
- Tang, H., Cuevas, H. E., Das, S., Sezen, U. U., Zhou, C., Guo, H., et al. (2013). Seed shattering in a wild sorghum is conferred by a locus unrelated to domestication. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15824–15829. doi: 10.1073/pnas.1305213110
- Van Soest, P., and Wine, R. (1967). Use of detergents in the analysis of fibrous feeds. IV. Determination of plant cell-wall constituents. *J. Assoc. Anal. Chem.* 50, 50–55.
- Weeden, N. F. (2007). Genetic changes accompanying the domestication of *Pisum sativum*: is there a common genetic basis to the “domestication syndrome” for legumes? *Ann. Bot.* 100, 1017–1025. doi: 10.1093/aob/mcm122

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

Copyright © 2017 Murgia, Attene, Rodríguez, Bitocchi, Bellucci, Fois, Nanni, Gioia, Albani, Papa and Rau. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Genomic and Phenomic Screens for Flower Related RING Type Ubiquitin E3 Ligases in *Arabidopsis*

Mirko Pavicic<sup>1,2</sup>, Katriina Mouhu<sup>1,2</sup>, Feng Wang<sup>1,2</sup>, Marcelina Bilicka<sup>1,2</sup>, Erik Chovancák<sup>1</sup> and Kristiina Himanen<sup>1,2\*</sup>

<sup>1</sup> Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland, <sup>2</sup> Viikki Plant Science Centre, University of Helsinki, Helsinki, Finland

## OPEN ACCESS

### Edited by:

Marcos Egea-Cortines,  
Universidad Politécnica de Cartagena,  
Spain

### Reviewed by:

Federico Valverde,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain  
Daniel Hofius,  
Swedish University of Agricultural  
Sciences, Sweden

### \*Correspondence:

Kristiina Himanen  
kristiina.himanen@helsinki.fi

### Specialty section:

This article was submitted to  
Technical Advances in Plant Science,  
a section of the journal  
*Frontiers in Plant Science*

Received: 30 September 2016

Accepted: 10 March 2017

Published: 28 March 2017

### Citation:

Pavicic M, Mouhu K, Wang F, Bilicka M, Chovancák E and Himanen K (2017) Genomic and Phenomic Screens for Flower Related RING Type Ubiquitin E3 Ligases in *Arabidopsis*. *Front. Plant Sci.* 8:416.  
doi: 10.3389/fpls.2017.00416

Flowering time control integrates endogenous as well as environmental signals to promote flower development. The pathways and molecular networks involved are complex and integrate many modes of signal transduction. In plants ubiquitin mediated protein degradation pathway has been proposed to be as important mode of signaling as phosphorylation and transcription. To systematically study the role of ubiquitin signaling in the molecular regulation of flowering we have taken a genomic approach to identify flower related Ubiquitin Proteasome System components. As a large and versatile gene family the RING type ubiquitin E3 ligases were chosen as targets of the genomic screen. The complete list of *Arabidopsis* RING E3 ligases were retrieved and verified in the *Arabidopsis* genome v11 and their differential expression was used for their categorization into flower organs or developmental stages. Known regulators of flowering time or floral organ development were identified in these categories through literature search and representative mutants for each category were purchased for functional characterization by growth and morphological phenotyping. To this end, a workflow was developed for high throughput phenotypic screening of growth, morphology and flowering of nearly a thousand *Arabidopsis* plants in one experimental round.

**Keywords:** *Arabidopsis*, flower, RING E3 ligase, ubiquitin, high throughput, image based phenotyping, phenomics data analysis

## INTRODUCTION

Flowering time control is a complex network that integrates many modes of signal transduction promoting transition from vegetative stage to reproduction and ultimately leading to the development of flower organs. The endogenous changes that signal the beginning of flowering are referred as autonomous pathways (Amasino and Michaels, 2010). Multiple studies have established the major role that photoperiod has in flowering (Piñeiro and Jarillo, 2013). Flowering in *Arabidopsis* is strongly promoted in long day (LD) conditions but will also ultimately occur under short day (SD) conditions (Steffen et al., 2014). Under LDs, flower induction is dependent on the expression and protein levels of CONSTANS (CO; Suárez-López et al., 2001). Light controls the CO transcription via the circadian clock system, inducing a CO mRNA peak during the latter part of the day (Suárez-López et al., 2001). CO transcription is repressed by CYCLING DOF FACTORs (CDFs; Fornara et al., 2009). Under LDs, the CO mRNA afternoon peak coincides with a blue-light activated complex containing FLAVIN-BINDING, KELCH REPEAT, F-BOX 1 (FKF1) and GIGANTEA (GI), which lead CO transcription repressors

CDFs to degradation (Sawa et al., 2007; Fornara et al., 2009; Song et al., 2012). Additionally, the FKF1-GI complex also stabilizes CO protein in the afternoon (Sawa et al., 2007; Song et al., 2012). CO protein degradation is promoted by at least two ubiquitin E3 ligases: HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 1 (HOS1) and CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1; Jang et al., 2008; Lazaro et al., 2012). In the morning, red light promotes HOS1 interaction with CO via phytochrome B activation (Lazaro et al., 2012). COP1 mediates CO protein degradation in a complex with SUPPRESSOR OF PHYA-105 (SPA; Laubinger et al., 2006). In the afternoon, blue light inhibits COP1-SPA-mediated CO degradation by activating CRYPTOCHROME 2 (CRY2) interaction with COP1 (Liu et al., 2008). Thus, both CO transcription is up-regulated and CO protein stabilized allowing up-regulation of the mobile flowering signal gene *FLOWERING LOCUS T* (*FT*) expression in the phloem during the afternoon under LDs, but not under SDs (Piñeiro and Jarillo, 2013). Also regulation of flower development is likely to involve Ubiquitin Proteasome System (UPS) components (Vierstra, 2009).

The UPS has emerged as a powerful regulatory mechanism that facilitates irreversible transitions between developmental stages, and responses to environmental stimuli by selectively degrading short-lived regulators, such as transcription factors and receptors (Sadanandom et al., 2012). Genetic analyses in plants have proposed that this pathway plays a vital role in hormone regulation, floral homeostasis, stress responses and pathogen defense; however, very few targets have been identified in plants apart from the hormone signaling components (Santner and Estelle, 2010). In the UPS system, the highly conserved 76-amino acid protein, ubiquitin, acts as a covalent molecular tag to signal target proteins for proteasome mediated degradation. Ubiquitin attachment requires three distinct enzymatic activities: E1, ubiquitin activating enzymes; E2, ubiquitin conjugating enzymes; and E3, ubiquitin ligase enzymes. Moreover, the UPS consists of accompanying proteins that modulate target recognition and degradation (such as RAD23, SPA1), deubiquitinating enzymes (DUB1) and the proteasome (26S and 20S structures). According Vierstra (2009) over 6% of the *Arabidopsis* proteome is potentially involved in UPS. However, the common strategy for functionally addressing the role of all UPS components is still evolving. The ubiquitin E3 ligases are the most abundant UPS components and mediate the important recognition of the target proteins for ubiquitination (Kosarev et al., 2002; Stone et al., 2005). The E3 ligases found in plants belong to one of four subtypes: single subunit Homology to E6-AP C-Terminus (HECT), U-box and Really Interesting New Gene (RING) or multisubunit cullin-RING ligases (Sadanandom et al., 2012). The RING-type E3 proteins are the most abundant among the single subunit E3 ligases (Kosarev et al., 2002; Stone et al., 2005).

To unravel the role of the RING type ubiquitin E3 ligase protein family, we took a reverse-genetics approach to identify the RING E3 ligases that could be involved in regulation of *Arabidopsis* flowering time and/or flower development. To this end, we first curated the RING E3 protein family,

earlier described by Stone et al. (2005), in the most recent *Arabidopsis* genome release. The *Arabidopsis* protein sequences were subjected to InterProScan for protein domain search and the number of ubiquitin E3 ligases containing RING domains was established to be 509. Association of these RING protein encoding genes with *Arabidopsis* flowering and floral organs was done through the Genevestigator transcriptome database (Hruz et al., 2008). To this end, the expression profiles were divided into categories based on their specificity, high expression or enrichment in flower organs and in the developmental stages of *Arabidopsis*. Several already characterized regulators were identified among these genes, such as the anther dehiscence regulating DAF gene family (Peng et al., 2013), flower size regulating DA2 (Xia et al., 2013) and FRG1 involved in flowering time related DNA methylation (Groth et al., 2014). The well-established flowering time regulator *COP1* fell just below the cut off criteria due to its wide expression profile. A representative mutant collection for each category was obtained from NASC stock center. Additional candidates were also selected based on literature. The genotypically verified mutant collection was subjected to systematic morphological and growth analysis using an automated imaging based plant phenotyping facility. After the thorough vegetative assessment, the flowering time parameters such as number of leaves at bolting and days to bolting were recorded together with morphological analysis of the flower structures. The phenotypic assessment indicated lines with altered growth, morphology, or flowering time. Furthermore, one of the lines showed growth defects in sepals and petals.

## MATERIALS AND METHODS

### Bioinformatic Screens

To curate putative RING-type ubiquitin E3 ligases in *Arabidopsis thaliana* genome version ARA11, classification made by Kosarev et al. (2002) and Stone et al. (2005) were used. To this end, the whole *Arabidopsis* proteome was downloaded from ARAPORT (<https://www.araport.org/downloads/>), and screened with InterProScan for protein families and domain architecture. To confirm that the newly identified RING domain containing protein sequences indeed represented ubiquitin E3 ligase type RING domains, InterProScan 5 (v5.16-55.0) Gene3D, SUPERFAMILY, ProSiteProfiles, SMART, Pfam, and ProSitePatterns signatures were used. Most of InterProScan tools use Hidden Markov Models (HMMs) to detect conserved domains along protein sequences. HMMs have been developed for conserved protein domains and they define for the software, which and where critical residues should be located along the analyzed protein sequence. From the protein domain collection, the ubiquitin E3 ligase type RING domains were filtered according to the criteria provided by Kosarev et al. (2002) and Stone et al. (2005) for canonical RING domains. Once the RING domains were identified, they were aligned with Jalview using ProbCons algorithm with two rounds of pre-training. The metal ligand binding residues were manually inspected and corrected, and small misalignments were edited. Sequences that failed to meet the criteria of InterProScan search engines were not considered in this study.

## Transcriptomic Database Screens

To associate the curated collection of 509 RING type ubiquitin E3 ligases with flowering the Genevestigator gene expression database software was used (Hruz et al., 2008). The experiments AT-00087, AT-00088, AT-00089, and AT-00090 containing developmental expression data of AtGenExpress initiative microarrays were selected for the analysis (Schmid et al., 2005). In the selected experiments, hybridization probes were available for 393 *RING E3* genes out of the 509. From these experiments, the linear expression data was extracted for the developmental stages of developed rosette, bolting, young flower, developed flower, and flower and siliques. For flower organs, the gene expression profiles were extracted for categories of shoot apical meristem (SAM), sepal, petal, stamen, and carpels. In these categories, genes were ranked for their at least 2-fold differential expression against the developed rosette. Their relative expression levels were obtained by  $\log_2(\text{FC}) = \log_2(\text{FL}) - \log_2(\text{R})$ , where FC is fold of change, FL is flower organ or development stage and R is rosette. The results for each category were sorted by their  $\log_2(\text{FC})$  and all genes with  $\log_2(\text{FC}) > 1$  were considered as up-regulated.

## Candidate Genes Selected by Literature

For the candidate approach, we used interaction networks from BioGRID (<http://thebiogrid.org/>) and cross-checked them with flowering pathway genes listed in the Flowering Interactive Database FLOR-ID (Bouché et al., 2016) to identify RING E3 ligases interacting with known flowering time regulators CONSTANS (CO), CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1), and TARGET OF EARLY ACTIVATION TAGGED (EAT) 2 (TOE2).

## Plant Material and Growth Conditions

For functional characterization of the identified top most differentially expressed genes and for the selected candidates, *Arabidopsis* mutant lines were obtained from the NASC stock center representing CATMA, SAIL, SALK, and GABI-Kat collections (Alonso et al., 2003; Rosso et al., 2003; Schmid et al., 2005; Kleinboelting et al., 2012). Altogether 49 lines were genotyped by combination of segregation analysis and T-DNA PCR with primers listed in Supplemental Table 1. From these, 43 lines represented 30 unique gene accessions (Supplemental Table 1). As a wild type control, Columbia (Col-0) ecotype of *Arabidopsis thaliana* was used.

For genotyping, plants were grown *in vitro* on MS media supplemented with the corresponding selection. For phenotyping, seeds were sown directly on soil with 50% peat and 50% vermiculite. Trays were covered with plastic wrap and cold stratified at +4 °C for three nights, after which they were transferred to the growth chamber (FytoScope, PSI, Czech Rep.). Seven days after stratification (DAS) the seedlings were transferred to their own pots, placed on the analysis trays and sand was added on top of the peat to prevent growth of any green algae. From the full water saturation of the soil, the water content was let to decrease until 70% and was kept at this level through daily weighing and watering. Growth conditions in the *Arabidopsis* growth chambers were 16 h light/8 h darkness and 22 °C. Relative air humidity of the growth chambers was targeted

at 60%. The light intensity was set and controlled at 130 μE (MS6610, Mastech, China).

## Genotyping of the Mutant Lines

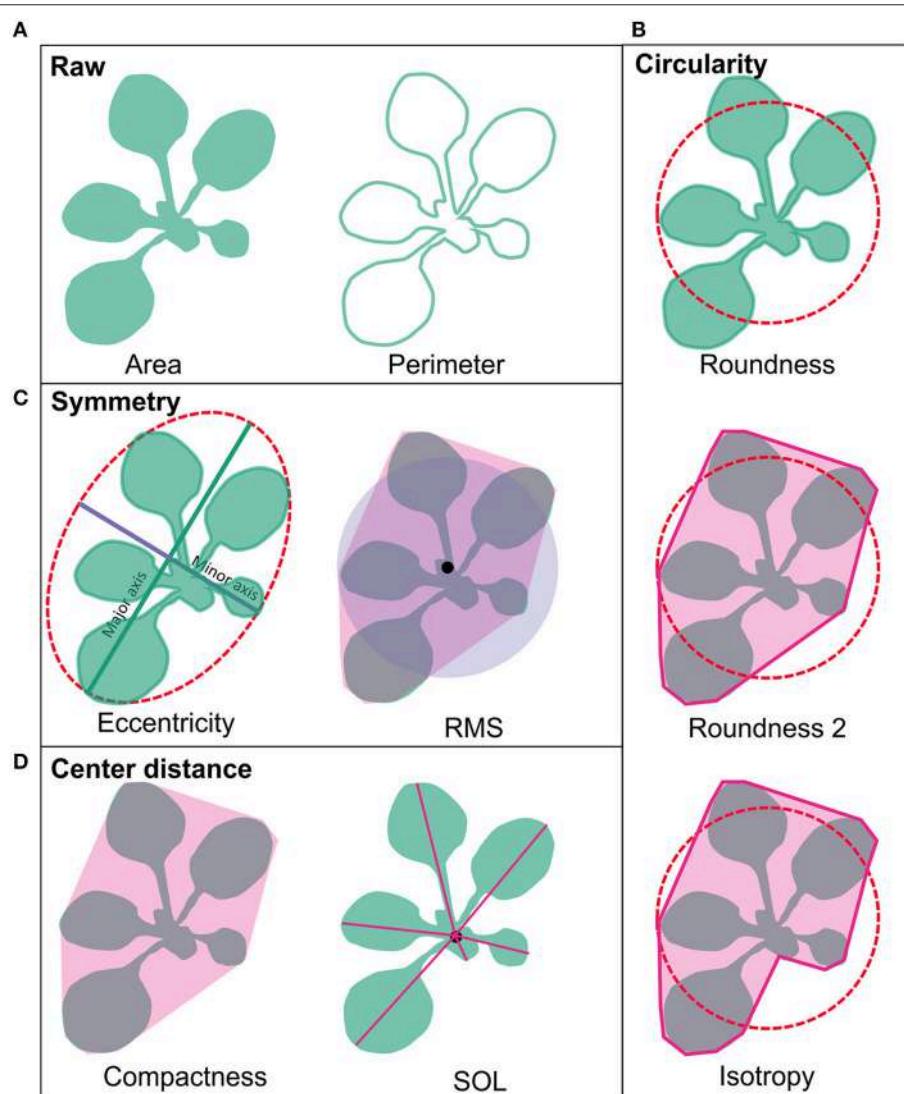
Homozygous one locus mutant lines were confirmed by segregation analysis and T-DNA specific PCRs. The PCR primers, T-DNA position and line information were summarized in Supplemental Table 1. The transcript levels of the T-DNA targeted genes were verified by quantitative real-time PCR (qPCR) analysis. The sample material for qPCR was harvested from the tissue indicated by *Arabidopsis* eFP Browser for each gene expression pattern: if the gene expression pattern indicated at least moderate expression in flower parts during floral development, tips of inflorescences with developing and open flowers were pooled from three to five individual plants. If the expression was in the seeds, young to mature siliques were pooled. If no expression data via eFP was found, new leaves were pooled with developing and open flowers. Three to four biological replications were harvested for each RNA preparation. RNA was extracted using InviTrap® Spin Plant RNA Kit (STRATEC Molecular), complementary DNA was prepared with SuperScript® IV Reverse Transcriptase (Thermo Fisher Scientific), and the qPCRs were performed using Roche Lightcycler® 480 Instrument II (Roche Diagnostics) using LightCycler® 480 SYBR Green I Master (Roche Diagnostics) with primers listed in Supplemental Table 1. Primers were primarily designed to locus downstream of the T-DNA. The fold up values (mutant line against Col-0) were calculated using the  $2^{-\Delta\Delta CT}$  method according to Livak and Schmittgen (2001). Reference genes used in this study were the most stable *Arabidopsis* genes according to Czechowski et al. (2005): *TIP41 LIKE* (AT4G34270, forward: GTGAAAAGCTTGGAGAGAAAGCAA, reverse: TCAACTGGATACCCCTTT^CGCA), *AP2M* (AT5G46630, forward: TTGAAAATTGGAGTACCGTACCAA, reverse: TCCCTCGTATACTGAGCCA) and *PTB1* (AT3G01150, forward: TTGAAGGAGTGGAAATCTCACG, reverse: ATGTGCGGAAAGCAGATACC). Significance level of the qPCR were set at 0.1–0.5 FC for knock-down; <0.1 FC for knock-out; and >2 FC for up-regulated (Supplemental Table 1).

## High Throughput Plant Phenotyping Platform

The plant phenotyping facility at the University of Helsinki Viikki campus (<http://blogs.helsinki.fi/nappi-blog/>) was used for the phenotypic characterization of the selected *Arabidopsis* mutant collection. The plants were imaged daily by overhead CCD camera for RGB images positioned in a PlantScreen™ analysis chamber with automated plant transportation between the imaging, weighing and watering stations. The RGB images were obtained for 20 plants at the time and stored in central database. The images were pre-processed online as described in Awlia et al. (2016) to allow collecting binary and RGB data for each plant. The obtained binary images were used for calculating growth parameters of area and perimeter. The obtained parameters of area, perimeter and the convex hull were then used for automatic online calculations of morphometric rosette parameters including: roundness1, roundness2, isotropy,

eccentricity, compactness, Rotational Mass Symmetry (RMS), and Slenderness of Leaves (SOL) (PlantScreen<sup>TM</sup> analyzer, PSI, Czech R.). To characterize the general morphology of the mutant lines these nine morphological parameters were grouped into four categories based on their type: raw, circularity, symmetry and center distance, and compared over time (**Figure 1**). Raw parameters were represented by area and perimeter of the rosette and they were calculated by counting pixels of a rosette picture and the edge pixels respectively and transformed to millimeters (**Figure 1A**). The parameters of roundness1 and roundness2 and isotropy represented the circular parameters (**Figure 1B**). The parameter roundness describes rosette area in comparison to perfect circle with same perimeter and is affected by leaf

slenderness, petiole length and leaf perimeter. For wild type plant, this parameter usually takes values between 0.1 and 0.5 while a perfect circle has value 1. Roundness value tends to decay overtime due to leaf development that at the same time increases the rosette perimeter. Roundness 2 uses rosette convex hull area and perimeter for its computation and for wild type plants this parameter appears to have values between 0.7 and 1.0 following an oscillating pattern with less steep peaks over time (**Figure 1B**). Isotropy uses the area of a drawn polygon on top of the rosette (**Figure 1B**). Thus, isotropy has a behavior similar to roundness 2 over time, but with less steep peaks and decreasing tendency similar to roundness. The eccentricity and RMS were symmetric parameters (**Figure 1C**). Eccentricity describes how



**FIGURE 1 | Rosette morphology parameters.** **(A)** Area and perimeter of the rosette are raw parameters and are calculated by counting pixels of rosette binary images and the edge pixels respectively and transformed to millimeters. **(B)** The parameters of roundness and roundness 2 and isotropy represented circular parameters. **(C)** Eccentricity and RMS were symmetric parameters. **(D)** Compactness and SOL were based on the center distance. Pink area around RMS, compactness and roundness 2 represent rosette convex hull, and for isotropy it represents rosette polygon. The characteristics of the parameters are described in detail in the Section Materials and Methods.

elliptical the plant rosette is, where a value close to 1 corresponds to a rosette with highly sharp elliptical shape, while a value close to 0 describes a circular shape. Wild type rosette shows a high eccentricity peak that decays over time with a second smaller peak by the end of growth, thus the rosette shape fluctuates between a round and an elliptical shape. On the other hand, RMS describes the symmetry of the plant rosette by making a ratio between the non-overlapping rosette convex hull area and a perfect circle of the same area centered in the plant centroid and the overlapping area of both. RMS shows a similar pattern as eccentricity, but with higher absolute values and a sharper peak. Compactness and SOL were based on the center distance (**Figure 1D**). Compactness is the ratio between the rosette area and the rosette convex hull area. This parameter tells about petiole length and leaf blade width. The parameter SOL explains how sharp the leaf blades are, but it is also affected by the leaf number. SOL was derived from the ratio between squared rosette skeleton and rosette area. Thus, SOL can take values greater than 0 and below 50 in dimensionless units for wild type plants (**Figure 1D**).

## Experimental Design

Ten days old (10 DAS) *Arabidopsis* plants were subjected to growth and morphological characterization by top view imaging for the following 10 days. One phenotyping round was designed to accommodate a maximum of 960 *Arabidopsis* plants representing 36 genotypes at a time in six consecutive experimental rounds called F006–F011. The total number of lines analyzed throughout the six rounds was 43. The maximum of 36 genotypes were divided in three batches that were rotating between the growth area and the PlantScreen™ analysis chamber. Each batch consisted of three experimental units of four mutant genotypes randomized with Col-0, each represented by 20 individual plants. One experiment consisted thus of five trays of altogether 100 plants. Each experimental unit had their own Col-0 wild type in randomized block design to normalize for any local differences in the microenvironments of the PlantScreen™ or the growth area. Each line showing any phenotypic responses was analyzed in at least three independent experimental rounds. Lines that did not show differences as compared to the Col-0 wild type were excluded from the subsequent rounds thus resulting in reduced numbers of genotypes included.

## Phenotypic Analysis of Flowering Time and Flower Structures

After the image based growth and morphological measurements of the 20 mutant and 20 Col-0 plants in the PlantScreen™ system, the flowering time parameters leaf numbers at bolting (LAB) and days to bolting (DTB), were manually counted for each of the plant individuals. The number of rosette leaves were counted at the appearance of the flower bud (developmental stage 5.10, Boyes et al., 2001) and the DTB was recorded at the same time. The flowering time phenotypes were observed in two to three independent experimental rounds. Finally, flowers of the main inflorescences were photographed and further dissected for floral organ analysis under stereomicroscope (SteREO Discovery.V20, Zeiss). Microscopic pictures of the

inflorescence tips, single flowers, sepals and petals were taken with the attached camera (AxioCam ICc3, Zeiss). The analyzed inflorescences and flowers originated from several independent experiments. Flower developmental stages were determined as in Smyth et al. (1990). To confirm pollen viability pollen grain staining according to the modified Alexander method was performed (Peterson et al., 2010). Anther images were captured using Leica DFC420 C camera attached to an optical microscope.

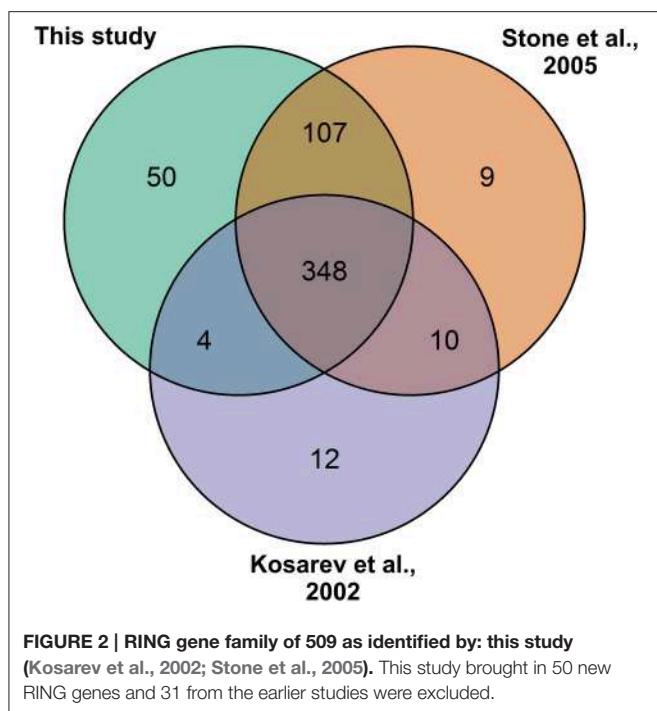
## Statistical Analysis

The significance of the differences between mutant lines and Col-0 was computed by contrasting two fitted models to the data points using several order polynomials (Mirman, 2014). First, a model was fitted to all data points and then a second model was fitted including the factor genotype (wild type and mutant). These two models were then compared using the Chi square test to determine if the second model explained more variance than the first one beyond the significance ( $\alpha = 0.05$ ). If the second model was statistically different from the first one, it implied that the compared genotypes were different. These statistical analyses were conducted in R software (<https://www.r-project.org/>). LAB and DTB analysis were performed by Analysis of Variance using GLM procedure and pairwise comparisons against Col-0 using option Dunnett in the MEANS statement using SAS/STAT® software version 9.4 (SAS Institute Inc., Cary, NC, USA).

## RESULTS

### Curation of the *Arabidopsis* RING-Type Ubiquitin E3 Ligase Protein Sequences

To re-confirm the published RING-type ubiquitin E3 ligase proteins encoded in *Arabidopsis* genome the 27,667 *Arabidopsis* proteins from the latest genome annotation release (ARA11) were scanned for RING domains. Through filtering the signatures in Gene3D, SUPERFAMILY, ProSiteProfiles, SMART, Pfam, ProSitePatterns altogether 509 putative RING domain containing protein sequences were obtained (Supplemental Table 2). RING gene names and descriptions were obtained from Araport using Thalemine tool (Supplemental Table 2). Araport used curated but also automatic gene annotation, therefore many RING domain containing proteins were annotated as RING/U-box protein although they did not contain U-box domain. Similarly, some were annotated as RING/FYVE/PHD zinc finger superfamily proteins. The 509 RING sequences were compared to the previously described RING-type protein sequences (Kosarev et al., 2002; Stone et al., 2005). From the 509 identified RING domain proteins 457 matched with the 490 previously described, thus resulting in 31 non-matching sequences (**Figure 2**). These non-matching sequences were thoroughly analyzed and, 6 of them were found to be merged with other gene models, 10 had no RING domain, 3 were not found in the database, 3 corresponded to pseudogenes, 7 were split and a new locus identifier had been assigned for them, and 2 were transposable elements (Supplemental Table 3A). The 50 additionally identified RING domain proteins were shown to represent diverse RING domains such as, 1 of D type, 4 of



**FIGURE 2 |** RING gene family of 509 as identified by: this study (Kosarev et al., 2002; Stone et al., 2005). This study brought in 50 new RING genes and 31 from the earlier studies were excluded.

C2 type, 20 of H2, 16 of HC, 2 of S/T, and 7 of V (CH) type, according to the Stone et al. (2005) classification (Supplemental Table 2).

## Differential Gene Expression Data Identifies 122 Flower Related RING Ubiquitin E3 ligases

To associate the RING domain proteins with flowering or flower development two approaches were followed: (1) identifying those with gene expression enhanced or enriched during flower development or in flower organs, and (2) by searching RING proteins interacting with known flower regulators. For the first approach Genevestigator (Hruz et al., 2008) tool was used to rank differentially expressed genes (DEG) of the identified RING genes over *Arabidopsis* developmental stages and in flower organs relative to their expression in developed rosette (Figures 3A,B). In the selected experiments in Genevestigator database, probes were available for 393 of the 509 RING E3 ligases analyzed. The cut off for DEGs was set at 2-fold to be included in the selection resulting in lists of genes of interest for each of the categories. This process was repeated to identify gene expression enrichment at each of the development stages of bolting, young flower, developed flower, and flower and siliques. For the developmental categories altogether 71 DEGs were identified (Figure 3C). In addition to the developmental stages, enrichment for shoot apical meristem, sepal, petal, stamen and pistil organs were retrieved and resulted in 109 DEGs (Figure 3D). Some of the RING genes were common between these two categories and in total 122 unique RING genes were up-regulated in the flower related processes. The gene identifying AGI codes of these 122 flower related candidates are provided in the Supplemental Table 3B.

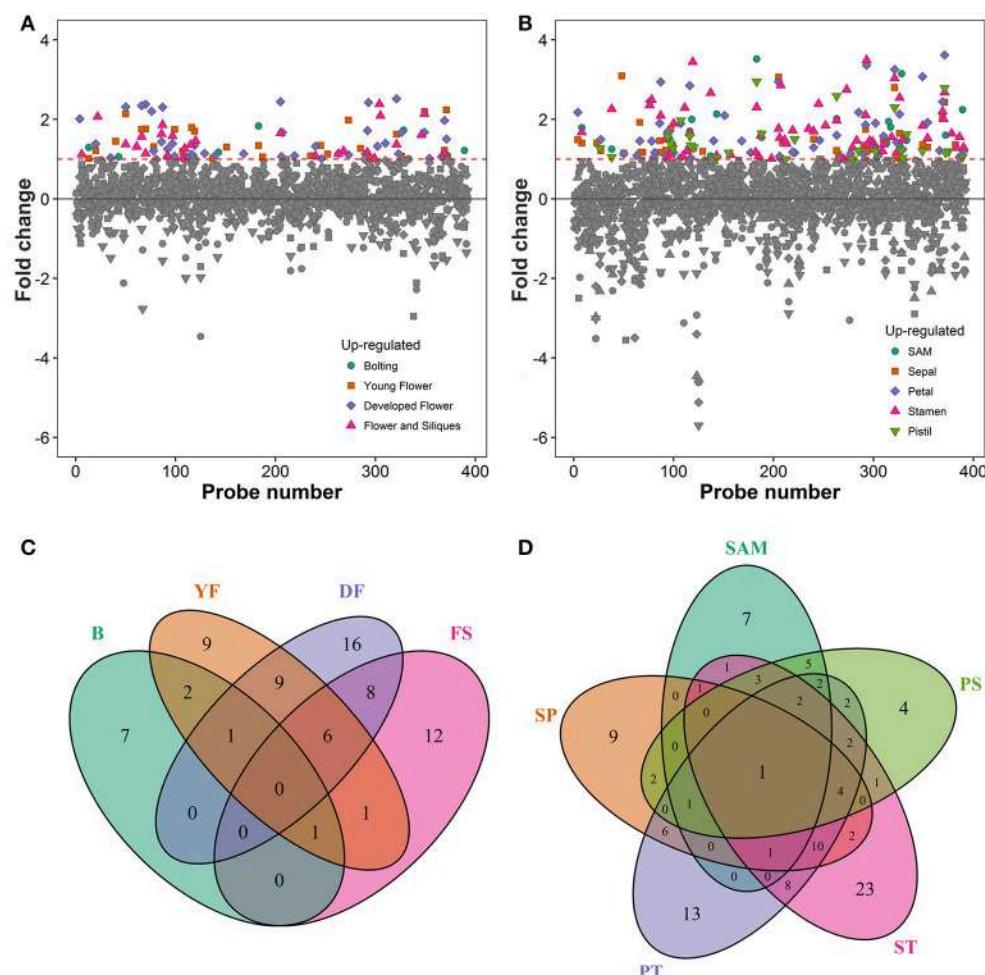
For the second approach we identified 6 additional genes of interest through literature study and from interaction networks of CO, COP1, and TOE2 from BioGRID (<http://thebiogrid.org/>). Based on these interaction screens 5 RING E3 ligases were selected to the study, represented by the following mutant lines; N656705 (AT5G65683), N686069 (AT1G61620), N372291 (AT3G29270), N2037522 and N67002 (AT4G17680), and N742646 (AT2G44410). In addition, a mutant line for COP1, *cop1-6*, and RED AND FAR-RED INSENSITIVE 2 (RFI2) for which a role in mediating red and far-red light signaling and ubiquitination activity has been shown *in vitro*, were included (Stone et al., 2005; Chen and Ni, 2006a). This E3 ligase was selected as a candidate since its expression is regulated by circadian clock and *rfi2-1* mutant flowers early (Chen and Ni, 2006b). Thus, one mutant allele for RFI2 (N878610) was included in the study. Mutants representing these genes were analyzed together with the flower up-regulated RINGs and were named flower related UPS candidates in the Supplemental Table 1.

## Representative Mutant Collection

For functional characterization of the 122 flower related UPS candidates and those selected based on literature, a mutant collection was obtained from the NASC stock center. The mutants represented lines from CATMA, SAIL, SALK, and GABI-Kat collections (Alonso et al., 2003; Rosso et al., 2003; Schmid et al., 2005; Kleinboelting et al., 2012). Altogether 43 lines were shown to contain T-DNA insertion in one locus, six were doubtful and were omitted from the analysis. To confirm that the T-DNA insertion had interrupted the gene of interest, their altered expression levels were confirmed by qPCR analysis with primers listed in Supplemental Table 1. For 43 accessions representing 30 unique loci from the 122 flower related UPS candidates and the selected candidates a differential gene expression pattern was analyzed. Altogether 19 lines were knock-outs, and 13 knock-down mutants, and for 10 lines up-regulation of the gene of interested was observed (Supplemental Table 1). For one line, no differential expression was confirmed and this was excluded from the phenotyping. For 14 lines alleles were available with similar or opposite gene expression patterns.

## Phenotypic Screen of the Mutant Accessions

From the genetically and qPCR confirmed T-DNA insertion mutant lines, 43 were subjected to phenotypic characterization by top view RGB imaging using the PlantScreen™ system. Image series of each analyzed line were collected daily allowing analyzing the growth and changes in morphology over time. For scoring those lines showing phenotypes, we fitted general additive models (GAM) to each parameter of each analyzed lines (data not shown). Most of the lines showed no differences to their corresponding Col-0 controls. However, three lines were consistently different across the experiments compared to Col-0 in both growth and rosette morphology: *csu1-4* (*cop1 suppressor 1-4*, N686069), *sinal7-2* (*seven in absentia like 7-2*, N833574) (Peralta et al., 2016) and *rha1a-1* (*ring-h2 finger a1a-1*, N2045046) (Table 1, Supplemental Tables 1, 4). The *csu1-4* mutant rosette was clearly smaller than Col-0 and showed a



**FIGURE 3 | (A)** Differential expression profiles of developmental stage enriched RING genes relative to developed rosette. **(B)** Differential expression profiles of flower organ enriched RING genes relative to rosette; SAM, Shoot apical meristem. **(C)** Venn diagram of RING genes expressed in the different developmental stages (B, bolting; YF, Young flower; DF, Developed flower; FS, Flowers and siliques). **(D)** Venn diagram of RING genes expressed in the different flower organs; SAM, Shoot apical meristem; SP, Sepal; PT, Petal; ST, stamen; PS, pistil.

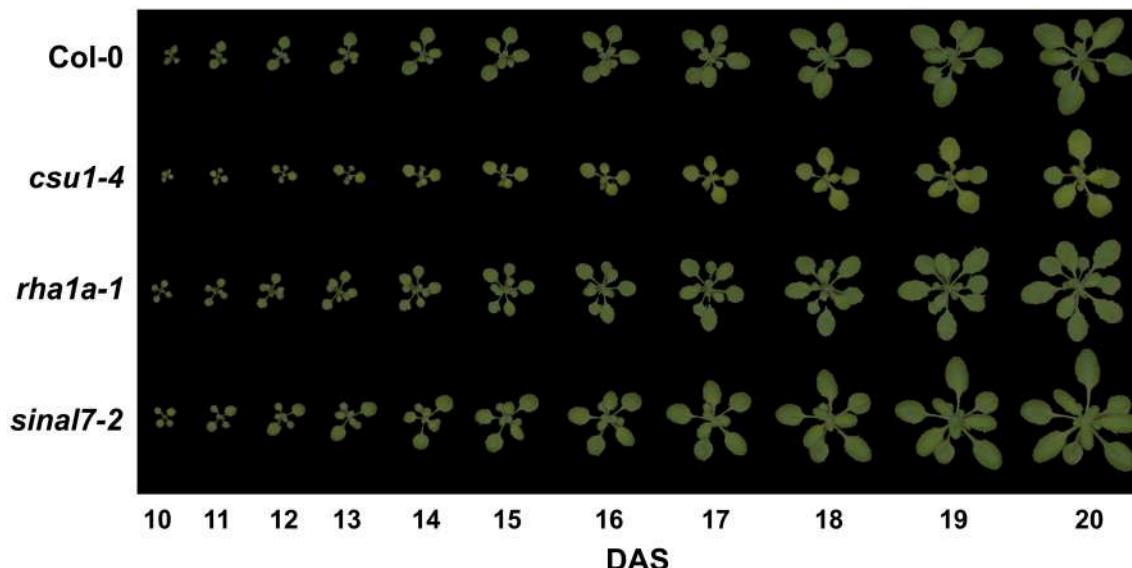
yellowish coloration (Figure 4). The mutant line *rha1a-1* seemed to have smaller leaves than Col-0, however, at the end of growth it appeared to have more leaves that resulted in similar final rosette area as compared to Col-0. This line also had shorter petioles and leaf serration. The third line *sinal7-2* rosette was clearly larger than Col-0 but did not show major differences in color, shape or number of leaves (Figure 4).

To further analyze these three lines, mixed non-linear models were fitted to their data using several order polynomials for parametric analysis of the models. This analysis confirmed the earlier observations of significant changes in growth and development for these lines over time (Table 1). Line *csu1-4* showed slower growth, reduced rosette area and perimeter compared to Col-0 along the complete measured period (Figures 5A,D). For line *rha1a-1* the rosette area was very similar to Col-0 being, however, slightly but significantly larger over time probably due to its higher number of leaves (Figure 5B, Table 1). Although the differences between *rha1a-1* and Col-0 were small

the statistical model was able to capture those. Conversely, *sinal7-2* showed both area and perimeter larger than Col-0 indicating more vigorous growth (Figures 5C,F).

## Circularity Parameters

Morphological data for parameters of circularity that include roundness, roundness 2 and isotropy were also evaluated for these lines (Figure 1). Line *csu1-4* showed increased roundness over the total period analyzed in comparison to Col-0 (Figures 6A–C, Table 1). However, *csu1-4* roundness curve had similar pattern to Col-0 but shifted to the right (Figure 6A). Similar situation was observed for *sinal7-2*, where the roundness curve shape was almost identical to Col-0 but in this case was shifted to the left, showing lower roundness along the total time period (Figure 6C, Table 1). Roundness curve of *rha1a-1* was neither shifted nor similar to Col-0 curve. This line showed a lower roundness than Col-0 at the beginning of the analysis,



**FIGURE 4 | Rosette growth of *csu1-4*, *rha1a-1* and *sinal7-2* mutants.** Representative rosette images are shown from days 10 to 20 after stratification. DAS, days after stratification.

**TABLE 1 | Polynomial order and their respective Chi square probability from ANOVA test for each parameter used in this study.**

Parameter	Polynomial order used	Chi square probability*		
		<i>csu1-4</i>	<i>rha1a-1</i>	<i>sinal7-2</i>
Area	3	<2.2e-16***	<2.2e-16***	<2.2e-16***
Perimeter	3	9.116e-15***	9.076e-11***	1.506e-10***
Compactness	4	<2.2e-16***	1.274e-10***	1.132e-06***
Roundness	5	<2.2e-16***	2.665e-15***	1.613e-06***
Roundness 2	5	<2.2e-16***	<2.2e-16***	8.43e-06***
Isotropy	6	0.0008271***	8.16e-08***	5.058e-05***
Eccentricity	6	<2.2e-16***	<2.2e-16***	5.102e-14***
RMS	6	<2.2e-16***	<2.2e-16***	4.53e-05***
SOL	3	0.0006792***	3.867e-07***	0.006651**

\*Comparison was performed using an ANOVA test between a base model and a model including the genetic background as factor. Significance codes: \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ .

Base model = Parameter ~ polynomial of Day + Random factor Day and Plant ID.

Model = Parameter ~ polynomial of Day \* genetic background (Col-0 or knockout line) + Random factor Day and Plant ID.

reaching a stabilization point around 16 DAS (Table 1). For Col-0 plants roundness continued decreasing until it became lower than *rha1a-1* (Figure 6B).

Line *csu1-4* showed a similar roundness 2 pattern as Col-0 that is shifted to the right by approximately 2 days (Figures 6D–F). Line *rha1a-1*, showed an oscillating pattern too, however, its roundness 2 values were constantly close to 0.9 with less steep peaks than Col-0, presenting the highest differences between days 12 and 16 (Figure 6E, Table 1). Similarly, to line *csu1-4*, line *sinal7-2* presented an oscillating pattern very similar to Col-0, however, this time the curve had shifted to the left by approximately 1 day (Figure 6F).

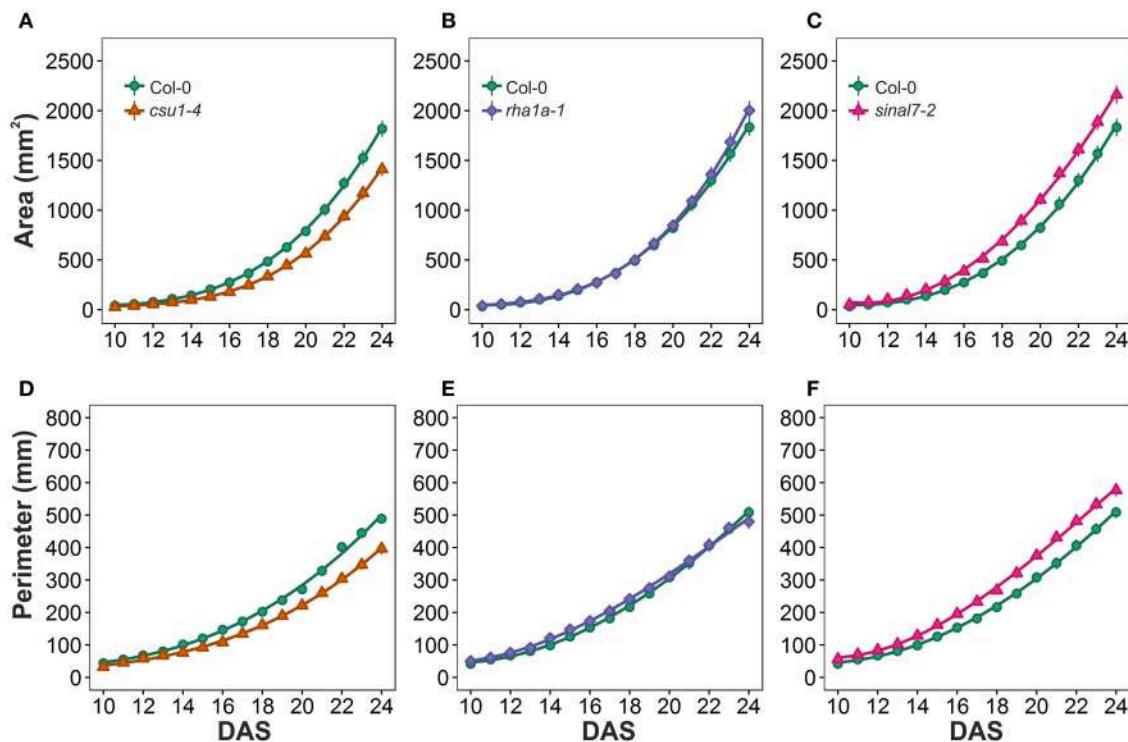
Isotropy showed similar results as roundness and roundness 2, where line *csu1-4* and *sinal7-2* had similar oscillating patterns as Col-0, but *csu1-4* curve is shifted to the right, while the curve for *sinal7-2* is shifted to the left (Figures 6G–I). Line *rha1a-1* showed a constant high isotropy value decreasing over time until reaching Col-0 pattern by day 23 (Figure 6H, Table 1).

## Symmetry Parameters

The morphological parameters describing symmetry were eccentricity and rotational mass symmetry (RMS) (Figure 1). For eccentricity, line *csu1-4* showed a similar pattern as Col-0 plants with a large and a small eccentricity peak, but shifted to the right (Figure 7A). Line *rha1a-1* presented no shift in its curve, but it showed a rather flat peak around days 11 and 15, remaining lower than Col-0 until the end of the analysis (Figure 7B, Table 1). This result shows that *rha1a-1* is less eccentric than Col-0 along the complete analysis. Line *sinal7-2* showed also a similar pattern to Col-0 plants with two eccentric peaks, but slightly shifted to the left (Figure 7C). For RMS line *csu1-4* showed similar pattern as Col-0 plants, but shifted again to the right about 1 day for the highest peak and remained higher than Col-0 in the last days of the analysis (Figure 7D). On the other hand, *rha1a-1* presented no shift in its curve, but it showed a decrease in the peak around days 11 and 15, decaying faster and remaining lower than Col-0 plants (Figure 7E, Table 1). Like in eccentricity, *sinal7-2* was almost indistinguishable from the Col-0 plants, except for a slight shift to the left captured by the model (Figure 7F).

## Center Distance Parameters

The last two morphological parameters analyzed were compactness and slenderness of the leaves (SOL), which



**FIGURE 5 | Growth measurements from day 10 to 20.** Area (A) *csu1-4*, (B) *rha1a-1*, and (C) *sinal7-2*. Perimeter (D) *csu1-4*, (E) *rha1a-1*, and (F) *sinal7-2*. Markers, daily mean; error bars, 95% confidence interval; Curves, fitted models;  $n = 20$  plants. The experiment was repeated at least 3 times with similar results. DAS, days after stratification.

were based on the center distance (**Figure 1**). Here line *csu1-4* showed a decay of compactness overtime in a similar way to Col-0 plants, but its curve was shifted to the right (**Figure 8A**). Lines *csu1-4* and *sinal7-2* presented quite normal compactness curves, while for *rha1a-1* the pattern that was less compact than Col-0 plants at the beginning of the analyzed period (**Figure 8B, Table 1**). The compactness later rises above Col-0, showing higher compactness values. Like for the previously described parameters, *sinal7-2* compactness curve showed slightly lower values than Col-0, except for the last 2 days where Col-0 plants reached *sinal7-2* compactness (**Figure 8C**).

Line *csu1-4* showed lower SOL values than Col-0, while *rha1a-1* and *sinal7-2* showed higher SOL values than Col-0 (**Figures 8E,F, Table 1**). The main differences in SOL could be observed during the exponential growing phase of the rosette and reaching a plateau at the end of the analyzed period where the differences to Col-0 plants become insignificant (**Figures 8D-F**).

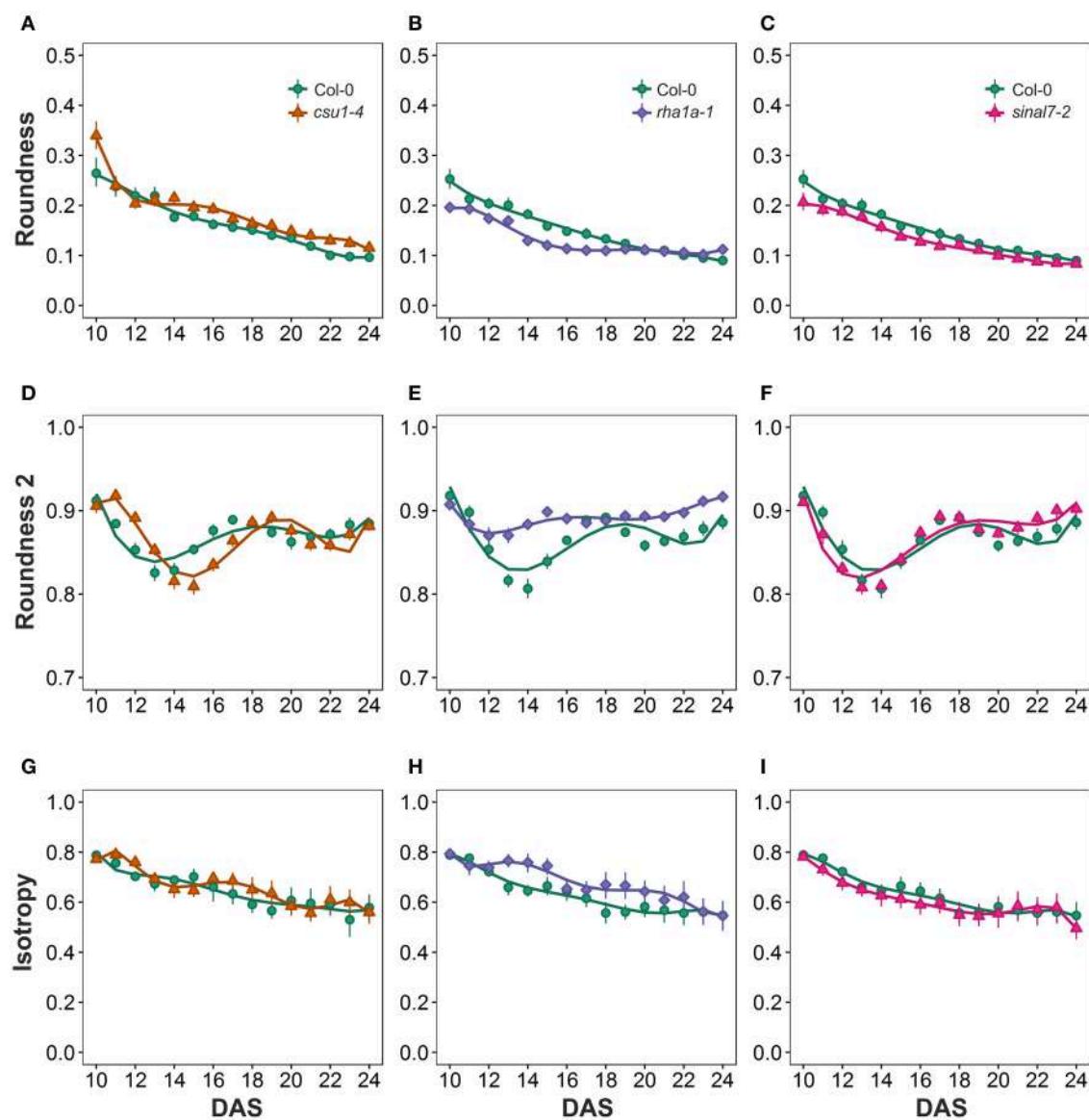
## Flowering Time Phenotypes

Flowering time mutants identified in the screen represented both with reduced and increased leaf numbers at bolting (**Table 2**). Line *csu1-4* (AT1G61620) was clearly early-flowering in both experimental replications. AT5G63970, a putative forkhead box protein, mutant line was early flowering in one

of two experimental replications. SBP (S-ribonuclease binding protein) family protein (AT4G17680) was late flowering in both experimental replications. As already shown by others, *cop1-6* mutant was early flowering in both LAB (7) and DTB (22). In most of the mutant lines, LAB did not differ from Col-0 in all experimental replications, but the trend was observed in both or all. LAB or DTB of *sinal7-2* did not differ from Col-0 in either of the experimental replications.

## Mutation in *SINAL7* Causes Flower Growth Phenotypes

Flower morphology of the analyzed mutants was observed under stereomicroscope. The mutant line *sinal7-2* was found to produce flower buds of abnormal shape, characterized by presence of cavities in the bud tips (**Figures 9A,B**). These openings were present at one or both sides of the affected buds and were caused by tips of the lateral sepals bending inwards (**Figures 9E,F**). Also medial sepals frequently showed altered morphology: their tips covered the buds to a lesser extent than in Columbia, resulting in their “blunt” appearance. Whereas these phenotypes were present in all 18 analyzed inflorescences of *sinal7-2* plants, regardless of the plant age—only two out of 13 analyzed wild type inflorescences showed similar sepal features, restricted to the first six flowers on the main stems. Scoring flowers stage late 12–15 (located between

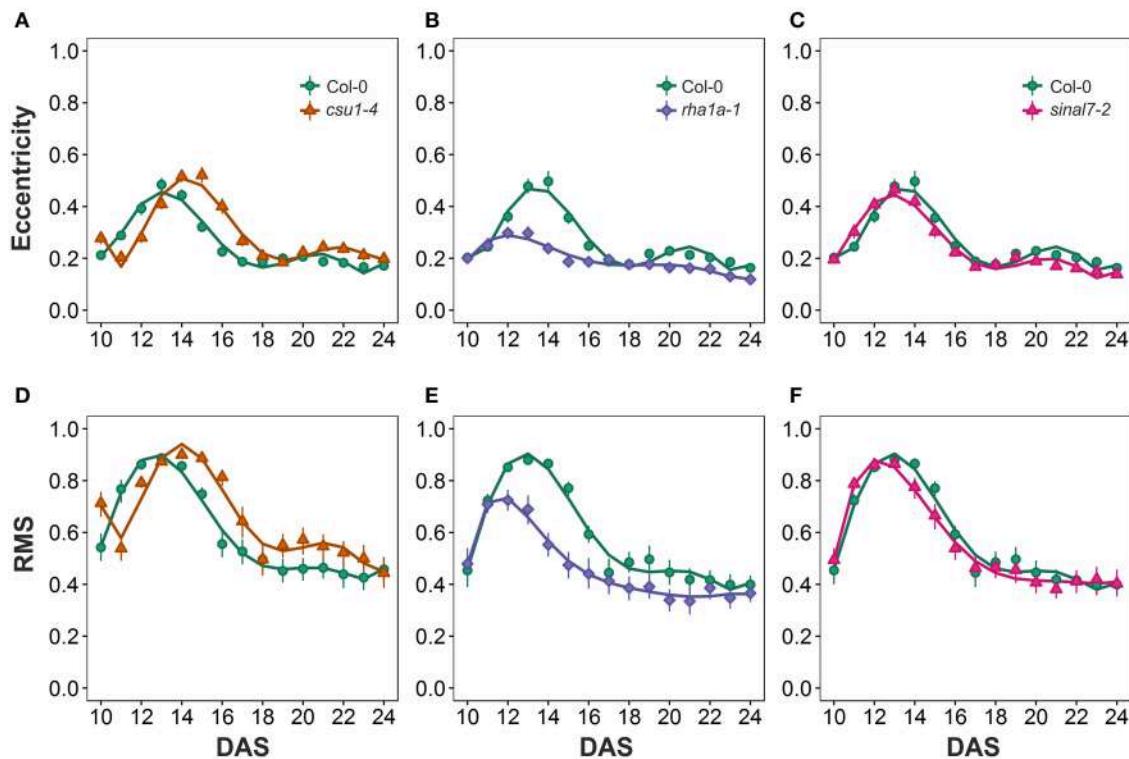


**FIGURE 6 |** Circularity measurements from day 10 to 20. Roundness (A) *csu1-4*, (B) *rha1a-1*, and (C) *sinal7-2*. Roundness2 (D) *csu1-4*, (E) *rha1a-1*, and (F) *sinal7-2*. Isotropy (G) *csu1-4*, (H) *rha1a-1*, and (I) *sinal7-2*. Markers, daily mean; error bars, 95% confidence interval; Curves, fitted models;  $n = 20$  plants. The experiment was repeated at least 3 times with similar results. DAS, days after stratification.

positions 1st and 20th on the main inflorescences) revealed that in 54% of the mutant flowers (43/80) at least one lateral sepal tip was bent inwards—as compared to 6/50, i.e., 12% in Col-0 (the analyzed flowers came from 13 to 9 individual plants, respectively).

Dissecting flower buds at the end of stage 12 revealed that the occurrence of ingrown lateral sepal tips was accompanied by petal wrinkling, as the sepal shape interfered with elongation of the petals (Figures 9G,H). Indeed, in some of the mature flowers with bent lateral sepal tips, the petal blades remained wrinkled; in several cases also pistil or stamen shape was affected (Figures 9C,D,I,J).

*SINAL7* has been shown to mediate ubiquitination of glyceraldehyde-3-phosphate dehydrogenase 1 (GAPC1) enzyme *in vitro* and to affect its enzymatic activity and subcellular localization in *Arabidopsis* (Peralta et al., 2016). In plants lacking GAPC1 male sterility was observed (Rius et al., 2008). To investigate whether deficiency of *SINAL7* impairs male fertility in the *sinal7-2* mutant, pollen viability was inspected according to the modified Alexander method (Peterson et al., 2010). Anthers of 12 mutant and 11 Col-0 flowers in the developmental stages late 12 and 13 were stained (early and late flowers, originating from at least five individual plants per line). However, no difference between the mutant and Col-0 pollen was observed:



**FIGURE 7 | Symmetry measurements from day 10 to 20.** Eccentricity (A) *csu1-4*, (B) *rha1a-1*, and (C) *sinal7-2*. Rotational Mass Symmetry (RMS) (D) *csu1-4*, (E) *rha1a-1*, and (F) *sinal7-2*. Markers, daily mean; error bars, 95% confidence interval; Curves, fitted models;  $n = 20$  plants. The experiment was repeated at least 3 times with similar results. DAS, days after stratification.

anthers of both lines contained almost exclusively viable pollen grains (Figures 9K,L).

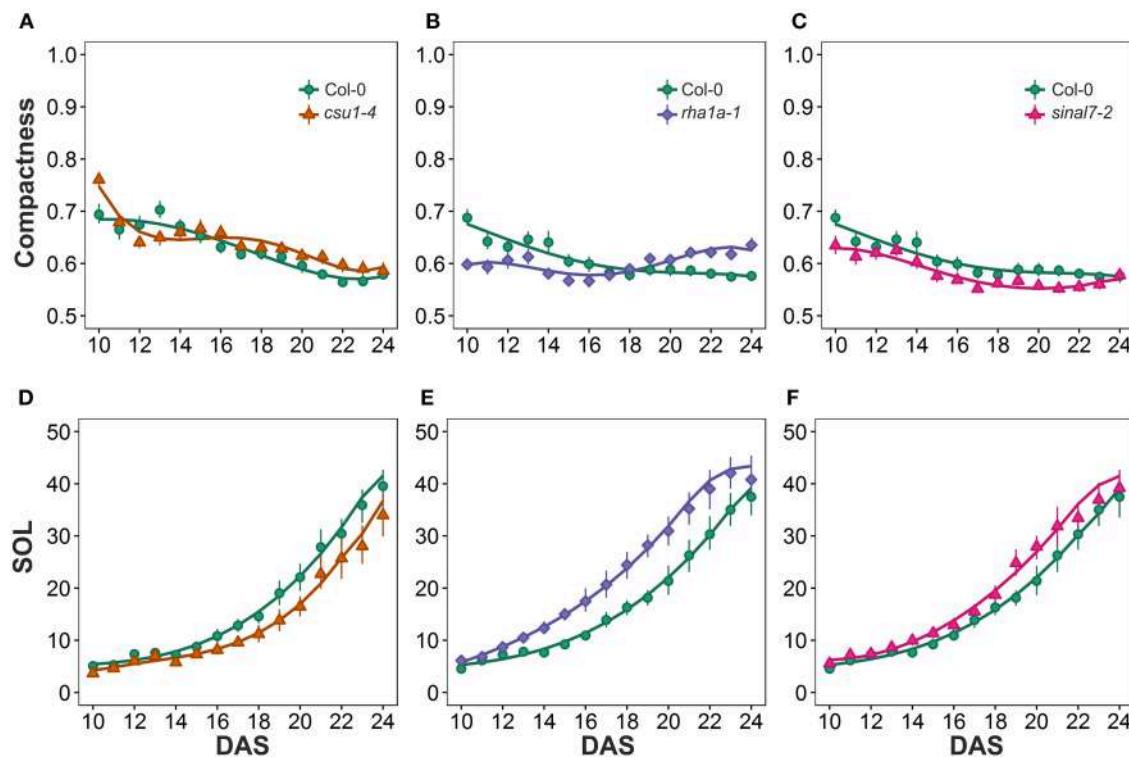
## DISCUSSION

Genomic knowledge in both model plants and crops is expanding at a fast pace. However, translating the knowledge from sequence to function and thereby from models to applications is hampered by bottlenecks in screening for the phenotypes associated with the genotypes. Here, we set out to conduct a reverse genetic approach (Bolle et al., 2011) by defining a proportion of the RING type ubiquitin E3 ligases to the developmental processes of flowering time control or flower development. To this end, the RING type ubiquitin E3 ligases were first curated in the most recent *Arabidopsis* genome annotation (ARA11) that had been improved, e.g., by the next generation sequencing techniques (Krishnakumar et al., 2015). Thereby, many gene models had indeed become obsolete, split, merged or their original sequence had changed. We also found that in the annotations there are a considerable number of RING domain containing proteins annotated as RING/U-box genes. RING and U-box share similar functions and are structurally and functionally similar, both are ubiquitin E3 ligases that work as scaffolds between the ubiquitin E2 conjugase and substrate. However, at the amino acid residual level RING and U-box domains are significantly different; in the RING domain the arrangement of cysteines and histidines

mediate binding of two zinc ions to stabilize the RING domain, while the U-box domains are stabilized by a set of hydrogen bonds and salt bridges (Wiborg et al., 2008).

Recent studies have revealed complex molecular networks that include ubiquitin E3 ligases in regulation of flowering (Lazaro et al., 2012; Peng et al., 2013; Xia et al., 2013). To start defining the genomic flower related Ubiquitin Proteasome System of RING E3 ligases, we first verified the gene expression patterns of the curated RING genes. RING E3 ligases work at protein level but are likely to be transcriptionally directed to their relevant tissues. From the 509 RING genes, 122 were indeed associated with flowering with enrichment of gene expression in flower organs or during flowering. This observation prompted us to obtain a representative mutant collection for phenotypic evaluation.

To screen for phenotypes associated with the mutant collection, an automated plant phenotyping facility was utilized. To facilitate a phenotypic screening of a large *Arabidopsis* mutant collection a phenomics workflow was established to analyze simultaneously up to 36 genotypes in a PlantScreen<sup>TM</sup> imaging system installed at the Viikki campus of the University of Helsinki (<http://blogs.helsinki.fi/nappi-blog/>). Although T-DNA insertion knock-out mutants do not always impair gene function (Bolle et al., 2011), the high-throughput phenomics screen of altogether 43 genotypes singled out three mutant lines with clear growth and morphology phenotypes, three mutant lines with flowering time phenotypes and only one with altered flower structure.



**FIGURE 8 |** Center distance measurements from day 10 to 20. Compactness (A) *csu1-4*, (B) *rha1a-1*, and (C) *sinal7-2*. Slenderness Of Leaves (SOL) (D) *csu1-4*, (E) *rha1a-1*, and (F) *sinal7-2*. Markers, daily mean; error bars, 95% confidence interval; Curves, fitted models;  $n = 20$  plants. The experiment was repeated at least 3 times with similar results. DAS, days after stratification.

For the *Arabidopsis* growth assessment, we analyzed rosette growth from 10 to 20 DAS. The analysis of such longitudinal data is challenging and demands automated statistical analysis and modeling steps. The rosette growth normally follows a sigmoid pattern showing a lag phase represented by slow growth around the first 10 days, accelerating in the middle and slowing down when getting close to the transition from the vegetative to reproductive phase. The best way to model data with sigmoid behavior is by fitting a three parameter logistic regression (3PL) to explain the three stages (Paine et al., 2012; Tessmer et al., 2013; Neilson et al., 2015). However, our analysis time window captured only the lag and the exponential phases, so a 3PL model was not suitable for our data. Therefore, we used polynomials for more flexibility and a better explanation of the data for all the parameters. This was particularly useful for the initial screening of the data of the tens of lines for the complex parameters like roundness, roundness 2, isotropy, compactness and RMS.

Typically, the parameters of roundness 2, isotropy and RMS increase and decrease over time. This behavior is due to the natural cycle of leaf initiation and expansion. At the beginning when the two first true leaves are developed, the rosette has an elliptical shape that becomes more circular when the leaves 3 and 4 appear and start to expand. Because leaves 3 and 4 keep on expanding, while the leaves 1 and 2 have already stopped expanding, the rosette takes an elliptical shape around day 12 (Figure 4). This process is repeated each time two new

leaves develop and expand, explaining the oscillating behavior of these parameters. The steepness of each peak decrease over time because previously generated leaves expand making the rosette more circular. Thus, recording fluctuations in these parameters allows establishing the developmental timing of leaf initiation and expansion.

Here, three lines showed consistently significant differences in growth and morphology compared to the wild type Col-0. The mutant lines *csu1-4* and *sinal7-2* showed similar growth curve shapes as Col-0, but shifted to the left or right, respectively, for all morphological parameters. This behavior was explained by their speed of growth over time. If two lines differ in their growth rate but were analyzed only on one particular day after germination, they could show high differences in morphological parameters. Therefore, longitudinal time course analysis of *Arabidopsis* rosette growth and shape became compulsory for making accurate conclusions about the effect of a mutation also on morphology. On the contrary, the *rha1a-1* mutant did not show major differences in growth, but did for morphology. The increased number and serration of rosette leaves in *rha1a-1* rendered the rosette perimeter and the skeleton longer, thereby, reducing the roundness and increasing SOL during all time points (Figures 6B, 8E). Furthermore, the increased number of leaves of *rha1a-1* prevented its rosette from taking overly elliptical shape, keeping it more circular than Col-0 plants over time (Figure 4). This characteristic was translated in

**TABLE 2 | Number of leaves and number of days to bolting in *Arabidopsis* mutant lines grown in LDs.**

AGI	NASC ID	Leaves at bolting			Days to bolting		
		Mutant line	Dunnett	Col-0	Mutant line	Dunnett	Col-0
AT1G61620	N686069	9.7 ± 2.1	*	13.5 ± 1.8	25.2 ± 1.6		25.7 ± 2.2
CSU1	csu1-4	9.9 ± 1.5	*	15.5 ± 1.1	23.6 ± 1.6	*	26.8 ± 0.7
		12.1 ± 0.4	*	15.1 ± 0.4	24.9 ± 0.6	*	26.2 ± 0.7
AT1G68820	N667194	15.8 ± 1.5	*	13.5 ± 2.0	27.2 ± 1.5	*	25.3 ± 1.6
<i>RING E3</i>		16.0 ± 1.2		16.6 ± 1.4	28.3 ± 1.2		28.0 ± 1.1
		14.3 ± 0.7	*	15.9 ± 0.3	24.9 ± 0.9	*	27.2 ± 0.6
	N659628	14.1 ± 1.8		13.6 ± 2.0	26.9 ± 1.6	*	25.5 ± 1.7
		16.5 ± 1.1		15.4 ± 0.9	28.2 ± 1.1	*	26.7 ± 1.3
		16.1 ± 1.3		16.6 ± 1.4	27.9 ± 1.1		28.0 ± 1.1
AT2G22680	N219963	16.4 ± 1.0	*	12.6 ± 1.7	27.9 ± 1.3	*	25.2 ± 1.4
WAHV1		16.6 ± 1.3		15.6 ± 1.7	27.7 ± 1.2	*	26.1 ± 1.2
	N653622	13.3 ± 2.0		12.5 ± 1.9	25.0 ± 1.5		25.0 ± 1.6
		15.5 ± 1.0		14.6 ± 1.4	26.2 ± 1.0		25.4 ± 1.5
AT2G32950		7.5 ± 0.6	*	12.6 ± 0.7	22.7 ± 1.4	*	28.4 ± 0.9
COP1	cop1-6	6.7 ± 0.7	*	13.7 ± 0.3	21.1 ± 0.6	*	27.3 ± 0.9
AT2G37150	N685421	15.3 ± 1.5	*	13.5 ± 2.0	27.2 ± 1.1	*	25.3 ± 1.6
<i>RING E3</i>		16.4 ± 1.9		15.4 ± 0.9	27.0 ± 1.3		26.7 ± 1.3
		16.6 ± 1.3		16.6 ± 1.4	28.3 ± 0.8		28.0 ± 1.1
AT2G47700	N878610	15.5 ± 2.1	*	13.0 ± 1.4	26.6 ± 1.5		25.6 ± 1.3
RFI2	rfi2-3	16.5 ± 1.4		15.5 ± 1.1	27.7 ± 0.9		26.8 ± 0.7
AT3G07200	N481270	15.3 ± 2.0	*	13.5 ± 1.8	27.7 ± 1.7	*	25.7 ± 2.2
STUBL3	stubb3	15.6 ± 2.0		15.5 ± 1.1	27.3 ± 1.7		26.8 ± 0.7
AT3G09760	N653280	13.0 ± 1.8		12.6 ± 1.7	24.8 ± 1.4		25.2 ± 1.4
<i>RING E3</i>		13.9 ± 1.5		14.6 ± 1.4	23.9 ± 1.3	*	25.4 ± 1.5
AT4G17680	N67002	15.2 ± 2.1	*	13.5 ± 1.8	27.2 ± 1.7	*	25.7 ± 2.2
SBP		16.8 ± 1.3	*	15.5 ± 1.1	27.9 ± 1.3	*	26.8 ± 0.7
AT5G37890	N596989	13.1 ± 2.1		13.0 ± 1.4	25.6 ± 1.9		25.6 ± 1.3
SINAL7	sinal7-2	14.4 ± 1.7		15.1 ± 1.3	25.7 ± 1.3		26.6 ± 1.4
	N833574	14.0 ± 0.7	*	15.1 ± 0.4	24.3 ± 0.5	*	26.2 ± 0.7
AT5G63970	N694155	13.3 ± 2.2		13.6 ± 2.0	25.6 ± 1.8		25.5 ± 1.7
<i>RING E3</i>		13.5 ± 1.3	*	15.4 ± 0.9	24.6 ± 1.5	*	26.7 ± 1.3
		14.4 ± 0.7		15.1 ± 0.4	24.8 ± 1.0	*	26.2 ± 0.7

Pairwise comparisons were performed against corresponding Col-0-line using Dunnett's test.

\*Indicates statistically significant difference ( $\alpha = 0.05$ ).

N = 19–20 in each row.

higher roundness 2, isotropy, compactness and lower eccentricity and RMS (Figures 6E–H, 7B–E, 8B). Thus, the morphological parameters can be used not only to record developmental timing but also to explain the plant architecture in a numeric manner.

The line showing an early flowering time phenotype was COP1 SUPPRESSOR1 (CSU1). *csu1-4* plants flowered three to six leaves earlier than Col-0 grown under LDs (Table 2). In addition to early flowering, *csu1-4* plants showed vegetative phenotypes: plants were smaller than Col-0 (Figure 5), the eccentricity, RMS and roundness2 development started later than Col-0 (Figures 6, 7), and SOL was smaller than in Col-0 (Figure 8). CSU1 has been shown to negatively regulate hypocotyls length in the dark, via ubiquitination of COP1 and repression of SPA1 (Xu et al., 2014). Our results indicate that CSU1 may regulate both vegetative and generative development. The line showing a late flowering

phenotype, SBP family protein (AT4G17680), flowered one to two leaves later than Col-0 (Table 2). This gene was selected for the phenotypic analysis based on its interaction with TOE2. *toe2* is late flowering, and *toe1 toe2* double mutant represses *FT* expression (Zhai et al., 2015). Our results suggest that this SBP family protein could be involved in regulation of flowering time possibly through TOE2. Some SBP family members are known to regulate flowering time. Four SBP proteins, BOTRYTIS SUSCEPTIBLE1 INTERACTOR (BOI) and its three homologous repress flowering by repressing *FT* expression in a CO dependent manner and a CO independent manner via DELLA proteins (Nguyen et al., 2015). This evidence suggests that there might be a connection between SBP proteins and flowering time control.

In the *sinal7-2* mutant defects in flower morphology were observed. SINAL7 has been shown to ubiquitinate



**FIGURE 9 | Flower phenotypes of the *sinal7-2* mutant.** Flower developmental stages assigned according to Smyth et al. (1990). Scale bars: 1 mm (**A–J**) and 100  $\mu$ m (**K,L**). (**A,B**) Representative inflorescences of Col-0 (**A**) and *sinal7-2* (**B**). All flowers and siliques older than stage 12 have been removed. Mutant flower buds contain cavities beneath the bud tip (indicated with white arrows). (**C,D**) Petals of a Col-0 (**C**) and a *sinal7-2* (**D**) flower at stage 15. White arrows pointing at the wrinkled mutant petals. (**E,F**) Adaxial surface of the sepals from a Col-0 (**E**) and a *sinal7-2* (**F**) flower at stage 15. White arrow pointing at the bending lateral sepal tip of *sinal7-2*. (**G,H**) Late stage 12 flower buds of Col-0 (**G**) and *sinal7-2* (**H**). The medial sepals have been removed to reveal the elongating and wrinkling petals blocked by the ingrown lateral sepals of the mutant. (**I,J**) Col-0 (**I**) and *sinal7-2* (**J**) flowers stage 15. (**K,L**) Representative anthers from Col-0 (**K**) and *sinal7-2* (**L**) flowers stage 12–13 stained for pollen viability.

glyceraldehyde-3-phosphate dehydrogenase 1 (GAPC1) and to regulate its enzymatic activity and movement to nucleus (Peralta et al., 2016). GAPC1 plays a role in glycolysis, thus regulating carbon metabolism, and it has been also associated with cytoskeleton and mitochondria (Giegé et al., 2003; Anderson et al., 2004). *SINAL7* gene is differentially expressed in the *gapc1* knockout mutant as well as in *u-ATP9* plants with mitochondrial dysfunction (Rius et al., 2008; Busi et al., 2011). Although both *gapc1* and *u-ATP9* lines showed defects in male fertility (Gómez-Casati et al., 2002; Rius et al., 2008), we did not observe increased number of aborted pollen grains in the *sinal7-2* mutant (**Figures 9K,L**), suggesting that *SINAL7*-mediated GAPC1 regulation does not impact pollen maturation. Although we have not tested if the *sinal7-2* mutation influences pollen germination and pollen tube growth, fertility of the mutant did not seem to be strongly compromised. Instead, we observed defects in *sinal7-2* flower morphology - cavities in flower buds and wrinkled petals. Sepal curvature is controlled by giant cells in the abaxial epidermis, in which cell expansion is connected with endoreduplication (Roeder et al., 2010, 2012). A couple of mutants have been identified in which reduction of giant cells was accompanied by their sepals bending inwards. Closer examination of *sinal7-2* sepal epidermis will show whether the observed bent sepal tips and resulting flower bud cavities (**Figures 9A,B,E–H**) originate from endoreduplication defects, which could suggest a novel role for the *SINAL7* protein. Other flower phenotypes of the mutant—wrinkling of petals as well as bending of stamens and pistils (**Figures 9C,D,G–J**)—seem to be a direct consequence of the abnormal shape of sepals posing an obstacle for the developing floral organs during their growth and release from the buds. Nevertheless, at this point it cannot be ruled out that the *SINAL7* ubiquitin E3 ligase could be involved in the development of these flower organs in other ways.

Here we showed that automated, imaging based phenotyping platform is an efficient tool to overcome the limiting factors of manual and visual phenotypic measurements of large plant collections. Imaging based platforms also allow deep resolution

of the phenotypes and thereby more precise association with the genotypes. Furthermore, the automated plant management and transportation to imaging, facilitates time course experiments. Thereby, recording longitudinal numeric values indicating changes in rosette size and morphology can be utilized in developmental timing of plant growth and development. Here the customized solution of the PSI PlantScreen™ system by top view CCD camera in combination with online data processing was used for high throughput phenotyping of an *Arabidopsis* mutant collection. The obtained resolution and high throughput, whereby hundreds of plants can be analyzed in the time that normally a handful would be analyzed, is an obvious advantage.

## AUTHOR CONTRIBUTIONS

MP conducted the genomic screens, performed the phenomic screens and statistical analysis of these data. KM performed the QPCR analysis and participated in phenotyping and statistical data analysis, FW designed and revised genotyping of the mutant collection. MB designed and revised the flower phenotype analysis. EC participated in the phenotyping assays and conducted flowering time experiments. KH designed the project as a whole, approved the data and wrote the manuscript.

## FUNDING

This project was supported by The Academy of Finland (Suomen Akatemia #283138, #256094, #250972) and Becas Chile from Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) Chile.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00416/full#supplementary-material>

## REFERENCES

- Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657. doi: 10.1126/science.1086391
- Amasino, R. M., and Michaels, S. D. (2010). The timing of flowering. *Plant Physiol.* 154, 516–520. doi: 10.1104/pp.110.161653
- Anderson, L. E., Ringenbergs, M. R., and Carol, A. A. (2004). Cytosolic glyceraldehyde-3-P dehydrogenase and the B subunit of the chloroplast enzyme are present in the pea leaf nucleus. *Protoplasma* 223, 33–43. doi: 10.1007/s00709-003-0030-6
- Awlia, M., Nigro, A., Fajkus, J., Schmockel, S., Negrao, S., Santelia, D., et al. (2016). High-throughput non-destructive phenotyping of traits that contribute to salinity tolerance in *Arabidopsis thaliana*. *Front. Plant Sci.* 7:1414. doi: 10.3389/fpls.2016.01414
- Bolle, C., Schneider, A., and Leister, D. (2011). Perspectives on systematic analyses of gene function in *Arabidopsis thaliana*: new tools, topics and trends. *Curr. Genomics* 12, 1. doi: 10.2174/138920211794520187
- Bouché, F., Lobet, G., Tocquin, P., and Périlleux, C. (2016). FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Res.* 44, D1167–D1171. doi: 10.1093/nar/gkv1054
- Boyes, D., Zayed, A., Ascenzi, R., McCaskill, A., Hoffman, N., Davis, K., et al. (2001). Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. *Plant Cell* 13, 1499–1510. doi: 10.1105/tpc.13.7.1499
- Busi, M., Gomez-Lobato, M., Rius, S., Turowski, V., Casati, P., Zabaleta, E., et al. (2011). Effect of mitochondrial dysfunction on carbon metabolism and gene expression in flower tissues of *Arabidopsis thaliana*. *Mol. Plant* 4, 127–143. doi: 10.1093/mp/ssq065
- Chen, M., and Ni, M. (2006a). RED AND FAR-RED INSENSITIVE 2, a RING-domain zinc finger protein, mediates phytochrome-controlled seedling deetiolation responses. *Plant Physiol.* 140, 457–465. doi: 10.1104/pp.105.073163
- Chen, M., and Ni, M. (2006b). RFI2, a RING-domain zinc finger protein, negatively regulates CONSTANS expression and photoperiodic flowering. *Plant J.* 46, 823–833. doi: 10.1111/j.1365-313X.2006.02740.x
- Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W.-R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.* 139, 5–17. doi: 10.1104/pp.105.063743
- Fornara, F., Panigrahi, K. C. S., Gissot, L., Sauerbrunn, N., Rühl, M., Jarillo, J. A., et al. (2009). *Arabidopsis* DOF transcription factors act redundantly

- to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Dev. Cell* 17, 75–86. doi: 10.1016/j.devcel.2009.06.015
- Giegé, P., Heazlewood, J. L., Roessner-Tunali, U., Millar, A. H., Fernie, A. R., Leaver, C. J., et al. (2003). Enzymes of glycolysis are functionally associated with the mitochondrion in *Arabidopsis* cells. *Plant Cell* 15, 2140–2151. doi: 10.1105/tpc.012500
- Gómez-Casati, D., Busi, M., Gonzalez-Schain, N., Mouras, A., Zabaleta, E., and Araya, A. (2002). A mitochondrial dysfunction induces the expression of nuclear encoded complex I genes in engineered male sterile *Arabidopsis thaliana*. *FEBS Lett.* 4, 70–74. doi: 10.1016/S0014-5793(02)03631-1
- Groth, M., Stroud, H., Feng, S., Greenberg, M., Vashisht, A., Wohlschlegel, J., et al. (2014). SNF2 chromatin remodeler-family proteins FRG1 and –2 are required for RNA-directed DNA methylation. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17666–17671. doi: 10.1073/pnas.1420515111
- Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., et al. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinformatics* 2008:420747. doi: 10.1155/2008/420747
- Jang, S., Marchal, V., Panigrahi, K. C., Wenkel, S., Soppe, W., Deng, X. W., et al. (2008). *Arabidopsis* COP1 shapes the temporal pattern of CO accumulation conferring a photoperiodic flowering response. *EMBO J.* 27, 1277–1288. doi: 10.1038/emboj.2008.68
- Kleinboelting, N., Huep, G., Kloetgen, A., Viehoever, P., and Weisshaar, B. (2012). GABI-Kat SimpleSearch: new features of the *Arabidopsis thaliana* T-DNA mutant database. *Nucleic Acids Res.* 40, D1211–D1215. doi: 10.1093/nar/gkr1047
- Kosarev, P., Mayer, K. F. X., and Hardtke, C. S. (2002). Evaluation and classification of RING-finger domains encoded by the *Arabidopsis* genome. *Genome Biol.* 3, 1. doi: 10.1186/gb-2002-3-4-research0016
- Krishnakumar, V., Hanlon, M. R., Contrino, S., Ferlanti, E. S., Karamycheva, S., Kim, M., et al. (2015). Araport: the *Arabidopsis* information portal. *Nucleic Acids Res.* 43, D1003–D1009. doi: 10.1093/nar/gku1200
- Laubinger, S., Marchal, V., Gentilhomme, J., Wenkel, S., Adrian, J., Jang, S., et al. (2006). *Arabidopsis* SPA proteins regulate photoperiodic flowering and interact with the floral inducer CONSTANS to regulate its stability. *Development* 133, 3213–3222. doi: 10.1242/dev.02481
- Lazaro, A., Valverde, F., Piñeiro, M., and Jarillo, J. A. (2012). The *Arabidopsis* E3 ubiquitin ligase HOS1 negatively regulates CONSTANS abundance in the photoperiodic control of flowering. *Plant Cell* 24, 982–999. doi: 10.1105/tpc.110.081885
- Liu, L.-J., Zhang, Y.-C., Li, Q.-H., Sang, Y., Mao, J., Lian, H.-L., et al. (2008). COP1-mediated ubiquitination of CONSTANS is implicated in cryptochrome regulation of flowering in *Arabidopsis*. *Plant Cell* 20, 292–306. doi: 10.1105/tpc.107.057281
- Livak, K., and Schmittgen, T. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta Ct}$  method. *Methods* 25, 40. doi: 10.1006/meth.2001.1262
- Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. CRC Press.
- Neilson, E. H., Edwards, A. M., Blomstedt, C. K., Berger, B., Möller, B. L., and Gleadow, R. M. (2015). Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal crop plant to nitrogen and water deficiency over time. *J. Exp. Bot.* 66, 1817–1832. doi: 10.1093/jxb/eru526
- Nguyen, K. T., Park, J., Park, E., Lee, I., and Choi, G. (2015). The *Arabidopsis* RING domain protein BOI inhibits flowering via co-dependent and co-independent mechanisms. *Mol. Plant* 8, 1725–1736. doi: 10.1016/j.molp.2015.08.005
- Paine, C. E. T., Marthews, T. R., Vogt, D. R., Purves, D., Rees, M., Hector, A., et al. (2012). How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. *Methods Ecol. Evol.* 3, 245–256. doi: 10.1111/j.2041-210X.2011.00155.x
- Peng, Y., Shih, C., Yang, J., Tan, C., Hsu, W., Huang, Y., et al. (2013). A RING-type E3 ligase controls anther dehiscence by activating the jasmonate biosynthetic pathway gene DEFECTIVE IN ANTER DEHISCENCE1 in *Arabidopsis*. *Plant J.* 74, 310–327. doi: 10.1111/tpj.12122
- Peralta, D. A., Araya, A., Busi, M. V., and Gomez-Casati, D. F. (2016). The E3 ubiquitin-ligase SEVEN IN ABSENTIA like 7 mono-ubiquitinates glyceraldehyde-3-phosphate dehydrogenase 1 isoform *in vitro* and is required for its nuclear localization in *Arabidopsis thaliana*. *Int. J. Biochem. Cell Biol.* 70, 48–56. doi: 10.1016/j.biocel.2015.11.007
- Peterson, R., Slovin, J. P., and Chen, C. (2010). A simplified method for differential staining of aborted and non-aborted pollen grains. *Int. J. Plant Biol.* 1:13. doi: 10.4081/pb.2010.e13
- Piñeiro, M., and Jarillo, J. A. (2013). Ubiquitination in the control of photoperiodic flowering. *Plant Sci.* 198, 98–109. doi: 10.1016/j.plantsci.2012.10.005
- Rius, S. P., Casati, P., Iglesias, A. A., and Gomez-Casati, D. F. (2008). Characterization of *Arabidopsis* lines deficient in GAPC-1, a cytosolic NAD-dependent glyceraldehyde-3-phosphate dehydrogenase. *Plant Physiol.* 148, 1655–1667. doi: 10.1104/pp.108.128769
- Roeder, A. H. K., Chickarmane, V., Cunha, A., Obara, B., Manjunath, B. S., and Meyerowitz, E. M. (2010). Variability in the control of cell division underlies sepal epidermal patterning in *Arabidopsis thaliana*. *PLoS Biol.* 8:e1000367. doi: 10.1371/journal.pbio.1000367
- Roeder, A. H. K., Cunha, A., Ohno, C. K., and Meyerowitz, E. M. (2012). Cell cycle regulates cell type in the *Arabidopsis* sepal. *Development* 139, 4416–4427. doi: 10.1242/dev.082925
- Rosso, M. G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003). An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. *Plant Mol. Biol.* 53, 247–259. doi: 10.1023/B:PLAN.0000009297.37235.4a
- Sadanandom, A., Bailey, M., Ewan, R., Lee, J., and Nelis, S. (2012). The ubiquitin-proteasome system: central modifier of plant signalling. *New Phytol.* 196, 13–28. doi: 10.1111/j.1469-8137.2012.04266.x
- Santner, A., and Estelle, M. (2010). The ubiquitin-proteasome system regulates plant hormone signaling. *Plant J.* 61, 1029–1040. doi: 10.1111/j.1365-313X.2010.04112.x
- Sawa, M., Nusinow, D. A., Kay, S. A., and Imaizumi, T. (2007). FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. *Science* 318, 261–265. doi: 10.1126/science.1146994
- Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., et al. (2005). A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* 37, 501–506. doi: 10.1038/ng1543
- Smyth, D. R., Bowman, J. L., and Meyerowitz, E. M. (1990). Early flower development in *Arabidopsis*. *Plant Cell* 2, 755–767. doi: 10.1105/tpc.2.8.755
- Song, Y. H., Smith, R. W., To, B. J., Millar, A. J., and Imaizumi, T. (2012). FKF1 conveys timing information for CONSTANS stabilization in photoperiodic flowering. *Science* 336, 1045–1049. doi: 10.1126/science.1219644
- Steffen, A., Fischer, A., and Staiger, D. (2014). Determination of photoperiodic flowering time control in *Arabidopsis* and barley. *Methods Mol. Biol.* 2, 285–295. doi: 10.1007/978-1-4939-0700-7\_19
- Stone, S. L., Hauksdóttir, H., Troy, A., Herschleb, J., Kraft, E., and Callis, J. (2005). Functional analysis of the RING-type ubiquitin ligase family of *Arabidopsis*. *Plant Physiol.* 137, 13–30. doi: 10.1104/pp.104.052423
- Suárez-López, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., and Coupland, G. (2001). CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* 410, 1116–1120. doi: 10.1038/35074138
- Tessmer, O. L., Jiao, Y., Cruz, J. A., Kramer, D. M., and Chen, J. (2013). Functional approach to high-throughput plant growth analysis. *BMC Syst. Biol.* 7, 1–13. doi: 10.1186/1752-0509-7-S6-S17
- Vierstra, R. D. (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* 10, 385–397. doi: 10.1038/nrm2688
- Wiborg, J., O’Shea, C., and Skriver, K. (2008). Biochemical function of typical and variant *Arabidopsis thaliana* U-box E3 ubiquitin-protein ligases. *Biochem. J.* 413, 447–457. doi: 10.1042/BJ20071568
- Xia, T., Li, N., Dumenil, J., Li, J., Kamenski, A., Bevan, M. W., et al. (2013). The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2

- to regulate seed and organ size in Arabidopsis. *Plant Cell* 25, 3347–3359. doi: 10.1105/tpc.113.115063
- Xu, D., Lin, F., Jiang, Y., Huang, X., Li, J., Ling, J., et al. (2014). The RING-finger E3 ubiquitin ligase COP1 SUPPRESSOR1 negatively regulates COP1 abundance in maintaining COP1 homeostasis in dark-grown Arabidopsis seedlings. *Plant Cell* 26, 1981–1991. doi: 10.1105/tpc.114.124024
- Zhai, Q., Zhang, X., Wu, F., Feng, H., Deng, L., Xu, L., et al. (2015). Transcriptional mechanism of jasmonate receptor COI1-mediated delay of flowering time in Arabidopsis. *Plant Cell* 27, 2814–2828. doi: 10.1105/tpc.15.00619

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

Copyright © 2017 Pavicic, Mouhu, Wang, Bilicka, Chovanček and Himanen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Automated Method to Determine Two Critical Growth Stages of Wheat: Heading and Flowering

Pouria Sadeghi-Tehran <sup>\*</sup>, Kasra Sabermanesh, Nicolas Virlet and Malcolm J. Hawkesford <sup>\*</sup>

Department of Plant Biology and Crop Sciences, Rothamsted Research, Harpenden, UK

## OPEN ACCESS

### Edited by:

John Doonan,  
Aberystwyth University, UK

### Reviewed by:

Yuhui Chen,  
Samuel Roberts Noble Foundation,  
USA

Andrew French,  
University of Nottingham, UK

### \*Correspondence:

Pouria Sadeghi-Tehran  
[pouria.sadeghi-tehran@rothamsted.ac.uk](mailto:pouria.sadeghi-tehran@rothamsted.ac.uk)  
Malcolm J. Hawkesford  
[malcolm.hawkesford@rothamsted.ac.uk](mailto:malcolm.hawkesford@rothamsted.ac.uk)

### Specialty section:

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

Received: 28 September 2016

Accepted: 09 February 2017

Published: 27 February 2017

### Citation:

Sadeghi-Tehran P, Sabermanesh K, Virlet N and Hawkesford MJ (2017)

Automated Method to Determine Two Critical Growth Stages of Wheat:  
Heading and Flowering.

Front. Plant Sci. 8:252.

doi: 10.3389/fpls.2017.00252

Recording growth stage information is an important aspect of precision agriculture, crop breeding and phenotyping. In practice, crop growth stage is still primarily monitored by-eye, which is not only laborious and time-consuming, but also subjective and error-prone. The application of computer vision on digital images offers a high-throughput and non-invasive alternative to manual observations and its use in agriculture and high-throughput phenotyping is increasing. This paper presents an automated method to detect wheat heading and flowering stages, which uses the application of computer vision on digital images. The bag-of-visual-word technique is used to identify the growth stage during heading and flowering within digital images. Scale invariant feature transformation feature extraction technique is used for lower level feature extraction; subsequently, local linear constraint coding and spatial pyramid matching are developed in the mid-level representation stage. At the end, support vector machine classification is used to train and test the data samples. The method outperformed existing algorithms, having yielded 95.24, 97.79, 99.59% at early, medium and late stages of heading, respectively and 85.45% accuracy for flowering detection. The results also illustrate that the proposed method is robust enough to handle complex environmental changes (illumination, occlusion). Although the proposed method is applied only on identifying growth stage in wheat, there is potential for application to other crops and categorization concepts, such as disease classification.

**Keywords:** image categorization, computer vision in agriculture, automated field phenotyping, automated growth stage observation, Field Scanalyzer, wheat heading stage, wheat flowering time

## 1. INTRODUCTION

An estimated doubling in required crop production is projected by 2,050 in order to meet the demand of the rapid growth human population (Tilman et al., 2011). To achieve this, an approximate 38% increase over current increases in annual crop production rates is required, and on not much more arable land. Further concerns exist around not only achieving this target in a changing climate, but also achieving it sustainably, whereby reducing agricultural inputs to reduce the environmental degradation caused by our agricultural footprint (Tester and Langridge, 2010). With wheat providing  $\geq 20\%$  of the world's calorie and protein intake (Braun et al., 2010), the requirement to increase yield and production is widely recognized.

Breeding and precision agriculture, including information-based management of agricultural systems, are fundamental for achieving sustainable increases in wheat productivity and production.

One component critical to both crop breeding and precision agriculture is the monitoring of developmental growth stages, as (i) it helps crop producers understand which phases of wheat development are most vulnerable to biotic and environmental stresses, and (ii) supports precision agriculture by helping making informed-decisions around which treatment should be applied, to what location and when to apply it. Two critical growth stages monitored in crops, including wheat, are heading date and flowering time, as cultivars with appropriate heading time to their target environment and life cycle duration will help maximize yield potential (Snape et al., 2001; Zhang et al., 2008).

The monitoring of heading and flowering stages are still primarily performed by human eye, which is labor-intensive and time-consuming, as these observations need to be performed on up to thousands of cultivars/varieties on a daily or bi-daily basis, given the importance in catching the starting date of these growth stages. Given that manual growth stage monitoring is also subjective, different observers may likely perceive the growth stage of the same plot differently, which introduces human-error into obtained data.

Computer vision offers an effective alternative for growth stage monitoring because of its low-cost (relative to man-hours invested in to manual observations) and the requirement for minimal human intervention. Computer vision has facilitated automation in high-throughput phenotyping, as well as areas of agriculture, such as disease detection (Pourreza et al., 2015), weed identification (Guerrero et al., 2012) and quality control (Alahi et al., 2012; Valiente-González et al., 2014). Despite the efforts of computer vision specialists over the past decades, developing reliable image-based model to identify and categorize images based on visual information is still difficult to achieve and remains an unsolved problem in the computer vision community. The visual recognition of object categories is a natural and trivial task for humans. Humans can recognize objects effortlessly even with changes in an object's appearance, such as viewing direction or a shadow being cast across the object. On the other hand, in computer vision it can be a challenging task to achieve such level of performance due to the difficulties inherent in the problem. Images are quite abstract and subjected to illumination, scale, deformation, background clutter, etc. Moreover, in computer vision, teaching a machine to distinguish and categorize objects is all about teaching it which differences in the image is matters and which don't, by scanning through diverse datasets, which is a computationally exhaustive process.

Computer vision has shown promise in detecting growth stages of crops. For seedling emergence, color segmentation approaches have been applied in maize (Yihang et al., 2014) and oilseed rape (Yu et al., 2013), using images acquired from a digital camera. Some approaches for observing later growth stages, such as heading date and flowering stage, have also been

developed. Zhu et al. (2016), developed a method to detect wheat heading stage from RGB images using a two-step coarse-to-fine detection approach. For flowering stages, Guo et al. (2015) used object-recognition to detect flowering stages from rice panicles. Although the approaches by Zhu et al. (2016) and Guo et al. (2015) were effective on a single variety, within small patches of whole canopies, applications that are more versatile and that also can be applied on different varieties on larger scale canopies are required.

This study utilizes a novel visual-based approach to monitor heading and flowering stage of field grown wheat, through the automated learning of the visual consistency between classes of canopy images, in order to identify the critical growing stages of wheat (e.g., whether ears are emerging in canopies). This method searches through an image database to identify and retrieve images containing emerged wheat ears and ears at flowering stages. This visual-based approach is:

- i. Not limited to specific wheat cultivars and is applicable to a variety of categorical wheat without implementing specific tuning for each category.
- ii. Robust to handle illumination changes and natural lightning conditions in the field.
- iii. Robust in distinguishing the early emerged ears, despite the color difference between ears and leaves being hardly distinguishable to the naked eye.

## 2. MATERIALS AND METHODOLOGY

The introduced technique is performed in four main steps (**Figure 1**):

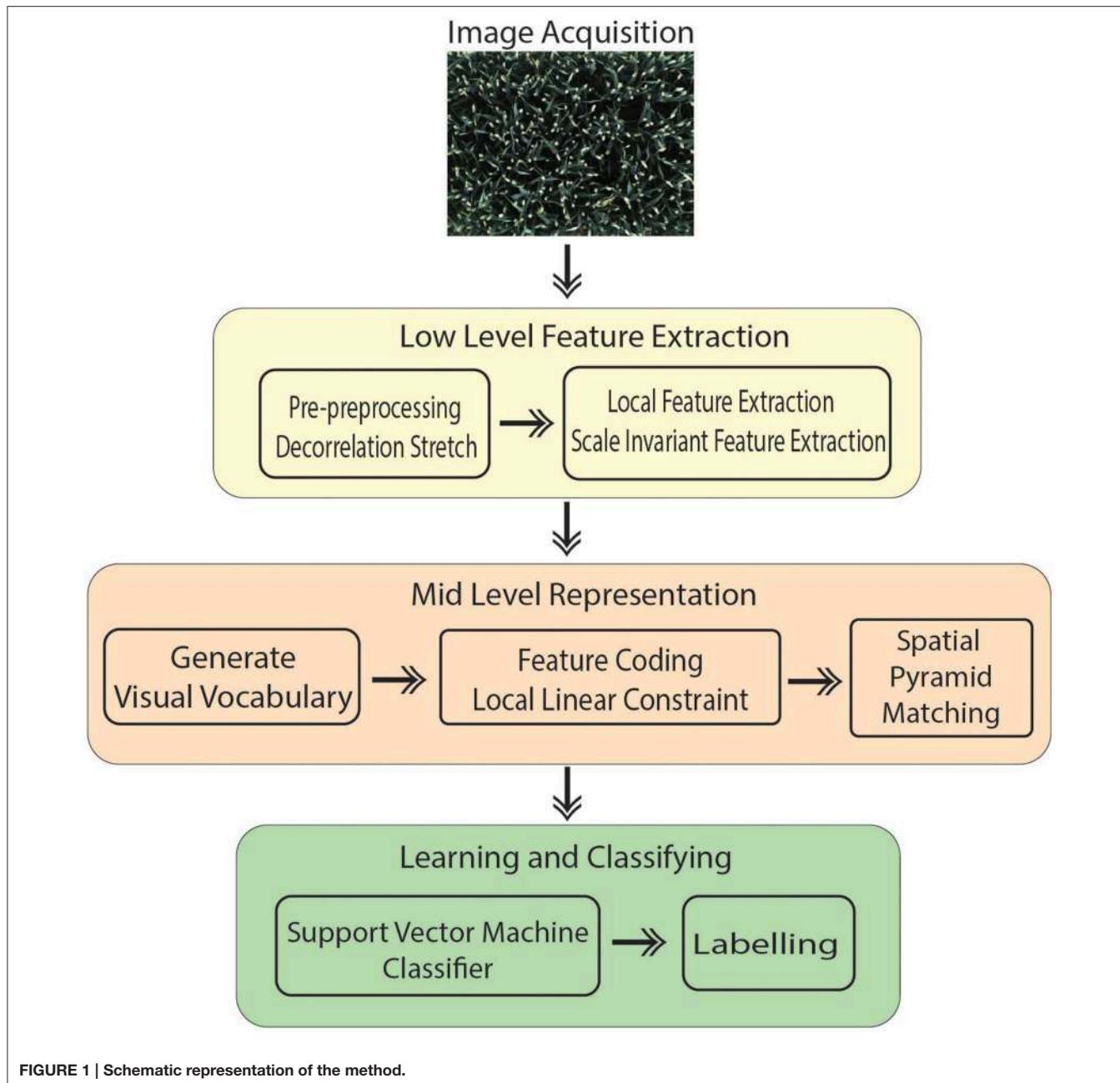
- Image acquisition: A RGB image is captured from 8 MP camera mounted inside the camera bay.
- Pre-processing of the images to improve the contrast.
- Extracting features that contain suitable information to discriminate images at the category level.
- Classification: Images classified in different categories as specified.

Bag of Visual Words (BoVW) proved to be the leading strategy in computer vision applications such as image retrieval and image categorization (Csurka et al., 2004); thus, it is being opted for the presented work. Categorizing digital images, embarks on extracting features and creating a visual vocabulary for the given dataset. It comprises of following states:

1. Extracting features.
2. Constructing visual vocabulary by clustering.
3. Using multi-class classifier for training using bags as feature vector.
4. For the testing image, obtain the nearest vocabulary based on the most optimum prediction of classifier.

However, in this study, several steps are integrated in the process to improve the overall performance compared to Csurka et al. (2004) described in Section 2.3. Our method treats canopy images acquired automatically in the field as a collection of unordered appearance descriptors extracted from local patches;

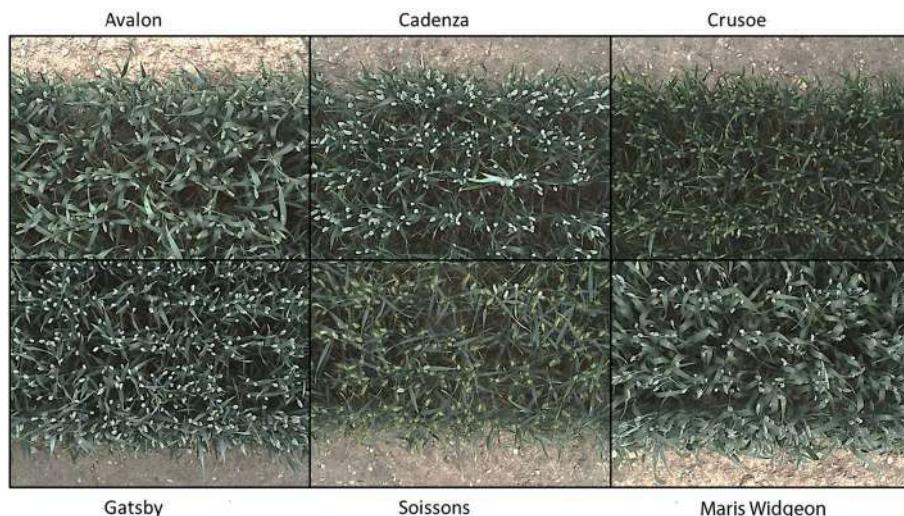
**Abbreviations:** BoVW, bag of visual words; SVM, support vector machine; SPM, spatial pyramid matching; RBF, radius basis kernel; LLC, local linear constraint; SIFT, scale invariant feature transform; DoG, difference-of-gaussian; LoG, laplacian-of-gaussian; PCA, principle component analysis; SURF, speeded up robust features; KNN, k-nearest-neighbors; UAV, unmanned aerial vehicle.



then, quantizes them into discrete visual words. Each image is defined by a feature vector listing the number of regions which belongs to each cluster and are later used to train a classifier. In addition, the location information is taken into account which is one of the important factors in object recognition scenarios. In the final step, a linear Support Vector Machine (SVM) classifier is used to determine pre-defined classes (e.g., ear emergence, flowering). The experimental results show that the introduced method is capable of automatically identifying key wheat growing stage with high accuracy and efficiency (Section 3).

## 2.1. Field Experiment and Image Acquisition

Six wheat cultivars (*Triticum aestivum* L. cv. Avalon, Cadenza, Crusoe, Gatsby, Soissons and Maris Widgeon) were grown in the field at Rothamsted Research, Harpenden, UK, sown in Autumn 2015 and maturing in 2016. These cultivars were selected as they had different properties visible to the naked-eye (awns/no awns, differing wax properties, straight/floppy leaves, different ear morphology) (Figure 2). All cultivars were sown 20 October 2015, at a planting density of 350 plants/m<sup>2</sup>. Nitrogen (N) treatments were applied as ammonium nitrate in the spring, at



**FIGURE 2 |** Digital images of the six contrasting wheat cultivars (*Triticum aestivum* L. cv. **Avalon**, **Cadenza**, **Crusoe**, **Gatsby**, **Soissons**, and **Maris Widgeon**) used, at growth stage Z5.9.

rates of 0 kg  $ha^{-1}$  (residual soil N; N1), 100 kg  $ha^{-1}$  (N2) and 200 kg  $ha^{-1}$  (N3) (Figure 4).

The Field Scanalyzer phenotyping platform (LemnaTec GmbH; Virlet et al., 2017) was used to acquire all images (Figure 3). The Field Scanalyzer is a fully-automated, high-throughput, fixed-field phenotyping platform, carrying multiple sensors for non-invasive monitoring of plant growth, morphology, physiology and health. The on-board visible camera (color 12 bit Prosilica GT3300) was used to acquire RGB images at high-resolution (3,296 × 2,472 pixels). The camera is positioned perpendicular to the ground, and automatically adjusts to ensure a 2.5 m distance is maintained between the camera and canopy. The camera is set up in auto-exposure mode, to compensate for outdoor light changes. Wheat canopies were imaged daily during three stages of ear emergence: Stage 1 (Zadoks scale Z5.0; 3–5 June 2016 Zadoks et al., 1974); Stage 2 (Z5.3–Z5.7; 7–10 June 2016) and Stage 3 ( $\geq$  Z5.9; 12–14 June 2016), as well as flowering stage (14–18 June 2016). In addition, illumination conditions were recorded during the image acquisition (Table 1). Manual growth stages were recorded daily or on alternating days during heading and flowering. The growth stage of the plot was defined manually by the stage of  $\geq 50\%$  of the plot. Videos and more information of the Field Scanalyzer platform can be accessed in our website: <http://www.rothamsted.ac.uk/field-scanalyzer>.

## 2.2. Image Pre-processing and Enhancement

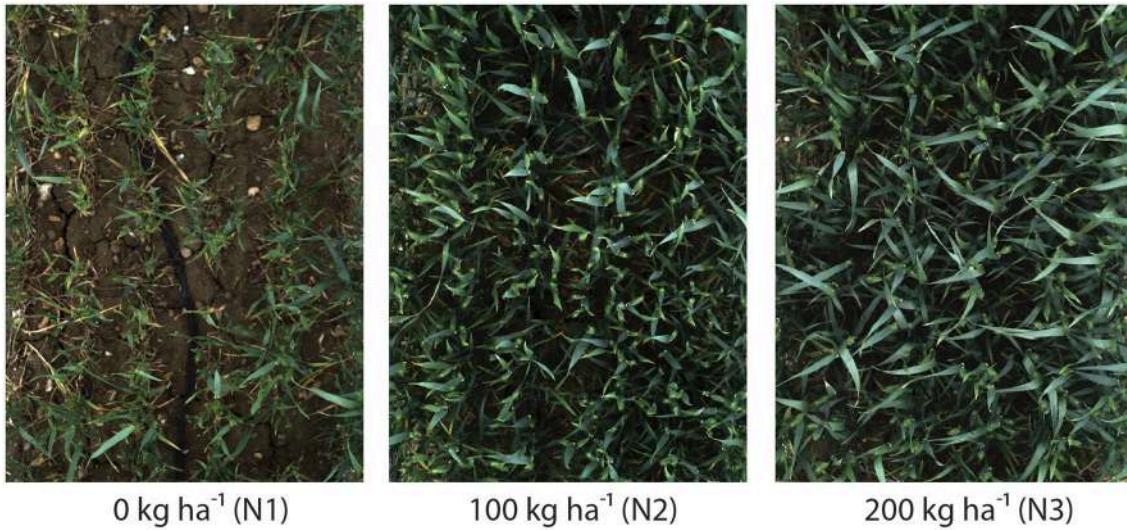
The color of ears at early development stages are very similar to leaves and hardly discernable with the naked-eye (Figures 5A,C). In order to make the ears stand out in canopies and discriminate them from the background more easily, a pre-processing method is applied on plot images before extracting features, known



**FIGURE 3 |** The Field Scanalyzer at Rothamsted research.

as decorrelation stretching (DS). The decorrelation stretching technique enhances the color differences and increasing the image contrast in each plot image by removing the inter-channel correlation found in the pixels (Gillespie et al., 1986). Therefore, it allows to see details such as ears that are otherwise too subtle for the naked-eye (Figures 5B,D). If the red, green, and blue values of pixels are treated coordinates in space, decorrelation stretch moves these points in space further apart, so they become much easier to see a difference between them.

The DS among the RGB channels is achieved through principle component analysis (PCA) to remove inter-channel correlation in an image. The application of PCA to the digital analysis of an image is based on first, calculating the covariance matrix between the three RGB bands. Then, obtaining eigenvectors and eigenvalues. Finally, rotating the original image vector to a new space by multiplying it by the eigenvectors



**FIGURE 4 | Digital images highlighting the impact of  $0 \text{ kg ha}^{-1}$  (N1),  $100 \text{ kg ha}^{-1}$  (N2) or  $200 \text{ kg ha}^{-1}$  (N3) nitrogen fertilizer application on canopy complexity. Images were acquired 2.5 m above *Triticum aestivum* L. cv. Soissons canopies.**

**TABLE 1 | Date, start/end time, and PAR values of each images acquisition periods during ears emergence and flowering stages.**

Date	Start	End	PAR ( $\mu\text{mol.m}^{-2.s}^{-1}$ )
03/06/2016	11:19:16	12:08:24	$404 \pm 58$
04/06/2016	11:53:11	12:42:41	$512 \pm 67$
05/06/2016	08:29:34	09:18:38	$287 \pm 43$
07/06/2016	13:33:40	14:25:07	$1,037 \pm 95$
08/06/2016	08:07:05	08:58:46	$315 \pm 19$
08/06/2016	18:05:53	18:38:33	$528 \pm 180$
09/06/2016	08:32:25	09:24:05	$530 \pm 215$
10/06/2016	07:43:46	08:37:33	$461 \pm 33$
12/06/2016	14:32:33	15:26:18	$703 \pm 304$
13/06/2016	09:49:04	10:41:12	$555 \pm 110$
14/06/2016	10:14:16	11:06:10	$800 \pm 196$
14/06/2016	15:01:19	15:33:55	$569 \pm 121$
16/06/2016	08:02:55	08:35:57	$363 \pm 31$
18/06/2016	10:51:31	11:44:17	$919 \pm 238$

PAR mean and standard deviation values are computed from the 54 scans collected during one acquisition periods.

(Equation 1) (Jolliffe, 2002; Cerrillo-Cuenca and Sepúlveda, 2015).

$$p_n = R^T i_n \quad (1)$$

where  $i_n$  is the image vector;  $n$  is the number of pixels; and  $R$  is the rotation matrix.

Campbell (1996) proposed a general framework consists of the following steps:

- (i) Calculating  $p_n$  from Equation (1), eigenvalues and eigenvectors are obtained from the correlation matrix or alternatively from the covariance matrix.

- (ii) Generating a stretch vector: diagonalize the covariance matrix composed by the inverse of the eigenvectors:

$$D = \begin{bmatrix} \frac{1}{\sqrt{\nu_1}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{\nu_2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{\nu_3}} \end{bmatrix} \quad (2)$$

where  $D$  is a diagonal matrix;  $\nu$  denotes each of the eigenvalues. Alternatively,  $D$  can be multiplied by an integer value that serves to achieve a higher contrast in the image (Alley, 1996). Finally, the resultant matrix is applied to  $p_n$  (Equation 3). At this step, the matrix is re-centered and stretched its values to a maximum.

$$w_n = D p_n \quad (3)$$

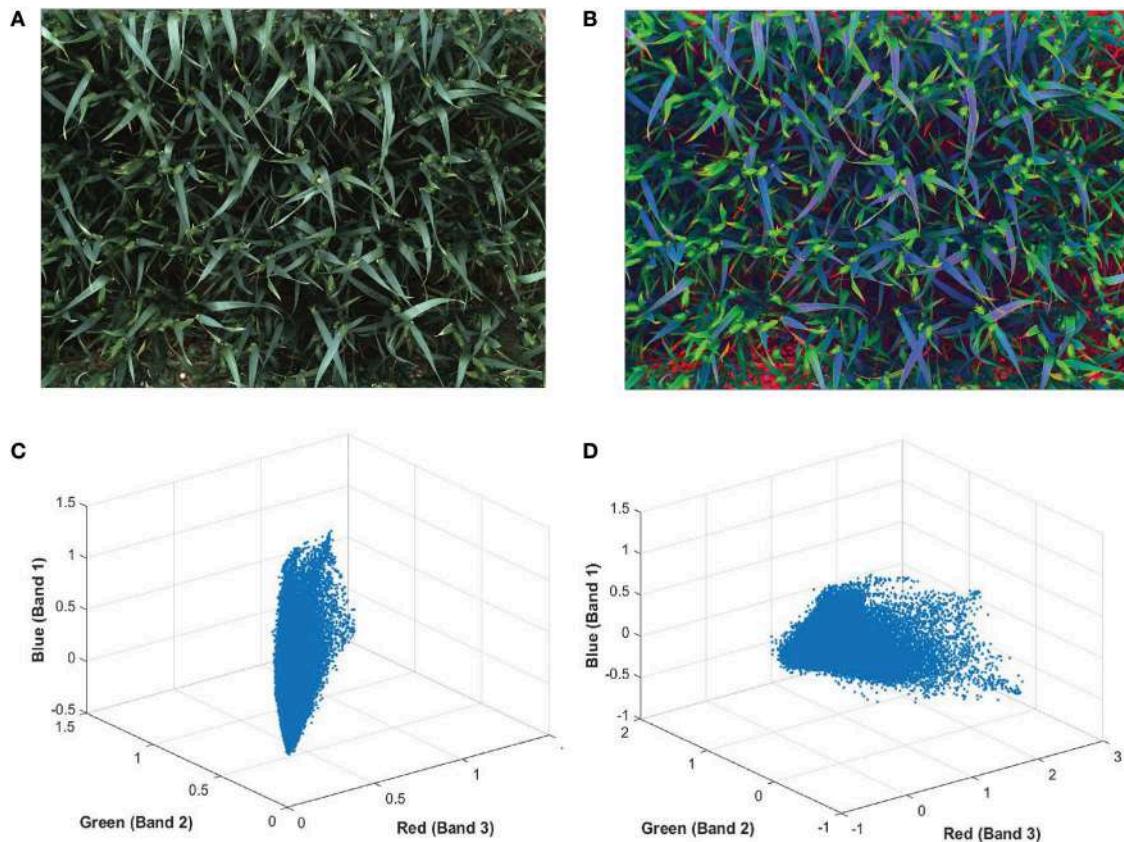
- (iii) The inverse transform is applied to map the colors back to the original space. The information is decorrelated into a new vector  $c_n$  composed of three matrices (RGB) (Equation 4)

$$c_n = R w_n = R D R^T i_n \quad (4)$$

- (iv) Finally, a standard deviation value is applied to visually increase the contrast (Alley, 1996).

### 2.3. Bag of Visual Words Construction

The first step of BoVW framework corresponds to feature extraction. Fixed length feature extraction techniques based on color (Swain and Ballard, 1991; Chen et al., 2010), texture (Duda et al., 2000), shape (Mehrotra and Gary, 1995), or a combination of two or more techniques, extract pixel values of an image only. These are excellent in comparing the overall image similarity (Angelov and Sadeghi-Tehran, 2016); however, they are not scale or rotation invariant. Moreover, they are very sensitive to noise



**FIGURE 5 |** Digital image of wheat (*Triticum aestivum* L. cv. Soissons) canopy (A) before and (B) after enhancement of image contrast and application of decorrelation stretching. Scatterplot of every pixels normalized red, green blue (RGB) values from (C) the original image and (D) after applying the decorrelation stretch and contrast increase.

and illumination changes; thus, are unable to describe the object-based properties of the image content.

As opposed to global feature extraction methods mentioned earlier, local extraction algorithms are robust to partial visibility and clutter. It is an ideal candidate for object recognition, template matching and image mosaicing. There are several feature detector methods, which are scale and rotation invariant. They are also robust enough to handle illumination changes and resistant to geometry (Bay et al., 2006; Leutenegger et al., 2011; Alahi et al., 2012). Among the proposed descriptors, Scale Invariant Feature Transform (SIFT) is selected due to its excellent performance attested in various applications (Mikolajczyk and Schmid, 2005). It returns an  $N \times 128$  dimension image descriptor, where  $N$  is the number of features.

SIFT consists of Lowe (2004):

- **Constructing a scale space:** in this stage, location and scales of each keypoint are identified. Laplacian-of-Gaussian (LoG) is calculated for an image with various  $\sigma$ . Due to change in  $\sigma$ , LoG detects blobs of various sizes, then the local maxima can be found across the scale and space with a list of  $(x, y, \sigma)$  values, which show there is a candidate keypoint at location  $x, y$  with scale of  $\sigma$ . However, in order to reduce the computational

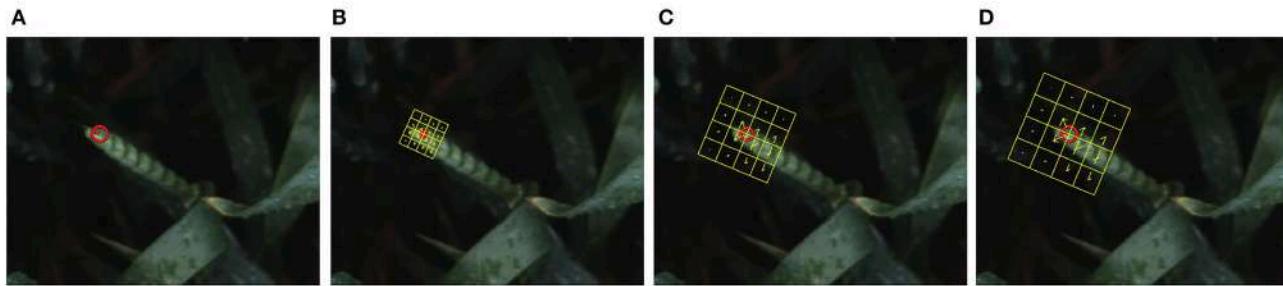
complexity, SIFT uses Difference-of-Gaussian (DoG) which is a convolved image in scale space separated by a constant factor  $k$ :

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (5)$$

where  $I(x, y)$  is an input image;  $L(x, y, k\sigma)$  is the scale space of an image;  $G(x, y, k\sigma)$  is variable-scale Gaussian.

$D$  is computed by simple image subtraction and the Gaussian image is sub-sampled by a factor of 2 and produces DoG for the sampled image. Once the DoG is computed, images are searched for local extrema over space and scale. For instance, one pixel is compared with its  $n \times n$  neighborhood ( $n = 3$  in our experiment) as well as 9 pixels in the next scale and 9 pixels in previous scales (Lowe, 2004).

- **Keypoint localization and filtering:** Once the location of keypoints candidates are found, they are refined and some are eliminated to get a more accurate location of extrema. For instance, if the intensity at the extrema is less than a certain threshold (threshold  $< 0.03$ ) it is rejected. In addition, edges



**FIGURE 6 | (A)** A single keypoint candidate in the image; **(B–D)** SIFT descriptor calculated at different scales of 4, 8, and 10; At each scale, the descriptor has  $4 \times 4$  patches (color coded in yellow), which are rotated to the dominant orientation of the feature point. Each patch is represented in gradient magnitudes of eight directions, represented by yellow arrows inside each bin.

and low contrast regions are considered as bad keypoints and will be rejected.

- **Orientation assignment:** The orientation of each keypoint is obtained based on image gradient and local image gradient directions to achieve rotation invariance. Depending on the scale a neighborhood is taken around the keypoint location and the gradient magnitude and direction is calculated in that region.
- **Keypoint descriptor:** In order to generate a keypoint descriptor, the local image descriptor is computed for each keypoint based on image gradient magnitude and orientation at each image sample point in a region centered at keypoint. These samples build a 3D histogram of gradient location and orientation; with a  $4 \times 4$  array location grid and 8 orientation bins in each sample, which creates 128 element dimensions of the keypoint descriptor, causing robustness against changes in scale and rotation (**Figure 6**).

The next step is to form clusters of similar features and assign them as visual words. The objective of constructing codebook is to relate features of testing images to the features previously extracted from the training image samples (**Figure 7**). Although in the field of unsupervised learning, clustering is a standard procedure, there is no single clustering algorithm that can be applied uniformly to all the application domains or address all related issues in a satisfactory manner. Here, a partition-based clustering approach known as  $K$ -means clustering is used to quantize each descriptor and generate a codebook. The process is iterative as follows (Lloyd, 1982):

#### Algorithm 1 $K$ -means clustering procedure

- 1: Select  $K$  points as initial centers
- 2: **repeat**
- 3: Assign each input data to its closest center
- 4: Re-compute the center of each cluster by averaging all the members in the clusters
- 5: **until**
- 6: convergence which means no pixel shifts from one cluster to another; centers do not change

In  $K$ -means the number of clusters is pre-defined beforehand and it should be large enough to identify relevant changes in each wheat cultivars. For an image having  $N$  features, the model will distribute the features with  $K$  clusters, which is the size of the visual vocabulary. We have been able to find the optimum numbers and get very good results with number of vocabulary (codebook)  $K = 2000$  (**Table 2**).

The codebook is used for quantizing features. A vector quantizer takes a feature vector and maps it to the index of the nearest code vector in a codebook. In our work, in order to project the descriptors onto the codebook elements, Local Linear Constraint (LLC) (Wang et al., 2010) is used to generate a final vector which represents an image. LLC reduces the computational complexity to  $O(K + K)$  (where  $K$  is the length of the codebook;  $K = 2000$  in this case) for each descriptor and can achieve acceptable image classification accuracy even with a linear SVM classifier (Wang et al., 2010).

The main drawback of BoVW is that it is unable to capture spatial relationships between images. In order to preserve the spatial relations of the code vector Spatial Pyramid Matching (SPM) is implemented where the entire image is divided into levels. Each image is divided into spatial sub-regions and computes histograms of features from each sub-region. Each level divides the image into  $2^l \times 2^{l-1}$ ; where  $l$  is level (Grauman and Darrell, 2005; Lazebnik et al., 2006). The features are computed locally for each grid and the spatial information is incorporated into histograms. A three level SPM is used with first, level 0 which comprises of a single histogram; level 1, comprising of 4 histograms, finally level 2, comprising of 16 histograms (**Figure 7**). The histogram from all the sub-region are concatenated together to generate the final representation of the image for classification. The result is a feature weighted histogram of  $21 \times K$  (number of words = 2000). Using such method will preserve the discriminative power of the descriptors; in addition, changes in the positioning of the objects and variations in the background will not affect the overall performance of the method.

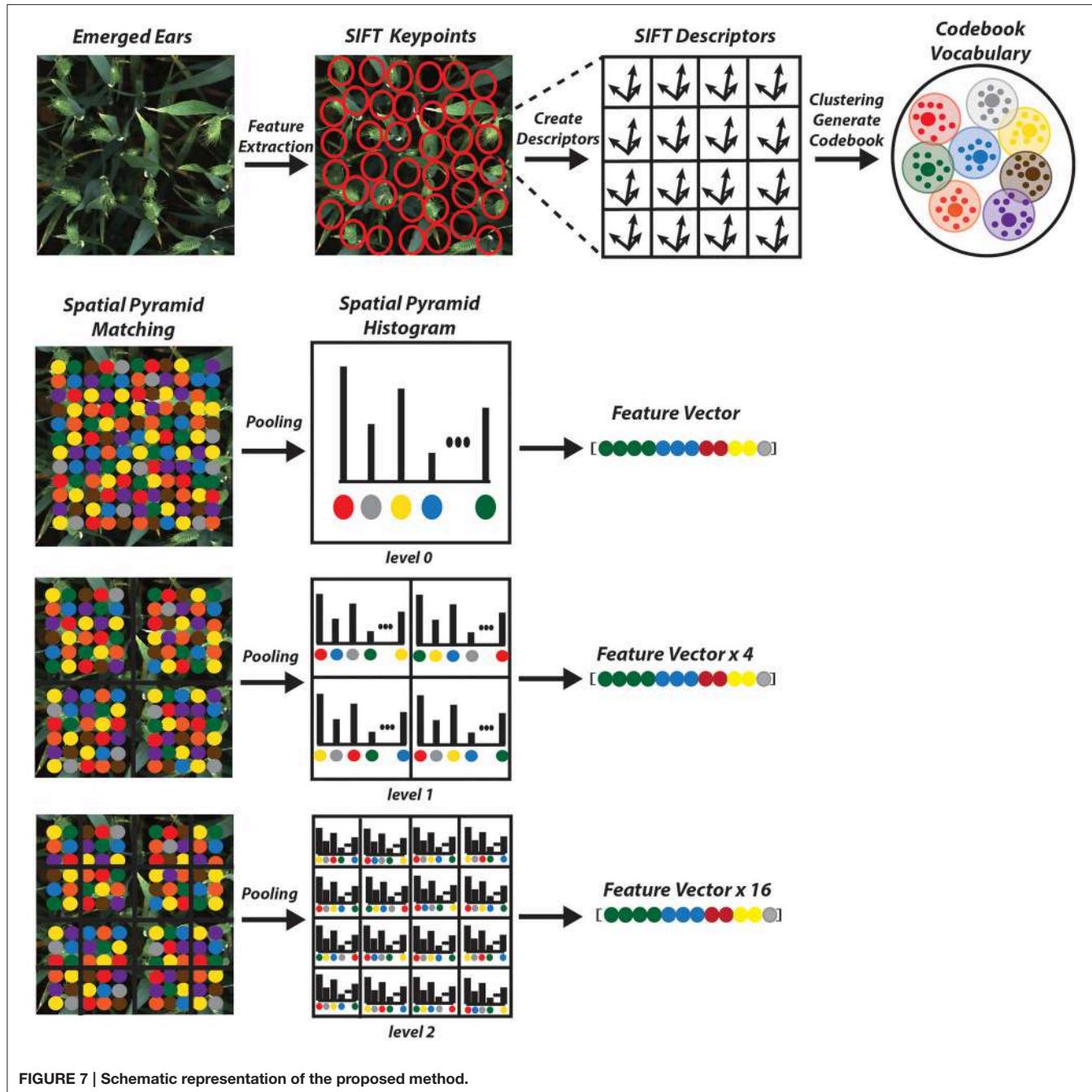


FIGURE 7 | Schematic representation of the proposed method.

## 2.4. Learning Model

The construction of the model for our image annotation is based on the supervised machine learning principle. Supervised learning can be thought as learning by examples represented by a set of training-testing samples. In order to classifying unknown testing images, a certain number of training images are used for each class to train the classifier. A classifier approximates the mapping between the images and correctly labels the training set, called the *training* phase. After the model is trained, it is able to classify unknown image, into one of the learned class labels.

In our model, the complexity of visual categorization is reduced to two-class with positive and negative training patches. The SVM classifier is used as our classifier of choice as it is fast and can handle the long feature vectors generated by the SPM. During the training phase, labeled images (ears and background) are fed to the classifier and used to adapt a statistical decision procedure. Among many available classifiers, linear SVM with Hellinger kernel is used to predict the unlabeled test images and retrieve as much of the data as possible in a high ranked position. Feature vectors generated from each image are normalized to a

**TABLE 2 | Comparison of different methods applied on the three heading stages.**

Decorrelation pre-processing	Feature extraction	Coding method	Spatial pyramid	Vocabulary length	Accuracy Z5.0 (%)	Accuracy Z5.3–Z5.7 (%)	Accuracy $\geq$ Z5.9 (%)
Yes	SIFT	LLC	Yes	2,000	95.24	97.79	99.59
No	SIFT	LLC	Yes	2,000	57.29	82.20	85.38
Yes	SIFT	k-NN	Yes	2,000	90.54	94.48	96.97
Yes	SIFT	LLC	No	2,000	92.90	96.54	96.90
Yes	SURF	LLC	Yes	2,000	56.84	71.55	78.45
Yes	SIFT	LLC	Yes	1,000	93.91	97.52	98.91
Yes	SIFT	LLC	Yes	1,500	94.61	97.64	99.24
Yes	SIFT	LLC	Yes	2,500	94.90	97.59	99.49

unit Euclidean norm and used for a linear SVM classifier with the Hellinger kernel to compute the feature map (Vedaldi and Zisserman, 2012).

$$K(n, n') = \sum_{m=1}^d \sqrt{n_m n'_m}; n = [n_1, \dots, n_d]; n' = [n'_1, \dots, n'_d] \quad (6)$$

where  $n$  and  $n'$  are normalized histograms;  $d = 42,000$

One-vs.-all strategy is chosen to train the SVM. Two classes are trained, each labels the sample inside one class as +1 and other samples (background) as -1. The SVM calculates the similarity of all trained classes and assigned the test image to the class with the highest similarity measure.

### 3. EXPERIMENTAL RESULTS AND DISCUSSION

The experiment is divided into two sections of identifying ear emergence and flowering stages from the digital images acquired in the field. In the first section, ear emergence was tested at different time points, from early stages where only few spikelets are visible, to a more advanced stage where ears are fully emerged (Section 3.1). In the second part of the experiment, the method was tested to identify flowering growth stage during anthesis (Section 3.2). The training dataset for the ear emergence experiment includes images with ears at different emergence stages (positive class) and leaves, soil, etc. (negative class), which are manually cropped and stored in the dataset. On the other hand, the training dataset for the flowering experiment contains ears at different flowering time points (positive class) and ears before and after flowering (negative class). The collected dataset focuses on different challenges regardless of light conditions in the field and to demonstrate the robustness of the method to environmental changes. In addition, the versatility of the proposed technique were also tested by minimizing the number of cultivars as training patches, and evaluating the method on more varieties.

The research was conducted with the following specifications. System comprised of 24 GB RAM, Intel quad core Processor (3.40 GHz) with Windows 10 OS. The models have been developed in MATLAB (Mathworks Inc.); however, to improve the processing

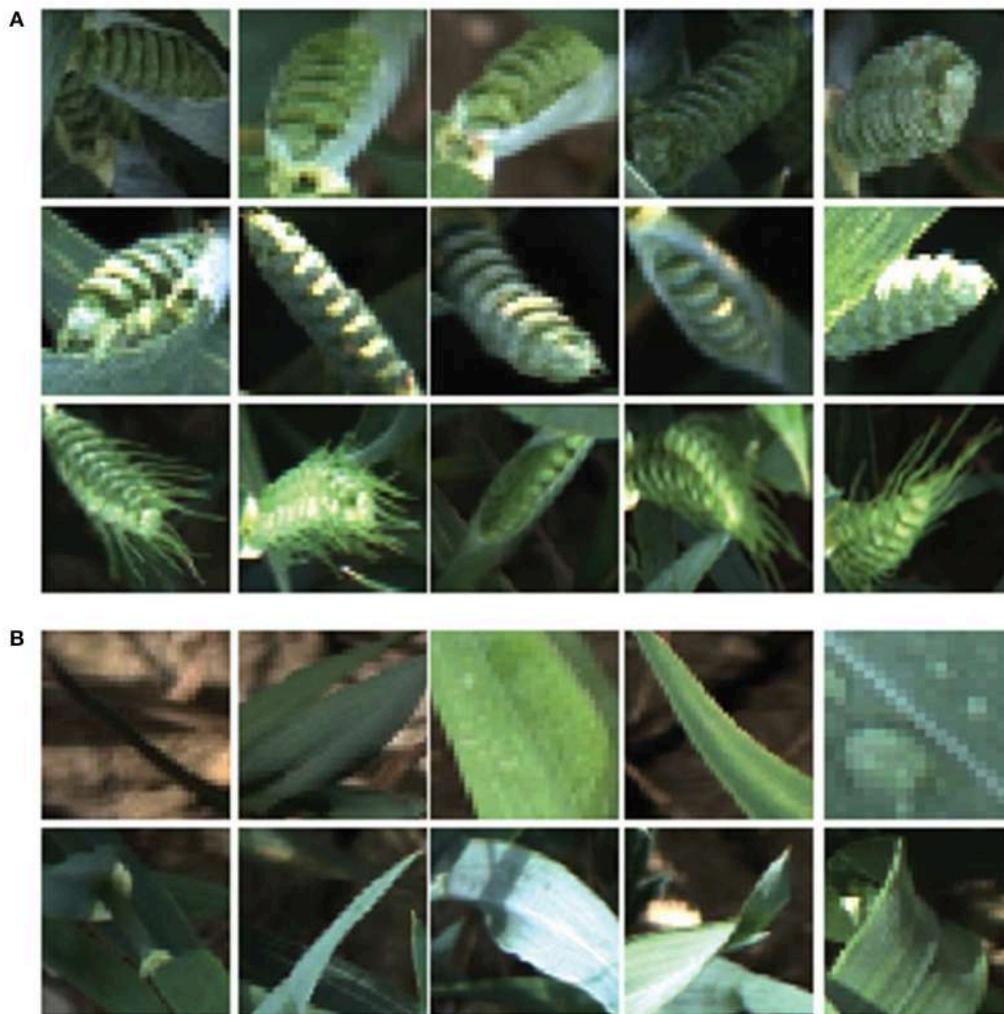
time, some of the algorithm, such as SIFT were written in C++ programming language. Utilities like VLFeat library (Vedaldi and Fulkerson, 2010) to extract features as well as LibLinear library (Fan et al., 2008) to train and test the SVM classifier. Using the above configured computer system, extracting features and generating code vectors from each training image approximately takes 0.45 s. However, the processing time increases to 5.4 s for each testing patch with resolution of  $3,298 \times 2,474$  pixels.

Precision (Pr) and Recall (Re) are the most commonly used measurements to evaluate the performance of image retrieval systems. Thus, it is used in our experiment to quantitatively assess the precision of the proposed approach in detecting the two main growing stages of ear emergence and flowering. Precision is defined as the ratio of the number of retrieved relevant images  $N_r$  to the total number of retrieved images  $N$  (Equation 7); on the other hand, Recall is defined as the number of retrieved relevant images  $N_r$  over the total number of positive images  $N_t$  available in the database. In an ideal scenario, both Pr and Re should have high values (1). Therefore, instead of using Pr and Re individually, usually accuracy curve is used to characterize the performance of the retrieval system.

$$\text{Pr} = \frac{N_r}{N}; \text{Re} = \frac{N_r}{N_t} \quad (7)$$

#### 3.1. Ear Emergence

The learning process starts with 1,000 training image samples divided into 500 ears (positive class), which are manually cropped from full size canopy image and 500 background images (negative class). **Figure 8** shows image samples randomly selected from training patches which are not necessarily the same dimensions. Moreover, to observe the field challenges during data acquisition, ears are selected from different positions and illumination conditions (with or without occlusions and overlapping; sunny or cloudy days). Three different wheat cultivars are used as a training dataset including Avalon, Cadenza, and Soissons. Cadenza can present short awnlettes/scurs at the ear tip, although most of the times no awns are present in contrast to Soissons which is an awned variety. Although three wheat cultivars were used as a training dataset, six cultivars including Maris Widgeon, Avalon, and Gatsby are tested to highlight the versatility of the proposed technique.



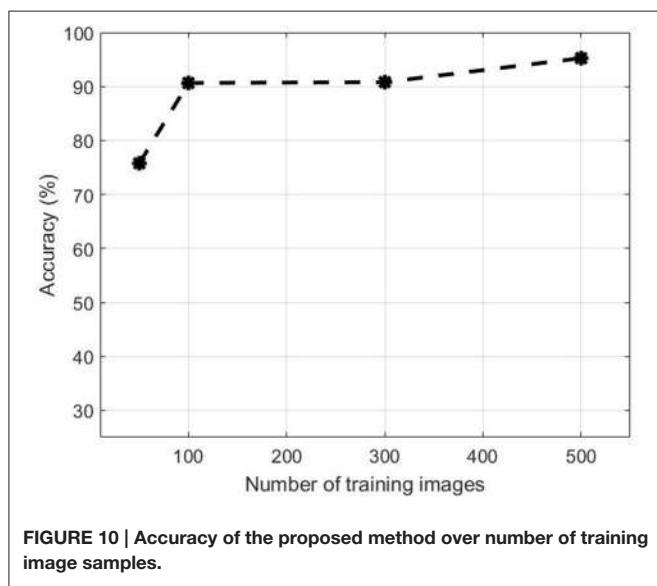
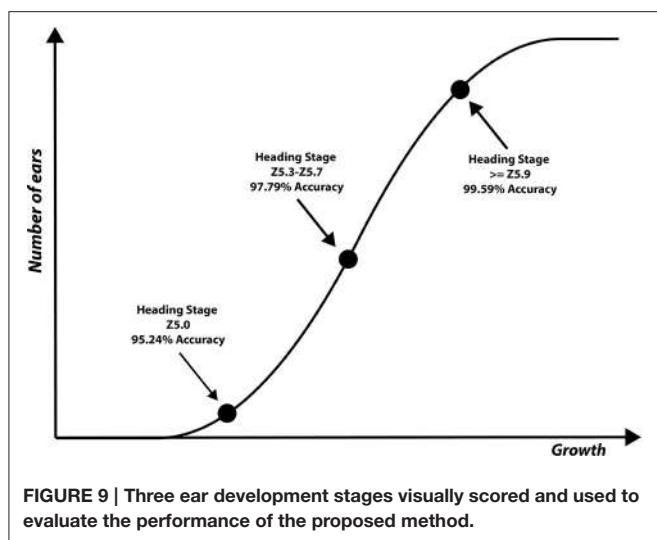
**FIGURE 8 | Example of ear emergence training patches.** Note that the training patches are not necessarily of equal size and resized for illustration purposes. **(A)** Examples of positive training patches of three different wheat cultivars (*Triticum aestivum* L. cv. Soissons, Avalon, and Cadenza). **(B)** Examples of negative training image patches.

Ear identification was evaluated at three different time points of the emergence period, (i) at Z5.0, when the ears start to be visible (first spikelet of inflorescence visible), (ii) between Z5.3–Z5.7, when 1/4 to 3/4 of the ears are emerged and (iii) at Z $\geq$ 5.9, when ears are fully emerged (Figure 8). Each time point was tested independently from datasets containing 80 images (40 with ears present and 40 without) of full size wheat canopies with the original resolution of 3,298  $\times$  2,474 pixels.

The results for each ear development stage are shown in Table 2. The accuracy of the method is evaluated using different techniques at different processing stages. (i) presence/absence of decorrelation processing, (ii) SIFT vs. SURF, (iii) LLC vs. KNN, (iv) presence/absence of spatial pyramid and (v) the vocabulary length. As shown in the Tables, the best performance was obtained using decorrelation pre-processing, SIFT, LLC coding, and a 2,000 entry codebook. The best performance at heading stage Z5.0 is 95.24%, and for heading stages Z5.3–Z5.7 and  $\geq$

Z5.9 are 97.79 and 99.59%, respectively (Figure 9). Out of the eight tested scenarios, we achieved accuracy of  $>$  90% at Z5.0 and  $>$  96% at Z5.9 in six scenarios. The impact of codebook size on the performance of the method was also investigated. It is clearly shown that the increasing number of codebook improves the accuracy; however, the accuracy plateaus at 2,000 visual words. Moreover, the low-level feature extraction and the decorrelation pre-processing technique has the biggest influence in the quality of results; especially in the early heading (Z5.0). The main conclusion is that mid-level feature coding and classification are highly impacted by the low level pre-processing and feature extraction techniques.

Figure 10 illustrates the performance evolution of the heading stage Z5.0 over the number of images in the training dataset, for both positive and negative data. Training patches of 50, 100, 300 were selected randomly apart from the full set when all 500 samples were used. The accuracy improves by increasing the



number of training samples. The accuracy increased from 75.77 to 90.65% when the training dataset increased from 50 to 100. On the other hand, there was no substantial change in accuracy between 100 and 200 samples. However, the performance jumped by more than 5% from 90.80 to 95.24% when the dataset increased to 500.

### 3.2. Flowering Time

Similarly to ear emergence identification, two training classes were created, which comprised of three wheat cultivars (Soissons, Maris Widgeon, and Cadenza). The first class (positive class) contained 140 manually cropped images at flowering stage while the second class (negative class) contained the same number of images as the positive class, but with ears before and after flowering.

**Figure 11** shows randomly selected samples from the training patches. All training images were collected without considering the environmental changes and positioning or occlusion. As

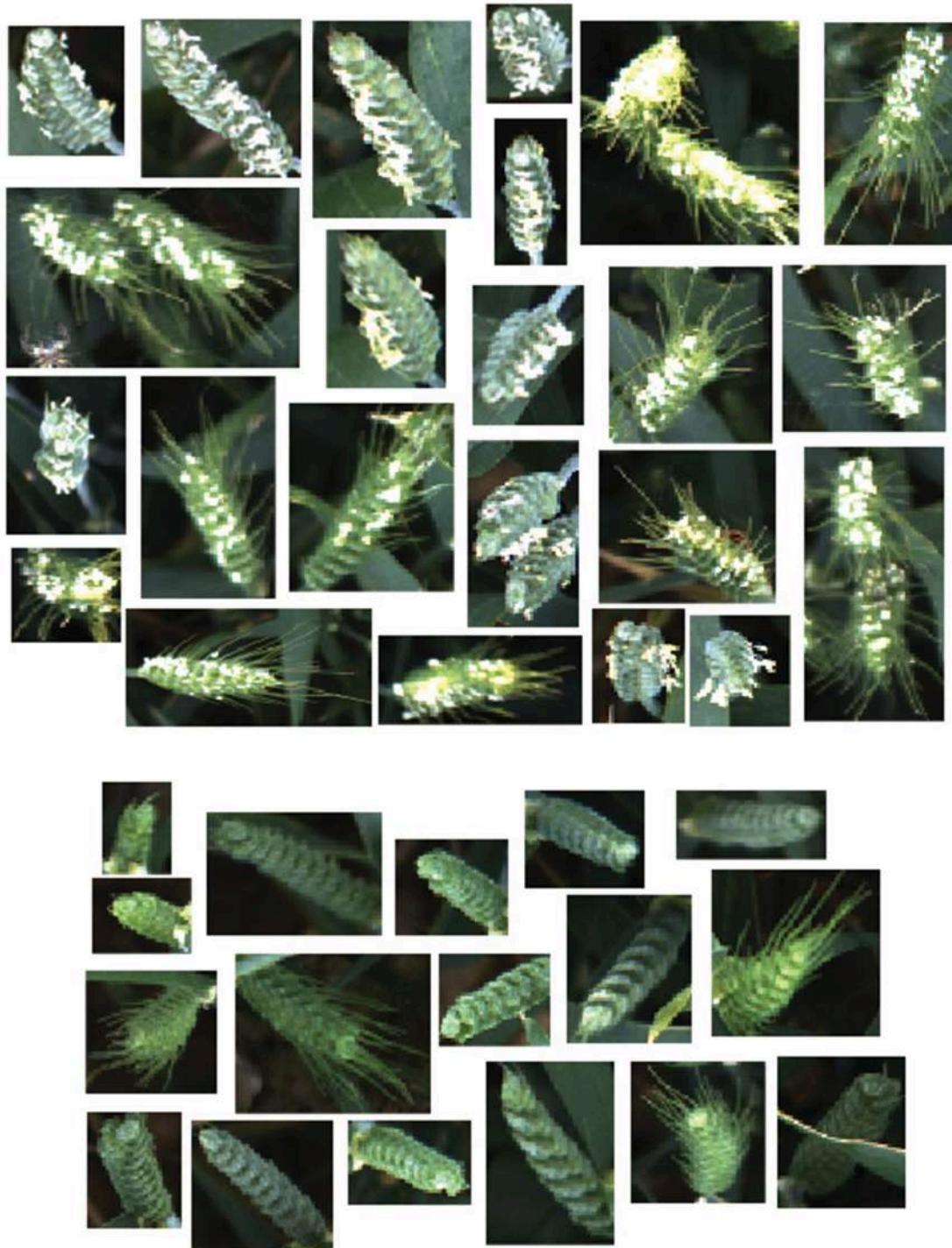
flowering development may be completed in only a few days, the beginning or intermediate stages can be easily missed. Therefore, all flowering images along the flowering duration were included. For the testing dataset, 108 full size canopy images were used with the original resolution, which includes 54 canopies with ears during flowering stage and 54 canopies with ears before or after flowering stage. The method selected to test the flowering stage was the one which produced the best result in the ear emergence experiment (decorrelation stretching, SIFT, LLC, and SPM algorithms with the vocabulary length of 2,000). The method was tested on each cultivar separately, as well as all three together. For all three cultivars, 38 images out of 54 images were retrieved correctly, which shows 82.54% accuracy. On the other hand, the accuracy when testing Soissons, Cadenza, and Widgeon individually was 76.72, 92.91, and 80.33%, respectively (**Table 3**).

### 3.3. Discussion

To the best of our knowledge, few efforts have been made to automate the detection of crop growth stage (Thorp and Dierig, 2011; Yu et al., 2013; Guo et al., 2015; Zhu et al., 2016). Furthermore, the published methods have only been applied to small sections of the crops and generally tested only on a single cultivar. Unlike alternative methods, such as Yu et al. (2013), which used color properties to determine growth stages of maize, our approach uses rich feature collection techniques, such as SIFT, which carry suitable information to discriminate images at the category level on the canopy scale. The technique used by Guo et al. (2015) was only tested on two rice varieties individually at flowering stage and obtained just over 80% accuracy. However, our method integrated statistical variables, such as vector coding and spatial pyramid matching, which improved the accuracy and general versatility of the growth stage identification. On the other hand, their training system contained only flowering rice as the positive class and leaves as the negative class; failing to define rice before and after the flowering stage. This may have likely made their dataset more challenging because more variables would be added to the training dataset and distinguishing between non-flowering and flowering panicles would have added difficulty, potentially detecting false positives, ultimately reducing the accuracy of their method.

In our case, the accuracy of flowering detection is less than heading. This could be due to the size and color of anthers. The color of anthers can range from yellow to white depending on the cultivar, and the pale color of the anther has increased the sensitivity to over/under exposure as a result of changes in ambient illumination. Moreover, anthers are far smaller objects compared to wheat ears and are prone to noise, adding difficulty to detection them accurately. Nevertheless, the proposed method yielded greater accuracy than the existing method (Guo et al., 2015).

Pre-processing is also an important factor in our method. Newly emerging ears are difficult to distinguish as they are nearly the same color as the canopy, making methods based on color features inadequate for this purpose. However, the use of color enhancement methods, such as decorrelation stretching, yields higher accuracy. In our case, the absence of decorrelation stretching, results a decrease in accuracy from 95.24% to 57.29%



**FIGURE 11 | Example of flowering training patches.** Examples of positive training patches of three different cultivars; (*Triticum aestivum* L. cv. Soissons, Maris Widgeon, and Cadenza), which contain flowering ears. Examples of background training patches which do not contain flowering ears.

and from 99.59 to 85.38% at earliest and latest stage of heading, respectively. Moreover, applying decorrelation stretching as a color enhancement tool early in the process minimize various ambient light conditions. The other important factor is the

low level feature extraction in the BoVW process. SIFT was replaced by SURF as an alternative technique; however, although SURF performs faster as a result of using integral images and Hessian Matrix (Bay et al., 2006); SIFT still outperformed

**TABLE 3 | Comparison of flowering accuracy between three wheat cultivars.**

Wheat cultivar	No. training images	No. testing images	Accuracy (%)
Cadenza	410	16	92.91
Soissons	410	23	76.72
Maris widgeon	410	15	80.33
All three	410	54	82.54

SURF (**Table 2**) in our experiment. It has also been examined that SIFT showed more stability on blurry images and more robust to rotation and scale invariants (Mikolajczyk and Schmid, 2005).

It should also be highlighted that the quality of the training dataset plays an important role in the overall performance. We aimed to define more scenarios for the system (e.g., ears at different positions, scales, and illumination conditions in the field, etc.). As shown in **Figure 11**, the accuracy of the ear emergence detection would increase by adding more training data. We would expect to improve the accuracy of the flowering experiment, by collecting data more frequently during the flowering period and increasing the size of the training dataset.

## 4. CONCLUSION

We proposed an automated observing system using computer vision to determine two key growth stages in wheat: ear

emergence and flowering time. The proposed method is capable of distinguishing the critical growth stages from the RGB images taken in the field. The approach demonstrated a high performance for identifying such development changes and was not affected by the environmental conditions or illumination invariants in the field.

In future work, we aim to test our proposed method on additional wheat genetic material and other species, and in addition, to investigate the effect of alternative computer vision techniques from features extraction to classification on the performance and overall accuracy. Finally, we aim to apply the proposed method on images acquired by Unmanned Aerial Vehicles (UAVs) to monitor large fields efficiently and believe it will dramatically accelerate the recording of such development stages.

## AUTHOR CONTRIBUTIONS

PS proposed, developed, and tested the method; KS and NV planned and conducted the experiment; MH contributed to the revision of the manuscript and supervised the experiment; PS, KS, and NV contributed to writing the manuscript; all authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

Rothamsted Research receives support from the Biotechnology and Biological Sciences Research Council (BBSRC) of the UK as part of the 20:20 Wheat® project.

## REFERENCES

- Alahi, A., Ortiz, R., and Vandergheynst, P. (2012). *FREAK: Fast Retina Keypoint*, 2012 IEEE Conference on Computer Vision and Pattern Recognition (Providence, RI: Providence Rhode Island Convention Center).
- Alley, R. (1996). *Algorithm Theoretical Basis Document For: Decorrelation Stretch*. Pasadena, CA: Jet Propulsion Laboratory.
- Angelov, P., and Sadeghi-Tehran, P. (2016). Look-a-Like: a fast content-based image retrieval approach using a hierarchically nested dynamically evolving image clouds and recursive local data density. *Int. J. Intell. Syst.* 32, 82–103. doi: 10.1002/int.21837
- Bay, H., Tuytelaars, T., and Van Gool, L. (2006). “SURF: Speeded Up Robust Features,” in *Computer Vision – ECCV 2006: 9th European Conference on Computer Vision Graz, Austria, May 7–13, 2006. Proceedings, Part I*, eds A. Leonardis, H. Bischof, and A. Pinz (Berlin, Heidelberg: Springer Berlin Heidelberg), 404–417.
- Braun, H. J., Atlin, G., and Payne, T. (2010). “Multi-location testing as a tool to identify plant response to global climate change,” in *Climate Change and Crop Production*, ed M. P. Reynolds. doi: 10.1079/9781845936334.0115
- Campbell, N. (1996). The decorrelation stretch transformation. *Int. J. Remote Sens.* 17, 1939–1949. doi: 10.1080/01431169608948749
- Cerrillo-Cuenca, E., and Sepúlveda, M. (2015). An assessment of methods for the digital enhancement of rock paintings: the rock art from the precordillera of Arica (Chile) as a case study. *J. Archaeol. Sci.* 55, 197–208. doi: 10.1016/j.jas.2015.01.006
- Chen, W.-T., Liu, W.-C., and Chen, M.-S. (2010). Adaptive color feature extraction based on image color distributions. *IEEE Trans. Image Process.* 19, 2005–2016. doi: 10.1109/TIP.2010.2051753
- Csurka, G., Dance, C., Fan, L., Willamowski, J., and Bray, C. (2004). *Visual Categorization with Bags of keypoints*, Workshop on Statistical Learning in Computer Vision, ECCV. Prague: Czech Technical University.
- Duda, R. O., Hart, P. E., and Stork, D. G. (2000). *Pattern Classification*, 2nd Edn., Wiley-Interscience.
- Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. (2008). LIBLINEAR: a library for large linear classification. *J. Mach. Learn. Res.* 9, 1871–1874.
- Gillespie, A. R., Kahle, A. B., and Walker, R. E. (1986). Color enhancement of highly correlated images. I. Decorrelation and HSI contrast stretches. *Remote Sens. Environ.* 20, 209–235. doi: 10.1016/0034-4257(86)90044-1
- Grauman, K., and Darrell, T. (2005). “The pyramid match kernel: discriminative classification with sets of image features,” in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. (Beijing).
- Guerrero, J. M., Pajares, G., Montalvo, M., Romeo, J., and Guijarro, M. (2012). Support vector machines for crop/weeds identification in maize fields. *Exp. Syst. Appl.* 39, 11149–11155. doi: 10.1016/j.eswa.2012.03.040
- Guo, W., Fukatsu, T., and Ninomiya, S. (2015). Automated characterization of flowering dynamics in rice using field-acquired time-series RGB images. *Plant Methods* 11:7. doi: 10.1186/s13007-015-0047-9
- Jolliffe, I. T. (2002). “Principal component analysis and factor analysis,” in *Principal Component Analysis* (New York, NY: Springer), 150–166.
- Lazebnik, S., Schmid, C., and Ponce, J. (2006). *Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories*, 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (New York, NY).
- Leutenegger, S., Chli, M., and Siegwart, R. Y. (2011). *BRISK: Binary Robust invariant scalable keypoints*, 2011 International Conference on Computer Vision (Barcelona).

- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Trans. Inform. Theory* 28, 129–137. doi: 10.1109/TIT.1982.1056489
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60, 91–110. doi: 10.1023/B:VISI.0000029664.99615.94
- Mehrotra, R., and Gary, J. E. (1995). Similar-shape retrieval in shape data management. *Computer* 28, 57–62. doi: 10.1109/2.410154
- Mikolajczyk, K., and Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Trans. Patt. Anal. Mach. Intell.* 27, 1615–1630. doi: 10.1109/TPAMI.2005.188
- Pourreza, A., Lee, W. S., Etxeberria, E., and Banerjee, A. (2015). An evaluation of a vision-based sensor performance in huanglongbing disease identification. *Biosyst. Eng.* 130, 13–22. doi: 10.1016/j.biosystemseng.2014.11.013
- Snape, J., Butterworth, K., Whitechurch, E., and Worland, A. J. (2001). “Waiting for fine times: genetics of flowering time in wheat,” in *Wheat in a Global Environment: Proceedings of the 6th International Wheat Conference, Budapest, Hungary* eds Z. Bedö and L. Láng (Dordrecht: Springer), 67–74.
- Swain, M. Z., and Ballard, D. H. (1991). Color indexing. *Int. J. Comput. Vis.* 7, 11–32. doi: 10.1007/BF00130487
- Tester, M., and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822. doi: 10.1126/science.1183700
- Thorp, K. R., and Dierig, D. A. (2011). Color image segmentation approach to monitor flowering in lesquerella. *Indust. Crops Products* 34, 1150–1159. doi: 10.1016/j.indcrop.2011.04.002
- 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
- Valiente-González, J. M., Andreu-García, G., Potter, P., and Rodas-Jordá, Á. (2014). Automatic corn (*Zea mays*) kernel inspection system using novelty detection based on principal component analysis. *Biosyst. Eng.* 117, 94–103. doi: 10.1016/j.biosystemseng.2013.09.003
- Vedaldi, A., and Fulkerson, B. (2010). “Vlfeat: an open and portable library of computer vision algorithms,” in *Proceedings of the 18th ACM International Conference on Multimedia* (Firenze), 1469–1472.
- Vedaldi, A., and Zisserman, A. (2012). Efficient additive kernels via explicit feature maps. *IEEE Trans. Patt. Anal. Mach. Intell.* 34, 480–492. doi: 10.1109/TPAMI.2011.153
- Virlet, N., Sabermanesh, K., Sadeghi-Tehran, P., and Hawkesford, M. (2017). Field Scanalyzer: an automated robotic field phenotyping platform for detailed crop monitoring. *Funct. Plant Biol.* 44, 143–153. doi: 10.1071/FP16163
- Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., and Gong, Y. (2010). *Locality-constrained Linear Coding for image classification*, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (San Francisco, CA).
- Yihang, F., Tingting, C., Ruifang, Z., and Xingyu, W. (2014). *Automatic recognition of rape seedling emergence stage based on computer vision technology*, 2014 The Third International Conference on Agro-Geoinformatics. Beijing: Friendship Hotel Beijing.
- Yu, Z., Cao, Z., Wu, X., Bai, X., Qin, Y., Zhuo, W., et al. (2013). Automatic image-based detection technology for two critical growth stages of maize: emergence and three-leaf stage. *Agricul. For. Meteorol.* 174–175, 65–84. doi: 10.1016/j.agrformet.2013.02.011
- Zadoks, J. C., Chang, T. T., and Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Res.* 14, 415–421. doi: 10.1111/j.1365-3180.1974.tb01084.x
- Zhang, K., Tian, J., Zhao, L., Liu, B., and Chen, G. (2008). Detection of quantitative trait loci for heading date based on the doubled haploid progeny of two elite Chinese wheat cultivars. *Genetica* 135, 257–265. doi: 10.1007/s10709-008-9274-6
- Zhu, Y., Cao, Z., Lu, H., Li, Y., and Xiao, Y. (2016). In-field automatic observation of wheat heading stage using computer vision. *Biosyst. Eng.* 143, 28–41. doi: 10.1016/j.biosystemseng.2015.12.015

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

Copyright © 2017 Sadeghi-Tehran, Sabermanesh, Virlet and Hawkesford. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Plant Recycling for Molecular Biofarming to Produce Recombinant Anti-Cancer mAb

Deuk-Su Kim<sup>1</sup>, Ilchan Song<sup>1</sup>, Jinhee Kim<sup>2</sup>, Do-Sun Kim<sup>2</sup> and Kisung Ko<sup>1\*</sup>

<sup>1</sup> Department of Medicine, College of Medicine, Chung-Ang University, Seoul, South Korea, <sup>2</sup> Vegetable Research Division, National Institute of Horticultural and Herbal Science, Rural Development Administration, Wanju-gun, South Korea

## OPEN ACCESS

**Edited by:**

Marcos Egea-Cortines,  
Universidad Politécnica de Cartagena,  
Spain

**Reviewed by:**

Yongzhen Pang,  
Institute of Botany, The Chinese  
Academy of Sciences, China  
Julia Christine Meitz-Hopkins,  
Stellenbosch University, South Africa

**\*Correspondence:**

Kisung Ko  
ksko@cau.ac.kr

**Specialty section:**

This article was submitted to  
Technical Advances in Plant Science,  
a section of the journal  
*Frontiers in Plant Science*

**Received:** 21 April 2016

**Accepted:** 01 July 2016

**Published:** 18 July 2016

**Citation:**

Kim D-S, Song I, Kim J, Kim D-S and Ko K (2016) Plant Recycling for Molecular Biofarming to Produce Recombinant Anti-Cancer mAb.  
*Front. Plant Sci.* 7:1037.  
doi: 10.3389/fpls.2016.01037

The expression and glycosylation patterns of anti-colorectal cancer therapeutic monoclonal antibody (mAb) CO17-1A recognizing the tumor-associated antigen GA733-2, expressed in human colorectal carcinoma cells, were observed in the leaf and stem tissues of primary (0 cycle), secondary (1 cycle), and tertiary (2 cycle) growths of seedlings obtained from the stem cut of T<sub>2</sub> plants. The bottom portion of the stem of T<sub>2</sub> seedlings was cut to induce the 1 cycle shoot growth, which was again cut to induce the 2 cycle shoot growth. In the 1 and 2 cycle growths, the periods for floral organ formation (35 days) was shorter than that (100 days) for the 0 cycle growth. The genes of heavy and light chains of mAb CO17-1A existed at the top, middle, and basal portions of the leaves and stem obtained from the 0, 1, and 2 cycle plants. The protein levels in the leaves and stem tissues from the 1 and 2 cycles were similar to those in the tissues from the 0 cycle. The glycosylation level and pattern in the leaf and stem did not alter dramatically over the different cycles. Surface plasmon resonance (SPR) confirmed that mAbs CO17-1A obtained from leaf and stem tissues of the 0, 1, and 2 cycles had similar binding affinity for the GA733-2 antigen. These data suggest that the shoot growth by bottom stem cutting is applicable to speed up the growth of plant biomass expressing anti-colorectal cancer mAb without variation of expression, glycosylation, and functionality.

**Keywords:** plant product system, axillary bud, biomass, recycling plant, anti colorectal monoclonal antibody

## BACKGROUND

Plants are well recognized as alternative hosts for production of highly valuable recombinant proteins, such as antibodies, vaccines, human blood products, hormones, and growth regulators (Fernandez-San Millan et al., 2003; Rigano and Walmsley, 2005; Schillberg et al., 2013). They offer mass production and safety advantages, with ease of seed storage, compared to the microbial and animal cell-based systems (Twyman et al., 2003; Fischer et al., 2013; Shanmugaraj and Ramalingam, 2014). Plant cultivation can easily be modulated in response to the demand for the recombinant protein by controlling the number of seeds sown (Fischer and Emans, 2000; Leckie and Stewart, 2011). The speed of obtaining plant biomass is essential in the application to save time for production of recombinant proteins. In general, however, more than 14 weeks are required for the development of a fully grown tobacco plant from the time of sowing of its seeds (Lim et al., 2015). A plant-based recombinant protein production system might have drawbacks, such as relatively long cultivation period for obtaining full biomass from the seeds, especially when there is

time-constraint for use of the field for cultivation of such plants. Therefore, rapid enhancement of the plant biomass is imperative for increasing the efficiency of plant-based systems for production of recombinant proteins.

Physiological mechanisms responsible for rapid growth of lateral shoots, without a corresponding increase in root growth, can promote biomass accumulation (Vysotskaya, 2005). Auxin, an important hormone of shoots, in particular, regulates apical dominance (Cline et al., 1997; Teale et al., 2006). The lower lateral buds can be induced by cutting the terminal bud to remove apical dominance (Skoog and Thimann, 1934; Leyser, 2003; Umehara et al., 2008). Moreover, the speed of plant shoot growth could be enhanced by lowering the shoot/root ratio (Vysotskaya, 2005), to overcome the space and time limitations. The cutting of lower branches could be a strategy for plant recycling system to speed up the biomass production without the need for further sowing of seeds. In this study, the effect of cutting the stem to remove apical dominance on shoot growth rate was determined. Furthermore, the expression, glycosylation, and function of a recombinant anti-colorectal cancer monoclonal antibody (mAb) CO17-1A, expressed in the newly induced shoots was determined to confirm whether secondary shoot growth would be applicable as a plant recycling system to obtain increased biomass for enhanced production of recombinant anti-cancer mAb proteins.

## MATERIALS AND METHODS

### Plant Material and Cultivation

Forty seeds of transgenic tobacco T<sub>2</sub> plants, with plant-derived anti-colorectal cancer mAb (mAb<sup>P</sup> CO17-1A) production capability (So et al., 2012), were sown in pots (18.5 cm × 18.5 cm × 14.5 cm) filled with steam-sterilized commercial soil mixture (Sun Gro Horticulture, Agawam, MA; **Figure 1A**), respectively. Forty seedlings were grown in greenhouse under simulated natural light conditions with an average 12 h light/12 h dark photoperiod. The growth of plants was measured immediately before flowering. The lower branch of the primary plant seedling (0 cycle) was cut to induce axillary buds for growth of lateral branches (1 cycle) on the remnant 10 cm long base stem (**Figure 1A**). The base stem of the 0 cycle plant in the pot was maintained to induce the 1 cycle for the growth of lateral branches until the appearance of inflorescence on the shoot obtained from the axillary bud (**Figure 1A**). The growth of tertiary (2 cycle) shoots was induced from the cut stem of the fully grown 1 cycle plant (**Figure 1A**).

### Polymerase Chain Reaction (PCR) Amplification from Genomic DNA of Leaf and Stem in 0, 1, and 2 Cycle Plants

Genomic DNA was isolated from approximately 100 mg of leaf and stem tissues from the plants (0, 1, and 2 cycles) using DNeasy kit (Qiagen, Hilden, Germany), according to the manufacturer's recommendations. The extracted DNA was amplified by polymerase chain reaction (PCR) to confirm the

presence of genes for mAb CO17-1A heavy chain (HC; 1,471 bp) and light chain (LC; 764 bp), by using the following forward and reverse primers: HC forward primer, 5'-GCGAATTCTATGGAA TGGAGCAGAGTCTTAT C-3'; HC reverse primer, 5'-GATTA ATCGATTTACCCGGAGTCCG-3'; LC forward primer, 5'-GC CTCG AGATGGGCATCAAGATGGAATCACAG-3'; LC reverse primer, 5'-GAGGTACCCTAACACTCATTCTGTTGAAGCTC-3'.

### Western Blot Analysis

Eighty milligram of fresh leaves and stems (from top, middle, and basal portions of plants) was crushed by cryo-milling to extract the total soluble proteins. The homogenized plant samples were mixed with 280 µL of sample buffer (1 M Tris-HCl, 50% glycerol, 10% SDS, 5% 2-mercaptoethanol, 0.1% bromophenol blue), and the homogenates were loaded on a sodium dodecyl sulfate polyacrylamide gel. The electrophoresed proteins were transferred on to a nitrocellulose membrane (Millipore Corp., Billerica, MA, USA), which was blocked with 5% skimmed milk (Sigma, St. Louis, MO, USA), prepared in 1× phosphate-buffered saline (PBS), for 2 h. The blot was subsequently probed with goat anti-murine IgG Fcγ and anti-murine IgG F(ab)'<sub>2</sub>, which recognize the HC and LC of mAb CO17-1A, respectively. The purified mAb<sup>P</sup> CO17-1A was used as a positive control (Ko et al., 2005).

### Purification of Recombinant mAb<sup>P</sup> CO17-1AK from Leaf and Stem of Plant from Each Cycle

For purification of mAb<sup>P</sup> CO17-1AK, the leaves and stem from the tobacco plants of the 0, 1, and 2 cycles were homogenized on ice in the extraction buffer (37.5 mM Tris-HCl pH 7.5, 50 mM NaCl, 15 mM EDTA, 75 mM sodium citrate, and 0.2% sodium thiosulfate) using a blender. After centrifugation at 8,800 × g for 30 min at 4°C, the supernatant was filtered through a Miracloth (Biosciences, La Jolla, CA, USA), and its pH was adjusted to 5.1 with acetic acid. The supernatant was further centrifuged at 10,200 × g for 30 min. The pH of the supernatant, thus obtained, was adjusted to neutral by addition of 3 M Tris-HCl. The total soluble protein was precipitated with ammonium sulfate after overnight incubation in a cold room followed by centrifugation at 4°C for 30 min. The pellet was resuspended in one-tenth of the starting volume of extraction buffer, and the obtained solution was centrifuged at 10,200 × g for 30 min at 4°C (Park et al., 2015). The mAb<sup>P</sup> CO17-1A protein was purified using protein A Sepharose 4 Fast Flow (GE Healthcare, Sweden, NJ, USA), according to the manufacturer's recommendations. The mAb<sup>P</sup> CO17-1A protein was dialyzed against 1× PBS (pH 7.4). The protein concentration was determined using a Nano-drop (Biotek, Highland, VT, USA) and the purified protein was visualized by SDS-PAGE. Aliquots of the purified protein were stored at -80°C for further studies.

### Glycan Analysis

The purified mAb<sup>P</sup> CO17-1A protein samples were treated twice with 1 µL pepsin in an incubator at 37°C for 16 h to

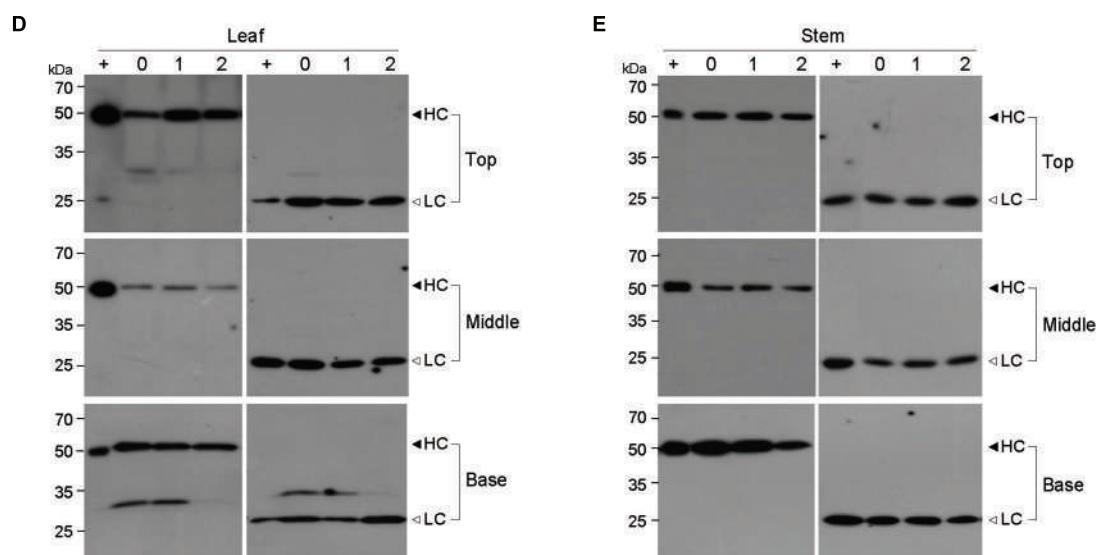
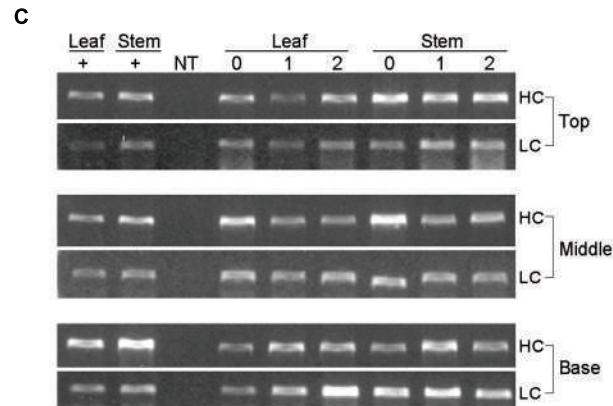
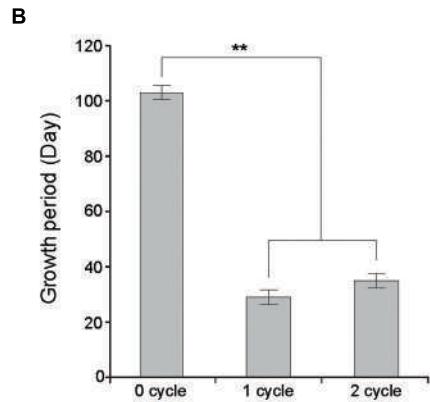
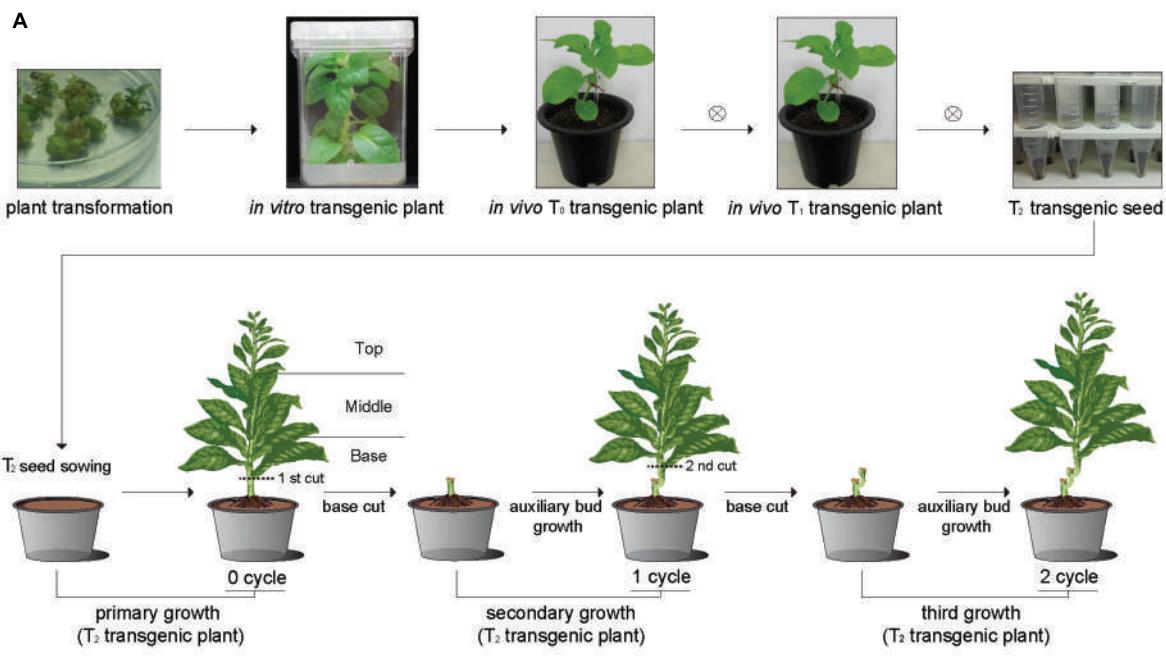


FIGURE 1 | Continued

**FIGURE 1 | Continued**

**Schematic diagram for the primary (0 cycle), secondary (1 cycle), and tertiary (2 cycle) growth of plants expressing the anti-colorectal cancer mAb CO17-1A.** **(A)** Schematic diagram showing primary (0 cycle), secondary (1 cycle), and tertiary (2 cycle) growths from T<sub>0</sub> transgenic plants. The base stem of 0 cycle plant was cut to induce axillary buds for secondary plant growth (1 cycle), and the base stem of the 1 cycle plant was cut for the tertiary plant growth (2 cycle). T, top of the whole plant; M, middle of the whole plant; B, base of the whole plant. **(B)** Comparison of plant growth period for flowering in 0, 1, and 2 cycles. The growth of 0 cycle plants was compared with 1 and 2 cycle plants obtained from base stem cutting. The asterisks indicate statistically significant differences (\*\*p < 0.01). **(C)** PCR analysis to confirm the existence of HC and LC genes in top, middle, and basal portions of both the leaves and stems of transgenic plant (0, 1, and 2 cycles). The genomic DNA fragments of mAb<sup>P</sup> CO17-1A were amplified and electrophoretically separated on a 1% agarose gel. NT, non-transgenic plant; HC, heavy chain of mAb<sup>P</sup> CO17-1A; LC, light chain of mAb<sup>P</sup> CO17-1A. **(D,E)** Western blot analysis to confirm the mAb<sup>P</sup> CO17-1A HC and LC expression in the leaves and stems of transgenic plants through 0, 1, and 2 cycles. The bands for HC (50 kDa) and LC (25 kDa) were detected with horse radish peroxidase-conjugated goat anti-mouse Fc and goat anti-mouse F(ab')<sub>2</sub>-specific antibody, respectively. +, purified mAb CO17-1A from plant (So et al., 2012); Top, top portion of plant; Middle, middle portion of plant; Base, basal portion of plant **(A).**

digest the protein into glycopeptides. The glycopeptides were collected using a C18 Sep-Pak cartridge (Waters, Lexington, MA, USA). The N-glycosidase (PNGase) A glycan enzyme was added to the collected glycopeptides to release the N-glycans, and the mixtures were incubated overnight at 37°C. The released N-glycans were purified from the samples by using a graphitized carbon resin from Carbograph (Alltech, Lexington, MA, USA). The purified glycans were 2-aminobenzamide (2-AB)-labeled using previously described methods (Bigge et al., 1995). The 2-AB-labeled glycans were separated on a TSK amide-80 column (5 μm, 4.6 mm × 250 mm; Tosoh Bioscience, Prussia, PA, USA) using a high performance liquid chromatography (HPLC) system with a fluorescence detector (330 nm excitation and 425 nm emission; Lim et al., 2015). The separation of the labeled glycans was achieved at a flow rate of 1.0 mL/min using a mixture of solvent A (100% acetonitrile) and solvent B (50 mM ammonium formate, pH 4.4). After the column was equilibrated using 30% solvent B, the sample was injected and then eluted by a linear gradient to 45% of solvent B for 60 min. HPLC analysis was repeated more than three times.

### Surface Plasmon Resonance Analysis

The surface plasmon resonance (SPR) analysis was performed to confirm the affinity of mAb<sup>P</sup> CO17-1A to GA733 antigen using a commercially available GLC chip on an XPR36 surface instrument (Bio-Rad, Hercules, CA, USA). The GA733 protein was immobilized on a GLC chip, and an acidic buffer at pH 6.0 was allowed to flow over the biochip surface at a rate of 50 μL/min. One microgram of the purified mAb<sup>P</sup> CO17-1A from leaf and stem (0, 1, and 2 cycle samples) was dissolved in 300 μL of 1× PBS, and the 300 μL was applied to immobilized receptors with a flow rate of 50 μL/min at 25°C and pH 6.0. After each measurement, the surface of the sensor chip was regenerated using phosphoric acid buffer.

## RESULTS

### Induction and Growth of Axillary Buds by Bottom Stem Cut

*Agrobacterium*-mediated tobacco plant transformation was conducted to generate transgenic plants expressing the anti-colorectal cancer mAb CO17-1A (Ko et al., 2005; So et al., 2012).

The seeds of T<sub>2</sub> transgenic plants were obtained by consecutive self-fertilization of the T<sub>0</sub> and T<sub>1</sub> plants. Two well-expanded true leaves appeared in the plantlets 21 days after sowing of T<sub>2</sub> seeds. The growth period of the T<sub>2</sub> plants until flower formation was around 100 days (Figures 1A,B). The lateral shoot was induced after retaining the root system by cutting the bottom stem of transgenic plants expressing the anti-colorectal cancer mAb CO17-1A (Figure 1A). Only a single lateral shoot was left to grow until just before the floral organ formation (Figure 1A). The bottom stem cutting was conducted in 2 cycles (Figure 1A). However, the growth period of the axillary shoot to the flowering stage from the cut stem was 30 and 35 days in the 1 and 2 cycles (Figure 1B). Overall, the growth period of the lateral shoot in 1 and 2 cycles was almost three times shorter compared to that of the primary shoot from the seedlings (0 cycle).

### Existence of mAb CO17-1A HC and LC Genes in 0, 1, and 2 Cycle Plants

The PCR analysis was conducted to confirm the existence of HC and LC genes of mAb CO17-1A in the leaf and stem tissues from top, middle, and base stem portions in 0, 1, and 2 cycles (Figure 1C). The HC and LC genes existed in all the portions of leaf and stem tissues of lateral shoot in all the cycles (Figure 1C). No HC or LC gene was amplified in the samples from the non-transgenic (NT) plants.

### HC and LC Protein Levels of mAb CO17-1A in the Leaf and Stem Tissues from Top, Middle, and Base Portions Through Recycling

The changes in HC and LC protein levels in top, middle, and basal leaves and stems in 0, 1, and 2 cycles were investigated by western blotting (Figures 1D,E). In the leaf tissue, the HC and LC protein levels were stable over the cycles (Figure 1D left and right panels, respectively). In the top leaves, the HC levels slightly increased with the cycles. The LC levels were steady over the cycles. In the stem tissue, the HC and LC levels were stable over the cycles (Figure 1E left and right panels, respectively). In the basal stems, the HC levels slightly decreased over the cycles (Figure 1E left panel). Overall, the HC and LC protein levels were similar in the samples from all the portions of stem through the cycles.

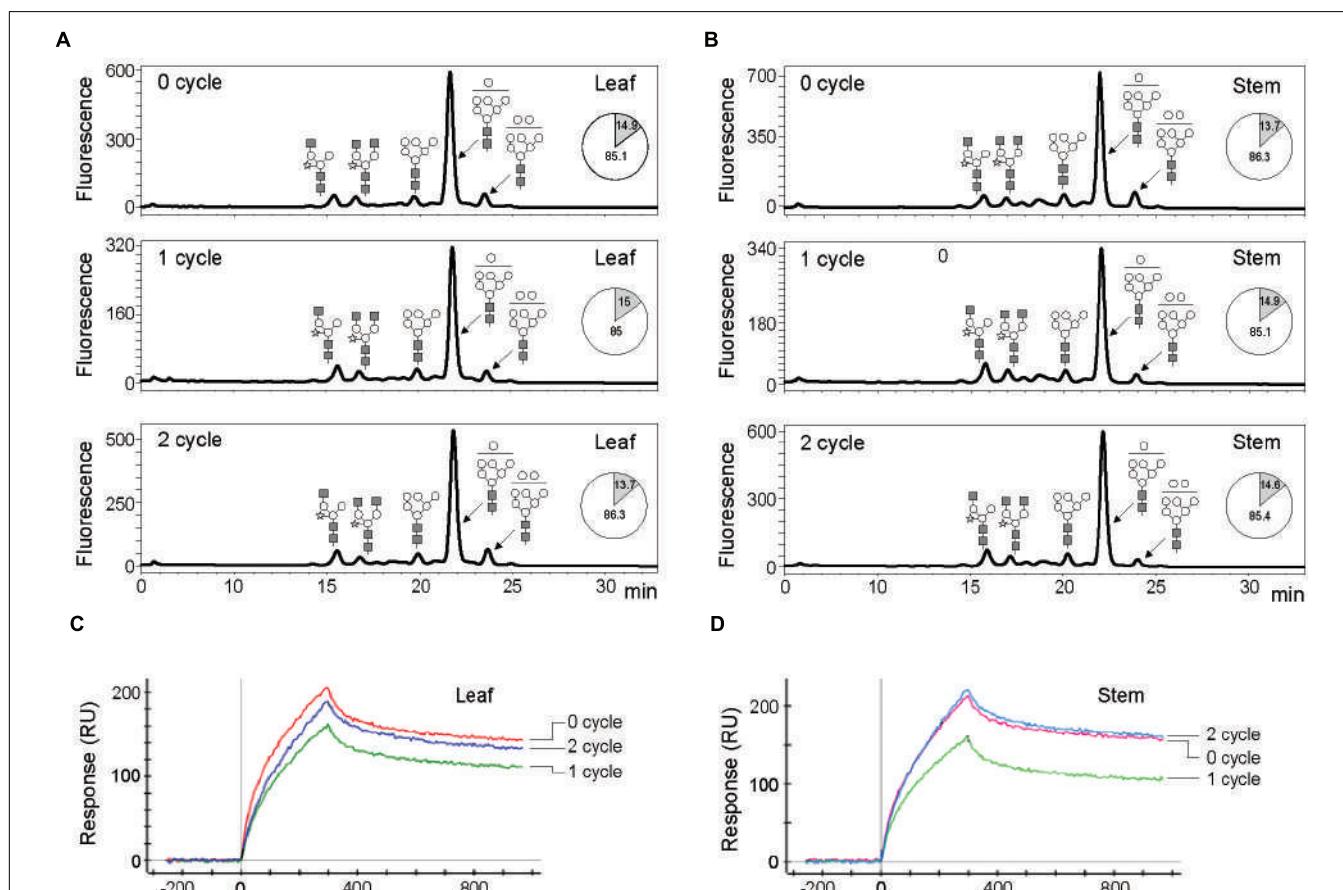
## Glycan Analysis of mAb<sup>P</sup> CO17-1A Purified from Leaf and Stem of 0, 1, and 2 Cycle Plants

The N-glycans of mAb<sup>P</sup> CO17-1A purified from the leaf and stem of 0, 1, and 2 cycle plants were analyzed by HPLC. The glycosylation patterns were analyzed in the leaf and stem (0, 1, and 2 cycles) tissues (Figures 2A,B). The glycan profile of leaf and stem from plants (0, 1 and 2 cycles) was similar and showed a high mannose-type glycan structure profile. The percentages (%) of oligomannose glycan in leaf and stem were ~15 and ~13.7–14.9, respectively, through the cycles (Figures 2A,B). The glycan structure profiles of mAb<sup>P</sup> CO17-1A in 0, 1, and 2 cycle plant leaves and stems were similar.

## SPR Analysis of mAb CO17-1A Purified from Leaf and Stem of 0, 1, and 2 Cycle Plants

SDS-PAGE analysis was performed to identify the HC and LC of the purified mAb CO17-1A in the plant leaf and stem

samples obtained in each cycle (0, 1, and 2; Data not shown). The purified mAb CO17-1A from 0, 1, and 2 cycle plants showed the same band sizes for HC and LC. Although the cycle number increased, there was no change in the quality of mAb CO17-1A, which remained undegraded. Expression and purity of mAb CO17-1A in stems was also confirmed in 0, 1, and 2 cycle plants. mAb<sup>P</sup> CO17-1A purified from the leaf and stem samples of 0, 1 and 2 cycle plants were compared for their binding activities (Figures 2C,D). All the mAb<sup>P</sup> CO17-1A purified from leaf and stem tissues collected in different cycles showed relatively similar interaction with the antigen GA733-Fc using SPR (Figures 2C,D). The binding affinities of the mAb CO17-1A purified from leaf and stem samples collected from different cycle plants were similar except for the 2 cycle where the mAb CO17-1A purified from stem (Figure 2D) showed slightly higher affinity for the antigen than the mAb purified from the leaf (Figure 2C). When the binding activities of mAb CO17-1A purified from the leaf and stem samples were compared among the cycles, the 1 cycle showed slightly lower affinity than the 0



**FIGURE 2 | Glycosylation and function analyses of mAb CO17-1A protein purified from leaves and stem of 0, 1, and 2 cycle plants. (A,B)** Profiles of N-glycan from mAb<sup>P</sup> CO17-1A were analyzed using high performance liquid chromatography of the 2-AB labeled glycans. **(A)** Glycan structure profiles of mAb<sup>P</sup> CO17-1A purified from leaf of 0, 1, and 2 cycle plants. **(B)** Glycan structure profiles of mAb<sup>P</sup> CO17-1A purified from stem of 0, 1, and 2 cycle plants. GlcNAc, mannose, and xylose are depicted using black square, white circle, and white star, respectively. The ratios of oligomannose (white) and plant-specific (gray) glycans of mAb<sup>P</sup> CO17-1A in leaves and stem of 0, 1, and 2 cycle plants were shown in a pie chart. Binding affinity of mAb CO17-1A purified from leaves **(C)** and stem **(D)** of 0, 1, and 2 cycle plants to GA733 antigen using surface plasmon resonance (SPR). Purified mAb<sup>P</sup> CO17-1AK from the leaves of 0, 1, and 2 cycle plants was incubated with the GA733 adsorbed biochip **(C)**. Purified mAb<sup>P</sup> CO17-1A from the stem of 0, 1, and 2 cycle plants was incubated with the GA733 adsorbed biochip **(D)**.

and 2 cycles (**Figures 2C,D**). In general, however, the peaks of mAb CO17-A from both the leaf and stem samples were similar between the 0 and 2 cycles.

## DISCUSSION

In the present study, we demonstrate that fresh stem and leaves regrow from axillary buds after cutting of the stem. The lateral shoots, thus generated, could stably express functional anti-colorectal cancer therapeutic antibody, mAb CO17-1A recombinant protein, without alterations in its glycosylation pattern.

The lateral shoots emerge from axillary meristems when the apical dominance is removed (Leyser, 2003). In the present study, the plants (1 and 2 cycles), induced to produce lateral branches from the axillary buds, grew faster (~30–35 days) to their full size and flowered than the plants grown from the seedlings (0 cycle), which required ~100 days to reach their full size.

In fact, the use of plant expression systems has been limited due to longer growth period to obtain a full-sized plant with high biomass (Hood et al., 2002; Horn et al., 2004; Teli and Timko, 2004). In our previous study, fully grown *Nicotiana tabacum* plants started to form the floral organs at 12 weeks after sowing (Lim et al., 2015), which is much longer than *N. benthamiana* (7 weeks; Conley et al., 2011). *N. benthamiana* is another host plant for production of recombinant proteins such as vaccine and antibody, and has been established for their transient expression (Gomez et al., 2013; Li et al., 2016). *N. benthamiana* needs to be transfected every time with expression vector inoculums to produce recombinant proteins. In addition, their left over biomass should be properly discarded for avoiding contamination, and the transfected plant can not be regrown for further transfection usage. Thus, the transgenic plant regrowth by axillary shoot induction with less than 4 weeks appears to be an easy method to quickly increase the full biomass for production of recombinant proteins even under limitations of space. Our results suggest that axillary bud induction from the base stem with root could be used in molecular biofarming strategies to overcome the constraints of space and time.

The existence of HC and LC genes in both the leaves and stem generated from the axillary buds during the regrowth cycles was confirmed using PCR analysis, which revealed that the genes were present in the top, middle, and basal portions of the leaves and stem of the plants (0, 1, and 2 cycles) without any deletion.

The expression of HC and LC of mAb CO17-1A in the leaves and stem from the top, middle, and basal positions obtained from 0, 1, and 2 cycle plants was confirmed by western blot. The HC and LC expression rates were not significantly different among the samples.

## REFERENCES

- Bigge, J. C., Patel, T. P., Bruce, J. A., Goulding, P. N., Charles, S. M., and Parekh, R. B. (1995). Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid. *Anal. Biochem.* 230, 229–238. doi: 10.1006/abio.1995.1468

The mAb CO17-1A purified from the plants had a mainly oligomannose structure profiles because of the C-terminus KDEL signal tagging of HC for ER retention (So et al., 2012). The glycosylation profiles were unmodified in leaf and stem in samples from the 0, 1, and 2 cycles.

The mAb CO17-1A purified from primary (0 cycle), secondary (1 cycle), and tertiary (2 cycle) plants showed similar binding affinity to the GA733 antigen in SPR analysis. Leaves and stem from the 1 cycle showed slightly lower binding activity than those from the 0 and 2 cycle plants. However, it is speculated that the slight fluctuation in the binding activity was due to the variation in sample preparation, and not due to an actual loss in the activity. The binding activities of mAb CO17-1A purified from both the leaves and stem of plants from the same cycle were similar. The results present notable evidence that plant recycling can be applied for efficient biomass enhancement without any variation in expression and function of the recombinant anti-cancer therapeutic mAb. The rapid plant regrowth using by the existing lateral buds in stem attached to the root is possible for the plant biomass production in a limited space.

Taken together, the leaf and stem of the secondary and tertiary cycles of plant growth (1 and 2 cycles) had similar mAb CO17-1A expression rate, and the antigen affinity as well as glycan structure profile of the purified mAb were comparable to the purified mAb samples obtained from the primary plant growth (0 cycle). This study shows that novel recycling plant system by using regrowth from axillary buds can effectively circumvent the space and time limitations for cultivation of plants. The strategy of recycling plant production could be exploited for obtaining increased transgenic plant biomass in less time and could be useful for producing highly valuable recombinant proteins for varied use.

## AUTHOR CONTRIBUTIONS

KK and D-SK conceived and designed the experiments. D-SK and IS performed the experiments. KK, D-SK, and IS analyzed the data. D-SK and JK contributed reagents/materials/analysis tools. KK and D-SK wrote the paper. All of the authors carefully checked and approved this version of the manuscript.

## ACKNOWLEDGMENTS

This research was supported by a grant (Code# PJ011110) from the Korean Rural Development Administration, National Research Foundation of Korea Grant funded by the Korean Government (MEST) (NRF-2014R1A2A1A11052922).

- Cline, M., Wessel, T., and Iwamura, H. (1997). Cytokinin/auxin control of apical dominance in *Ipomoea nil*. *Plant Cell Physiol.* 38, 659–667. doi: 10.1093/oxfordjournals.pcp.a029218  
Conley, A. J., Zhu, H., Le, L. C., Jevnikar, A. M., Lee, B. H., Brandle, J. E., et al. (2011). Recombinant protein production in a variety of *Nicotiana* hosts:

- a comparative analysis. *Plant Biotechnol. J.* 9, 434–444. doi: 10.1111/j.1467-7652.2010.00563.x
- Fernandez-San Millan, A., Mingo-Castel, A., Miller, M., and Daniell, H. (2003). A chloroplast transgenic approach to hyper-express and purify Human Serum Albumin, a protein highly susceptible to proteolytic degradation. *Plant Biotechnol. J.* 1, 71–79. doi: 10.1046/j.1467-7652.2003.00008.x
- Fischer, R., and Emans, N. (2000). Molecular farming of pharmaceutical proteins. *Transgenic Res.* 9, 279–299. doi: 10.1023/A:1008975123362
- Fischer, R., Schillberg, S., Buyel, J. F., and Twyman, R. M. (2013). Commercial aspects of pharmaceutical protein production in plants. *Curr. Pharm. Des.* 19, 5471–5477. doi: 10.2174/1381612811319310002
- Gomez, E., Soledad Lucero, M., Chimeno Zoth, S., Manuel Carballeda, J., Jose Gravisaco, M., and Berinstein, A. (2013). Transient expression of VP2 in *Nicotiana benthamiana* and its use as a plant-based vaccine against infectious bursal disease virus. *Vaccine* 31, 2623–2627. doi: 10.1016/j.vaccine.2013.03.064
- Hood, E. E., Woodard, S. L., and Horn, M. E. (2002). Monoclonal antibody manufacturing in transgenic plants – myths and realities. *Curr. Opin. Biotechnol.* 13, 630–635. doi: 10.1016/S0958-1669(02)00351-8
- Horn, M. E., Woodard, S. L., and Howard, J. A. (2004). Plant molecular farming: systems and products. *Plant Cell Rep.* 22, 711–720. doi: 10.1007/s00299-004-0767-1
- Ko, K., Steplewski, Z., Glogowska, M., and Koprowski, H. (2005). Inhibition of tumor growth by plant-derived mAb. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7026–7030. doi: 10.1073/pnas.0502533102
- Leckie, B. M., and Stewart, C. N. Jr. (2011). Agroinfiltration as a technique for rapid assays for evaluating candidate insect resistance transgenes in plants. *Plant Cell Rep.* 30, 325–334. doi: 10.1007/s00299-010-0961-2
- Leyser, O. (2003). Regulation of shoot branching by auxin. *Trends Plant Sci.* 8, 541–545. doi: 10.1016/j.tplants.2003.09.008
- Li, J., Stoddard, T. J., Demorest, Z. L., Lavoie, P. O., Luo, S., Clasen, B. M., et al. (2016). Multiplexed, targeted gene editing in *Nicotiana benthamiana* for glyco-engineering and monoclonal antibody production. *Plant Biotechnol. J.* 14, 533–542. doi: 10.1111/pbi.12403
- Lim, C. Y., Lee, K. J., Oh, D. B., and Ko, K. (2015). Effect of the developmental stage and tissue position on the expression and glycosylation of recombinant glycoprotein GA733-FcK in transgenic plants. *Front. Plant Sci.* 5:778. doi: 10.3389/fpls.2014.00778
- Park, S.-R., Lim, C.-Y., Kim, D.-S., and Ko, K. (2015). Optimization of ammonium sulfate concentration for purification of colorectal cancer vaccine candidate recombinant protein GA733-FcK isolated from plants. *Front. Plant Sci.* 040. doi: 10.3389/fpls.2015.01040
- Rigano, M. M., and Walmsley, A. M. (2005). Expression systems and developments in plant-made vaccines. *Immunol. Cell Biol.* 83, 271–277. doi: 10.1111/j.1440-1711.2005.01336.x
- Schillberg, S., Raven, N., Fischer, R., Twyman, R. M., and Schiermeyer, A. (2013). Molecular farming of pharmaceutical proteins using plant suspension cell and tissue cultures. *Curr. Pharm. Des.* 19, 5531–5542. doi: 10.2174/1381612811319310008
- Shanmugaraj, B. M., and Ramalingam, S. (2014). Plant expression platform for the production of recombinant pharmaceutical proteins. *Austin J. Biotechnol. Bioeng.* 1, 1–4.
- Skoog, F., and Thimann, K. V. (1934). Further experiments on the inhibition of the development of lateral buds by growth hormone. *Proc. Natl. Acad. Sci. U.S.A.* 20, 480–485. doi: 10.1073/pnas.20.8.480
- So, Y., Lee, K.-J., Kim, D.-S., Lee, J.-H., Oh, D.-B., Hwang, K.-A., et al. (2012). Glycomodification and characterization of anti-colorectal cancer immunotherapeutic monoclonal antibodies in transgenic tobacco. *Plant Cell Tiss. Organ Cult.* 113, 41–49. doi: 10.1007/s11240-012-0249-z
- Teale, W. D., Paponov, I. A., and Palme, K. (2006). Auxin in action: signalling, transport and the control of plant growth and development. *Nat. Rev. Mol. Cell Biol.* 7, 847–859. doi: 10.1038/nrm2020
- Teli, N. P., and Timko, M. P. (2004). Recent developments in the use of transgenic plants for the production of human therapeutics and biopharmaceuticals. *Plant Cell Tiss. Organ Cult.* 79, 125–145. doi: 10.1007/s11240-004-0653-0
- Twyman, R. M., Stoger, E., Schillberg, S., Christou, P., and Fischer, R. (2003). Molecular farming in plants: host systems and expression technology. *Trends Biotechnol.* 21, 570–578. doi: 10.1016/j.tibtech.2003.10.002
- Umeshara, M., Hanada, A., Yoshida, S., Akiyama, K., Arite, T., Takeda-Kamiya, N., et al. (2008). Inhibition of shoot branching by new terpenoid plant hormones. *Nature* 455, 195–200. doi: 10.1038/nature07272
- Vysotskaya, L. B. (2005). Mechanisms coordinating wheat seedling growth response as affected by shoot/root ratio. *Russian J. Plant Physiol.* 52, 679–684. doi: 10.1007/s11183-005-0101-y

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

Copyright © 2016 Kim, Song, Kim, Kim and Ko. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read for greatest visibility and readership



## FAST PUBLICATION

Around 90 days from submission to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative, and constructive peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers acknowledged by name on published articles



## 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