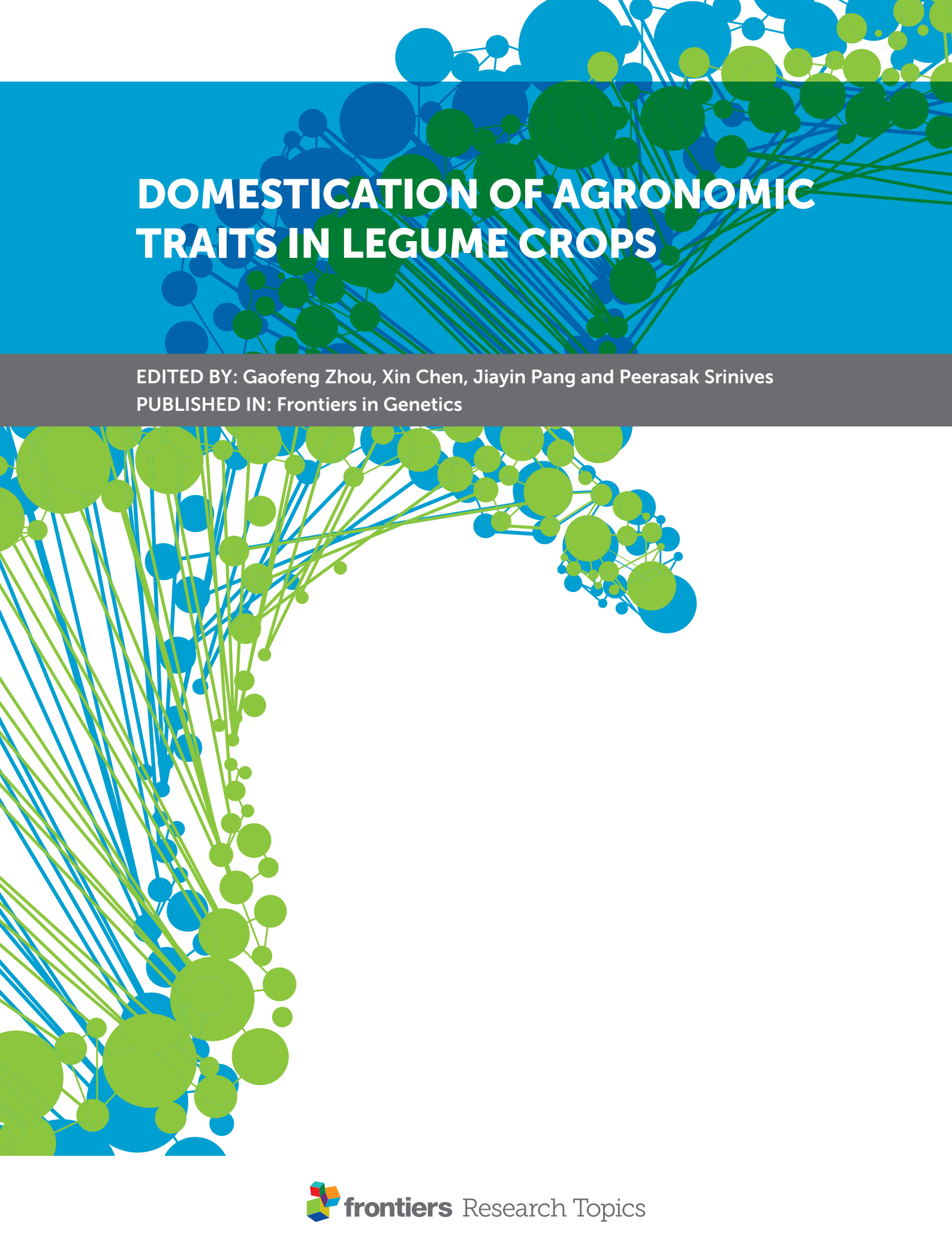




DOMESTICATION OF AGRONOMIC TRAITS IN LEGUME CROPS

EDITED BY: Gaofeng Zhou, Xin Chen, Jiayin Pang and Peerasak Srinives
PUBLISHED IN: *Frontiers in Genetics*



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88971-368-4

DOI 10.3389/978-2-88971-368-4

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

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

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

What are Frontiers Research Topics?

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

DOMESTICATION OF AGRONOMIC TRAITS IN LEGUME CROPS

Topic Editors:

Gaofeng Zhou, Department of Primary Industries and Regional Development of Western Australia (DPIRD), Australia

Xin Chen, Institute of Industrial Crops, China

Jiayin Pang, University of Western, Australia

Peerasak Srinives, Kasetsart University, Thailand

Citation: Zhou, G., Chen, X., Pang, J., Srinives, P., eds. (2021). Domestication of Agronomic Traits in Legume Crops. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88971-368-4

Table of Contents

- 05 Editorial: Domestication of Agronomic Traits in Legume Crops**
Xin Chen, Gaofeng Zhou, Jiayin Pang and Peerasak Srinives
- 08 Marker-Trait Association Analysis of Seed Traits in Accessions of Common Bean (*Phaseolus vulgaris* L.) in China**
Lei Lei, Lanfen Wang, Shumin Wang and Jing Wu
- 20 QTL Mapping for Agronomic and Adaptive Traits Confirmed Pleiotropic Effect of *mog* Gene in Black Gram [*Vigna mungo* (L.) Hepper]**
Prakit Somta, Jingbin Chen, Tarika Yimram, Chutintorn Yundaeng, Xingxing Yuan, Norihiko Tomooka and Xin Chen
- 29 Genetic Diversity of Common Bean (*Phaseolus vulgaris* L.) Germplasm Resources in Chongqing, Evidenced by Morphological Characterization**
Juechen Long, Jijun Zhang, Xiaochun Zhang, Jing Wu, Hong Chen, Ping Wang, Qiang Wang and Chengzhang Du
- 38 Same Locus for Non-shattering Seed Pod in Two Independently Domesticated Legumes, *Vigna angularis* and *Vigna unguiculata***
Yu Takahashi, Alisa Kongjaimun, Chiaki Muto, Yuki Kobayashi, Masahiko Kumagai, Hiroaki Sakai, Kazuhito Satou, Kuniko Teruya, Akino Shiroma, Makiko Shimoji, Takashi Hirano, Takehisa Isemura, Hiroki Saito, Akiko Baba-Kasai, Akito Kaga, Prakit Somta, Norihiko Tomooka and Ken Naito
- 49 Fine Mapping, Candidate Gene Identification and Co-segregating Marker Development for the *Phytophthora* Root Rot Resistance Gene *RpsYD25***
Chao Zhong, Suli Sun, Xuecui Zhang, Canxing Duan and Zhendong Zhu
- 61 The Impacts of Domestication and Breeding on Nitrogen Fixation Symbiosis in Legumes**
Jinge Liu, Xiaocheng Yu, Qiulin Qin, Randy D. Dinkins and Hongyan Zhu
- 70 Soybean *Endo-1,3-Beta-Glucanase* (*GmGLU*) Interaction With Soybean mosaic virus-Encoded *P3* Protein May Contribute to the Intercellular Movement**
Feifei Shi, Ying Wang, Fang Zhang, Xingxing Yuan, Huatao Chen, Xuehao Chen, Xin Chen and Xiaoyan Cui
- 82 Genome- and Transcriptome-Wide Identification of *C3Hs* in Common Bean (*Phaseolus vulgaris* L.) and Structural and Expression-Based Analyses of Their Functions During the Sprout Stage Under Salt-Stress Conditions**
Qi Zhang, Wen-jing Zhang, Zhen-gong Yin, Wei-jia Li, Hao-hao Zhao, Shuo Zhang, Lin Zhuang, Yu-xin Wang, Wen-Hui Zhang and Ji-Dao Du
- 96 The Effects of Domestication on Secondary Metabolite Composition in Legumes**
Yee-Shan Ku, Carolina A. Contador, Ming-Sin Ng, Jeongjun Yu, Gyuhwa Chung and Hon-Ming Lam
- 116 Identification of QTLs for Domestication-Related Traits in Zombi Pea [*Vigna vexillata* (L.) A. Rich], a Lost Crop of Africa**
Kitiya Amkul, Prakit Somta, Kularb Laosatit and Lixia Wang

- 129 ***The Impacts of Domestication and Agricultural Practices on Legume Nutrient Acquisition Through Symbiosis With Rhizobia and Arbuscular Mycorrhizal Fungi***
Ailin Liu, Yee-Shan Ku, Carolina A. Contador and Hon-Ming Lam
- 140 ***The Modification of Circadian Clock Components in Soybean During Domestication and Improvement***
Man-Wah Li and Hon-Ming Lam
- 152 ***Construction of a High-Density Genetic Map and Its Application for QTL Mapping of Leaflet Shapes in Mung Bean (*Vigna radiata* L.)***
Jie Wang, Jianling Li, Zhenxing Liu, Xingxing Yuan, Suhua Wang, Honglin Chen, Xin Chen, Xuzhen Cheng and Lixia Wang
- 161 ***Enhanced Chlorophyll Degradation Triggers the Pod Degreening of "Golden Hook," a Special Ecotype in Common Bean (*Phaseolus vulgaris* L.)***
Bo Hu, Jinlong Zhu, Hongyan Wu, Kun Xu, Hong Zhai, Ning Guo, Yi Gao, Jiayin Yang, Danhua Zhu and Zhengjun Xia
- 179 ***Molecular Basis Underlying Common Cutworm Resistance of the Primitive Soybean Landrace Peking***
Ryu Nakata, Mariko Yano, Susumu Hiraga, Masayoshi Teraishi, Yutaka Okumoto, Naoki Mori and Akito Kaga
- 190 ***The Impact of Domestication on Aboveground and Belowground Trait Responses to Nitrogen Fertilization in Wild and Cultivated Genotypes of Chickpea (*Cicer* sp.)***
Edward Marques, Christopher P. Krieg, Emmanuel Dacosta-Calheiros, Erika Bueno, Emily Sessa, R. Varma Penmetsa and Eric von Wettberg
- 201 ***Wild and Domestic Differences in Plant Development and Responses to Water Deficit in *Cicer****
Jens Berger, Raju Pushpavalli, Christiane Ludwig, Sylvia Parsons, Fatma Basdemir and Kelly Whisson
- 218 ***Targeting Low-Phytate Soybean Genotypes Without Compromising Desirable Phosphorus-Acquisition Traits***
Mireadili Kuerban, Wenfeng Jiao, Jiayin Pang, Jingying Jing, Li-Juan Qiu, Wenli Ding, Wen-Feng Cong, Fusuo Zhang and Hans Lambers
- 228 ***Identification of a Locus Controlling Compound Raceme Inflorescence in Mungbean [*Vigna radiata* (L.) R. Wilczek]***
Eunsoo Lee, Xuefei Yang, Jungmin Ha, Moon Young Kim, Keum Yong Park and Suk-Ha Lee
- 240 ***Single-Molecule Real-Time and Illumina-Based RNA Sequencing Data Identified Vernalization-Responsive Candidate Genes in Faba Bean (*Vicia faba* L.)***
Xingxing Yuan, Qiong Wang, Bin Yan, Jiong Zhang, Chenchen Xue, Jingbin Chen, Yun Lin, Xiaoyan Zhang, Wenbiao Shen and Xin Chen
- 252 ***Identification of Gray Leaf Spot Disease Candidate Gene in Narrow-Leafed Lupin (*Lupinus angustifolius* L.)***
Gaofeng Zhou, Huaan Yang, Daniel Renshaw, Meilin Zou, Geoff Thomas and Chengdao Li



Editorial: Domestication of Agronomic Traits in Legume Crops

Xin Chen^{1*}, Gaofeng Zhou², Jiayin Pang³ and Peerasak Srinives⁴

¹ Institute of Industrial Crops/Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, Jiangsu Academy of Agricultural Sciences, Nanjing, China, ² Department of Primary Industries and Regional Development, Government of Western Australia, South Perth, WA, Australia, ³ The University of Western Australia Institute of Agriculture, University of Western Australia, Perth, WA, Australia, ⁴ Department of Agronomy, Kasetsart University, Bangkok, Thailand

Keywords: legume crops, agronomic traits, domestication, genomics techniques, breeding

Editorial on the Research Topic

Domestication of Agronomic Traits in Legume Crops

Legumes (*Fabaceae*) play an active role in food security, including maintain human health, mitigate climate change, and increase biodiversity. Legume crops are important source of human foods, animal feeds, and industrial materials. In human diets, legumes are among the primary sources of plant-based protein, and they are high in dietary fiber, as well as vitamins and minerals (Mudryj et al., 2014). Therefore, FAO emphasized and pointed out that eating legumes as part of a healthy diet can contribute to addressing the national nutritional problems, from undernutrition and micronutrient deficiencies to obesity and diet-related diseases, which exist in many countries. Leguminous crops are grown worldwide, making them important economically. Ensuring the grain yield and improving the quality of legumes are research goals of plant scientists. Evidence from archaeological and crop diversity research indicates that the domestication of major crops has multiple origin centers (Turrill, 1926). The domestication of legume crops greatly enriched the dietary structure of human beings and provided food security guaranteed for the reproduction and development of human civilization. Enhancing our understanding of the legume crops domestication and improvement process can help to better boost legume breeding efforts. This Research Topic included 19 original or review articles focused on multiple aspects of legume domestication, including but not limited to quantitative trait loci (QTL) mapping and association mapping of domestication-related genes, the function and interaction of domesticated genes, RNA-seq analysis of genes associated with agronomic and adaptive traits.

Domestication is a process in which people use existing experience or knowledge to turn wild animals and plants into controllable, and ultimately harvest products according to human needs. Therefore, crop domestication can also be seen as an evolutionary process of human intervention, and domestication genetics is an important branch of evolutionary biology research. Legume crops have been domesticated and used as parts of the human diet for more than 10,000 years ago, among which lentils (*Lens culinaris* Medik.), chickpeas (*Cicer arietinum* L.), and peas (*Pisum sativum* L.) are Neolithic founder crops. They were among the plant species domesticated by early farmers in the Fertile Crescent region (Calles et al., 2019). In the long cultivation history, the domesticated legumes gradually lost parts of the characteristics of their wild ancestors. While the characteristics addressing the needs of people have been accumulated and strengthened, and finally formed the modern cultivated species with high grain yield and good quality. It is documented that the domesticated legumes emphasized the selection of favorable traits, including but not limited to the reduced seed dormancy, larger seed size, palatability, root architecture, and root exudate

OPEN ACCESS

Edited and reviewed by:

Genlou Sun,
Saint Mary's University, Canada
Rodomiro Ortiz,
Swedish University of Agricultural
Sciences, Sweden

*Correspondence:

Xin Chen
cx@jaas.ac.cn

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 10 May 2021

Accepted: 28 June 2021

Published: 19 July 2021

Citation:

Chen X, Zhou G, Pang J and
Srinives P (2021) Editorial:
Domestication of Agronomic Traits in
Legume Crops.
Front. Genet. 12:707600.
doi: 10.3389/fgene.2021.707600

composition (Abbo et al., 2014; Kim et al., 2021; Liu, Ku et al.). In this Research Topic, Lei et al. have investigated the genetic architecture of seed traits (100-seed weight, seed length, seed width, and seed height) of common bean (*Phaseolus vulgaris* L.) in China, and identified QTLs that could be used in marker-assisted breeding. The nitrogen fixation symbiosis in legumes is discussed in this topic (Liu, Yu et al. Marques et al.), showing that it is important for researchers to engineer or screen for elite rhizobia that are both competitive for nodulation and capable of high rates of nitrogen fixation. Another research in this topic reports soybean genotypes to have extremely low seed phytate concentrations, combined with important root traits for efficient phytate-phosphorus acquisition, providing material for soybean breeding (Kuerban et al.).

The domestication process has reduced the genetic diversity of the cultivated germplasm and self-sustaining capability, but increased susceptibility to pests, abiotic stresses and nutrient depletion in cultivated legumes, which can be illustrated by the research of soybean and common bean in this Research Topic (Nakata et al., Shi et al., Zhang et al. and Zhong et al.). The genetics of the traits, which are important for domestication, are studied in this topic. Wang et al. construct a single nucleotide polymorphism (SNP) genetic linkage map, which is good for precisely mapping genes, and is useful for investigating the mechanism of leaf development in mung bean or legumes. Hu et al. perform transcriptome and metabolome analyses, showing that the pod degreening of the golden hook in common bean resulting from chlorophyll degradation. Berger et al. reveal that the large wild-domestic differences (e.g., water extraction, water use efficiency and harvest index) between wild and domestic *Cicer* are indicative of evolution under contrasting selection pressures. In the study of Lee et al., a locus controlling compound raceme inflorescence in mung bean is identified, which provides valuable genetic information for understanding the architecture of the compound raceme in mung bean. Ku et al. discussed the effects of domestication on secondary metabolite composition in legumes. The authors pointed out that the genes responsible for determining the secondary metabolite composition that might have been lost due to domestication, and understanding these genes would enable breeding programs and metabolic engineering to produce legume varieties with favorable secondary metabolite profiles. Another study in this topic assesses the genetic diversity of common bean in Chongqing, China, providing support for the development of excellent germplasm and genetic resources (Long et al.).

At present, molecular genetics has been used to trace the evolutionary origin and domestication history of legume crops. Analytical research on the genome segments of crop domestication was undertaken mainly through population genetics analysis approaches based on genome resequencing and quantitative trait loci analysis based on linkage disequilibrium theory. The breakthroughs of legume crops made in recent years are mainly concentrated in soybean [*Glycine max* (L.)

Merr.], common bean (*Phaseolus vulgaris* L.), rice bean (*Vigna umbellata*), and cowpea [*Vigna unguiculata* (L.)] (Jia et al., 2019). In this topic, QTL mapping, association mapping, and cloning of domesticated genes have given insight into domestication-related plant agronomic traits. Somta et al. and Amkul et al. identify QTLs controlling the agronomic traits in black gram [*Vigna mungo* (L.) Hepper] and zombi pea [*Vigna vexillata* (L.) A. Rich], respectively. Takahashi et al. performed fine-mapping in adzuki bean and yard-long bean, showing that independent domestication on the two legumes has selected the same locus for the non-shattering phenotypes. Another paper in this topic summarizes the known modifications in soybean circadian clock components as a result of domestication and improvement (Li and Lam).

In recent years, genome editing technology represented by CRISPR/Cas9 has been widely used in crops. At present, this technology is used to improve the wild relatives of crops in a targeted way to accelerate their domestication. With the deepening of the research, more domestication-related genes show high value in legume breeding. In addition, genomics techniques have been widely used in the study on the mechanism of legume crop domestication. Quantitative genetic parameters based on high-density SNP markers, such as diversity (π) and likelihood ratio (CLR), are found helpful in selection analysis. Significant diversity reduction parameter ($\pi_{\text{wild}}/\pi_{\text{cultivar}}$), population differentiation statistics *Fst*, and cross-population complete likelihood ratio (XP-CLR) approach were developed and used to identify selective sweeps between two groups (Shi and Lai, 2015). Nevertheless, the understanding of legume domestication is still very limited. The main problems to be solved are as follows: (1) The number of domestication genes cloned in legume crops is still very limited, and a large number of selection regions for domestication revealed by population genetic research still need further research. (2) Intensive genetic analysis of the cloned domesticated genes was undertaken, but research on the molecular regulation mechanism and regulation network pathway remains weak. (3) The research on legume crop domestication process mostly focuses on solving theoretical understanding problems, while the exploration of breeding and utilization approaches of related genes is still lacking.

AUTHOR CONTRIBUTIONS

All authors co-edited the Research Topic, wrote, edited, and approved the final version of editorial.

ACKNOWLEDGMENTS

The Editors would like to thank all Research Topic Authors and additional editors for their contributions.

REFERENCES

- Abbo, S., Van-Oss, R. P., Gopher, A., Saranga, Y., and Peleg, Z. (2014). Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends Plant Sci.* 19, 351–360. doi: 10.1016/j.tplants.2013.12.002
- Calles, T., Castello, R. D., Baratelli, M., Xipsiti, M., and Navarro, D. K. (2019). *The International Year of Pulses: Final Report*. Food and Agriculture Organization of the United Nations.
- Jia, G., Meng, Q., Tang, S., and Zhang, R. (2019). Current advances and future perspectives on crop domestication. *J. Plant Genet. Resour.* 20, 1355–1371. doi: 10.13430/j.cnki.jpgr.20190822001
- Kim, M. S., Lozano, R., Ji, H. K., Dong, N. B., and Jeong, S. C. (2021). The patterns of deleterious mutations during the domestication of soybean. *Nat. Commun.* 12:97. doi: 10.1038/s41467-020-20337-3
- Mudryj, A. N., Yu, N., and Aukema, H. M. (2014). Nutritional and health benefits of pulses. *Appl. Physiol. Nutr. Metab.* 39, 1–8. doi: 10.1139/apnm-2013-0557
- Shi, J., and Lai, J. (2015). Patterns of genomic changes with crop domestication and breeding. *Curr. Opin. Plant Biol.* 24, 47–53. doi: 10.1016/j.pbi.2015.01.008
- Turrill, W. B. (1926). Studies on the origin of cultivated plants. *Nature* 118, 392–393. doi: 10.1038/118392a0

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

Copyright © 2021 Chen, Zhou, Pang and Srinives. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Marker-Trait Association Analysis of Seed Traits in Accessions of Common Bean (*Phaseolus vulgaris* L.) in China

Lei Lei^{1,2}, Lanfen Wang¹, Shumin Wang¹ and Jing Wu^{1,3*}

¹ Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China, ² Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China, ³ CAAS-CIAT Joint Laboratory in Advanced Technologies for Sustainable Agriculture, Beijing, China

OPEN ACCESS

Edited by:

Xin Chen,
Jiangsu Academy of Agricultural
Sciences, China

Reviewed by:

Andrea Coppi,
University of Florence, Italy
Peerasak Srinives,
Kasetsart University, Thailand

*Correspondence:

Jing Wu
wujing@caas.cn;
wj12345@126.com

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 20 January 2020

Accepted: 08 June 2020

Published: 30 June 2020

Citation:

Lei L, Wang L, Wang S and Wu J
(2020) Marker-Trait Association
Analysis of Seed Traits in Accessions
of Common Bean (*Phaseolus vulgaris*
L.) in China. *Front. Genet.* 11:698.
doi: 10.3389/fgene.2020.00698

Seed weight and seed size are the key agronomic traits that determine yield in common bean. To investigate the genetic architecture of four seed traits (100-seed weight, seed length, seed width, and seed height) of common bean in China, marker-trait association analysis of these seed traits was performed in a nationwide population of 395 common bean accessions using 116 polymorphic SSR markers. The four seed traits were evaluated in six trials across three environments. Seed size varied among the environments. Population structure was evaluated based on SSR markers and phaseolin, which divided the accessions into two main subpopulations representing the two known gene pools. Seed weight and seed size had a strong relationship with population clustering. In addition, in a Genome-wide association studies (GWAS), 21 significantly associated markers were identified for the four seed traits with two models, namely, general linear model (GLM) and mixed linear model (MLM). Some markers had pleiotropic effects, i.e., controlled more than one trait. The significant quantitative trait loci identified in this study could be used in marker-assisted breeding to accelerate the genetic improvement of yield in common bean.

Keywords: common bean, population structure, seed weight, seed size, GWAS

INTRODUCTION

Common bean (*Phaseolus vulgaris* L.) is a crop of major societal importance with high levels of nutrients and dietary protein (Welch et al., 2000; Kutoš et al., 2003; Krupa, 2008; Montoya et al., 2010). Common bean is grown worldwide, with production exceeding 26 million tons (Tg), and China is a large producer of common bean, with the sixth highest production (1 Tg) in the world (FAO, 2018¹). The species originated in two centers of diversity and, through parallel domestication events, formed two gene pools: the Mesoamerican and Andean gene pools (Gepts et al., 1986; Gepts and Debouck, 1991; Mamidi et al., 2011, 2013; Bitocchi et al., 2013; Gaut, 2014; Schmutz et al., 2014; Rendón-Anaya et al., 2017). As a result of the domestication process, the two gene pools show differences in agronomic traits, such as seed size, seed storage protein (phaseolin) type, bracteole shape, and growth habit, among others. Among these traits, seed size is the most

¹ <http://www.fao.org/faostat/en/>

obvious differentiator of the two gene pools, with the Mesoamerican gene pool producing medium [25–40 g 100-seed weight (100SW)] and small seeds (<25 g 100SW) and the Andean gene pool producing medium and large seeds (>40 g 100SW) (Singh et al., 1991a,b,c; McClean et al., 2002, 2004; Broughton et al., 2003; Díaz and Blair, 2006). In addition, seed size is related to phaseolin type (Koenig et al., 1990). China, thought to be the secondary center of genetic diversity for common bean, was reported to include materials from the two gene pools, with seeds being mainly small to medium in size but occasionally large (Zhang et al., 2008; Bellucci et al., 2014). However, compared with soybean, rice and maize (Wang et al., 2015; Xu et al., 2015, 2018; Sangiorgio et al., 2016), common bean (especially Chinese common bean germplasm) has been the subject of few studies on the genetic control of seed traits such as seed weight and seed size.

Seed weight and seed size are the key agronomic traits that determine yield in crops (Xu et al., 2018). It has been reported that yield-related traits such as seed weight and size are typical quantitative traits controlled by a complex of genes since the Danish plant scientist Wilhelm Johannsen concluded that seed size in self-fertilizing beans is influenced by a genetic effect (Johannsen, 1911; Sax, 1923; Motto et al., 1978; Vallejos and Chase, 1991). With the construction of a linkage map based on various types of molecular markers (Vallejos et al., 1992; Nodari et al., 1993; Freyre et al., 1998; Beebe et al., 2006; Hanai et al., 2010; Galeano et al., 2011, 2012) and the completion of genome-wide sequencing of typical common bean materials from the two gene pools (Schmutz et al., 2014; Vlasova et al., 2016), a large number of quantitative trait loci (QTLs) for yield-related traits have been identified (González et al., 2017). To date, a total of approximately 200 QTLs related to seed characteristics have been reported (Nodari et al., 1993; Johnson et al., 1996; Koinange et al., 1996; Freyre et al., 1998; Park et al., 2000; Tar'an et al., 2000; Johnson and Gepts, 2002; Beattie et al., 2003; Guzmán-Maldonado et al., 2003; Blair et al., 2006; Pérez-Vega et al., 2010; Wright and Kelly, 2011; Checa and Blair, 2012; Mukeshimana et al., 2014; Yuste-Lisbona et al., 2014; Qi, 2015; Geng et al., 2017). Among these QTLs, more than 100 have been identified for seed weight, seed length (SL), seed width (SWI) and seed height (SH), and these QTLs are distributed on 11 chromosomes of the common bean genome (**Supplementary Table S1**). It is worth noting that 16 QTLs are related to at least two of these seed traits (Park et al., 2000; Geng et al., 2017). For example, SL and SH appeared to correspond to a QTL for seed weight (Park et al., 2000). In addition, a study also found the “one cause multieffect phenomenon” for seed traits on Chr02, Chr04, Chr06, and Chr07 of common bean (Geng et al., 2017). In addition, seed size QTLs were mapped near the upper end of linkage groups (LGs) 02 and 06, the lower end of LGs 03, 07, 08, and 10, and the center of LGs 06 and 08 (Park et al., 2000; Blair et al., 2006; Pérez-Vega et al., 2010). Therefore, the markers located in the center of LG B8 and near the upper end of LG B6 could be good candidates for assisted selection for traits related to seed size (Pérez-Vega et al., 2010). However, most of the loci for seed weight or size in these studies were identified in a single environment or using a family group for seed phenotype

mapping, which may have caused the parental polymorphism level to be low, in which case the accuracy of QTL mapping would have been affected to some extent.

Important goals in the current study were to (1) investigate the diversity and population structure of 395 common bean accessions with wide geographical distributions in China; (2) carry out marker-trait association analysis in six different environments using general linear model (GLM) and mixed linear model (MLM) approaches, associating phenotypic and genotypic data, with the aim of obtaining stable QTLs for seed traits; and (3) obtain significantly associated markers and provide a theoretical reference for marker-assisted selection (MAS) in the breeding of common bean in China.

MATERIALS AND METHODS

Plant Materials

A total of 395 accessions of common bean were evaluated in this study, including 307 accessions collected from the main production areas of China and 84 accessions introduced from abroad, together with four control genotypes from the Andean and Mesoamerican gene pools. The control genotypes were DRK134 and DRK139, the Andean control genotypes, as well as BAT93 and Turrialba 1, the Mesoamerican controls. The accessions were predominantly landraces; however, the study also included a few modern varieties. All of the materials were selected from the National Crop Gene Bank at the Institute of Crop Science, Chinese Academy of Agriculture Sciences, Beijing, China. The list of accessions and their passport information and source of collection are given in the **Supplementary Table S2**.

Phenotypic Evaluation

Field trials were conducted in Harbin (HRB, 45°50' N and 126°51' E), Heilongjiang Province, and in Bijie (BJ, 27°18' N and 105°18' E), Guizhou Province, in 2014 and 2016, for a total of six environments (2014_HRB, 2015_HRB, 2016_HRB, 2014_BJ, 2015_BJ, and 2016_BJ). Twenty individual plants of each accession were cultivated in two consecutive rows at both locations, and the plots were 4.0 m in length, with 0.5 m between rows.

Five plants from twenty individual plants of each accession were selected randomly, and their seeds were phenotyped. Four quantitative seed traits, namely, (1) 100SW, determined by using 100 dry seeds per plot; (2) SL, defined as the longest dimension of the seed; (3) SWI, measured as the distance between the 2 lateral sides of the seed; and (4) SH, measured as the longest distance from one side to the other at the hilum (Yuste-Lisbona et al., 2014). 100SW, SL and SH were measured by an SC-G automatic seed testing system (Hangzhou WSeen Detection Technology Co., Ltd., China), with 5 replicates of 20 seeds each, which were harvested from 5 individual plants; then, the average of the 5 replicates was calculated. Analysis of variance (ANOVA) was carried out with SPSS 19.0. Pearson correlation coefficients of the traits in each environment were calculated in SAS 9.2 (Cary software, North Carolina SAS Institute Inc., 2004) software.

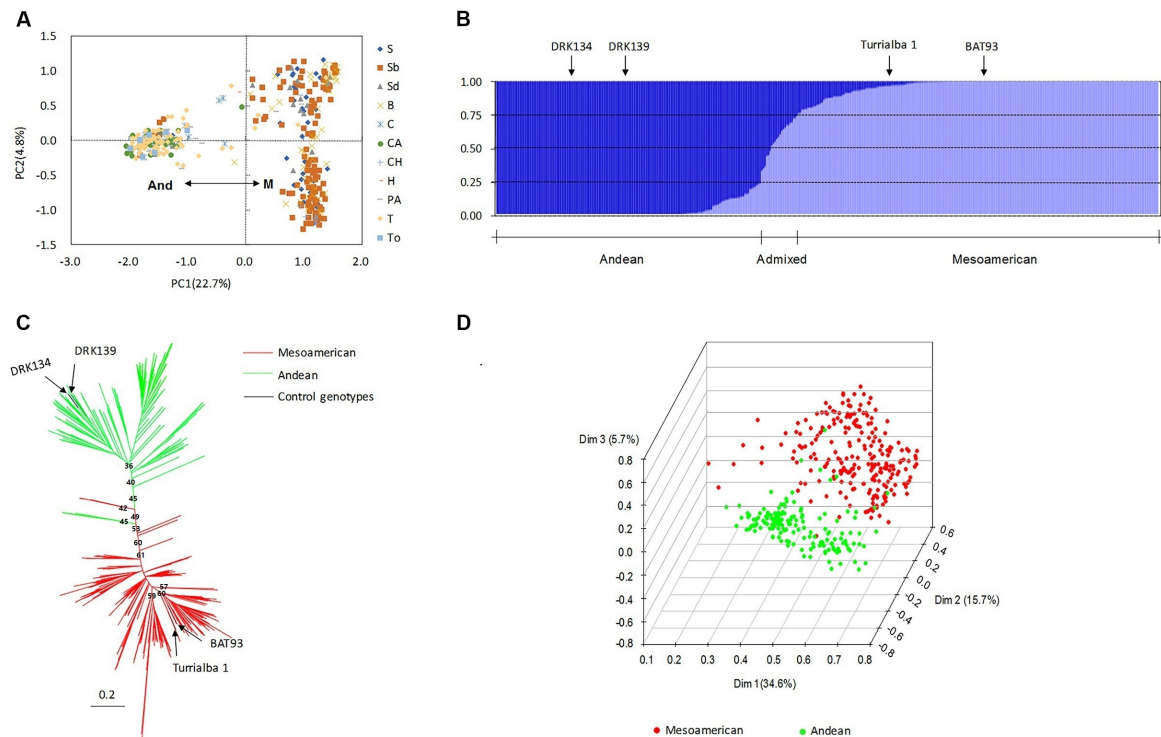


FIGURE 1 | Population structure of 395 accessions. **(A)** Principal component analysis (PCA) of the accessions based on 116 microsatellite loci and phaseolin. The grouping results obtained with phaseolin were almost the same as those obtained with the simple sequence repeat (SSR) markers. And, Andean gene pool; M, Mesoamerican gene pool. **(B)** Population structure of the accessions based on STRUCTURE analysis with $K = 2$. A total of 54.43% of the accessions were assigned to the Mesoamerican gene pool (Q1-values above 0.75), 40.00% were assigned to the Andean gene pool (Q2-values above 0.75), and 5.57% were considered admixed between the two subgroups (Q-values from 0.25 to 0.75). **(C)** Neighbor-joining (NJ) method for determining population structure with Nei's (1972) genetic distances based on SSR markers. **(D)** Three-dimensional principal coordinate analysis (PCoA) based on all 116 microsatellite markers. DRK134 and DRK139 are the Andean control genotypes, and BAT93 and Turrialba 1 are the Mesoamerican controls.

SSR Marker and Phaseolin Evaluation

To ensure an adequate sample size for DNA extraction, 10 seeds were randomly selected from each accession, germinated and grown in a greenhouse in Beijing. The first trifoliate leaves of 3-week-old seedlings were collected for total genomic DNA extraction using a CTAB extraction method (Afanador et al., 1993). DNA quality was evaluated with 1% agarose gels, and then DNA was diluted to 10 ng/ μ L for use in PCRs. PCR amplifications were carried out on an A-300 Fast Thermal Cycler using 96-well plates with 10 μ L final reaction volumes that included 1.5 μ L of genomic DNA, 1.0 μ L of each simple sequence repeat (SSR) primer (**Supplementary Material**) at a concentration of 2 pmol/L, 4.5 μ L of PCR 2 $\times$ Master Mix (Mg^{2+} , dNTP, Taq polymerase; Beijing Qingke Xinye Biotechnology Co., Ltd.) and 3.0 μ L of ddH₂O. The amplification conditions were as follows: 5 min at 95°C, followed by 35 cycles of 40 s at 95°C, 45 s at 54°C, and 40 s at 72°C and a final extension at 72°C for 10 min. The amplification products were evaluated with silver-stained 8% polyacrylamide gels with 1 $\times$ TBE buffer (89 mM Tris, 89 mM boric acid, and 2 mM EDTA). A total of 116 SSR markers distributed over all 11 chromosomes of *P. vulgaris* were used based on the whole genome sequence (Schmutz et al., 2014; Vlasova et al., 2016).

Phaseolin was extracted from 10 dry seeds of each accession, removing the testa and embryo manually from cotyledons before grinding, following the method of Igrejas et al. (2009). The treated seeds were ground into a fine powder in liquid nitrogen. The dry powder (0.1 g) was weighed and placed into 2 mL centrifuge tubes. A total of 1,000 μ L of sample extract (1% (w/v) NaCl, 0.4% (w/v) Tris-base, 0.2% (w/v) Tris-HCl, pH of 8.5) and 10 μ L of 5% (v/v) β -mercaptoethanol were added for protein extraction. Samples (3.5 μ L) were subjected to one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) following the method of Ma and Bliss (1978) modified by Igrejas et al. (2009). Electrophoresis was carried out in a 1-mm-thick, 6.38% stacking gel under 40 mA at loading; thereafter, a 12% gel at 75 mA was used until the separation was complete, requiring approximately 5–7 h. The gels were stained with Coomassie brilliant blue R-250. Phaseolin patterns were evaluated with the following standard panel of 14 phaseolin types: S, Sb, Sd, B, M₁₃, C, CA, T, PA, To, Ko, CH, H, and H₁.

Population Diversity and Structure Analyses

PowerMarker v.3.25 was used to evaluate the number of alleles, gene diversity and polymorphism information content (PIC)

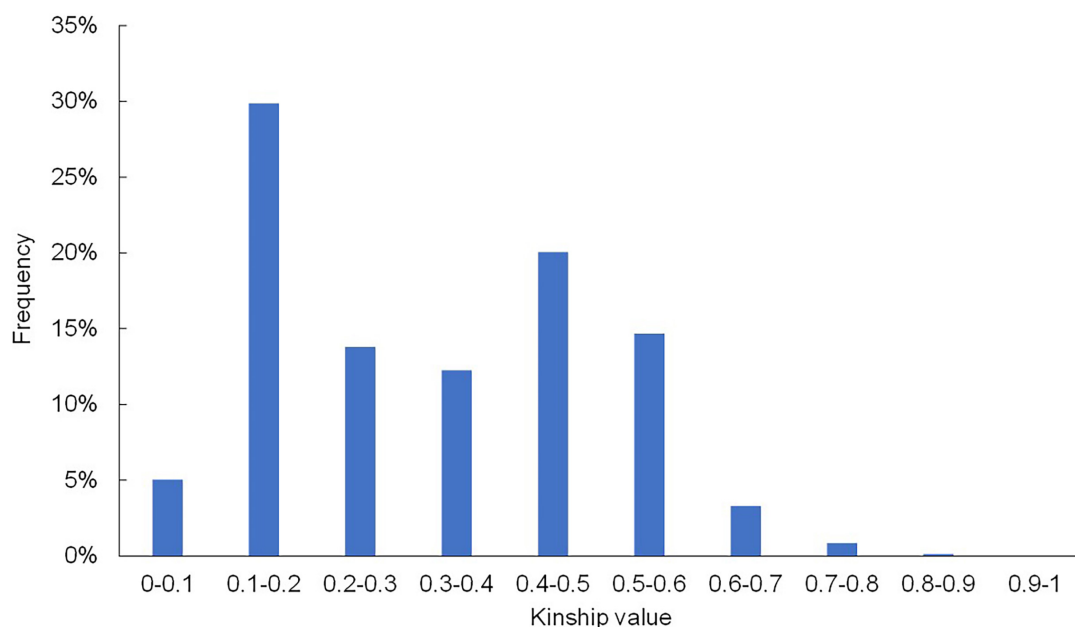


FIGURE 2 | Histogram of the frequency distribution of pairwise relative kinship coefficients. The frequency of kinship values 0–0.5 was greater than 80%, and only 0.15% of the kinship values were above 0.8.

for each marker, and clusters were analyzed to construct a dendrogram with PowerMarker² using Nei's coefficient and the neighbor-joining (NJ) algorithm (Liu and Muse, 2005).

The frequency of each phaseolin type was calculated using Microsoft Office Excel 2016. Two-dimensional principal component analysis (PCA) was performed with GenAlEx 6 based on groups of phaseolin types (Peakall and Smouse, 2006).

Population structure and the number of subpopulations (K) were evaluated with STRUCTURE 2.3.4 software (Pritchard et al., 2000; Falush et al., 2007). The assumed number of subpopulations was simulated from $k = 1$ to $k = 10$, and the ideal number of subpopulations was assessed with a burn-in period of 10,000 steps and 20,000 Markov chain Monte Carlo repetitions after the burn-in. The most likely number of subpopulations (K) was determined by examining the optimal ΔK value (Evanno et al., 2005) in STRUCTURE HARVESTER (Earl and vonHoldt, 2012). A graph (bar plot) of the population structure was generated using the R package pophelper 2.2.9³. In addition, to further analyze population structure, the relationships between subpopulations were graphed in three dimensions using NTSYSpc 2.1e (Exeter Software, Rohlf, 2000).

Association Study

Association analysis between SSR markers and various seed traits was conducted using TASSEL 2.1 (Bradbury et al., 2007). First, a GLM incorporating population genetic structure was fit, with a Q matrix derived from structure analysis as a covariate (GLM + Q). Second, an MLM incorporating a finer-scale relative

kinship matrix (K) and population genetic structure (Q) was fit to perform the association analysis (MLM + Q + K), which had more statistical power than the model including only "Q." Relatedness was determined by calculating the kinship coefficient matrix (K) in TASSEL. Meanwhile, the heritability of seed traits in six environments were calculated by MLM model using the same software program.

RESULTS

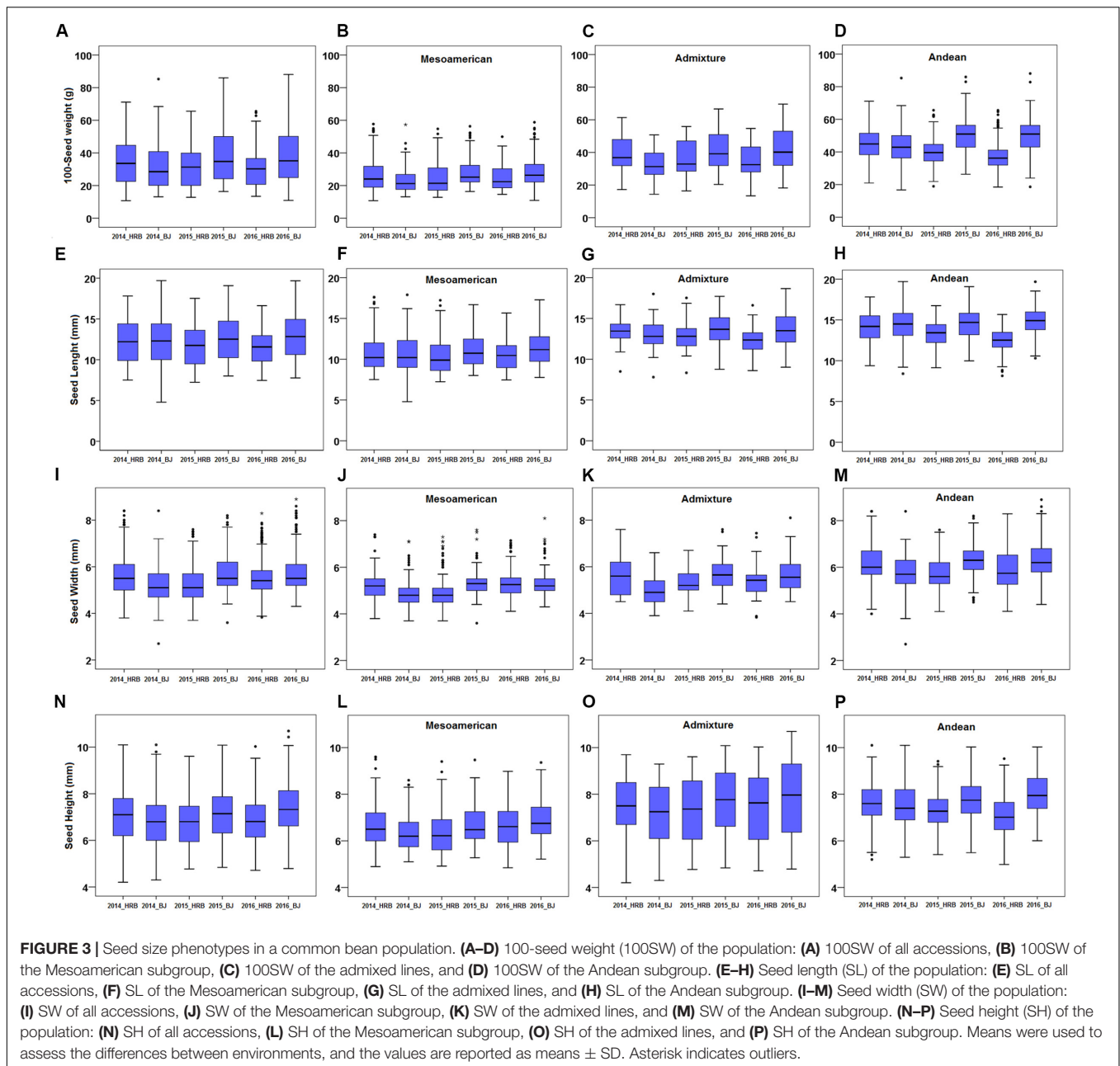
Genetic Variability

First, we evaluated the genetic diversity of the common bean accessions with phaseolin markers. Among these accessions, a total of 11 phaseolin patterns were identified: S, Sb, Sd, B, C, CA, CH, H, PA, T, and To (**Supplementary Figure 1**). The most frequent pattern was Sb (30.9%), followed by T (21.3%), which were the predominant types in the Mesoamerican and Andean gene pools, respectively. Among these eleven types of phaseolin, the S, Sb, Sd, and B types were Mesoamerican types, while C, CA, CH, H, PA, T, and To were Andean types. Therefore, this variation in protein bands could help distinguish the origins of the accessions. The results indicated that 58.2% of all the accessions belonged to the Mesoamerican gene pool while 41.8% belonged to the Andean gene pool.

Then, we explored the polymorphisms of the same accessions based on SSR markers. A total of 116 SSR markers revealed 917 different alleles, with an average number of alleles per locus of 7.9. The number of alleles identified for each SSR marker varied from 2 (CBS69, CBS170, CBS200, CBS306, P9S39, and CBS369) to 27 (CBS88). The gene diversity of individual SSR markers varied

²<http://www.powermarker.net>

³<http://royfrancis.github.io/pophelper/>



from 0.0051 (CBS200) to 0.9098 (CBS206), with an average of 0.5936 per locus.

Population Structure

Population structure is an important covariate used in association analysis to account for differentiation among panel groups and to avoid or minimize type-I errors. It is also important to study population structure in the context of genetic diversity and breeding to examine the genetic composition and relatedness of the individuals within the group. In this article, four methods were used to estimate the number of subgroups. First, PCA based on 116 SSR markers and phaseolin type was performed. Clear subpopulation structure was observed among the individuals,

and the two resulting subpopulations corresponded to the Mesoamerican gene pool (M) and the Andean gene pool (And) (Figure 1A). The percentages of genetic diversity explained by the two coordinates of the PCA were 22.7% (PC1) and 4.8% (PC2). Phaseolin types and SSR markers were able to identify the source of the accessions from the two gene pools with almost consistent results. Second, the STRUCTURE software results indicated that the value of Evanno's ΔK showed an obvious spike at $K = 2$. This result suggested that the population could be divided into two subgroups, with a few admixed individuals between these two subgroups (Figure 1A). According to the location of the control genotypes in the group, 54.43% of the accessions were assigned to the Mesoamerican gene pool (Q1

TABLE 1 | ANOVA analysis of trait in different environments.

Traits	Environment	Mean $\pm$ SD	Subgroup			Heritability (%)
			Mesoamerican	Admixture	Andean	
100SW (g)	2014_HRB	34.51 $\pm$ 13.30Bb	26.12 $\pm$ 8.99BCbc	39.40 $\pm$ 12.92Aa	45.24 $\pm$ 9.80Bb	61.29
	2014_BJ	31.46 $\pm$ 13.04Cc	22.82 $\pm$ 6.95De	31.67 $\pm$ 9.74Ab	43.25 $\pm$ 10.34Bb	53.65
	2015_HRB	31.35 $\pm$ 11.83Cc	24.45 $\pm$ 9.02CDd	36.10 $\pm$ 12.18Aa	40.09 $\pm$ 8.67Cc	59.93
	2015_BJ	37.51 $\pm$ 14.08Aa	28.00 $\pm$ 7.77ABa	40.44 $\pm$ 13.17Aa	50.05 $\pm$ 10.43Aa	54.88
	2016_HRB	30.25 $\pm$ 10.14Cc	24.95 $\pm$ 7.38Ccd	33.62 $\pm$ 11.54Aa	36.99 $\pm$ 8.95Dd	52.08
	2016_BJ	37.8 $\pm$ 14.29Aa	28.57 $\pm$ 8.40Aa	40.86 $\pm$ 13.76Aa	49.93 $\pm$ 11.19Aa	54.82
SL (mm)	2014_HRB	12.3 $\pm$ 2.63Bb	10.84 $\pm$ 2.26BCbc	13.30 $\pm$ 1.94Aab	14.13 $\pm$ 1.86Bb	59.03
	2014_BJ	12.24 $\pm$ 2.77Bb	10.67 $\pm$ 2.19BCc	12.93 $\pm$ 2.21Aab	14.29 $\pm$ 2.10ABab	58.72
	2015_HRB	11.68 $\pm$ 2.42Cc	10.38 $\pm$ 2.17Ccd	12.89 $\pm$ 2.08Aab	13.27 $\pm$ 1.65Cc	66.56
	2015_BJ	12.57 $\pm$ 2.56ABab	11.12 $\pm$ 2.06ABab	13.58 $\pm$ 2.01Aa	14.41 $\pm$ 1.92ABab	65.12
	2016_HRB	11.47 $\pm$ 2.03Cc	10.58 $\pm$ 1.92Ccd	12.32 $\pm$ 1.90Ab	12.56 $\pm$ 1.56De	64.31
	2016_BJ	12.83 $\pm$ 2.55Aa	11.38 $\pm$ 2.09Aa	13.68 $\pm$ 2.24Aa	14.69 $\pm$ 1.79Aa	68.27
SWI (mm)	2014_HRB	5.61 $\pm$ 0.86ABa	5.19 $\pm$ 0.55Ab	5.66 $\pm$ 0.93ABa	6.17 $\pm$ 0.87Aa	61.08
	2014_BJ	5.21 $\pm$ 0.77Cc	4.84 $\pm$ 0.51Bc	5.01 $\pm$ 0.76Bb	5.75 $\pm$ 0.76Bb	49.75
	2015_HRB	5.28 $\pm$ 0.79Cc	4.91 $\pm$ 0.61Bc	5.35 $\pm$ 0.73ABab	5.77 $\pm$ 0.76Bb	56.63
	2015_BJ	5.73 $\pm$ 0.80Aa	5.29 $\pm$ 0.51Aa	5.73 $\pm$ 0.86Aa	6.33 $\pm$ 0.74Aa	55.30
	2016_HRB	5.53 $\pm$ 0.77Bb	5.26 $\pm$ 0.56Aab	5.45 $\pm$ 0.93ABab	5.91 $\pm$ 0.83Bb	59.28
	2016_BJ	5.73 $\pm$ 0.85Aa	5.31 $\pm$ 0.53Aa	5.69 $\pm$ 0.89Aa	6.31 $\pm$ 0.86Aa	62.74
SH (mm)	2014_HRB	7.06 $\pm$ 1.04Bb	6.61 $\pm$ 0.88Bb	7.48 $\pm$ 1.36a	7.62 $\pm$ 0.89Bbc	24.98
	2014_BJ	6.87 $\pm$ 1.08Cc	6.37 $\pm$ 0.85Cc	7.08 $\pm$ 1.51a	7.51 $\pm$ 0.93BCcd	53.13
	2015_HRB	6.80 $\pm$ 1.02Cc	6.34 $\pm$ 0.86Cc	7.39 $\pm$ 1.42a	7.33 $\pm$ 0.84CDd	59.06
	2015_BJ	7.19 $\pm$ 1.03Bb	6.71 $\pm$ 0.82ABb	7.75 $\pm$ 1.54a	7.76 $\pm$ 0.87Bb	58.21
	2016_HRB	6.87 $\pm$ 0.96Cc	6.63 $\pm$ 0.88Bb	7.43 $\pm$ 1.49a	7.11 $\pm$ 0.87De	59.36
	2016_BJ	7.41 $\pm$ 1.07Aa	6.91 $\pm$ 0.81Aa	7.87 $\pm$ 1.66a	8.04 $\pm$ 0.91Aa	57.34

Significance of the differences in the table are based on the comparison of differences between different years of the same trait or the comparison among different years of the same sub-group of the same trait. The same traits with different capital letters in the six environments mean have extremely significant differences ($p < 0.01$), with different lowercase letters mean have significant differences ($p < 0.05$).

values above 0.75), 5.57% were considered admixed between the two subgroups (Q values from 0.25 to 0.75), and 40.00% were assigned to the Andean gene pool (Q2 values above 0.75) (**Figure 1B**). Third, a NJ dendrogram based on Nei's (1972) genetic distance also grouped the 395 accessions into two clusters (**Figure 1C**). Finally, in the three-dimensional principal coordinate analysis (PCoA) based on SSR markers showing the spatial distribution of the 395 accessions, each dimension explained 34.6% (Dim1), 15.7% (Dim2), or 5.7% (Dim3) of the variation (**Figure 1D**). The three dimensions together explained 56.0% of the total variation present in the data set. Overall, the consistent results from PCA, STRUCTURE analysis, the NJ tree and PCoA confirmed that there were two subpopulations of Chinese common bean germplasm.

In addition to population structure, the kinship (K) matrix is another important factor for association analysis. The frequency distribution values for relative kinship revealed that the genetic relatedness ranged from 0 to 0.98 (**Figure 2**), and the average pairwise relative kinship coefficient was 0.323. Pairwise relative kinship values from 0.1 to 0.2 accounted for 29.9% of all kinship coefficients. In addition, kinship values from 0 to 0.5 accounted for more than 80% of all pairwise kinship coefficients. Only 16.957% of the pairwise relative kinship coefficients were greater than 0.5, and only 0.15% of the kinship values were above 0.8.

This result suggested that the majority of the 395 accessions were genetically diverse in this study.

Phenotypic Variability in Seed Traits

Differences in growth environments and years affected the weight and size of common bean seeds. Hundred seed weight, SL, SWI and SH were evaluated in six environments, namely, 2014_HRB, 2014_BJ, 2015_HRB, 2015_BJ, 2016_HRB, and 2016_BJ. Each trait had a high degree of variation (**Figure 3**), especially SWI (**Figure 3I–M**).

There were significant differences among the environments (**Table 1**). For the four traits, the differences between years in the same location were smaller than the differences between locations, and there was almost no difference in the same location, such as between 2015_BJ and 2016_BJ or 2015_HRB and 2016_HRB. However, differences between locations were significant. In addition, the heritability of these four traits in each year and environment was analyzed, and the heritability were in the range of 50–70%, so it showed that the environment has a relatively high level of phenotypic plasticity in seed traits. These findings indicated that the environment had a strong impact on traits and the necessity of multipoint phenotypic investigations.

Furthermore, ANOVA was carried out in different subgroups in the same environment (**Table 2**). Mesoamerican, admixed and

TABLE 2 | ANOVA analysis of trait in different subgroups.

Environment	Subgroup	Trait			
		100SW (g)	SL (mm)	SWI (mm)	SH (mm)
2014_HRB	Mesoamerican	26.12 ± 8.99C	10.84 ± 2.26Bb	5.19 ± 0.55C	6.61 ± 0.88Bb
	Admixture	39.40 ± 12.92B	13.30 ± 1.94Aa	5.66 ± 0.93B	7.48 ± 1.36Aa
	Andean	45.24 ± 9.80A	14.13 ± 1.86Aa	6.17 ± 0.87A	7.62 ± 0.89Aa
2014_BJ	Mesoamerican	22.82 ± 6.95C	10.67 ± 2.19C	4.84 ± 0.51Bb	6.37 ± 0.85Bc
	Admixture	31.67 ± 9.74B	12.93 ± 2.21B	5.01 ± 0.76Bb	7.08 ± 1.51Ab
	Andean	43.25 ± 10.34A	14.29 ± 2.10A	5.75 ± 0.76Aa	7.52 ± 0.93Aa
2015_HRB	Mesoamerican	24.45 ± 9.02Bb	10.38 ± 2.17Bb	4.91 ± 0.61C	6.34 ± 0.86Bb
	Admixture	36.10 ± 12.18Aa	12.89 ± 2.08Aa	5.35 ± 0.73B	7.39 ± 1.42Aa
	Andean	40.09 ± 8.67Aa	13.27 ± 1.65Aa	5.77 ± 0.76A	7.33 ± 0.84Aa
2015_BJ	Mesoamerican	28.00 ± 7.77C	11.12 ± 2.06Bb	5.29 ± 0.51C	6.71 ± 0.82Bb
	Admixture	40.44 ± 13.17B	13.58 ± 2.01Aa	5.73 ± 0.86B	7.75 ± 1.54Aa
	Andean	50.05 ± 10.43A	14.41 ± 1.92Aa	6.33 ± 0.74A	7.76 ± 0.87Aa
2016_HRB	Mesoamerican	24.95 ± 7.38Bb	10.58 ± 1.92Bb	5.26 ± 0.56Bb	6.63 ± 0.88Bb
	Admixture	33.62 ± 11.54Aa	12.32 ± 1.90Aa	5.45 ± 0.93Bb	7.43 ± 1.49Aa
	Andean	36.99 ± 8.95Aa	12.56 ± 1.56Aa	5.91 ± 0.83Aa	7.11 ± 0.87Aa
2016_BJ	Mesoamerican	28.57 ± 8.40C	11.38 ± 2.09Bc	5.31 ± 0.53Bc	6.91 ± 0.81Bb
	Admixture	40.86 ± 13.76B	13.68 ± 2.24Ab	5.69 ± 0.89Bb	7.87 ± 1.66Aa
	Andean	49.93 ± 11.19A	14.69 ± 1.79Aa	6.32 ± 0.86Aa	8.04 ± 0.91Aa

Significance of the differences in the table are based on the comparison of differences subgroups. The same traits with different capital letters in the six environments mean have extremely significant differences ($p < 0.01$), with different lowercase letters mean have significant differences ($p < 0.05$).

Andean groups were assembled based on the grouping results from STRUCTURE. Among these three subgroups, the Andean group always produced large seeds, while the Mesoamerican group produced small seeds and the admixed group produced medium seeds. These three subgroups showed significant differences, especially the Mesoamerican and Andean subgroups.

Further exploring the correlation among the traits, we found that 100SW, SL, SWI, and SH were all significantly positively correlated with each other in all six environments (Table 3). Thus, these four seed traits were closely related to each other.

and Supplementary Table S5), and these four markers were detected in only a single environment.

Similarly, 61 significant SSR markers were associated with SL in the GLM + Q model (Figure 4), and 44 markers were detected in all six environments (Supplementary Table S4). However, only four significant markers were detected with the MLM + Q + K model, and all were located on Chr05 (Supplementary Table S5). The markers CBS162 and CBS178

Marker-Trait Association Analysis of Seed Traits

Association analysis Genome-wide association studies (GWAS) was conducted using SSR markers and the phaseolin locus (*Phs*) with six phenotypic data sets and two models, namely, GLM + Q and MLM + Q + K. A QTL for *Phs* was identified in LG 07 that spanned a region reported to code for phaseolin seed protein (Nodari et al., 1992; Koinange et al., 1996; Park et al., 2000). *Phs* has been demonstrated to be significantly associated with 100SW, SWI, and SL (Blair et al., 2009). We were interested in confirming the associations identified in previous studies of QTLs for these traits. All the results met the strict threshold of 4.3×10^{-4} and had a significance level of 0.05 after Bonferroni correction for both models.

Regarding 100SW, a total of 60 significant SSR markers were detected using the GLM + Q model (Figure 4). Among these markers, 23 were detected in all six environments (Supplementary Table S4). In contrast, only four significant SSR markers were detected with the MLM + Q + K model (Figure 4

TABLE 3 | Pearson correlation coefficients for seed traits of common bean.

Year	Traits	100SW	SL	SWI	SH
2014	100SW		0.832***	0.742***	0.787***
	SL	0.794***		0.498***	0.653***
	SWI	0.768***	0.395***		0.731***
	SH	0.666***	0.470***	0.588***	
2015	100SW		0.849***	0.784***	0.830***
	SL	0.822***		0.450***	0.662***
	SWI	0.806***	0.471***		0.697***
	SH	0.903***	0.661***	0.733***	
2016	100SW		0.814***	0.740***	0.837***
	SL	0.777***		0.655***	0.385***
	SWI	0.639***	0.312***		0.683***
	SH	0.755***	0.591***	0.604***	

The data at upper right corner in same year mean trait phenotype were collected from the locations of BJ, and at lower left corner mean trait phenotype were collected from the locations of HRB. ***indicated the significance at the 0.001 level, **indicated the significance at the 0.01 level, *indicated the significance at the 0.05 level.

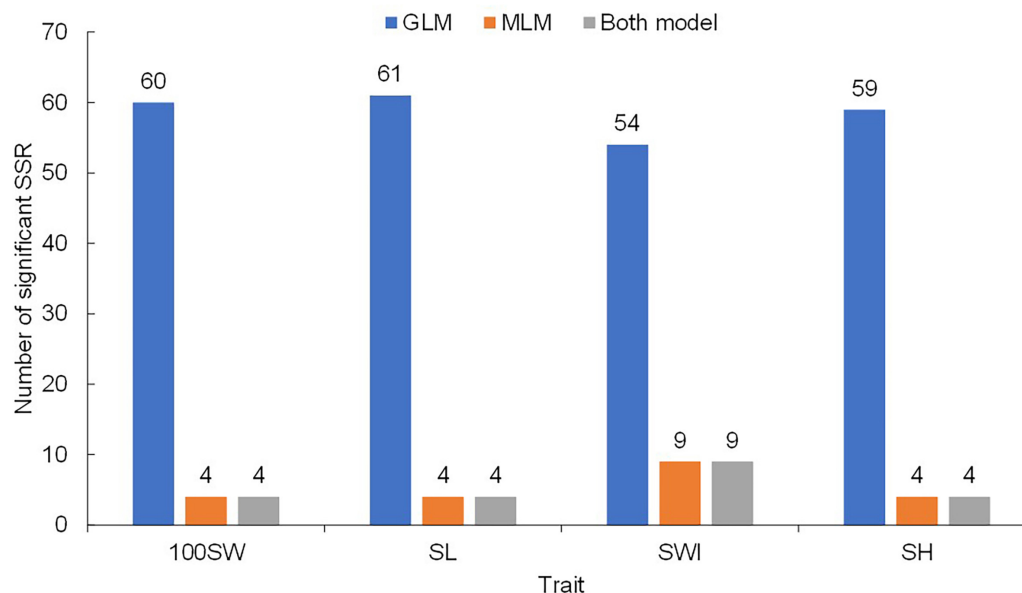


FIGURE 4 | Simple sequence repeat (SSR) markers detected by using different models.

were detected simultaneously in two environments with the MLM + Q + K model.

Similarly, 54 significant markers were associated with SWI based on the GLM + Q model, and 12 significant markers were detected in all six environments (**Supplementary Table S4**). In addition, under the MLM + Q + K model, 10 significant markers were detected (**Supplementary Table S5**). Among the 10 markers, CBS149 and CBS345 were simultaneously detected in three environments.

For SH, 59 significant markers were detected with the GLM + Q model, and 31 significant markers were detected in all six environments (**Supplementary Table S4**). However, only four significant markers were detected with the MLM + Q + K model (**Supplementary Table S5**), and these four markers were detected in a single environment.

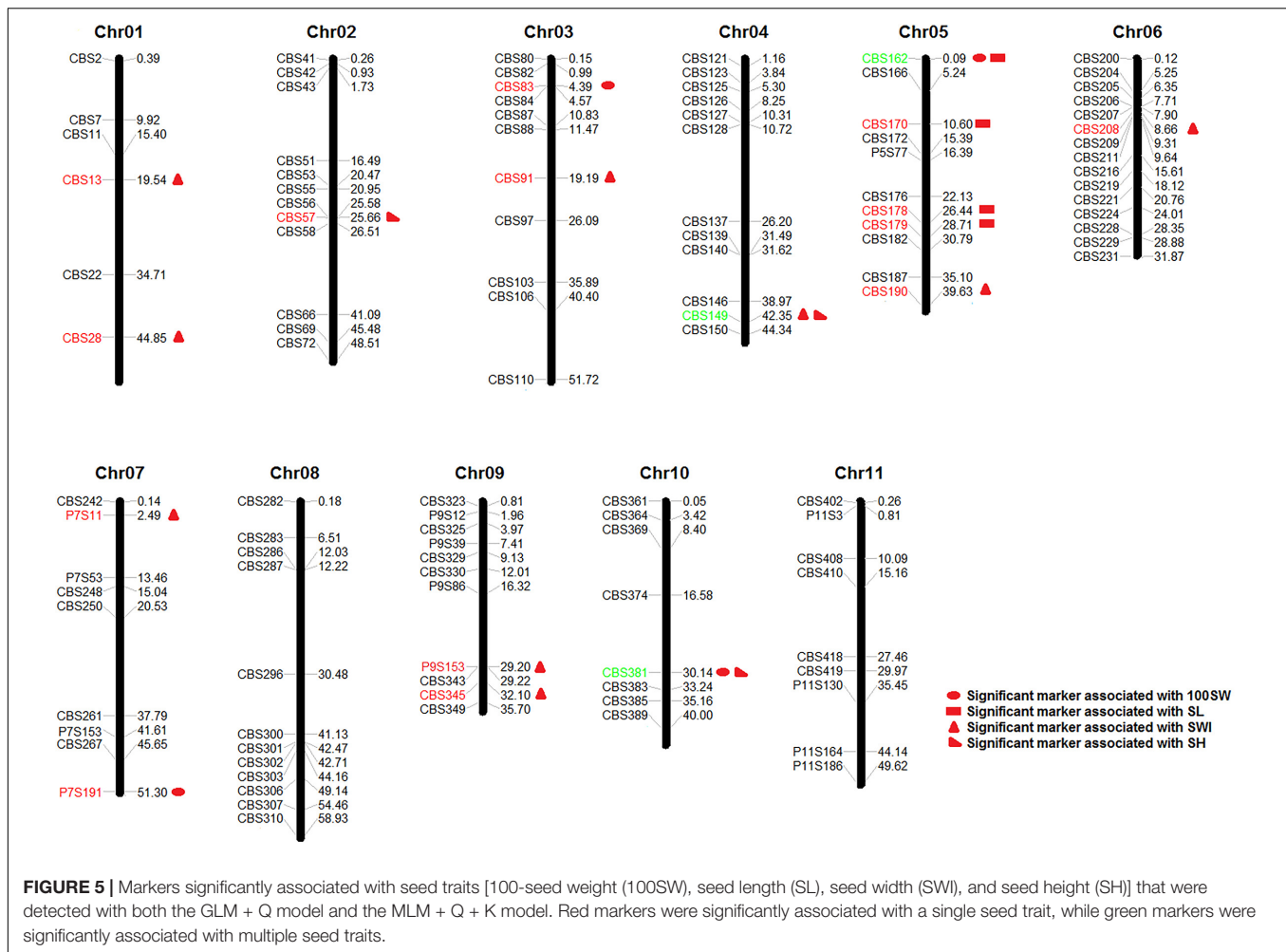
In summary, a total of 20 significant SSR markers were detected for four seed traits with both models (**Figure 5**). Some markers were simultaneously significantly associated with multiple traits, such as the marker CBS162, which was significantly associated with 100SW and SL; CBS381, which was significantly associated with 100SW and SH; and CBS149, which was significantly associated with SWI and SH. This result once again proved that there were correlations among seed traits or that the traits may have overlapping genetic control regions. Moreover, the *Phs* was found to be associated with SWI based on the two models used in our study (**Supplementary Table S6**).

DISCUSSION

In this study, we evaluated the genetic diversity of common bean accessions with phaseolin and SSR markers, and phaseolin diversity revealed that the genetic diversity of Chinese common

bean was higher than that in other countries or areas (Rodríguez et al., 2001, 2006; Blair et al., 2003; Ocampo et al., 2005; Negahi et al., 2014). We even detected several Chinese landraces with phaseolin types such as CH, To, and H, which suggested that the landraces of China have rich variation. Many studies have shown a correlation between phaseolin type and seed weight, seed size, low soil pH, growth habit, precocity and antiparasitic traits (Hartana and Bliss, 1984; Gepts and Bliss, 1986; Koenig et al., 1990; Johnson et al., 1996; Escribano et al., 1998). Therefore, it is very important to identify the phaseolin type of China's preserved common bean germplasm, which will facilitate yield and insect resistance breeding.

Regarding population structure, Phaseolin and SSR markers from the present study both reveal Chinese common bean germplasms containing two gene pools materials, and the main group of Chinese accessions were of Mesoamerican origin, with fewer of Andean origin. This is different from European countries, where Andean genotypes were more prevalent than Mesoamerican (Raggi et al., 2013; Maras et al., 2015; Caroviae-Stanko et al., 2017; Pipan and Megliè, 2019). This different distribution of two gene pools genotypes in countries might be related to the time of germplasm introduction, adaptation abilities of germplasm, ecological types, the political regulation within these countries in the past, and consumer preferences. Additionally, distance-based analyses cluster analysis based on microsatellite markers in congruence with the results of phaseolin type analysis up to 88.6% in this study. These results showed that Phaseolin and SSR markers have sufficient power to distinguish between the Mesoamerican and Andean germplasms, and clear about the background of Chinese germplasm. The old and elite common bean landraces of China also has a diverse genetic background based on SSR markers (**Supplementary Figure 2**). Therefore, it is important to deeply explore the



excellent germplasm resources of Chinese common bean and to tap the potential of Chinese common bean to increase yield and improve quality.

In the association analysis, the phaseolin marker *Phs* was significantly associated with SWI (Supplementary Table S6), which agreed with the result from a previous seed size association study (Blair et al., 2006). For the significant SSR markers, we selected a suitable linkage disequilibrium (LD) decay distance of approximately 100 Kb as the confidence interval for screening target trait candidate genes (Rossi et al., 2009). The marker P7S191, which was significantly associated with 100SW, was located near the QTL SW7' in a previous study (Geng et al., 2017). We also identified significantly associated markers on *Chr03* and *Chr10* that a previous study reported as QTLs for 100SW (Nodari et al., 1993; Johnson et al., 1996; Freyre et al., 1998; Park et al., 2000; Guzmán-Maldonado et al., 2003; Blair et al., 2006). The confidence intervals of the two SSR markers CBS178 and CBS179 on *Pv05*, which were significant for SL, were (24014961, 24215231) and (29909932, 30110144), respectively. Both of these markers were included in the interval of a QTL (20.89–36.47 Mb on *Pv05*) previously reported for SL based on linkage analysis (Geng et al., 2017). These results are more helpful for confirming

QTLs location related to seed traits. Nevertheless, the candidate QTLs for seed traits require further verification with improved accuracy such as SNP analysis due to the limitations of this study. Apart from these, we also found the phenomenon of marker one cause multiple effects among seed traits. There have been similar reports in previous studies on seed size of common bean (Blair et al., 2009; Geng et al., 2017). For example, Blair et al. (2009) reported the marker BM183 significantly associated with seed weight, width and length also in both genepools accessions. For this pleiotropism, how to balance the effect of the seed traits is essential to improve the yield. It is proposed that the ratio between the length, width and height of seed be correlated with genetic markers to obtain a trade-off between the two traits (Geng et al., 2017). Inspired by this, we can further obtain high yields by associated the ratios of seed length/width, length/height, width/height and length × width/height, length × height/width with SSR marker to balance each seed trait.

CONCLUSION

This study provides a comprehensive picture of genetic diversity and structure of Chinese common bean accessions. Out of 395

accessions, 54.43% were of Mesoamerican origin, 40.00% of Andean, while 5.57% of accessions represented putative hybrids between gene pools. For the most part, the classification of common bean accessions according to phaseolin type analysis was in congruence with the results of distance-based analyses of SSR marker. Based on the population structure, 20 significant SSR markers and 1 significant phaseolin marker were associated with the 4 seed traits by GLM and MLM in this study. Association data of seed traits can be used for common bean breeding, especially in terms of its adaptation to different environments. Thus, by combining and genomic selection, we can effectively select the effective QTL alleles for seed weight and size in breeding populations. The results of this study provide a valuable resource to dissect the role of candidate QTL locus regions of the genome governing seed weight/size in common bean.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

REFERENCES

- Afanador, L. K., Haley, S. D., and Kelly, J. D. (1993). Adoption of a “mini prep” DNA extraction method for RAPD marker analysis in common bean (*Phaseolus vulgaris* L.). *Annu. Rep. Bean Improv. Coop.* 36, 10–11.
- Beattie, A. D., Larsen, J., Michaels, T. E., and Pauls, K. P. (2003). Mapping quantitative trait loci for a common bean (*Phaseolus vulgaris* L.) ideotype. *Genome* 46, 411–422. doi: 10.1139/g03-015
- Beebe, S. E., Rojas-Pierce, M., Yan, X., Blair, M. W., Pedraza, F., Munoz, F., et al. (2006). Quantitative trait loci for root architecture traits correlated with phosphorus acquisition in common bean. *Crop Sci.* 46, 413–423. doi: 10.2135/cropsci2005.0226
- Bellucci, E., Nanni, L., Biagetti, E., Bitocchi, E., Giardini, A., Rau, D., et al. (2014). “Common bean origin, evolution and spread from America,” in *Phaseolus: A New World Gift to mankind Why Common Beans are so Common?*, ed. D. Rubiales (Córdoba: International Legume Society), 12–16.
- Bitocchi, E., Bellucci, E., Giardini, A., Rau, D., Rodriguez, M., Biagetti, E., et al. (2013). Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytol.* 197, 300–313. doi: 10.1111/j.1469-8137.2012.04377.x
- Blair, M. W., Díaz, L. M., Buendía, H. F., and Duque, M. C. (2009). Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). *Theor. Appl. Gene.* 119, 955–972. doi: 10.1007/s00122-009-1064-8
- Blair, M. W., Giraldo, M. C., Duran, L., Beaver, J., and Nin, J. C. (2003). Phaseolin characterization of Caribbean common bean germplasm. *Annu. Rep. Bean Improv. Coop.* 46, 63–64.
- Blair, M. W., Iriarte, G., and Beebe, S. (2006). QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean × wild common bean (*Phaseolus vulgaris* L.) cross. *Theor. Appl. Gene* 112, 1149–1163. doi: 10.1007/s00122-006-0217-2
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Broughton, W. J., Hernandez, G., Blair, M., Beebe, S., Gepts, P., and Vanderleyden, J. (2003). Beans (*Phaseolus* spp.)—model food legumes. *Plant Soil* 252, 55–128. doi: 10.1023/A:1024146710611

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the National Key R&D Program of China (2018YFD1000700/2018YFD1000704), the Ministry of Agriculture of China [the earmarked fund for the China Agriculture Research System (CARS-08)], the Agricultural Science and Technology Innovation Program of CAAS, and the Fundamental Research Funds for the Chinese Academy of Agricultural Sciences (Y2020GH14-2).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00698/full#supplementary-material>

- Caroviae-Stanko, K., Liber, Z., Vidak, M., Barešić, A., Grdiša, M., Lazarević, B., et al. (2017). Genetic diversity of croatian common bean landraces. *Front. Plant Sci.* 8:604. doi: 10.3389/fpls.2017.00604
- Checa, O. E., and Blair, M. W. (2012). Inheritance of yield-related traits in climbing beans (*Phaseolus vulgaris* L.). *Crop Sci.* 52, 1998–2013. doi: 10.2135/cropsci2011.07.0368
- Díaz, L. M., and Blair, M. W. (2006). Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. *Theor. Appl. Gene.* 114, 143–154. doi: 10.1007/s00122-006-0417-9
- Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conser. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-95487
- Escribano, M. R., Santalla, M., Casquero, P. A., and De Ron, A. M. (1998). Patterns of genetic diversity in landraces of common bean (*Phaseolus vulgaris* L.) from Galicia. *Plant Breed.* 117, 49–56. doi: 10.1111/j.1439-0523.1998.tb01447.x
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Falush, D., Stephens, M., and Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7, 574–578. doi: 10.1111/j.14718286.2007.01758.x
- Freyre, R., Skroch, P. W., Geffroy, V., Adam-Blondon, A. F., Shirmohamadali, A., Johnson, W. C., et al. (1998). Towards an integrated linkage map of common bean. 4. Development of a core linkage map and alignment of RFLP maps. *Theor. Appl. Gene* 97, 847–856. doi: 10.1007/s001220050964
- Galeano, C. H., Cortés, A. J., Fernández, A. C., Soler, Á., Franco-Herrera, N., Makunde, G., et al. (2012). Gene-based single nucleotide polymorphism markers for genetic and association mapping in common bean. *BMC Genet.* 13:48. doi: 10.1186/1471-2156-13-48
- Galeano, C. H., Fernandez, A. C., Franco-Herrera, N., Cichy, K. A., McClean, P. E., Vanderleyden, J., et al. (2011). Saturation of an intra-gene pool linkage map: towards a unified consensus linkage map for fine mapping and synteny analysis in common bean. *PLoS One* 6:e28135. doi: 10.1371/journal.pone.0028135
- Gaut, B. S. (2014). The complex domestication history of the common bean. *Nat. Genet.* 46, 663–664. doi: 10.1038/ng.3017
- Geng, Q. H., Wang, L. F., Wu, J., and Wang, S. M. (2017). QTL mapping for seed size and shape in common bean. *Acta Agron. Sin.* 43, 1149–1160.

- Gepts, P., and Bliss, F. A. (1986). Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. *Econ. Bot.* 40, 469–478. doi: 10.1007/bf02859660
- Gepts, P., and Debouck, D. (1991). “Origin, domestication, and evolution of the common bean (*Phaseolus vulgaris* L.),” in *Common Beans: Research for Crop Improvement*, eds A. Schoonhoven and O. Voysest (Colombia: CIAT press), 7–53.
- Gepts, P., Osborn, T. C., Rashka, K., and Bliss, F. A. (1986). Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. *Econ. Bot.* 40, 451–468. doi: 10.1007/BF02859659
- González, A. M., Yuste-Lisbona, F. J., Fernández-Lozano, A., Lozano, R., and Santalla, M. (2017). “Genetic mapping and QTL analysis in common bean,” in *The Common Bean Genome*, eds S. Pérez de la Vega, M. Santalla, and F. Marsolais (Switzerland: Springer International Publishing AG press), 69–107. doi: 10.1007/978-3-319-63526-2_4
- Guzmán-Maldonado, S. H., Martínez, O., Acosta-Gallegos, J. A., Guevara-Lara, F., and Paredes-Lopez, O. (2003). Putative quantitative trait loci for physical and chemical components of common bean. *Crop Sci.* 43, 1029–1035. doi: 10.2135/cropsci2003.1029
- Hanai, L. R., Santini, L., Camargo, L. E. A., Fungaro, M. H. P., Gepts, P., Tsai, S. M., et al. (2010). Extension of the core map of common bean with EST-SSR, RGA, AFLP, and putative functional markers. *Mol. Breed.* 25, 25–45. doi: 10.1007/s11032-009-9306-7
- Hartana, A., and Bliss, F. A. (1984). Genetic variability in seed protein levels associated with two phaseolin protein type in common bean (*Phaseolus vulgaris* L.). *Annu. Rep. Bean Improv. Coop.* 27:159.
- Igrejas, G., Carnide, V., Pereira, P., Mesquita, F., and Guedes-Pinto, H. (2009). Genetic diversity and phaseolin variation in Portuguese common bean landraces. *Plant Genet. Resour.* 7, 230–236. doi: 10.1017/s1479262109264124
- Johannsen, W. (1911). The genotype conception of heredity. *Am. Nat.* 45, 129–259.
- Johnson, W. C., and Gepts, P. (2002). The role of epistasis in controlling seed yield and other agronomic traits in an Andean×Mesoamerican cross of common bean (*Phaseolus vulgaris* L.). *Euphytica* 125, 69–79. doi: 10.1023/A:1015775822132
- Johnson, W. C., Menéndez, C., Nodari, R., Koinange, E. M., Magnusson, S., Singh, S. P., et al. (1996). Association of a seed weight factor with the phaseolin seed storage protein locus across genotypes, environments, and genomes in *Phaseolus*-*Vigna* spp.: Sax (1923) revisited. *J. Agric. Genomics* 2, 1–19.
- Koenig, R. L., Singh, S. P., and Gepts, P. (1990). Novel phaseolin types in wild and cultivated common bean (*Phaseolus vulgaris*, Fabaceae). *Econ. Bot.* 44, 50–60. doi: 10.1007/BF02861066
- 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
- Krupa, U. (2008). Main nutritional and antinutritional compounds of bean seeds—a review. *Pol. J. Food Nutr. Sci.* 58, 149–155. doi: 10.7589/0090355844.3.731
- Kutoš, T., Golob, T., Kač, M., and Plestenjak, A. (2003). Dietary fibre content of dry and processed beans. *Food Chem.* 80, 231–235. doi: 10.1016/S0308-8146(02)00258-3
- Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282
- Ma, Y., and Bliss, F. A. (1978). Seed proteins of common bean 1. *Crop Sci.* 18, 431–437. doi: 10.2135/cropsci1978.0011183X001800030018x
- Mamidi, S., Rossi, M., Annam, D., Moghaddam, S., Lee, R., Papa, R., et al. (2011). Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. *Funct. Plant Biol.* 38, 953–967. doi: 10.1071/FP11124
- Mamidi, S., Rossi, M., Moghaddam, S. M., Annam, D., Lee, R., Papa, R., et al. (2013). Demographic factors shaped diversity in the two gene pools of wild common bean *Phaseolus vulgaris* L. *Heredity*. 110, 267–276. doi: 10.1038/hdy.2012.82
- Maras, M., Pipan, B., Utar-Vozli, J., Todorovi, V., and Megliè, V. (2015). Examination of genetic diversity of common bean from the western Balkans. *J. Am. Soc. Hortic. Sci.* 140, 308–316. doi: 10.21273/JASHS.140.4.308
- McClean, P., Gepts, P., and Kamir, J. (2004). Genomics and genetic diversity in common bean. *Legume Crop Genomics* 61–82.
- McClean, P. E., Lee, R. K., Otto, C., Gepts, P., and Bassett, M. J. (2002). Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean (*Phaseolus vulgaris* L.). *J. Hered.* 93, 148–152. doi: 10.1093/jhered/93.2.148
- Montoya, C. A., Lallès, J. P., Beebe, S., and Leterme, P. (2010). Phaseolin diversity as a possible strategy to improve the nutritional value of common beans (*Phaseolus vulgaris*). *Food Res. Int.* 43, 443–449. doi: 10.1016/j.foodres.2009.09.040
- Motto, M., Soressi, G. P., and Salamini, F. (1978). Seed size inheritance in a cross between wild and cultivated common beans (*Phaseolus vulgaris* L.). *Genetica* 49, 31–36. doi: 10.1007/BF00187811
- Mukeshimana, G., Butare, L., Cregan, P. B., Blair, M. W., and Kelly, J. D. (2014). Quantitative trait loci associated with drought tolerance in common bean. *Crop Sci.* 54, 923–938. doi: 10.2135/cropsci2013.06.0427
- Negahi, A. Z. A. M., Bihamta, M. R., and Negahi, Z. A. H. R. A. (2014). Diversity in Iranian and exotic common bean (*Phaseolus vulgaris* L.) using seed storage protein (Phaseolin). *Agric. Commun.* 2, 34–40.
- Nei, M. (1972). Genetic distance between populations. *Am. Nat.* 106, 283–292. doi: 10.1086/282771
- Nodari, R. O., Koinange, E. M. K., Kelly, J. D., and Gepts, P. (1992). Towards an integrated linkage map of common bean. *Theor. Appl. Gene* 84, 186–192. doi: 10.1007/bf00223999
- Nodari, R. O., Tsail, S. M., Gilbertson, R. L., and Gepts, P. (1993). Towards an integrated linkage map of common bean 2. Development of an RFLP-based linkage map. *Theor. Appl. Gene* 85, 513–520. doi: 10.1007/BF00220907
- Ocampo, C. H., Martín, J. P., Sánchez-Yélamo, M. D., Ortiz, J. M., and Toro, O. (2005). Tracing the origin of Spanish common bean cultivars using biochemical and molecular markers. *Genet. Resour. Crop Evol.* 52, 33–40. doi: 10.1007/s10722-005-1931-3
- Park, S. O., Coyne, D. P., Jung, G., Skroch, P. W., Arnaud-Santana, E., Steadman, J. R., et al. (2000). Mapping of QTL for seed size and shape traits in common bean. *J. Am. Soc. Hortic. Sci.* 125, 466–475. doi: 10.21273/JASHS.125.4.466
- Peakall, R. O. D., and Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6, 288–295. doi: 10.1111/j.1471-8286.2005.01155.x
- Pérez-Vega, E., Pañeda, A., Rodríguez-Suárez, C., Campa, A., Giraldez, R., and Ferreira, J. J. (2010). Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Gene* 120, 1367–1380. doi: 10.1007/s00122-010-1261-5
- Pipan, B., and Megliè, V. (2019). Diversification and genetic structure of the western-to-eastern progression of European *Phaseolus vulgaris* L. germplasm. *BMC Plant Biol.* 19:442. doi: 10.1186/s12870-019-2051-0
- Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. (2000). Association mapping in structured populations. *Am. J. Hum. Genet.* 67, 170–181. doi: 10.1086/302959
- Qi, Y. (2015). *Characterization of a Putative Yield-Related Gene in Common Bean (Phaseolus vulgaris L.)*. Ph.D. dissertation, The University of Guelph, Canada.
- Raggi, L., Tiranti, B., and Negri, V. (2013). Italian common bean landraces: diversity and population structure. *Genet. Resour. Crop Evol.* 60, 1515–1530. doi: 10.1007/s10722-012-9939-y
- Rendón-Anaya, M., Montero-Vargas, J. M., Saburido-Álvarez, S., Vlasova, A., Capella-Gutiérrez, S., Ordaz-Ortiz, J. J., et al. (2017). Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biol.* 18:60. doi: 10.1186/s13059-017-1190-6
- Rodiño, A. P., Santalla, M., González, A. M., De Ron, A. M., and Singh, S. P. (2006). Novel genetic variation in common bean from the Iberian Peninsula. *Crop Sci.* 46, 2540–2546. doi: 10.2135/cropsci2006.02.0104
- Rodiño, A. P., Santalla, M., Montero, I., Casquero, P. A., and De Ron, A. M. (2001). Diversity of common bean (*Phaseolus vulgaris* L.) germplasm from Portugal. *Genet. Resour. Crop Evol.* 48, 409–417.
- Rossi, M., Bitocchi, E., Bellucci, E., Nanni, L., Rau, D., Attene, G., et al. (2009). Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evol. Appl.* 2, 504–522. doi: 10.1111/j.1752-4571.2009.00082.x
- Sangiorgio, S., Carabelli, L., Gabotti, D., Manzotti, P. S., Persico, M., Consonni, G., et al. (2016). A mutational approach for the detection of genetic factors affecting

- seed size in maize. *Plant Reprod.* 29, 301–310. doi: 10.1007/s00497-016-0294-6
- Sax, K. (1923). The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8, 552–560. doi: 10.1007/BF02983088
- 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., Gepts, P., and Debouck, D. G. (1991a). Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Econ. Bot.* 45, 379–396. doi: 10.1007/BF02887079
- Singh, S. P., Gutierrez, J. A., Molina, A., Urrea, C., and Gepts, P. (1991b). Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. *Crop Sci.* 31, 23–29. doi: 10.2135/cropsci1991.0011183X003100010005x
- Singh, S. P., Nodari, R., and Gepts, P. (1991c). Genetic diversity in cultivated common bean: I. Allozymes. *Crop Sci.* 31, 19–23. doi: 10.2135/cropsci1991.0011183X00310001000-4x
- Tar'an, B., Michaels, T. E., and Pauls, K. P. (2000). Genetic mapping of agronomic traits in common bean (*Phaseolus vulgaris* L.). *Annu. Rep. Bean Improv. Coop.* 43, 60–61.
- Vallejos, C. E., and Chase, C. D. (1991). Linkage between isozyme markers and a locus affecting seed size in *Phaseolus vulgaris* L. *Theor. Appl. Gene* 81, 413–419. doi: 10.1007/BF00228685
- Vallejos, C. E., Sakiyama, N. S., and Chase, C. D. (1992). A molecular marker-based linkage map of *Phaseolus vulgaris* L. *Genetics* 131, 733–740. doi: 10.1002/elps.1150130187
- Vlasova, A., Capella-Gutiérrez, S., Rendon-Anaya, M., Hernández-Oñate, M., Minoche, A. E., Erb, I., et al. (2016). Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. *Genome Biol.* 17, 1–18. doi: 10.1186/s13059-016et-0883-6
- Wang, X., Li, Y., Zhang, H., Sun, G., Zhang, W., and Qiu, L. (2015). Evolution and association analysis of GmCYP78A10 gene with seed size/weight and pod number in soybean. *Mol. Biol. Rep.* 42, 489–496. doi: 10.1007/s11033-014-3792-3
- Welch, R. M., House, W. A., Beebe, S., and Cheng, Z. (2000). Genetic selection for enhanced bioavailable levels of iron in bean (*Phaseolus vulgaris* L.) seeds. *J. Agric. Food Chem.* 48, 3576–3580. doi: 10.1021/jf0000981
- Wright, E. M., and Kelly, J. D. (2011). Mapping QTL for seed yield and canning quality following processing of black bean (*Phaseolus vulgaris* L.). *Euphytica* 179, 471–484. doi: 10.1007/s10681-011-0369-2
- Xu, C., Liu, Y., Li, Y., Xu, X., Xu, C., Li, X., et al. (2015). Differential expression of GS5 regulates grain size in rice. *J. Exp. Bot.* 66, 2611–2623. doi: 10.1093/jxb/erv058
- Xu, R., Duan, P., Yu, H., Zhou, Z., Zhang, B., Wang, R., et al. (2018). Control of grain size and weight by the OsMKKK10-OsMCK4-OsMAPK6 signaling pathway in rice. *Mol. Plant* 11, 860–873. doi: 10.1016/j.molp.2018.04.004
- Yuste-Lisbona, F. J., González, A. M., Capel, C., García-Alcázar, M., Capel, J., De Ron, A. M., et al. (2014). Genetic analysis of single-locus and epistatic QTLs for seed traits in an adapted × nuña RIL population of common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Gene* 127, 897–912. doi: 10.1007/s00122-014-2265-3
- Zhang, X., Blair, M. W., and Wang, S. (2008). Genetic diversity of Chinese common bean (*Phaseolus vulgaris* L.) landraces assessed with simple sequence repeat markers. *Theor. Appl. Gene* 117, 629–640. doi: 10.1007/s00122-008-0807-2

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

Copyright © 2020 Lei, Wang, Wang and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



QTL Mapping for Agronomic and Adaptive Traits Confirmed Pleiotropic Effect of *mog* Gene in Black Gram [*Vigna mungo* (L.) Hepper]

OPEN ACCESS

Edited by:

Badri Padhukasahasram,
Illumina, United States

Reviewed by:

Dawei Xin,
Northeast Agricultural University,
China
Fa Cui,
University of Chinese Academy
of Sciences, China
Liu Jin Dong,
Chinese Academy of Agricultural
Sciences, China

*Correspondence:

Prakit Somta
agrpks@ku.ac.th
Xin Chen
cx@jaas.ac.cn

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 12 February 2020

Accepted: 26 May 2020

Published: 30 June 2020

Citation:

Somta P, Chen J, Yimram T,
Yundaeng C, Yuan X, Tomooka N and
Chen X (2020) QTL Mapping
for Agronomic and Adaptive Traits
Confirmed Pleiotropic Effect of *mog*
Gene in Black Gram [*Vigna mungo* (L.)
Hepper]. *Front. Genet.* 11:635.
doi: 10.3389/fgene.2020.00635

Prakit Somta^{1,2,3*}, Jingbin Chen¹, Tarika Yimram², Chutintorn Yundaeng¹,
Xingxing Yuan¹, Norihiko Tomooka⁴ and Xin Chen^{1*}

¹ Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China, ² Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Nakhon Pathom, Thailand, ³ Center of Excellence on Agricultural Biotechnology: (AG-BIO/PERDO-CHE), Bangkok, Thailand, ⁴ Genetic Resources Center, Gene Bank, National Agriculture and Food Research Organization, Tsukuba, Japan

Organ size and architecture of plants are important traits affecting crop yield and agronomic practices. An induced mutant, multiple-organ gigantism (MOG), of black gram (*Vigna mungo*) has been obtained, which shows gigantic leaves, fruit, seed, and architecture (plant height) but lower number of pods per plant. These traits are a pleiotropic effect of a single recessive gene, *mog*. In this study, we investigated variation of 16 agronomic and adaptive traits in a recombinant inbred line (RIL) population derived from a cross between the MOG mutant (*V. mungo* var. *mungo*) and wild black gram (*V. mungo* var. *silvestris*) accession TC2210 and identified quantitative trait loci (QTLs) controlling those traits to gain a better understanding of the effect of the *mog* gene on breeding. The results showed that most of the traits (100-seed weight, leaf size, and plant height) showed moderate narrow-sense heritability (h^2) (45–65%), while pod size and seed length (SDL) showed high h^2 (>75%) and pod dehiscence (shattering), and seed width (SDW) and days to flowering showed low h^2 (<35%). The QTLs for the traits were mapped onto a high-density linkage map developed for the RIL population. Inclusive composite interval mapping identified 42 QTLs in total for the 16 traits with number of QTLs per trait ranging from one to six. Major QTLs for the MOG phenotypes were clustered on linkage group (LG) 6, confirming the pleiotropic effect of the *mog* gene. Effect of the *mog* gene/QTL for the MOG phenotypic variations was not high, ranging from about 15% in plant height to 40% in leaf size. For 100-seed weight, which is the most interesting trait, the *mog* gene/QTL contributed about 30% of the total trait variation and showed an additive effect of only 0.51 g, which is only about 1.5-fold higher than that of the other five QTLs detected for this trait. These results

indicated that *mog* gene expression is highly affected by environment and the effect of the gene toward organ size and plant height is not extraordinarily high. Implications of the findings of this study and exploiting of the MOG mutant in breeding were also discussed.

Keywords: black gram, *Vigna mungo*, domestication, QTL, pleiotropic effect, pleiotropy

INTRODUCTION

Black gram [*Vigna mungo* (L.) Hepper] is an important legume crop in Asia, especially South Asia and Southeast Asia. The total planting area of black gram is about 5.5 million hectares, concentrated mainly in India, Myanmar, Pakistan, Bangladesh, Nepal, and Thailand. Because of its early maturity (about 80–90 days) and relative drought tolerance, black gram fits well in several cropping systems, although it is generally grown after rice, maize, or wheat. Owing to its ability in association with soil bacteria to fix atmospheric nitrogen to the soil, growing black gram improves soil fertility and can increase the yield of the crops grown after it. In South Asia, dry black gram seeds are cooked into a thick soup and consumed as a staple dish, while seed flour is used to prepare several dishes and sweets. In some countries such as Thailand and Japan, black gram seeds are preferred in the sprout industry (Somta et al., 2019). India and Myanmar are major producers of black gram cultivating about 3.5 and 1.0 million hectares of black gram each, respectively. Like several legume crops, the average seed yield of black gram is very low, e.g., about 650–800 kg/ha in India and Thailand.

Seed yield is a complex quantitative trait controlled by several genes with small genetic effects. Seed yield per plant is mainly determined by yield-related traits, such as seed size (weight), number of pods per plants, and number of seeds per pods. In general, seed size of improved black gram cultivars is relatively small; 100-seed weight (SD100WT) is about 4.00–5.50 g, which is lower than that of other species in the same genus such as mungbean [*Vigna radiata* (L.) Wilczek; SD100WT = 6.5–7.5 g], adzuki bean [*Vigna angularis* (Ohwi) Ohwi and Ohashi; SD100WT > 15.0 g], and cowpea [*Vigna unguiculata* (L.) Walp.; SD100WT > 15.0 g] (Tomooka et al., 2002). Breeding for increased seed size and yield is one of the major objectives in black gram. Mutagenesis is a fundamental method to artificially accelerating natural mutation of gene(s) in organisms. Induced mutation is the basis for the development of new plant cultivars (Ahloowalia and Maluszynski, 2001) and the processes of super-domestication of plants (“processes that lead to a domesticate with dramatically increased yield that could not be selected in natural environments from naturally occurring variation without recourse to new technologies”) (Vaughan et al., 2007). In Thailand, a black gram mutant showing a gigantic organ size was obtained by gamma-irradiation of the cultivar “Phitsanulok 2” (Chinchest and Nakeerak, 1990) and named the multiple-organ-gigantism (MOG) mutant (Tomooka et al., 2010). Compared with the wild type, the MOG mutant significantly expressed larger seeds (69%), higher biomass (40%), and bigger leaves (72%), albeit with a lower number of seeds per pod (75%), number of pods per plants (76%), and number of seeds per

plant (57%) (Tomooka et al., 2010). A map-based cloning study of the *mog* locus revealed that an 8-bp deletion in exon 6 in *VmPPD* (*V. mungo* *PEAPOD*) results in the MOG phenotypes (Naito et al., 2017). Knocking down the *VmPPD* orthologs in soybean resulted in larger seeds and leaves, but lower number of pods and seeds and lower seed yield per plant than a control soybean (Naito et al., 2017). However, to efficiently exploit useful traits such as increased seed size in the MOG mutant in black gram breeding, additional information on the pleiotropic effect of *VmPPD* on traits in other background genomes needs to be determined.

In this study, we report quantitative trait locus (QTL) mapping of agronomic and adaptive traits in a recombinant inbred line (RIL) population derived from the cross between the *mog* mutant and wild black gram (*V. mungo* var. *silvestris* Lukoki, Maréchal and Otoul). The aim of this study was to determine the effect of the *mog* gene on agronomic and adaptive traits, especially yield-related and plant architectural traits.

MATERIALS AND METHODS

Mapping Population

In this study, a RIL population of 150 lines generated from a cross between the MOG mutant and TC2210 by a single-seed descent (SSD) method was used. The SSD was started at the F₃ generation. This population was the same population used by Somta et al. (2019) to map QTL for bruchid resistance. The MOG mutant was named “BC48” in their study and “JP219132” in a study by Naito et al. (2017). TC2210 is a wild black gram from India, which shows several contrasting traits to the *mog* mutant.

Evaluation of Agronomic and Adaptive Traits

The F₁₀ RILs and their parents were grown in a randomized complete block design with two replicates under field conditions at Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, Thailand, during June to August 2017. Spacing between rows was 0.5 m, and the spacing between plants in the same row was 0.25 m. In each replicate, each line comprising 10 plants. Six plants between the first and last plants in the same line were randomly chosen for data recording. Sixteen agronomic and adaptive traits were recorded (Table 1). Leaf width (LFW), leaf length (LFL), and leaf area (LFA) were measured at 50 days after sowing using five fully expanded leaves. LFA was measured using an LI-3100C area meter (LI-COR Biosciences, United States). The number of days from sowing to first flowering (FLD) and the number of days from sowing to harvesting of the first pod (PDDM) were recorded.

TABLE 1 | Agronomic and adaptive traits evaluated in a black gram RIL population derived from a cross between *mog* mutant (cultivated) and TC2210 (wild).

Organ/attribute	Traits	Trait abbreviation	QTL	Evaluation
Leaf	Width (cm)	LFW	<i>Lfw</i>	Maximum width of fully expanded leaf
	Length (cm)	LFL	<i>Lfl</i>	Maximum length of fully expanded leaf
	Area (mm ²)	LFA	<i>Lfa</i>	Area of fully expanded leaf
Stem	Thickness (mm)	STT	<i>Stt</i>	Stem diameter under the primary leaf (measured at flowering stage)
	Number of branches per plant	BRNPP	<i>Brnpp</i>	Number of branches along the stem (measured at harvest)
	Plant height (cm)	PLH	<i>Plh</i>	Length from ground to terminal shoot
Pod	Number of twists of pods (count)	PDT	<i>Pdt</i>	Number of twists along the length of shattered pod
	Length (cm)	PDL	<i>Pdl</i>	Length of straight pod
	Width (mm)	PDW	<i>Pdw</i>	Maximum width
	Number of seeds per pod	SDNPP	<i>Sdnpp</i>	Number of seeds per pods
Seed	100-Seed weight (g)	SD100WT	<i>Sd100wt</i>	Weight of 100 seeds
	Length (mm)	SDL	<i>Sdl</i>	Maximum distance from top to bottom of the seed
	Width (mm)	SDW	<i>Sdw</i>	Maximum distance from hilum to its opposite side
	Water absorption (%)	SDWA	<i>Sdwa</i>	Seeds that absorbed water at 7 days after sowing at room temperature
Phenology	Days to first flower (day)	FLD	<i>Fld</i>	Number of days from planting to first flowering
	Days to maturity of first pod (day)	PDDM	<i>Pddm</i>	Number of days from first flowering to harvesting of first pod

Stem thickness (STT) was measured at harvest. Seed length (SDL) and seed width (SDW) were recorded as average values of five seeds. The 100-seed weight (SD100WT) was weighted using intact seeds. The number of seeds per pod (SDNPPD) and number of twists along the pod when kept at room temperature (PDT) were counted using five pods. PDT was used as an index of pod dehiscence. Pod length (PDL) and pod width (PDW) were recorded using five pods. Seed water absorption (SDWA) was measured as an index of seed dormancy; for this, 20 intact seeds were placed on wet filter paper and incubated in the dark at 25°C for 9 days, and the number of seeds that imbibed water was recorded. SDWA was determined although the parents showed similar expression; the population showed segregating for the trait.

Statistical Analysis and Estimation of Heritability for Traits

Correlation (r^2) between traits was calculated using R Program 10.2.0 (R Core Team, 2010). Analysis of variance of each trait was carried out to estimate narrow-sense heritability (h^2). The h^2 of each trait was subsequently calculated using the following formula: $h^2 = \sigma_g^2 / [\sigma_g^2 + (\sigma_e^2/r)]$, where σ_g^2 and σ_e^2 are the variances of the RILs and experimental error, respectively, and r is the number of replicates.

Genetic Linkage Map and QTL Analysis

A genetic linkage map previously constructed for the RIL population using SNP markers (Somta et al., 2019) generated using the specific locus amplified fragment sequencing (SLAF-seq) technique (Sun et al., 2013) was used for QTL analysis. The map comprised 3,675 markers clustered into a linkage map of 11 linkage groups (LGs) corresponding to the haploid chromosome number of black gram. The map had a total length of 1,588.7 cM with an average distance between adjacent markers of 0.57 cM. The number of markers per LG varied between 134 on LG10 and 665 on LG3 with an average of 334.1 markers. LG4 was the

shortest LG (83.4 cM), whereas LG7 was the longest (185.8 cM). The average length of the LGs was 144.43 cM. The marker density was 2.3 markers per cM. Since several markers were mapped to the same position, only one marker at such positions was selected for QTL analysis.

Quantitative trait locus analysis was performed using the ICIM method (Li et al., 2007) by QTL IciMapping 4.1 (Meng et al., 2015). ICIM was performed at every 0.1 cM with the probability in stepwise regression (PIN) of 0.001. A 1,000-permutation test at $P = 0.05$ was carried out to determine the significant LOD threshold for QTLs of each trait.

RESULTS

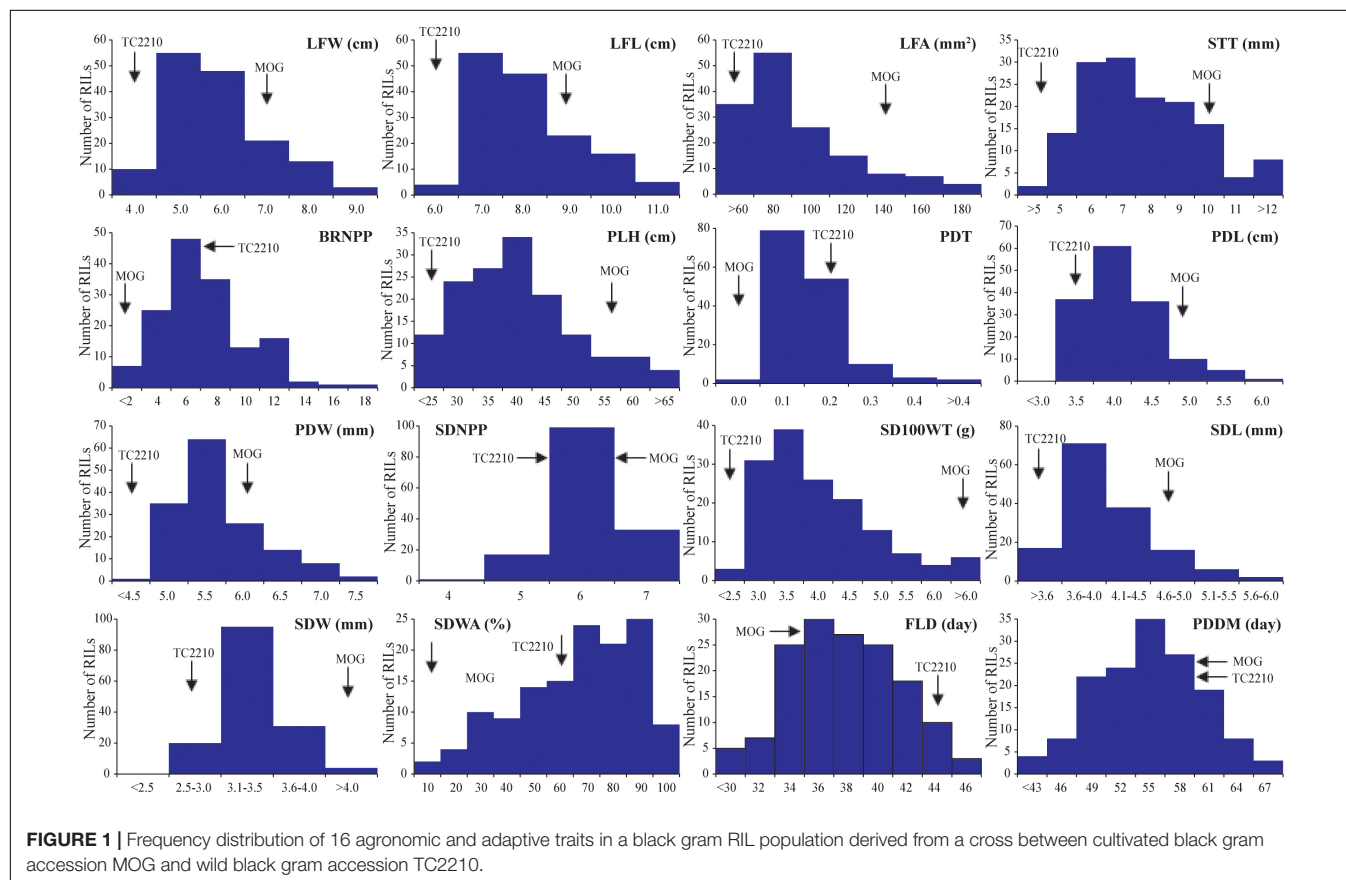
Trait Variations in Parents and RIL Population

The means of the parents and the means and range of traits of the RIL population are summarized in **Table 2**. The MOG mutant showed much higher values compared with TC2210 for all the traits measured apart from BRNPP, PDT, and FLD for which the MOG mutant showed lower values than TC2210. The parents all showed similar SDNPPD, SDWA, and PDDM values. Apart from PDDM, the means of all the traits of the RIL population were between the means of the parents. Analysis of variance revealed that the RILs were statistically different for all the traits (**Supplementary Table S1**). All the traits showed a continuous distribution (**Figure 1**). Clear transgressive segregation was found in LFA, STT, BRNPP, PLH, PDT, SDWA, and PDDM (**Figure 1** and **Table 2**), suggesting that both the MOG mutant and TC2210 possessed alleles that increased or decreased the values of these traits. It is worth noting that none of the RILs possessed a similar SD100WT to that of the MOG mutant. The RIL with the largest SD100WT was 6.5 g, about 1.5 g smaller than the MOG mutant (**Table 2**). Correlation between traits is shown in **Supplementary Table S2**. Significant

TABLE 2 | Basic descriptive statistics for agronomic and adaptive traits in a RIL population derived from a cross between cultivated black gram accession MOG and wild black gram accession TC2210.

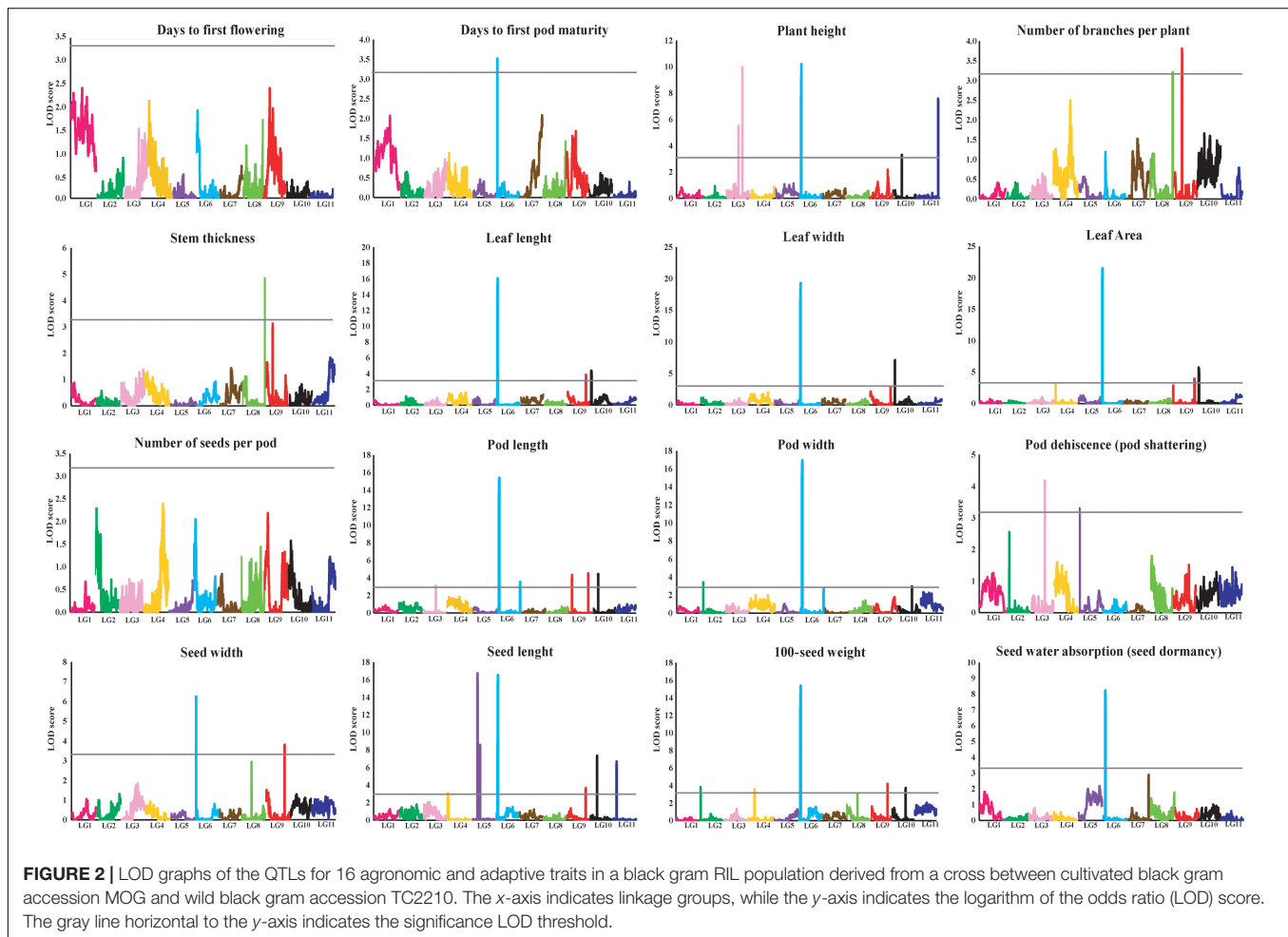
Trait*	MOG mutant	TC2210	RIL population		
			Min-max	Mean $\pm$ SE	h^2
LFW (cm)	7.07	3.63	3.40–8.67	5.36 $\pm$ 1.09	66.47
LFL (cm)	9.50	5.80	5.82–10.68	7.54 $\pm$ 1.09	49.48
LFA (mm ²)	130.43	41.01	40.97–173.19	82.51 $\pm$ 29.38	65.19
STT (mm)	10.31	4.23	3.35–14.38	7.23 $\pm$ 2.09	66.83
BRNPP	1.97	5.60	0.58–18.00	6.26 $\pm$ 3.00	57.23
PLH (cm)	53.28	20.50	17.75–76.17	37.47 $\pm$ 10.21	58.29
PDT	0.01	0.24	0.00–0.54	0.11 $\pm$ 0.08	34.86
PDL (cm)	5.75	3.40	3.08–5.69	3.87 $\pm$ 0.50	83.09
PDW (mm)	7.13	4.58	4.38–5.69	3.87 $\pm$ 0.58	84.82
SDNPP	5.66	6.13	3.76–6.63	5.60 $\pm$ 0.48	47.40
SD100WT (g)	7.99	2.03	2.31–6.54	3.78 $\pm$ 0.96	64.49
SDL (mm)	5.00	2.83	3.12–5.84	4.02 $\pm$ 0.51	78.88
SDW (mm)	4.89	2.83	2.51–4.26	3.32 $\pm$ 0.29	24.26
SDWA (%)	75.20	77.00	5.13–97.33	63.30 $\pm$ 22.44	58.93
FLD	36.00	43.00	28.50–46.00	37.02 $\pm$ 3.67	14.30
PDDM	58.00	58.00	42.00–67.00	53.61 $\pm$ 5.38	48.36

Narrow-sense heritability (h^2) of the traits in the RIL population is also included. *See description for trait abbreviation in Table 1.



and positive correlations were found among seed-related traits and leaf-related traits. Seed-related traits also showed positive correlation with leaf-related traits. Interestingly, seed-related,

pod-related, and leaf-related traits showed correlation with SDWA, PLH, LFD, and PDMM. STT and BRNPP showed correlation with FLD and PDMM. These correlations suggested



that these traits are possibly under the control of common genetic factor(s).

Narrow-Sense Heritability (h^2) of Traits in the RIL Population

In general, h^2 values estimated for all the traits in the RIL population were moderate or high (Table 2). Most of the traits showed moderate h^2 values of about 45–65%. Only PDL, PDW, and SDL showed high h^2 values, >75%, while PDT, SDW, and FLD showed low h^2 values, <35%.

QTLs for Traits

The results of the QTL analysis of each trait are summarized in Table 3. QTL LOD graphs of all the traits are shown in Figure 2. In total, 42 QTLs were identified for the 16 traits. The number of QTLs detected per trait ranged from one to six. Nonetheless, no significant QTL was detected for SDNPP, FLD, and PDM. Major QTLs of several traits were mapped to a small region on LG 6 (Table 3 and Figure 3).

In general, the MOG mutant had much larger organs than the wild black gram TC2210, with larger leaves, pods, and seeds in particular (Table 3). One to seven QTLs of traits related to

seed size, SD100WT, SDL, and SDW, were located on LGs 2, 5, 6, 8, 9, 10, and 11. At all QTLs except *qSdl5.2* on LG 5, alleles from the MOG mutant increased the size of each trait. The QTLs with the largest phenotypic variance explained (PVE) for SD100WT, SDL, and SDW were located on LG 6. These QTLs were close to each other. The QTL on LG6 of SD100WT showed the largest effect with a PVE of about 30%. Six QTLs on LGs 3, 6, 9, and 10 were detected for PDL, while two QTLs on LGs 6 and 10 were detected for PDW (Table 3). At all these QTLs for pod size, alleles from the MOG mutant increased pod size. The QTLs with the largest PVE for PDL and PDW were located on LG 6 and were in the same region. Three QTLs of LFL were identified on LGs 6, 9, and 10 (Table 3). Three QTLs of LFW were found on LGs 6 and 10 (Table 3). The QTLs on LG 6 and 10 for both traits were located at similar positions. The QTLs on LG6 showed the largest effect toward leaf size with a PVE of about 40%.

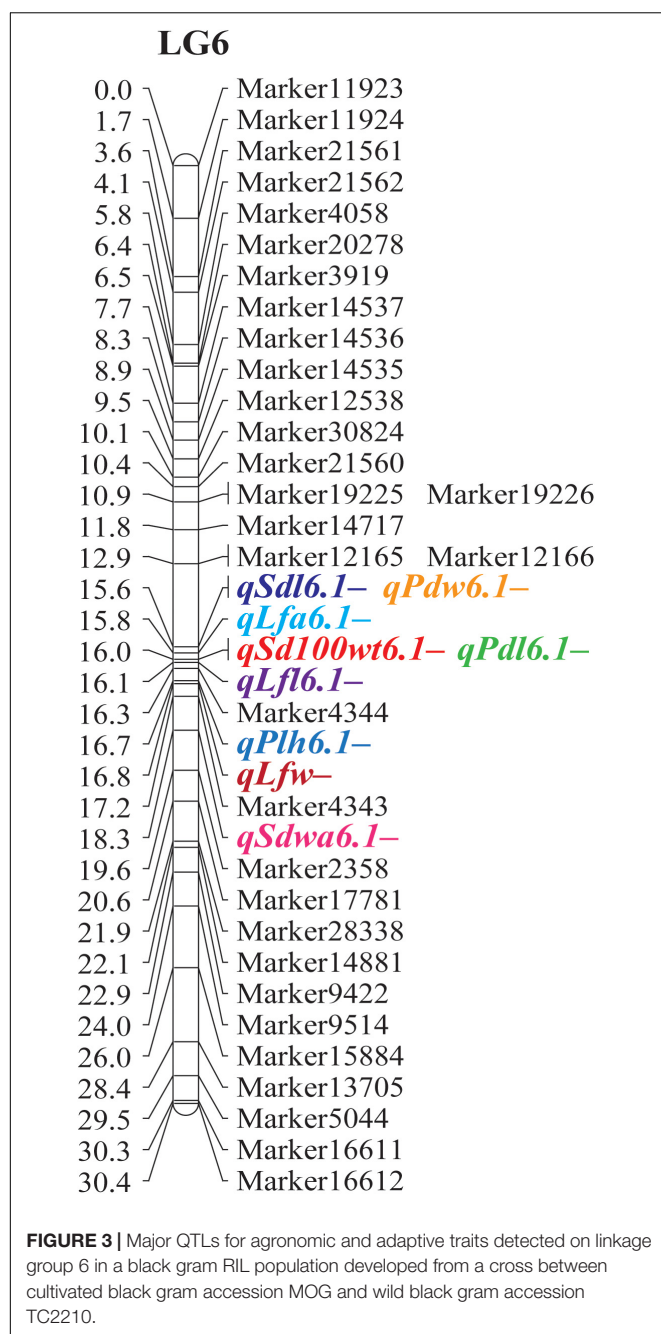
The MOG mutant showed higher STT, PLH, and BRNPP than the wild black gram parent TC2210. Two QTLs of STT were detected on LGs 3 and 6 (Table 3). Interestingly, the alleles from the MOG mutant at both QTLs decreased the STT. Four QTLs of PLH were found on LGs 3, 6, and 10 (Table 3). The alleles from the MOG mutant increased PLH at all these QTLs

TABLE 3 | Details of the QTLs detected for agronomic and adaptive traits in a black gram RIL population developed from a cross between cultivated black gram accession MOG and wild black gram accession TC2210.

Trait	Linkage group (LG)	QTL name	Position (cM)	Flanking markers		LOD score	Phenotypic variance explained (PVE, %)	Additive effect
				Left marker	Right marker			
PLH	3	<i>qPlh3.1+</i>	67.2	Marker2035	Marker2666	5.50	7.05	3.16
		<i>qPlh3.2-</i>	88.0	Marker15495	Marker18446	9.95	13.77	-4.32
	6	<i>qPlh6.1+</i>	16.7	Marker4344	Marker4343	10.21	14.09	4.70
	10	<i>qPlh3.4+</i>	50.0	Marker6869	Marker2731	3.32	4.09	2.36
BRNPP	8	<i>qBrnpp8.1-</i>	174.8	Marker3986	Marker5791	3.21	8.28	-0.88
	9	<i>qBrnpp9.1-</i>	59.8	Marker3713	Marker14454	3.81	9.88	-0.97
STT	8	<i>qStt8.1-</i>	178.1	Marker544	Marker3459	4.85	12.96	-0.75
	9	<i>qStt9.1-</i>	49.3	Marker6958	Marker5572	3.1 + 3	8.13	-0.59
LFL	6	<i>qLfl6.1+</i>	16.1	Marker12165	Marker4344	16.26	36.59	0.66
	9	<i>qLfl9.1+</i>	139.0	Marker4365	Marker7073	3.85	6.96	0.27
	10	<i>qLfl10.1+</i>	7.4	Marker22647	Marker20892	4.37	8.25	0.30
LFW	6	<i>qLfw6.1+</i>	16.8	Marker4344	Marker4343	19.33	40.11	0.69
	10	<i>qLfw10.1+</i>	6.3	Marker22876	Marker5816	7.12	12.08	0.36
LFA	4	<i>qLfa4.1+</i>	7.2	Marker3989	Marker10031	3.20	4.41	6.05
	6	<i>qLfa6.1+</i>	15.8	Marker12165	Marker4344	21.44	39.90	19.17
	9	<i>qLfa9.1+</i>	142.8	Marker26537	Marker8648	3.97	5.65	6.77
	10	<i>qLfa10.1+</i>	6.4	Marker5816	Marker22647	5.72	8.06	8.11
PDL	3	<i>qPdl3.1+</i>	70.4	Marker3330	Marker9631	3.09	4.94	0.10
	6	<i>qPdl6.1+</i>	16.0	Marker12165	Marker4344	15.41	30.33	0.26
		<i>qPdl6.2</i>	155.0	Marker13751	Marker13752	3.55	5.69	0.11
	9	<i>qPdl9.1+</i>	24.6	Marker1826	Marker2920	4.37	7.12	0.12
		<i>qPdl9.2</i>	136.7	Marker2793	Marker2792	4.55	7.41	0.12
	10	<i>qPdl10.1+</i>	29.9	Marker4712	Marker662	4.47	7.37	0.12
PDW	2	<i>qPdw2.1+</i>	12.6	Marker1902	Marker1901	3.45	6.38	0.13
	6	<i>qPdw6.1+</i>	15.6	Marker12165	Marker4344	16.97	40.21	0.35
PDT	3	<i>qPdt3.1-</i>	78.9	Marker23015	Marker22081	4.18	10.29	-0.03
	5	<i>qPdt5.1-</i>	3.6	Marker9103	Marker9104	3.32	8.02	-0.02
SDW	9	<i>qSdw9.1+</i>	132.1	Marker10591	Marker10592	3.81	7.09	0.09
SDL	4	<i>qSdl4.1+</i>	4.5	Marker15323	Marker12765	3.07	2.73	0.10
	5	<i>qSdl5.1+</i>	30.4	Marker10757	Marker9932	16.74	18.52	0.25
		<i>qSdl5.2-</i>	44.3	Marker5405	Marker17861	8.58	8.47	-0.18
	6	<i>qSdl6.1+</i>	15.6	Marker12165	Marker4344	16.55	18.88	0.27
	9	<i>qSdl9.1+</i>	139.0	Marker4365	Marker7073	3.65	3.26	0.11
	10	<i>qSdl10.1+</i>	39.4	Marker1947	Marker6202	7.34	6.95	0.15
	11	<i>qSdl11.1+</i>	21.0	Marker17291	Marker4235	6.69	6.28	0.15
	2	<i>qSd100wt2.1+</i>	4.4	Marker16001	Marker10353	3.83	6.18	0.22
	4	<i>qSd100wt4.1+</i>	22.8	Marker21597	Marker11828	3.59	5.80	0.21
	6	<i>qSd100wt6.1+</i>	16.0	Marker12165	Marker4344	15.39	30.68	0.51
SD100WT	8	<i>qSd100wt8.1+</i>	84.0	Marker22513	Marker6575	3.05	4.88	0.19
	9	<i>qSd100wt9.1+</i>	137.4	Marker2792	Marker4365	4.19	6.81	0.23
	10	<i>qSd100wt10.1+</i>	79.4	Marker14003	Marker14004	3.73	6.02	0.21
	6	<i>qSdwa6.1-</i>	18.3	Marker4343	Marker2358	8.24	20.57	-11.30
SDNPPD				No QTL found				
FLD				No QTL found				
PDMM				No QTL found				

except *qPlh3.2-*. Two QTLs of BRNPP were found on LGs 8 and 9 (Table 3). At both QTLs, the alleles from the MOG mutant decreased the number of branches. The QTLs on LG8 were at a similar position. All of the QTLs of these traits showed a PVE smaller than 15%.

Loss or reduction of seed dormancy and pod dehiscence (shattering) are important adaptive traits of cultivated crops grown in agricultural fields. Domestication of wild species of legume and cereal crops generally results in reduced seed dormancy and pod shattering, which are useful for timing and



uniform germination and ease of reap, respectively. A single QTL, *qSdwa6.1+* on the LG6, was found for SDWA (Table 3). At this QTL, alleles from the MOG mutant increased seed water permeability. Two QTLs of PDT were detected on LGs 3 and 5, *qPdt3.1-* and *qPdt5.1-*, respectively (Table 3). At both QTLs, alleles from the MOG mutant decreased pod shattering.

DISCUSSION

In this study, we investigated phenotypes of the MOG mutant black gram, which shows gigantism in seed, pod, and leaf,

in a RIL population derived from crossing the MOG mutant and a wild progenitor of black gram, accession TC2210, and successfully mapped the QTL controlling the MOG phenotypes to the linkage map.

The MOG mutant showed more gigantism in all the agronomic traits than the wild black gram and especially in seed weight and LFA (Table 2). The seed weight and LFA of the MOG mutant were about fourfold and threefold larger than those of the wild black gram, respectively. The difference in size of organs between the MOG mutant and TC2210 is considered a super-domestication trait (Vaughan et al., 2007), a combination of domestication (natural mutation) and induced mutation. The *mog* gene has pleiotropic effects on seed size, stem size, leaf size, and pod size (Chinchest and Nakeerak, 1990; Tomooka et al., 2010; Naito et al., 2017). This was supported by the significant and moderate or high correlation among plant height, leaf size, pod size, and seed size in the segregating RIL population (Supplementary Table S2). However, we found that STT did not correlate with leaf-size-related traits, although it correlated with pod-size-related and seed-size-related traits. This suggested that increase in stem size may not be a result of the *mog* gene.

Although the *mog* gene showed a marked effect toward gigantism of plant organs of black gram (Chinchest and Nakeerak, 1990; Tomooka et al., 2010; Naito et al., 2017) (see also Table 3), estimated h^2 values for most of the traits in the RIL population were moderate (Table 2). As well as suggesting the polygenic inheritance nature of the traits, it also indicated that expression of the *mog* gene is highly affected by environment. The narrow-sense heritability value estimated for the seed weight in our study (64%) was even higher than the average broad-sense heritability value of 47% reported for the same trait in black gram (Fery, 1980). Large environmental effect(s) toward expression of the MOG phenotypes is possibly due to the fact that the gigantic phenotypes are caused by increases in cell numbers caused by the mutation in the MOG (*VmPPD*) gene (Naito et al., 2017). An increase in cell number is a direct result of an increase in cell division, a process that needs optimum and suitable conditions/environments (Sablowski and Dornelas, 2013).

Quantitative trait locus analysis revealed that size of seed, leaf, and pod was controlled by a common QTL mapped on LG 6 (Table 3 and Figure 3). *qSd100wt6.1+*, *qSdl6.1+*, and *qSdwa6.1+* for seed size, *qLfl6.1+*, *qLfw6.1+*, and *qLfa6.1+* for leaf size, and *qPdl6.1+* and *qPdw6.1+* for pod size were located near to one another at around 16.0 cM. These QTLs were between the markers Marker17781 to Marker2358. Based on the reference genome of mungbean (Kang et al., 2014), the most closely related species of black gram, Marker17781 to Marker2358, are on chromosome 8 and are 804.2 Mb apart (Supplementary Table S3). There are 83 annotated genes in the 804.2-kb region. The *mog* gene corresponds to the annotated gene *LOC106772031* in mungbean (Supplementary Table S4). The *mog* gene locates on chromosome 8 and is not far from Marker4344, being about 259.0 kb away (Supplementary Table S3). These further confirmed the pleiotropic effects of the *mog* gene toward seed, leaf, pod, and plant (height) size. Among the gigantic traits

exerted by the effect of the *mog* gene, increased seed weight is the most interesting and valuable in plant breeding because it directly contributes to seed yield in black gram. However, it is worth mentioning that in this study the additive effect of the *mog* gene toward 100-seed weight was 0.51 g, which is only about onefold higher than that of the other loci controlling this trait (Table 3). In addition, the *mog* gene accounted for only 30% of the variation in seed weight in the RIL population (Table 3). These results contrast with previous findings that the additive effect of the *mog* gene for 100-seed weight was as high as 2.00 g and the gene explained up to 66% of the seed weight variation in an F₂ population (MOG mutant × TC2210) (Naito et al., 2017). In fact, the RIL population used in our study was developed from the same parents used by Naito et al. (2017). The numbers of QTLs identified for the seed weight in the two studies are similar, being 6 and 8, respectively. Regardless of the population types used, the large difference in the gene effect of the *mog* gene estimated in the two studies is possibly due to one or two factors: (i) linkage map used for gene mapping and (ii) environments. The linkage map used in our study was highly dense comprising more than 2,600 markers, while that used in Naito et al. (2017) comprised only 116 markers. LG 6 where the *mog* gene was mapped to in our study contained 167 markers. In contrast, LG 8 where the *mog* gene was located onto in their study contained only 13 markers. Moreover, the genetic distance between markers flanking the *mog* gene in our study was less than 4 cM, while that in their study was about 10 cM. A genetic map with a high marker density gives better precision of effect estimates of detected QTL than a map with low marker density (Stange et al., 2013). In our study, the RIL mapping population was replicated in an experiment under field conditions, whereas the F₂ population used by Naito was grown in a non-replicated experiment in a greenhouse condition. The large difference in environmental conditions may cause a contrasting expression of the trait and thus the estimated effects of the *mog* gene.

As discussed above, the pleiotropic effect of the *mog* gene results in larger seeds, leaves, and pods and higher plant height in black gram. Although the gene may be useful in improving these traits, the pleiotropic effect of the *mog* gene should be considered carefully before using it in breeding program(s). This is because if the *mog* gene is introgressed into an improved cultivar, such introgression line(s) would have not only larger seeds but also larger leaves and higher plant heights. If such an improved cultivar already has optimum and suitable leaf size and plant height, increasing these would change the plant architecture and possibly be detrimental to seed yield in such a cultivar. In fact, we have observed that the gigantic architecture and leaves of the *mog* mutant when grown under field conditions are easily damaged by winds and rains. Other effects of the *mog* gene that should be considered are number of pods and number of seeds per pods. The *mog* mutant showed lower values of these traits than its wild type, Phitsanulok 2 (Tomooka et al., 2010; Naito et al., 2017). Similar results have been reported in soybean where the *ln* gene showed pleiotropic effects toward seed weight, leaf size (shape), and number of seeds per pods (Dinkins et al., 2002;

Sayama et al., 2017). However, the gigantic effect of the *mog* gene will be useful for breeding black gram cultivar(s) with high biomass for animal feeds.

In this study, QTLs for pod dehiscence and seed dormancy, two key traits related to domestication/adaption, of black gram were also mapped. Pod dehiscence is important for seed dispersal and survival of wild plants in nature but causes significant yield loss of cultivated plants, especially in hot and dry conditions. Two minor QTLs, *qPdt3.1*- and *qPdt5.1*- (PVE ≤ 10%), were detected on LGs 3 and 5 for this trait in black gram (Table 3). The QTLs detected for pod dehiscence in other *Vigna* species closely related to black gram including mungbean (Isemura et al., 2012), adzuki bean (Kaga et al., 2008), moth bean [*Vigna aconitifolia* (Jacq.) Maréchal] (Yundaeng et al., 2019), and cowpea [*V. unguiculata* (L.) Walp.] (Kongjaimun et al., 2012) were consistently mapped on the same LGs: LG7 and/or LG1. Nevertheless, the two QTLs found for the trait in black gram were not common to the QTLs identified in the other four *Vigna* species. In terms of seed dormancy, only a single QTL with large effect (PVE about 20%) was detected on LG 6 (Table 3). The number of QTLs detected for this trait was less than that identified in mungbean, adzuki bean, and rice bean where five QTLs were identified in each of these species (Kaga et al., 2008; Isemura et al., 2010, 2012) and in moth bean where three QTLs were found (Yundaeng et al., 2019). However, none of the QTLs for seed dormancy were common among these four *Vigna* species. In fact, we expected that black gram shares some common QTLs for pod indehiscence and seed dormancy with mungbean because the two species both originated and are domesticated in India and show very similar morphological characteristics and are cultivated and used in the same fashion (Tomooka et al., 2002; Kaewwongwal et al., 2015). However, the results indicated that different mechanisms control seed dormancy in these five *Vigna* species.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

PS and XC conceptualized and supervised the study. TY, PS, and CY conducted the data collection and analysis. TY, NT, CY, and PS developed the population. JC, CY, XY, and PS performed the molecular analysis and constructed the linkage map. PS wrote and edited the manuscript.

FUNDING

This research was funded by (1) National Natural Science Foundation of China (Grant No. 31871696), (2) China Agricultural Research System CARS-08-G15, and (3) High-End Expert Program, Jiangsu Academy of Agricultural Sciences, China.

ACKNOWLEDGMENTS

We are thankful to the Joint Legume Research Center between Jiangsu Academy of Agriculture Sciences and Kasetsart University for molecular lab facilities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00635/full#supplementary-material>

REFERENCES

- Ahloowalia, B. S., and Maluszynski, M. (2001). Induced mutations – A new paradigm in plant breeding. *Euphytica* 118, 167–173. doi: 10.1023/A:1004162323428
- Chinchest, A., and Nakeerak, P. (1990). "Mutation breeding on blackgram (*Vigna mungo*)," in *Proceedings of the Mungbean Meeting 90*, ed. C. Thavarasook (Bangkok: Tropical Agriculture Research Center), 43–46.
- Dinkins, R. D., Keim, K. R., Farno, L., and Edwards, L. H. (2002). Expression of the narrow leaflet gene for yield and agronomic traits in soybean. *J. Hered.* 93, 346–351. doi: 10.1093/jhered/93.5.346
- Fery, R. L. (1980). Genetics of *Vigna*. *Hort. Rev.* 2, 311–394.
- Isemura, T., Kaga, A., Konishi, S., Ando, T., Tomooka, N., and Vaughan, D. A. (2010). The genetics of the domestication of rice bean, *Vigna umbellata*. *Ann. Bot.* 106, 927–944. doi: 10.1093/aob/mcq188
- Isemura, T., Kaga, A., Tabata, S., Somta, P., Srinives, P., Shimizu, T., et al. (2012). Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). *PLoS One* 7:e41304. doi: 10.1371/journal.pone.0041304
- Kaewwongwal, A., Kongjaimun, A., Somta, P., Chankaew, K., Yimram, T., and Srinives, P. (2015). Genetic diversity of the black gram [*Vigna mungo* (L.) Hepper] gene pool as revealed by SSR markers. *Breed. Sci.* 65, 127–137. doi: 10.1270/jsbbs.65.127
- Kaga, A., Isemura, T., Toomoka, N., and Vaughan, D. A. (2008). The genetics of domestication of azuki bean (*Vigna angularis*). *Genetics* 178, 1013–1036. doi: 10.1534/genetics.107.078451
- Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B. K., et al. (2014). Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat. Commun.* 5:5443. doi: 10.1038/ncomms6443
- Kongjaimun, A., Kaga, A., Tomooka, N., Somta, P., Vaughan, D. A., and Srinives, P. (2012). The genetics of domestication of yardlong bean, *Vigna unguiculata* (L.) Walp. ssp. *unguiculata* cv.-gr. *sesquipedalis*. *Ann. Bot.* 109, 1185–1200. doi: 10.1093/aob/mcs048
- Li, H., Ye, G., and Wang, J. (2007). A modified algorithm for the improvement of composite interval mapping. *Genetics* 175, 361–374. doi: 10.1534/genetics.106.066811
- Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001
- Naito, N., Takahashi, Y., Chaiteng, B., Hirano, K., Kaga, A., Takagi, K., et al. (2017). Multiple organ gigantism caused by mutation in *VmPPD* gene in blackgram (*Vigna mungo*). *Breed. Sci.* 67, 151–158. doi: 10.1270/jsbbs.16184
- R Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Sablowski, R., and Dornelas, M. C. (2013). Interplay between cell growth and cell cycle in plants. *J. Exp. Bot.* 65, 2703–2714. doi: 10.1093/jxb/ert354
- Sayama, T., Tanabata, T., Saruta, M., Yamada, T., Anai, T., Kaga, A., et al. (2017). Confirmation of the pleiotropic control of leaflet shape and number of seeds per pod by the *Ln* gene in induced soybean mutants. *Breed. Sci.* 67, 363–369. doi: 10.1270/jsbbs.16201
- Somta, P., Chen, J., Yundaeng, C., Yuan, X., Yimram, T., Tomooka, N., et al. (2019). Development of an SNP-based high-density linkage map and QTL analysis for bruchid (*Callosobruchus maculatus* F.) resistance in black gram (*Vigna mungo* (L.) Hepper). *Sci. Rep.* 9:3930. doi: 10.1038/s41598-019-40669-5
- Stange, M., Utz, H. F., Schrag, T. A., Melchinger, A. E., and Würschum, T. (2013). High-density genotyping: an overkill for QTL mapping? Lessons learned from a case study in maize and simulations. *Theor. Appl. Genet.* 126, 2563–2574. doi: 10.1007/s00122-013-2155-0
- Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., et al. (2013). SLAF-seq: an efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. *PLoS One* 8:e58700. doi: 10.1371/journal.pone.0058700
- Tomooka, N., Kaga, A., Isemura, T., Naito, K., and Vaughan, D. A. (2010). Neo-Domestication of *Vigna* (*Leguminosae*) species and identification of a multiple organ gigantism (mog) mutant of cultivated black gram (*Vigna mungo*). *Gamma Field Symp.* 49, 41–46.
- Tomooka, N., Vaughan, D. A., Moss, H., and Maxted, N. (2002). *The Asian Vigna: Genus Vigna Subgenus Ceratotropis Genetic Resources*. Dordrecht: Kluwer Academic Publishers.
- Vaughan, D. A., Balázs, E., and Heslop-Harrison, J. S. (2007). From crop domestication to super domestication. *Ann. Bot.* 100, 893–901. doi: 10.1093/aob/mcm224
- Yundaeng, C., Somta, P., Amkul, K., Kongjaimun, A., Kaga, A., and Tomooka, N. (2019). Construction of genetic linkage map and genome dissection of domestication-related traits of moth bean (*Vigna aconitifolia*), a legume crop of arid areas. *Mol. Genet. Genomics* 294, 621–635. doi: 10.1007/s00438-019-01536-0

TABLE S1 | Analysis of variance of 16 agronomic and adaptive traits in recombinant inbred line (RIL) population derived from a cross between MOG mutant and wild black gram accession TC2210. The population was grown under field condition.

TABLE S2 | Correlation coefficient (r) between traits evaluated in a black gram RIL population derived from a cross between MOG mutant and wild black gram accession TC2210. Numbers in parentheses indicate probability.

TABLE S3 | Position of SNP markers locating nearby major QTLs controlling agronomic and adaptive traits and *VmPPD* gene of black gram on mungbean chromosome 8 based on the reference genome of mungbean cultivar VC1973A (Kang et al., 2014).

TABLE S4 | Annotated genes locating between markers Marker17781 to Marker2358 on the mungbean reference genome.

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

Copyright © 2020 Somta, Chen, Yimram, Yundaeng, Yuan, Tomooka and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genetic Diversity of Common Bean (*Phaseolus vulgaris* L.) Germplasm Resources in Chongqing, Evidenced by Morphological Characterization

Juechen Long^{1†}, Jijun Zhang^{1†}, Xiaochun Zhang¹, Jing Wu², Hong Chen¹, Ping Wang¹, Qiang Wang¹ and Chengzhang Du^{1*}

¹ Institute of Specialty Crop, Chongqing Academy of Agricultural Sciences, Chongqing, China, ² Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China

OPEN ACCESS

Edited by:

Gaofeng Zhou,
Department of Primary Industries
and Regional Development
of Western Australia (DPIRD),
Australia

Reviewed by:

Daolong Dou,
Nanjing Agricultural University, China
Yong Jia,
Murdoch University, Australia

*Correspondence:

Chengzhang Du
10695299@qq.com

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 26 February 2020

Accepted: 08 June 2020

Published: 08 July 2020

Citation:

Long J, Zhang J, Zhang X, Wu J,
Chen H, Wang P, Wang Q and Du C
(2020) Genetic Diversity of Common
Bean (*Phaseolus vulgaris* L.)
Germplasm Resources in Chongqing,
Evidenced by Morphological
Characterization.
Front. Genet. 11:697.
doi: 10.3389/fgene.2020.00697

In this study, the genetic diversity of 115 common bean germplasm resources collected from 27 counties in Chongqing over 3 years (2015–2017) was assessed. The results showed that the genetic diversity of the common bean germplasm resources was high, with an average diversity index of 1.447. The diversity of the qualitative traits of each organ was ranked as seed ($H' = 1.40$) > pod ($H' = 1.062$) > plant ($H' = 0.64$) > leaf ($H' = 0.565$), while the diversity of the quantitative traits of each organ was ranked as seed ($H' = 2.021$) > pod ($H' = 1.989$) > phenology ($H' = 1.941$) > plant ($H' = 1.646$). In the cluster analysis, 115 accessions were clustered into four groups. The accessions in the first and fourth group were characteristic of the Andean gene pool, while the accessions in the second group were characteristic of the Mesoamerican gene pool. However, the accessions of the third group possessed the combined characteristics of both gene pools and were thus classified as introgressed type. Four principal components represented 40.30% of the morphological diversity, with the first principal component representing the traits associated with plant growth; the second principal component representing flower characteristics; the third principal component being composed of seed characters; and the fourth principal component representing pod characteristics. The common bean germplasm resources are widely distributed in Chongqing, and the introgressed types are more prevalent in this region than elsewhere, with a rich genetic diversity and high utilization value in genetic breeding.

Keywords: common bean, genetic diversity, cluster analysis, principal component analysis, morphological characterization

INTRODUCTION

Common bean (*Phaseolus vulgaris* L.) is one of the most important legumes and is high in protein, low in fat, and rich in vitamins and dietary fiber. The regular consumption of common bean can reduce coronary heart disease, type II diabetes, and cancer (Kutoš et al., 2003; Krupa, 2008). Common bean is grown and produced all over the world. More than 120 countries or regions grow

common bean, encompassing a total area of 28.78 million hm^2 (FAO, 2018). The total production in China is 1.14 million tons and the cultivation area is 749,860 hm^2 (FAO, 2018). Common bean originated in central and south America (Delgadosalinas et al., 1999, 2006; Bitocchi et al., 2012; Desiderio et al., 2013; Bellucci et al., 2014). Over the course of its evolution and domestication, common bean eventually formed two separate gene pools, namely the Mesoamerican gene pool and the Andean gene pool (Debouck and Smartt, 1995). During its global domestication (Gepts et al., 1991; Kami et al., 1995), several changes in its morphological characteristics occurred, including seed and leaf enlargement, alterations in growth habits and photoperiodic response, and changes in seed color and seed coat markings (Singh et al., 1991; McClean et al., 2004). The two gene pools have been divided into seven races. The Mesoamerican gene pool contains four races: Durango (D), Jalisco (J), Central America (M), and Guatemala (G), and the Andean gene pool contains three races: Chile (C), New Granda (N), and Peru (P) (Beebe et al., 2001). Common bean was introduced into Europe in the early 16th century and spread rapidly from Europe to the Middle East, West Asia, and other regions in the 16th and 17th centuries (Zheng, 1997). Ultimately, centers of origin were formed in Europe (Santalla et al., 2002; Angioi et al., 2010, 2011), Brazil (Burle et al., 2010), South Africa (Martin and Adams, 1987a,b; Asfaw et al., 2009; Blair et al., 2010), and China (Zhang et al., 2008). Common bean was introduced in China more than 400 years ago (Zheng, 1997), and long periods of domestication and artificial selection have led to an increase in its genetic diversity.

Chongqing is located in the middle and upper reaches of the Yangtze River in southwest China and the eastern edge of the Szechwan Basin. The terrain of Chongqing is mainly mountainous and hilly, with a large elevation range, a subtropical monsoon humid climate, and an obvious three-dimensional climate. It has been divided into four natural ecological regions (Luo et al., 2006). In Chongqing, the tender bean pods of common bean are stir-fried, stewed, and boiled, while the seeds are typically cooked in soup. The people from the mountainous areas often stew potatoes with the tender pods as their staple food. The mountainous areas of Chongqing are inhabited by local minorities such as tujia, miao, yi, and gelao (Chongqing Ethnic and Religious Affairs Commission, 2002). The common bean germplasm resources also differ between populations from different ethnic cultural backgrounds. Common bean germplasm resources have been selected and domesticated according to the specific production and life activities of people, resulting in the formation of a unique distribution structure of common bean germplasm resources in Chongqing. In this study, 115 common bean germplasm resources were collected from 27 counties in Chongqing over 3 years (2015–2017). Their distribution was investigated and genetic diversity analysis of the agronomic traits was carried out in order to clarify the distribution structure and genetic diversity of common bean in Chongqing, providing support for the development of excellent germplasm and genetic resources.

MATERIALS AND METHODS

Plant Materials

Based on the third national survey of crop germplasm resources, this study investigated the common bean germplasm resources in different natural ecological regions of Chongqing from 2015 to 2017. A total of 115 common bean germplasm resources were collected, all of which originated from four natural ecological regions in Chongqing (natural ecological region I, Szechwan Basin agricultural natural ecological region; natural ecological region II, Three Gorges region and the parallel ridge valley compound natural ecological region; natural ecological region III, Qinling-Daba Mountain evergreen, broadleaf, and deciduous forest natural ecological region; natural ecological region IV, Southeast Chongqing montane evergreen broadleaf forest natural ecological region) (Luo et al., 2006). There were eight accessions distributed in natural ecological region I, 45 accessions in natural ecological region II, 28 accessions in natural ecological region III, and 34 accessions in natural ecological region IV (Figure 1).

Field Experiments and Phenotyping

Agronomic traits were investigated at the experimental base in Wu long (29.52°N, 107.65°E) at 1280 m above sea level. The average annual temperature was 14.8°C in 2017 and 2018, and the soil is sandy loam with good drainage. The experimental design was a randomized complete block with three repetitions. The experimental plots consisted of six rows with a planting distance of 0.75 m between the rows and 0.25 m within the rows, and with two seeds sown per planting station. The agronomic practices were the same for all plots and are typically used by most farmers.

Twenty-nine morpho-agronomic characters were measured for the 115 accessions based on the Descriptors and Date Standard for Common Bean (*Phaseolus vulgaris* L.) (Wang et al., 2006). Hypocotyl pigmentation (HP), color of standard (CS), color of wings (CW), growth habit (GH), leaf shape (LS), podding habit (PHA), stem type (ST), and phenological characters were recorded in the field. Pod color (PC), plant height (PH), node number of the main stem (NS), number of the primary branches (NB), number of pods per plant (NP), shape of the pod (SP), shape of the pod apex (SPA), pod surface (PS), pod length (PL), pod width (PW), number of seeds per pod (NSP), seed shape (SS), seed coat color (SCC), speckle of the seed coat (SSC), color speckle of the seed coat (CSSC), hilum color (HC), seed length (SL), seed width (SW), hundred-seed weight (HSW), and length/width of the seed (LWS) were recorded at harvest. There were 16 qualitative traits and 13 quantitative traits. All the traits evaluated are listed in Tables 1, 2.

The 13 quantitative traits in this study included agronomic and morphological traits. All of the recording standards and methods of the traits are listed in Table 1. The 16 qualitative traits in this study included agronomic and morphological traits, and the differences among all qualitative traits were numerically assigned (Table 2).

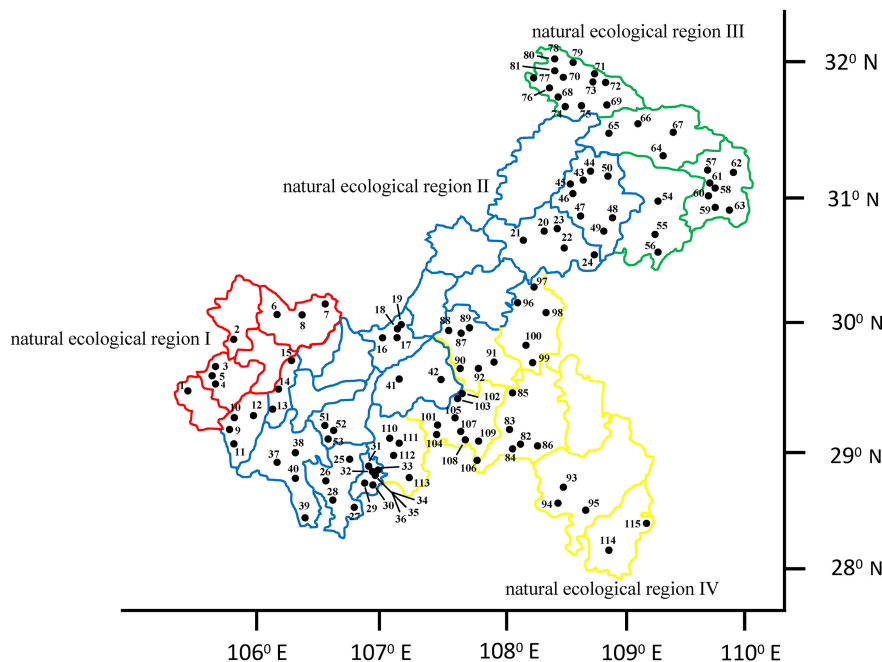


FIGURE 1 | Distribution of the common bean germplasm resources in the four natural ecological regions of Chongqing.

Phenotypic Variation Estimates

Statistical analyses were conducted using SPSS 16.0 (SPSS Inc., Chicago, IL, United States), and the average value ($\bar{X}$) and standard deviation (s) of each trait were calculated. The coefficient of variation (CV) was used to evaluate the degree of phenotypic variation of each trait as follows: $CV = s/\bar{X}$. Each quantitative trait of all the accessions was assigned to classes ranging from class 1: $[X_i < (\bar{X} - 2s)]$ to class 10: $[X_i > (\bar{X} + 2s)]$. Each $0.5s$ was assigned as a class, where X_i was the value in class i . The Shannon–Weaver index (H') was used to calculate the genetic diversity index (Khanjari et al., 2008) of each accession:

$$H' = - \sum_{i=1}^n P_i \ln P_i$$

where n is the number of phenotypic classes for a trait, and P_i is the genotypic frequency or the proportion of the total number of entries in the i th class.

Principal Component and Cluster Analyses

Principal component and cluster analysis were performed using SPSS 16.0. The spatial relationships among entries

TABLE 1 | Variable sets and 13 observed common bean quantitative characters, codes, measurement units, and measurement procedures.

Variable set (plant organ)	Character	Code	Measurement unit and measurement/sampling procedure
Phenology	Days to flowering	DF	No. of days from emergence to the time that 80% of plants within the center rows were flowering
	Days to physiological maturity	DM	No. of days from sowing to physiological maturity (80% of pods dry within the two center rows)
Plant	Plant height	PH	cm—measured at harvest; 10 plants within plot center
	Node number of the main stem	NS	No. of stem section; measured at harvest; 10 plants within plot center
	Number of primary branches	NB	No. of stem branches; measured at harvest; 10 plants within plot center
Pod	Number of pods per plant	NP	No. of pods per plant; measured at harvest; 10 plants within plot center
	Pod length	PL	cm—measured at harvest; 10 pods per plant from 10 plants within plot center
	Pod width	PW	cm—measured at harvest; 10 pods per plant from 10 plants within plot center
	Number of seeds per pod	NSP	Counted at harvest, for 10 pods per plant from 10 plants within plot center
Seed	Seed length	SL	cm—measured after 6 days sun drying, postharvest, 10 plants within plot center (5 pods per plant and 3 seed per pod)
	Seed width	SW	cm—measured after 6 days sun drying, postharvest, 10 plants within plot center (5 pods per plant and 3 seed per pod)
	Length/width of seed	LWS	Seed length - Seed width ratio
	Hundred-seed weight	HSW	g—measured at physiological maturity, on two samples of 100 seeds

TABLE 2 | Variable sets and 16 observed common bean qualitative characters, codes, measurement units, and measurement procedures.

Variable set (plant organ)	Character	Code	Measurement unit and measurement/sampling procedure
Plant	Hypocotyl pigmentation	HP	1 = green, 2 = purple assessed at seedling stage
	Growth habit	GH	1 = erect, 2 = prostrate
	Podding habit	PHA	1 = determinate, 2 = indeterminate
	Stem type	ST	1 = normal stem, 2 = clasp stem
Leaf	Leaf shape	LS	1 = ovate, 2 = rhombic ovate
Flower	Color of standard	CS	1 = white, 2 = pinkish-white, 3 = pinkish-red, 4 = purple, 5 = light purple
	Color of wings	CW	1 = white, 2 = pinkish-white, 3 = pinkish-red, 4 = purple, 5 = light purple
Pod	Pod color	PC	1 = stripe, 2 = light brown, 3 = yellowish white, 4 = brown
	Shape of pod	SP	1 = round curved, 2 = short flat strip, 3 = long flat strip, 4 = sickle-shaped, 5 = sword-shaped, 6 = round curved stick-shaped, 7 = short round stick-shaped, 8 = long round stick-shaped
	Shape of pod apex	SPA	1 = acute apex, 2 = obtuse apex
	Pod surface	PS	1 = tiny protruding, 2 = protruding, 3 = flat
Seed	Seed shape	SS	1 = long elliptic, 2 = ovate, 3 = flat round, 4 = square, 5 = short cylinder, 6 = elliptic, 7 = kidney-shaped, 8 = round
	Seed coat color	SCC	1 = stripe, 2 = yellow, 3 = white, 4 = black, 5 = brown, 6 = yellowish-white, 7 = pinkish-red, 8 = milk white, 9 = deep yellow, 10 = red
	Speckle of seed coat	SSC	0 = free, 1 = punctiform, 2 = Stripe, 3 = reticulate
	Color speckle of seed coat	CSSC	0 = free, 1 = brown, 2 = pink, 3 = black, 4 = light brown, 5 = red
	Hilum color	HC	1 = yellowish-white, 2 = white, 3 = light brown

(accessions) were visualized by plotting the scores of the first and second principal components in two-dimensional space. For the cluster analysis, all the traits of each accession were standardized, and the Euclidian distances were calculated using the unweighted group average linkage method (UPGMA).

RESULTS

Phenotypic Variation

According to the analysis of the 13 quantitative traits and 16 qualitative traits, the common bean germplasm resources in Chongqing are rich in genetic diversity.

The quantitative traits are shown in **Figure 2**. The genetic diversity of the quantitative traits of each organ and phenological was ranked as: seed ($H' = 2.021$) > pod ($H' = 1.989$) > phenological ($H' = 1.941$) > plant ($H' = 1.646$).

For the seeds, the genetic diversity index was ranked: SL ($H' = 2.083$) > SW ($H' = 2.022$) > HSW ($H' = 2.013$) > LWS ($H' = 1.968$). The length/width of seed (LWS) of 61.7% of the accessions was at the middle level, indicating that most seeds had a more oval SS. In addition, 87% of the accessions were large to medium seeds, while only 13% were small seeds, suggesting that consumers in Chongqing prefer large and oval seeds.

For the pods, the genetic diversity index was ranked as: NSP ($H' = 2.066$) > NP ($H' = 1.997$) > PL ($H' = 1.950$) > PW ($H' = 1.943$). More than 50% of NP and NSP were medium, and the numbers at the high level were significantly higher than that at the low level, which indicated that most of the accessions have the potential to achieve high yield.

For the phenological traits, the genetic diversity index was ranked as: DM ($H' = 2.023$) > DF ($H' = 1.858$). The materials

of the short growth period were 20.9%, while those of the long growth period were only 14.02%.

In terms of the plant traits, the genetic diversity index was ranked as: NB ($H' = 1.972$) > NS ($H' = 1.760$) > PH ($H' = 1.206$). Due to differences in GH, the variation coefficient of PH was the highest, while the diversity index was the lowest. In addition, more than half of the accessions were prostrate.

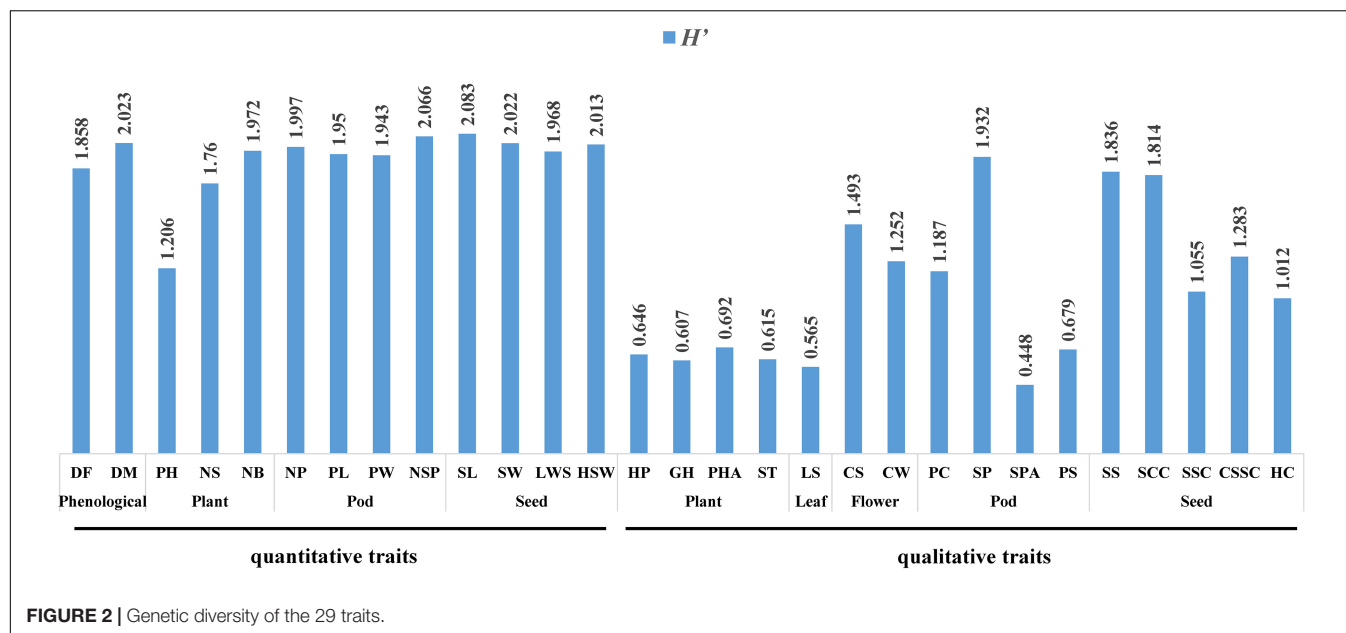
The qualitative traits are shown in **Figure 2**, where the characteristic frequency of each trait was counted and the genetic diversity coefficient was calculated. The genetic diversity of the qualitative traits of each organ was ranked as: seed ($H' = 1.40$) > pod ($H' = 1.062$) > plant ($H' = 0.64$) > leaf ($H' = 0.565$).

In terms of the plant traits, the order of the genetic diversity index was ranked as: PHA ($H' = 0.692$) > HP ($H' = 0.646$) > ST ($H' = 0.615$) > GH ($H' = 0.607$). The HP was mainly green, accounting for 65.2% of the accessions, while the remainder were purple. The GH was mainly prostrate type. There were more determinate pods than indeterminate pods in PHA, which indicated that farmers in Chongqing have a preference for prostrate-type common bean.

The LS of the accessions was mainly oval, and its genetic diversity index was only 0.565, which was the lowest of all the traits.

For the flower traits, the genetic diversity index was ranked as follows: CS ($H' = 1.493$) > CW ($H' = 1.252$). Both the CS and the CW were predominantly white.

For the pod traits, the genetic diversity index was as follows: SP ($H' = 1.932$) > PC ($H' = 1.187$) > PS ($H' = 0.679$) > SPA ($H' = 0.448$). The PC was mainly light brown, whereas only six accessions were least speckled. Most accessions belonged to the short flat strip type, while only four accessions were of the long round stick-shaped type. In terms of the shape of pod apex (SPA),



83.5% had an acute apex, while the remainder had an obtuse apex. PS was mainly characterized by tiny protrusions.

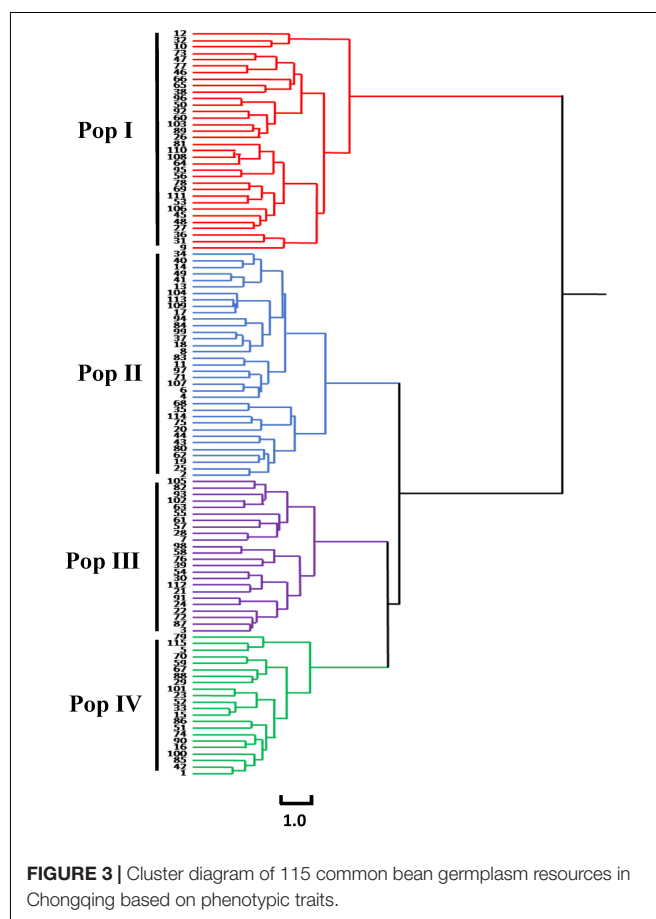
The seed traits and genetic diversity were ranked as follows: SS ($H' = 1.836$) > SCC ($H' = 1.814$) > CSSC ($H' = 1.283$) > SSC ($H' = 1.055$) > HC ($H' = 1.012$). The SS of most accessions was kidney-shaped, with 29 accessions. The majority of accessions (50) had a striped SCC, which indicated that the farmers in Chongqing have a preference for both the striped seed coats and large seeds.

Cluster Analysis

The results of the cluster analysis showed that the 115 accessions could be clustered into four groups (Figure 3).

The first group (Pop I) contained 34 accessions, and the agronomic characteristics were as follows: the DM was 70.2 days, which was the shortest of the four groups. The HP was mainly green. The GH of the accessions in Pop I was erect. The ST was mainly normal, and the PHA was mainly determinate, with a few indeterminate accessions. The CS was predominantly light purple, while the CW was mainly white. The NP was the lowest among the four groups (7.32 pods). The SP was mainly short flat strips, while the SS was largely long ellipse. Most of the seeds had punctiform pink stripes. The LWS was the largest, indicating that the seeds of Pop I were the most elongated among the four groups. The HSW was 38.66 g.

The second group (Pop II) contained 35 accessions, and the agronomic characteristics were as follows: the DM was the longest among the four groups at 79.24 days. The HP was mainly green, and the GH of the accessions was prostrate. The PHA of the accessions was mainly indeterminate, with a few that were determinate. The CS and CW were mainly light purple. The NP was the highest among the four groups, with 10.34 pods. The pods of most of the accessions were sickle-shaped, while SS was



mainly kidney-shaped. There were no stripes on the seeds of the accessions in Pop II, and the SCC was mainly white. The HSW was 34.87 g.

The third group (Pop III) contained 24 accessions, and the main agronomic traits were as follows: the DM was 77.6 days. The HP was mainly purple, and the GH of the accessions in this group was prostrate. The PHA of the accessions was mainly indeterminate. The CS and CW were mainly light purple and purple. The LS of Pop III was mainly rhombic ovate, which is unique among the four groups. The SP was mainly round curved stick-shaped, and the SS was mainly kidney-shaped, and most seeds had punctiform black stripes. The HSW was 33.03 g.

The fourth group (Pop IV) contained 22 accessions, and the main agronomic traits were as follows: the DM was 78.64 days. The GH of the accessions in Pop IV was mainly prostrate, and the PHA of the accessions was mainly indeterminate, with a few that were determinate. The CS and CW were mainly pink. The SP was mainly short flat strip with oval seeds, and most of the seeds possessed red punctiform speckles. The HSW was 40.38 g, and the seeds of Pop IV were the largest and roundest of the four groups.

The above results indicate that the characteristics of these four groups are quite different. The accessions of Pop I were more similar to the germplasm resources from the Andean gene pool due to the erect GH, white flowers, pink speckled seed coats, and large seeds. The accessions of Pop IV were more similar to the germplasm resources from the Peru (P) race in the Andean gene pool, due to the prostrate GH and larger and oval seeds. We thus classified the accessions of Pop I and Pop IV as being derived from the germplasm resources of the Andean gene pool. The accessions of Pop II were more similar to the germplasm resources of the Mesoamerican gene pool, as the seeds were small and free of speckles, and the seed coat was black, yellow, and white. Therefore, the accessions from Pop II were identified as being derived from the germplasm resources of the Mesoamerican gene pool. The accessions of Pop III were mainly prostrate, the seed coat was striped and medium in size, the flowers were mostly colorful, and there was both an indeterminate and determinate PHA, thus combining the characteristics of the Andean gene pool and the Mesoamerican gene bank. Therefore, Pop III was classified as introgressed.

Principal Component Analysis

The results of the Principal component analysis (PCA) (Table 3) of the 29 traits of the 115 accessions indicated that the first four principal components explained 40.298% of the total variation. The first component was highly correlated with PH, HP, ST, and NS, thus representing the traits associated with plant growth. The second component was highly correlated with CS and CW, suggesting that the second principal component was related to floral traits. The third component was highly correlated with SCC, SSC, and CSSC, thus representing traits related to the seeds. The fourth principal component was composed of PL and NSP, thus representing traits related to the pods.

Based on the cluster analysis, the first and second principal component scores of each of the 29 phenotypic traits were used to construct a two-dimensional scatter plot (Figure 4), and the genetic structure of the common bean germplasm resources in Chongqing was further analyzed. Most of the accessions in

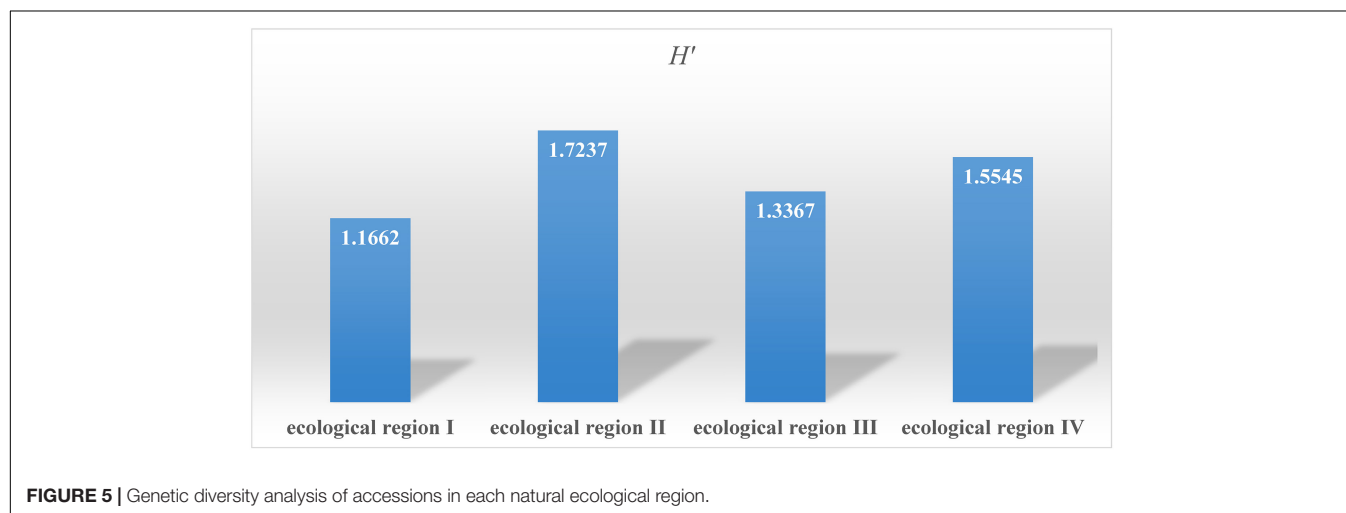
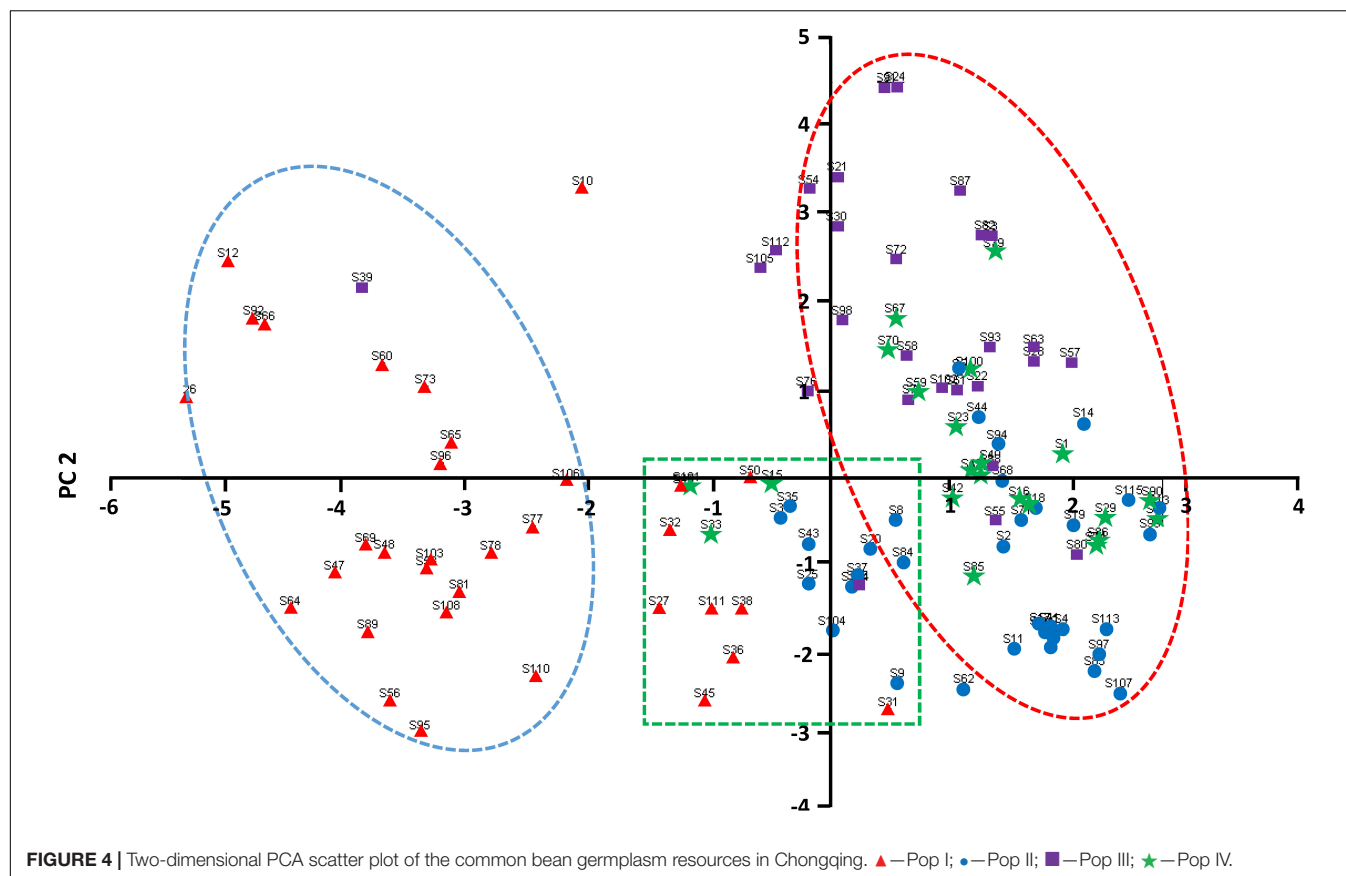
group I were obviously concentrated on the left side of the plot, with a clear boundary with the other groups. Most of the accessions in group III were concentrated in the upper right of the plot, while most of the material in group II was concentrated in the lower right of the plot. However, the spatial distribution of the accessions in group IV was between group II and group III. In addition, there were a few outliers in the four groups.

Geographic Distribution of the Target Germplasm

Combined with the results of the cluster analysis, there were 52 accessions derived from the Andean gene pool, which was the most widely distributed gene pool in Chongqing, accounting for 45.2%. The accessions derived from the Mesoamerican gene pool accounted for 34.8%, while the introgressed-type accessions accounted for 20%. The accessions derived from the

TABLE 3 | The first four principal component roots and eigenvectors.

Variable set (plant organ)	Trait	PC1	PC2	PC3	PC4
Phenology	DF	0.2288	0.0015	0.0489	-0.0901
	DM	0.1813	0.0402	-0.1016	-0.3027
Plant	PH	0.3856	0.0785	-0.0465	0.1488
	HP	0.0487	0.2300	0.1761	-0.1595
	GH	-0.3884	-0.1576	-0.0562	0.1821
	ST	0.3934	0.1693	0.0152	-0.1798
	NS	0.3683	0.0905	-0.0550	0.1530
	NB	-0.1939	0.0013	0.0171	-0.3114
	PHA	0.2605	0.1355	-0.0051	-0.2478
Flower	CS	-0.0668	0.4828	0.1771	0.1306
	CW	-0.0823	0.4885	0.1895	0.1302
Leaf	LS	-0.0954	0.2202	0.0575	0.2209
Pod	PC	-0.0207	-0.0073	-0.0670	-0.0506
	NP	0.1405	-0.0024	0.1206	0.0830
	SP	-0.1296	-0.0259	-0.0926	0.0961
	SPA	0.0572	-0.0552	-0.0327	-0.1491
	PS	-0.0024	0.2159	0.1124	0.2084
	PL	0.1728	-0.1718	0.0714	0.4245
	PW	0.1873	-0.0958	-0.3082	0.0762
	NSP	0.1492	-0.1323	0.1324	0.3396
	SS	0.0772	-0.0129	0.2462	0.1154
	SCC	0.0651	-0.1833	0.4235	-0.1187
	SSC	-0.0259	0.2630	-0.4607	0.0442
	CSSC	-0.0773	0.2282	-0.4483	0.0704
	SL	-0.0312	0.1507	0.0990	-0.0934
	SW	0.1097	-0.0183	-0.0504	0.1665
	LWS	-0.1097	0.1680	0.1041	-0.2171
Seed	HC	0.0116	-0.1048	0.0893	-0.1166
	HSW	0.1647	-0.1229	-0.1831	-0.0733
	Eigen-value	4.3895	2.6272	2.5903	2.0796
	Contribution rate (%)	15.1364	9.0592	8.9321	7.1709
	Accumulative contribution rate (%)	15.1364	24.1956	33.1276	40.2985



Mesoamerican gene pool were more widely distributed than the other germplasm types in ecological region I. However, natural ecological regions II, III, and IV were the opposite, and the accessions derived from the Andean gene pool were more widely distributed. The distribution of introgressed-type accessions differed among the four ecological regions, and the proportion of introgressed-type accessions in ecological region III and IV was greater than that in the other ecological regions.

DISCUSSION AND CONCLUSION

Phenotypic Variation

The results of the diversity analysis showed that the average diversity index of the common bean germplasm resources in Chongqing was 1.447, which is higher than that in Shaanxi ($H' = 1.344$), Sichuan ($H' = 1.390$), and Yunnan ($H' = 1.330$) provinces, and lower than that in Guizhou ($H' = 1.499$) and Shanxi ($H' = 1.450$) provinces (Zhang et al., 2005).

Among the four natural ecological regions in Chongqing, ecological region II has great variation in longitude, latitude, and elevation, and thus the diversity index of the accessions in this ecological zone was the highest at 1.737, followed by the accessions from ecological region IV, ecological region III, and ecological region I (Figure 5). The results of this study further supplement the diversity data of common bean germplasm resources in China.

The common bean germplasm resources in Honduras, Nicaragua, and Mexico have a short DF (García et al., 1997; Gómez et al., 2004; Meza et al., 2013); however, 73% of the DF in the Chongqing accessions was medium, which is more similar to the common bean germplasm resources in northern Portugal (Coelho et al., 2009). Only 29.6% of the Chongqing accessions are erect in GH, which is similar to the common bean germplasm resources in Mexico, while the germplasm resources of Honduras have a more erect GH (74%) (Meza et al., 2013). The seeds of the accessions in Chongqing are mainly speckled and medium to large in size, while those in Honduras and Mexico are mainly red and small, and those in Guatemala are mainly black, bright, and small (García et al., 1997; Meza et al., 2013). This indicates that the common bean germplasm resources in Chongqing have a unique genetic diversity resulting from long-term selection and domestication.

Cluster Analysis and Principal Component Analysis

Diversity analysis of common bean germplasm resources in Slovenia and its vicinity, the Iberian Peninsula, and central Africa has confirmed the existence of gene introgression between the two gene pools (Maras et al., 2006; Rodiño et al., 2006; Blair et al., 2010). Gene introgression also exists in the common bean germplasm resources in China (Lei, 2018). According to the cluster analysis, there were many introgressed-type accessions (20%), suggesting that introgression between the two gene pools has occurred very frequently in Chongqing. This also suggests that the farmers in Chongqing have been selecting common bean germplasm resources for a long time to suit the agricultural conditions of the region. The results of the PCA showed that the accessions could be roughly divided into three areas in the scatter plot by the first two principal components. The first principal component separated most of the erect accessions, which are derived from the Andean gene pool. The second principal component distinguished the accessions derived from Mesoamerican gene pool and the introgressed-type accessions. Notably, the scatter of the prostrate accessions in the Andean gene pool was somewhere in between, suggesting that the genetic heritage of the introgressed-type accessions was closely related to the prostrate accessions derived from the Andean gene pool.

Geographic Distribution of the Target Germplasm

The distribution of common bean germplasm resources depends on the unique agricultural production preferences of farmers, and it has been suggested that seed selection is closely related to seed size and color (Papa and Gepts, 2003). The diversity

index of seed-related traits was the highest in this study, thus confirming this view. In addition, farmers in Chongqing also tend to favor erect accessions when sowing common bean. As the terrain of the II, III, and IV natural ecological regions is dominated by steep hills and mountains, most of the common bean has been planted on slopes, and thus the farmers in these regions prefer the erect accessions in the Andean gene pool because of the reduced labor intensity. Additionally, as farmers in the mountainous areas lack means of transportation, they have been selecting and domesticating crop germplasm resources according to the specific requirements of the region for a long time, resulting in the continuous enrichment of the genetic diversity of the common bean germplasm resources. Therefore, the most introgressed-type accessions were distributed in natural ecological region III and IV.

The genetic background of the introgressed-type accessions was between the Andean gene pool and the Mesoamerican gene pool, providing new material and a bridge between the two gene pools for breeders. This is of great significance for improving the yield, quality, and disease resistance in common bean cultivation (Zhang, 2007). Breeders should thus intensify the collection of germplasm resources in ecological regions III and IV.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

CD was responsible for the study design. HC and QW performed to collect all germplasm resources. JZ, XZ, and PW analyzed the data. JW and JL contributed to the writing of the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by grants-in-aid from the Species Resource Conservation Project (111721301354052036), the Ministry of Agriculture of China the earmarked fund for the China Agriculture Research System (CARS-08), the Agricultural Development Fund Project of Chongqing Academy of Agricultural Sciences (NKY-2020AC011), and the Chongqing Science and Technology Bureau Project (cstc2018jxjl80041).

ACKNOWLEDGMENTS

We would like to thank the Chongqing Science and Technology Bureau for its financial assistance, as well as the National Industrial Technology System for its financial assistance in this research. We also thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

REFERENCES

- Angioi, S. A., Rau, D., Attene, G., Nanni, L., and Belluci, E. (2010). Beans in Europe: origin and structure of the European landraces of *Phaseolus vulgaris* L. *Theor. Appl. Genet.* 121, 829–843. doi: 10.1007/s00122-010-1353-2
- Angioi, S. A., Rau, D., Nanni, L., Belluci, E., Papa, R., and Attenne, G. (2011). The genetic make-up of the European landraces of the common bean. *Plant Genet. Res.* 9, 197–201. doi: 10.1017/S1479262111000190
- Asfaw, A., Blair, M. W., and Almekinders, C. (2009). Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands. *Theor. Appl. Genet.* 120, 1–12. doi: 10.1007/s00122-009-1154-7
- Beebe, S., Rengifo, J., Gaitan, E., Duque, M. C., and Tohme, J. (2001). Diversity and origin of Andean landraces of common bean. *Crop Sci.* 41, 854–862. doi: 10.2135/cropsci2001.413854x
- Bellucci, E., Bitocchi, E., Rau, D., Rodriguez, M., Biagetti, E., Giardini, A., et al. (2014). “Genomics of origin, domestication and evolution of *Phaseolus vulgaris* L.” in *Genomics of Plant Genetic Resources*, ed. R. Tuberosa (Cham: Springer), 483–507. doi: 10.1007/978-94-007-7572-5_20
- Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., and Giardini, A. (2012). Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl. Acad. Sci. U.S.A.* 109, 788–796. doi: 10.1073/pnas.1108973109
- Blair, M. W., González, L. F., Kimani, M., and Butare, L. (2010). Genetic diversity, inter-gene pool introgression and nutritional quality of common beans (*Phaseolus vulgaris* L.) from Central Africa. *Theor. Appl. Genet.* 121, 237–248. doi: 10.1007/s00122-010-1305-x
- Burle, M. L., Fonseca, J. R., Kami, J. A., and Gept, P. (2010). Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary center of diversity. *Theor. Appl. Genet.* 121, 801–813. doi: 10.1007/s00122-010-1350-5
- Chongqing Ethnic and Religious Affairs Commission (2002). *Ethnography of Chongqing*. Chongqing: Chongqing Publishing House.
- Coelho, R. C., Faria, M. A., Rocha, J., Reis, A., Oliveira, M. P., and Nunes, E. (2009). Assessing genetic variability in germplasm of *Phaseolus vulgaris* L. collected in Northern Portugal. *Sci. Hortic.* 122, 333–338. doi: 10.1016/j.scienta.2009.05.035
- Debouck, D. G., and Smartt, J. (1995). “Beans,” in *Evolution of Crop Plants*, ed. J. Smartt (Harlow: Longman), 287–294.
- Delgadosalinas, A., Bibler, R., and Lavin, M. (2006). Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. *Syst. Bot.* 31, 779–791. doi: 10.1600/036364406779695960
- Delgadosalinas, A., Turley, T., Richman, A., and Lavin, M. (1999). Phylogenetic analysis of the cultivated and wild species of *Phaseolus* (Fabaceae). *Syst. Bot.* 24, 438–460. doi: 10.1016/j.scienta.2009.05.035
- Desiderio, F., Bitocchi, E., Bellucci, E., Rau, D., and Rodriguez, M. (2013). Chloroplast microsatellite diversity in *Phaseolus vulgaris* L. *Front. Plant. Sci.* 3:312. doi: 10.3389/fpls.2012.00312
- FAO (2018). *Statistical Data Base For Agriculture*. Rome: Food and Agriculture Organization.
- García, E. H., Valdivia, C. B., Aguirre, J. R., and Muruaga, J. S. (1997). Morphological and agronomic traits of a wild population and an improved cultivar of common bean (*Phaseolus vulgaris* L.). *Ann. Bot.* 79, 207–213. doi: 10.1006/anbo.1996.0329
- Gepts, P., Debouck, D., and Schoonhoven, A. V. (1991). *Common Beans Research for Crop Improvement*. Wiltshire: Red Wood Press.
- Gómez, O. J., Blair, M. W., Frankow-Lindberg, B. E., and Gullberg, U. (2004). Molecular and phenotypic diversity of common bean landraces from Nicaragua. *Crop Sci.* 44, 1412–1418. doi: 10.2135/cropsci2004.1412
- Kami, J., Velásquez, V. B., Debouck, D. G., and Gepts, P. (1995). Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. *Proc. Natl. Acad. Sci. U.S.A.* 92, 1101–1104. doi: 10.1073/pnas.92.4.1101
- Khanjari, S., Filatenko, A. A., Hammer, K., and Buerkert, A. (2008). Morphological spike diversity of Omani wheat. *Genet. Resour. Crop Evol.* 55, 1185–1195. doi: 10.1007/s10722-008-9319-9
- Krupa, U. (2008). Main nutritional and antinutritional compounds of bean seeds a review. *Pol. J. Food Nutr. Sci.* 58, 149–155. doi: 10.7589/0090-3558-44.3.731
- Kutoš, T., Golob, T., Kač, M., and Plestenjak, A. (2003). Dietary fibre content of dry and processed beans. *Food Chem.* 80, 231–235. doi: 10.1016/s0308-8146(02)00258-3
- Lei, L. (2018). *Population Structure And Genetic Diversity Of A Core Collection Of Common Beans (Phaseolus vulgaris L.)*. Master's thesis, Chinese Academy of Agricultural Sciences, Beijing.
- Luo, H. L., Zhu, B., Liu, D. S., and He, X. B. (2006). Regionalization of ecological function in Chongqing City, China. *Acta Ecol. Sin.* 26, 3144–3151.
- Maras, M., Sušnik, S., and Šuštar-Vozlič, J. (2006). Temporal changes in genetic diversity of common bean (*Phaseolus vulgaris* L.) accessions cultivated between 1800 and 2000. *Russian J. Genet.* 42, 775–782. doi: 10.1134/s102279540607012x
- Martin, G. B., and Adams, M. W. (1987a). Landraces of *Phaseolus vulgaris* (Fabaceae) in northern Malawi. I. *Econ. Bot.* 41, 190–203. doi: 10.1007/BF02858965
- Martin, G. B., and Adams, M. W. (1987b). Landraces of *Phaseolus vulgaris* (Fabaceae) in northern Malawi. II. Generation and maintenance of variability. *Econ. Bot.* 41, 204–215. doi: 10.1007/BF02858967
- McClellan, P., Gepts, P., and Kamir, J. (2004). *Genomics and Genetic Diversity In Common Bean*. Urbana, IL: Aocs Press.
- Meza, N., Rosas, J. C., Martin, J. P., and Ortiz, M. J. (2013). Biodiversity of common bean (*Phaseolus vulgaris* L.) in Honduras, evidenced by morphological characterization. *Genet. Resour. Crop Evol.* 60, 1329–1336. doi: 10.1007/s10722-012-9922-7
- Papa, R., and Gepts, P. (2003). Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from mesoamerica. *Theor. Appl. Genet.* 106, 239–250. doi: 10.1007/s00122-002-1085-z
- Rodiño, A. P., Santalla, M., González, A. M., Antonio, M., Ron, D., and Shree, P. (2006). Novel genetic variation in common bean from the Iberian Peninsula. *Crop Sci.* 46, 2540–2546. doi: 10.2135/cropsci2006.02.0104
- Santalla, M., Rodio, A. P., and De, R. M. (2002). Allozyme evidence supporting southwest Europe as a secondary center of genetic diversity for common bean. *Theor. Appl. Genet.* 104, 934–944. doi: 10.1007/s00122-001-0844-6
- Singh, S. P., Gutierrez, J. A., Molina, A., Urrea, C., and Gepts, P. (1991). Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. *Crop Sci.* 31, 23–29. doi: 10.2135/cropsci1991.0011183X003100010005x
- Wang, S. M., Zhang, Y. Z., and Wei, S. H. (2006). *Descriptors and Data Standard For Common Bean (Phaseolus vulgaris L.)*. Beijing: China Agriculture Press.
- Zhang, C. H., Cao, Y. S., Zong, X. X., and Wang, S. M. (2005). Morphological diversity and classification of common bean (*Phaseolus vulgaris* L.) germplasm resource in China. *Sci. Agric. Sin.* 38, 27–32. doi: 10.3321/j.issn:0578-1752.2005.01.006
- Zhang, X. Y. (2007). *Genetic Source Analysis And Genetic Diversity Study On Common Bean Germplasm Resources In China*. Dissertation, Chinese Academy of Agricultural Sciences, Beijing.
- Zhang, X. Y., Blair, M. W., and Wang, S. M. (2008). Genetic diversity of Chinese common bean (*Phaseolus vulgaris* L.) landraces assessed with simple sequence repeats markers. *Theor. Appl. Genet.* 117, 629–640. doi: 10.1007/s00122-008-0807-2
- Zheng, Z. J. (1997). *Chinese Food Legumes*. Beijing: China Agriculture Press.

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

Copyright © 2020 Long, Zhang, Zhang, Wu, Chen, Wang, Wang and Du. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Same Locus for Non-shattering Seed Pod in Two Independently Domesticated Legumes, *Vigna angularis* and *Vigna unguiculata*

Yu Takahashi¹, Alisa Kongjaimun², Chiaki Muto¹, Yuki Kobayashi¹, Masahiko Kumagai³, Hiroaki Sakai³, Kazuhito Satou^{4,5}, Kuniko Teruya⁴, Akino Shiroma⁴, Makiko Shimoji⁴, Takashi Hirano^{4†}, Takehisa Isemura¹, Hiroki Saito⁶, Akiko Baba-Kasai¹, Akito Kaga⁷, Prakit Somta^{8,9}, Norihiko Tomooka¹ and Ken Naito^{1*}

OPEN ACCESS

Edited by:

Jiayin Pang,
The University of Western Australia,
Australia

Reviewed by:

Jinfeng Chen,
University of California, Riverside,
United States
Jie Qiu,
Shanghai Normal University, China

*Correspondence:

Ken Naito
knaito@affrc.go.jp

† Present address:

Takashi Hirano,
Okinawa Laboratory, SENTAN
Pharma, Uruma, Japan

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 03 May 2020

Accepted: 23 June 2020

Published: 23 July 2020

Citation:

Takahashi Y, Kongjaimun A,
Muto C, Kobayashi Y, Kumagai M,
Sakai H, Satou K, Teruya K,
Shiroma A, Shimoji M, Hirano T,
Isemura T, Saito H, Baba-Kasai A,
Kaga A, Somta P, Tomooka N and
Naito K (2020) Same Locus
for Non-shattering Seed Pod in Two
Independently Domesticated
Legumes, *Vigna angularis* and *Vigna*
unguiculata. *Front. Genet.* 11:748.
doi: 10.3389/fgene.2020.00748

¹ Genetic Resources Center, National Agriculture and Food Research Organization (NARO), Tsukuba, Japan, ² Faculty of Animal Sciences and Agricultural Technology, Silpakorn University, Cha-Am, Thailand, ³ Advanced Analysis Center, NARO, Tsukuba, Japan, ⁴ Okinawa Institute of Advanced Sciences, Uruma, Japan, ⁵ Department of Genome Medicine, National Center for Child Health and Development, Setagaya, Japan, ⁶ Tropical Agriculture Research Front, Japan International Research Center for Agricultural Sciences, Ishigaki, Japan, ⁷ Institute of Crop Science, NARO, Tsukuba, Japan, ⁸ Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Nakhon Pathom, Thailand, ⁹ Center of Excellence on Agricultural Biotechnology (AG-BIO/PERDO-CHE), Bangkok, Thailand

Loss of pod shattering is one of the most important domestication-related traits in legume crops. The non-shattering phenotypes have been achieved either by disturbed formation of abscission layer between the valves, or by loss of helical tension in sclerenchyma of endocarp, that split open the pods to disperse the seeds. During domestication, azuki bean (*Vigna angularis*) and yard-long bean (*Vigna unguiculata* cv-gr. Sesquipedalis) have reduced or lost the sclerenchyma and thus the shattering behavior of seed pods. Here we performed fine-mapping with backcrossed populations and narrowed the candidate genomic region down to 4 kbp in azuki bean and 13 kbp in yard-long bean. Among the genes located in these regions, we found MYB26 genes encoded truncated proteins in azuki bean, yard-long bean, and even cowpea. As such, our findings indicate that independent domestication on the two legumes has selected the same locus for the same traits. We also argue that MYB26 could be a target gene for improving shattering phenotype in other legumes, such as soybean.

Keywords: legume, azuki bean, yard-long bean, cowpea, secondary wall thickening, MYB26, pod shattering, domestication

INTRODUCTION

One of the most important phenotypic changes during crop domestication is loss of seed shattering (Doubly et al., 2006). In legumes, this has been achieved by loss of pod shattering via malformation of dehiscence zone (Dong et al., 2014) or loss of helical shape change in the pod (Suanum et al., 2016; Murgia et al., 2017). However, this trait still needs to be improved as shattering problems sometimes force farmers to lose more than 20% of their annual yield (Philbrook and Oplinger, 1989). Thus, elucidating genetic mechanisms of shattering is important to reduce harvest loss.

Among the angiosperms, legumes have evolved a unique and sophisticated manner of shattering (Fahn and Zohary, 1955). Legume seeds are enclosed in seed pods, which explosively split open with its two valves twisting helically away from each other upon maturity (Armon et al., 2011). This mechanism is much more complicated compared to cereals, where matured seeds freely fall by abscission layer developed in pedicels (stalk of individual flower) (reviewed by Dong and Wang, 2015). *Brassica* and *Arabidopsis* are more similar to legumes in that the seeds are embedded in the siliques that also spring open upon maturity (Spence et al., 1996). However, their siliques do not exhibit any helical shape change as legume pods do. As such, knowledge obtained from rice and *Arabidopsis* cannot be simply applied to understand the shattering of legumes.

The key of helical shape change of legume seed pods is the development of a thick sclerenchyma with a bilayer structure on the endocarp (inner surface of the pod). In this bilayer structure, cellulose microfibrils of the outer and inner layers run at $\pm 45^\circ$ from the longitudinal axis of seed pods (Erb et al., 2013). As the matured seed pods dry, the microfibrils shrink in perpendicular directions and generate the helical force to blow the seeds off (Erb et al., 2013). In contrast, the endocarp of *Arabidopsis*, which develops cell layers with thickened secondary walls, is inflexible (Spence et al., 1996). As the matured siliques dry, the pericarp tissues shrink but the endocarp cell layers do not, generating tension to spring open the siliques from the dehiscence zone at the valve margins (Spence et al., 1996).

However, no specific genes have been identified for the sclerenchyma formation, despite several quantitative trait loci (QTL) analyses and genome-wide association studies locating several loci involved in legumes' pod shattering (Dong et al., 2014; Suanum et al., 2016; Murgia et al., 2017; Lo et al., 2018; Rau et al., 2019). In soybean, the domestication-type allele of *SHAT1-5* disturbs only dehiscence zone formation and does not affect helical shape change (Dong et al., 2014). Fiber content in seed pods, which is related to sclerenchyma formation, is reduced in common bean (Murgia et al., 2017; Parker et al., 2019; Rau et al., 2019), but the responsible gene is not cloned yet. An exception is *Pdh1* gene in soybean, which reduces the helical force of the sclerenchyma when disrupted (Funatsuki et al., 2014). However, the plants with non-functional PDH1 protein still initiate sclerenchyma formation, so PDH1 seems involved in later steps in sclerenchyma development.

Contrary to soybean, azuki bean [*Vigna angularis* (Willd.) Ohwi et Ohashi] and yard-long bean [*Vigna unguiculata* (L.) Walp. cv-gr. *Sesquipedalis* E. Westphal] seem disrupted in initiating sclerenchyma formation. In azuki bean, an important legume crop in East Asia, the seed pods of domesticated accessions normally form a dehiscence zone, but the pericarps are thinner and tenderer and show little helical shape change (Isemura et al., 2007; Kaga et al., 2008). In yard-long bean, which is often cultivated in Southeast Asia, pods are very long (60–100 cm) and do not show any helical shape change at all (Kongjaimun et al., 2012). In addition, the pods are extremely tender, and thus young pods are favored as vegetables

(Kongjaimun et al., 2013; Suanum et al., 2016). Therefore, we consider azuki bean and yard-long bean are the best materials to isolate the gene for initiating sclerenchyma formation.

Thus, in this study, we performed fine-mapping to identify the responsible locus for pod shattering and pod tenderness in azuki bean and yard-long bean, respectively. Interestingly, we have previously revealed that the QTLs controlling pod shattering in azuki bean and pod tenderness in yard-long bean are co-localized around 8–10 cM on the linkage group 7 (LG7) (Kongjaimun et al., 2013), which corresponds to Chr7 in azuki bean (Sakai et al., 2015) and Chr5 in cowpea (Lonardi et al., 2019). To narrow down the candidate region, we developed backcrossed populations and DNA markers based on the genome sequence of azuki bean (Sakai et al., 2015), cowpea (Lonardi et al., 2019), and the wild cowpea (sequenced in this study). Our efforts identified MYB26 transcription factor as the only candidate.

MATERIALS AND METHODS

Plant Materials and Growth Condition

All the accessions used in this study were provided by NARO Genebank (Tsukuba, Japan) (Figure 1 and Supplementary Figure 1).

For azuki bean, we started from the BC₁F₂ plants derived from a cross between domesticated azuki bean (JP81481, the recurrent parent) and a wild relative, *Vigna nepalensis* Tateishi & Maxted (JP107881) (Isemura et al., 2007). Of them, we selected one BC₁F₂ plant where the locus for pod shattering is fixed with the wild-type allele but most of other loci are fixed with domestication-type alleles and crossed again to the recurrent parent. We further backcrossed the obtained BC₂F₁ plants to obtain BC₃F₁ plants. We selfed them and obtained BC₃F₂ seeds from those with pod shattering phenotypes. For BC₃F₃, BC₃F₄, and BC₃F₅ populations, we kept selecting and selfing those with recombination within the candidate region. We also included plants that were heterozygous throughout the candidate region for obtaining more recombination and some that are homozygous in the same region to reconfirm the relationship of genotype and phenotype.

For yard-long bean, we started from BC₁F₂ plants derived from a cross between yard-long bean (JP89083, the recurrent parent) and a wild cowpea, *Vigna unguiculata* subsp. *dekindtiana* (Harms) Verdc. (JP81610) (Kongjaimun et al., 2013). Of them, we selected one BC₁F₂ plant that are fixed with wild-type allele at the locus for pod tenderness but are fixed with domestication-type at most other loci. The selected BC₁F₂ plants were backcrossed to the recurrent parent and further backcrossed to obtain BC₃F₁ plants. We selfed them and obtained BC₃F₂ seeds from those with hard pods phenotype. We kept selecting and selfing up to BC₃F₅ populations as done in azuki bean.

All the plants were grown from July through November in a field located in Tsukuba city, Japan, except BC₃F₅ plants of yard-long bean that were grown in a greenhouse from September through January.

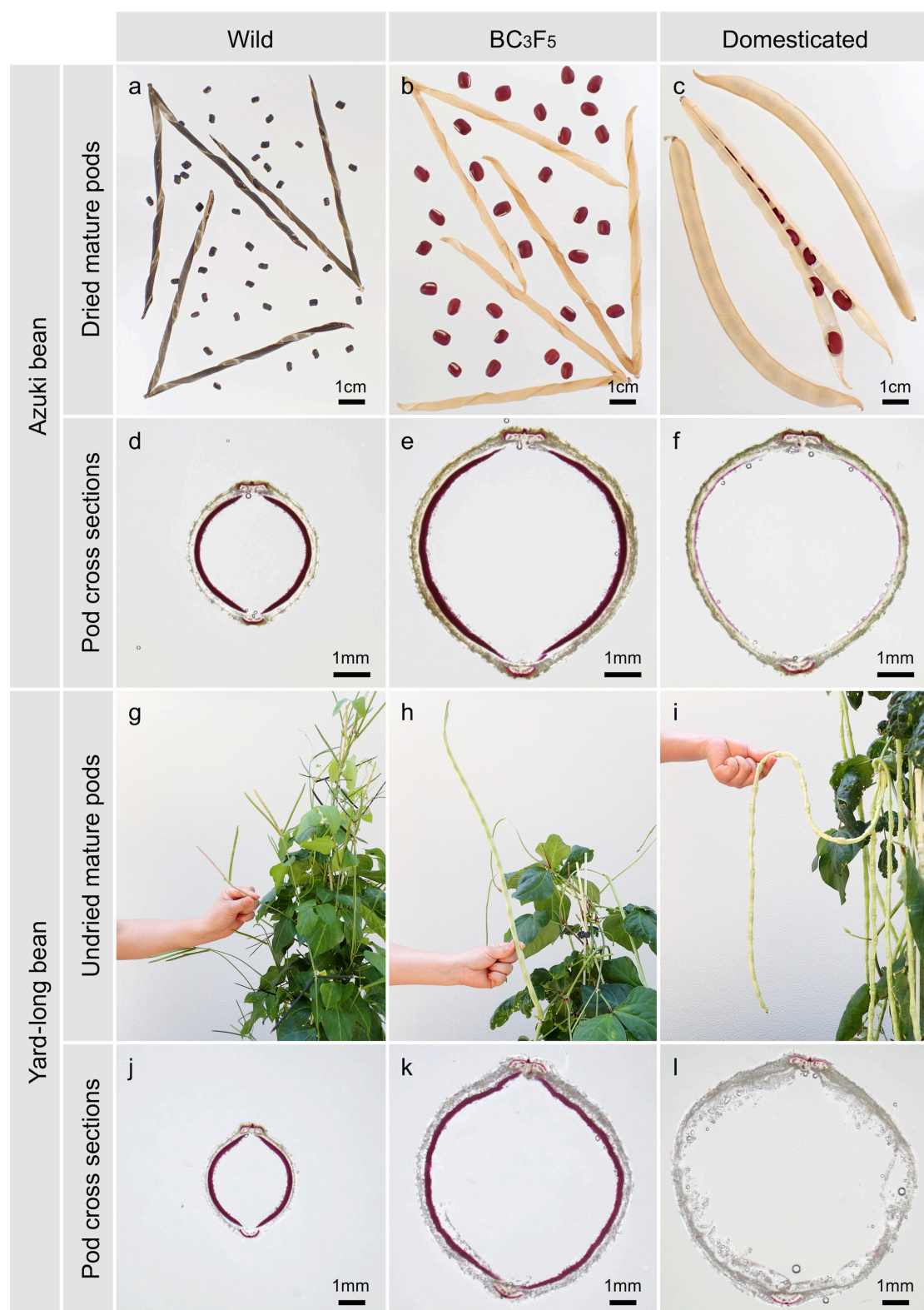


FIGURE 1 | Pod phenotypes of parental accessions and BC₃F₅ plants with wild-type phenotypes in genetic background of domesticated genotypes. **(a,d)** Wild azuki bean (JP107881). **(b,e)** BC₃F₅ plant with shattering phenotype. **(c,f)** Domesticated azuki bean (JP81481). **(g,j)** Wild cowpea (JP81610). **(h,k)** BC₃F₅ plants with hard pod phenotype. **(i,l)** Yard-long bean (JP89093). **(a–c)** Dried mature pods. **(d–f,j–l)** Cross sections of pods stained with phloroglucinol. **(g–i)** Young pods.

Observation of Sclerenchymal Tissue

We observed sclerenchymal tissues in seed pods by staining lignin with the Wiesner (phloroglucinol-HCl) reaction (Pormar et al., 2002). We harvested three undried mature seed pods per plant, sliced them to 200- μ m sections with a Plant Microtome MTH-1 (Nippon Medical & Chemical Instruments Co., Ltd., Osaka, Japan), incubated in phloroglucinol-HCl solution for 1 min and observed with stereoscopic microscope SZX7 (OLYMPUS, Tokyo) for BC₃F₅ plants of azuki bean and yard-long bean. Phloroglucinol-HCl solution was prepared by dissolving 1 g of phloroglucin in 50 ml ethanol and adding 25 ml of concentrated hydrochloric acid.

Phenotyping

For azuki bean populations, we evaluated pod shattering by counting the number of twists in a pod, as described in Isemura et al. (2007). For BC₃F₂ and BC₃F₃, we harvested five pods per plant before shattering and measured the length. We then incubated the pods at 60°C to let them totally dry. Unless the pods dehisced, we manually opened the pods and counted the number of twists and calculated the number of twists/cm. For BC₃F₄ and BC₃F₅, we harvested 10 pods per plant, let them dry, and evaluated the rate of shattering pods before counting the number of twists.

For yard-long bean, we manually evaluated the tenderness of five young pods per plant by binarizing hard or soft.

We also measured 100 seed weight on BC₃F₂ and BC₃F₅ plants of the azuki bean population and BC₃F₄ and BC₃F₅ plants of the yard-long bean population.

Sequencing and Assembly of the Wild Cowpea Genome

We sequenced the whole genome of wild cowpea (JP81610) with RSII sequencer (Pacific Biosciences, Menlo Park, CA, United States), as we have done previously for azuki bean (Sakai et al., 2015). DNA was isolated from 1 g of unexpanded leaves with the cetyl trimethylammonium bromide (CTAB) method and purified with Genomic Tip 20/G (Qiagen K. K. Tokyo). The extracted DNA was sheared into 20-kb fragments using g-TUBE (Covaris, MA, United States) and converted into 20-kb SMRTbell template libraries. The library was size selected for a lower cutoff of 10 kb with BluePippin (Sage Science, MA, United States). Sequencing was performed on the PacBio RS II using P6 polymerase binding and C4 sequencing kits with 360 min acquisition. In total, 21 SMRT cells were used to obtain ~26.4 Gb of raw reads.

In total, 4 million PacBio reads were used for *de novo* assembly with Canu v1.6 under the default settings (corOvlErrorRate = 0.2400, obtOvlErrorRate = 0.0450, utgOvlErrorRate = 0.0450, corErrorRate = 0.3000, obtErrorRate = 0.0450, utgErrorRate = 0.0450, cnsErrorRate = 0.0750). About 23.5x error corrected and trimmed reads longer than 1,000 bp were assembled to contigs. Assembled contigs were polished by PacBio raw reads by using the arrow in GenomicConsensus v2.3.2 (Pacific Biosciences of California, Inc.).

Repeat detection was conducted by RepeatMasker ver. 4.0 (Smit et al., 2013)¹. A *de novo* repeats library of wild cowpea genome constructed by RepeatModeler ver. 1.0.11 (see footnote 1) and the Fabaceae repeats library in RepBase24.02 (Smit and Hubley, 2008)² were used for the prediction.

Ab initio gene prediction was done with AUGUSTUS (version 3.3.2) (Stanke et al., 2008). A set of gene annotation information of recently published cowpea genome was used for training AUGUSTUS (Lonardi et al., 2019). We trained a new model twice using 1,000 high-confidential genes selected based on abundance of annotations of domains, pathway networks, and gene ontology information. BUSCO v3 (Waterhouse et al., 2017) was used to evaluate protein sequences of annotated genes.

Single-nucleotide polymorphisms (SNPs) and short insertions/deletions (indels) were detected by using MUMmer v3.23 (Kurtz et al., 2004). Genome alignment of our wild cowpea assembly and the reference cowpea genome (Lonardi et al., 2019) was conducted by the nucmer command with the following options: -maxgap = 500 -mincluster = 100. Then one-to-one alignment was extracted with delta-filter command with an option: -1. Based on the alignments, SNPs and indels were reported by the show-snps command.

Raw sequence data, the assembled sequences, and gene annotations are all available from DNA Data Bank of Japan (DDBJ)³ under the BioProject ID PRJDB8129.

Genotyping

We extracted DNA from the seeds as described by Kamiya and Kiguchi (2003) and genotyped them by fragment analysis with capillary electrophoresis for simple sequence repeats (SSRs) and INDELs as described by Isemura et al. (2007) or by directly sequencing SNP sites (see section “Direct Sequencing” below). The information of the markers we used is summarized in **Supplementary Table 1**.

We designed primers by Primer3 (Untergasser et al., 2012) to amplify polymorphic sites between the domesticated azuki bean and *V. nepalensis* that are available in *Vigna* Genome Server (VigGS⁴; Sakai et al., 2016) and those between the domesticated cowpea genome in Legume Information System⁵ and the wild cowpea genome sequenced in this study. Parameters for designing primers on Primer3⁶ were 20–30 bp in length, 55°C–65°C in annealing temperature, and 100–350 bp and >700 bp in expected length of amplified fragments for fragment analysis and direct sequencing, respectively.

For azuki bean, the markers we used were as follows:

BC₃F₂: CEDG064 and SPD01-SPD07

BC₃F₃: CEDG064, SPD03, SPD04, SPD08, SPD09

BC₃F₄: CEDG064 and SPD08-SPD11

BC₃F₅: CEDG064, SPD10, SPD11 and the sequencing primers for Vigan.07G034400

¹<http://repeatmasker.org>

²<https://www.girinst.org>

³<https://www.ddbj.nig.ac.jp/index.html>

⁴<https://viggs.dna.affrc.go.jp>

⁵<https://legumeinfo.org/>

⁶<http://primer3.ut.ee>

For yard-long bean, the markers we used were as follows:

BC₃F₄: VuPT01–VuPT06

BC₃F₅: VuPT03, VuPT04, VuPT07–VuPT13 and the sequencing primers for Vigun05g273300, Vigun05g273400, and Vigun05g0723500.

Direct Sequencing

To sequence the (potentially) polymorphic sites, we amplified the template DNA with AmpliTaq Gold 360 Master Mix (Thermo Fisher Scientific K. K., Tokyo), performed sequencing reaction with BigDye Terminator v3.1 (Thermo Fisher Scientific K. K., Tokyo), and sequenced with ABI Genetic Analyzer 3130xl (Thermo Fisher Scientific K. K., Tokyo), according to the provider's protocol.

Determining Transcribed Sequences of Vigan.07g34400 and Vigun05g273500

To sequence both the domesticated and the wild alleles of Vigan.07g34400.01 and Vigun05g273500 genes, we sequenced the transcribed sequences of both loci. We extracted total RNA from 100 mg of flower buds right before flowering of all the parental accessions using RNeasy Plant Mini Kit (QIAGEN) with RNase-free DNase I (Invitrogen). Total RNA of 1 µg was converted into first-strand cDNA with Super Script II Reverse Transcriptase (Invitrogen) and Oligo(dT)12–18 Primer (Invitrogen) following the manufacturer's instructions. The cDNA sequences of Vigan.07g34400.01 and Vigun05g273300.01 were then amplified and sequenced with the primers (**Supplementary Table 1**) and then transferred to the direct-sequencing protocol described above.

RESULTS

Anatomical Analysis of Pod Sclerenchyma

To confirm that azuki bean and yard-long bean have lost or reduced sclerenchyma in seed pods, we observed cross sections of seed pods stained with phloroglucinol–HCl and found a clear-cut difference between the wild and domesticated accessions. In the wild azuki bean and cowpea, which have shattering phenotypes, thick layers of sclerenchyma (~0.3 mm) were formed on the endocarps of seed pods (**Figure 1**). However, in the domesticated azuki bean and yard-long bean, which are both non-shattering, the sclerenchyma layer was slightly formed (<0.1 mm) or not formed at all, respectively (**Figure 1**).

We also evaluated the correlation between the thickness of pod sclerenchyma and helical shape change in seed pods (number of twists/cm) in the BC₃F₅ plants of azuki bean and yard-long bean. As a result, all the shattering plants in azuki bean population formed thicker sclerenchyma (0.15–0.20 mm) and showed a stronger helical shape change (0.30–0.43) than the non-shattering plants did (0.00–0.08 mm and 0.00–0.05, respectively) (**Figure 1** and **Supplementary Figure 1**). The correlation coefficient between the sclerenchyma thickness and the helical shape change was 0.92. In the yard-long bean

population, all the plants with hard pod phenotype formed sclerenchyma (0.16–0.22 mm) and showed little helical shape change (0.09–0.12), whereas those with soft pod phenotype did not form sclerenchyma or show helical shape change at all (**Figure 1** and **Supplementary Figure 1**). The correlation coefficient between the sclerenchyma thickness and the helical shape change was 0.99.

In addition, we observed cross sections of seed pods in a wild soybean and three domesticated soybean cultivars and found thick sclerenchyma layers in all the accessions (**Supplementary Figure 2**).

Fine-Mapping Pod Shattering in Azuki Bean

We previously mapped the QTL for pod shattering in between CEDG182 and CEDG174 on LG7, which is around 1.4–4.0 Mbp in Chr7 of the reference azuki bean genome (Sakai et al., 2015) (**Figure 2**). To more finely locate the genetic factor for pod shattering, we genotyped 1,049 BC₃F₂ plants and selected 238 for phenotyping. The obtained phenotype and genotype data revealed the pod shattering factor was completely linked with CEDG064 and located in between SPD03 and SPD04 (**Figure 2** and **Supplementary Table 2**). We selected 46 plants out of the 238 (**Supplementary Table 2**), selfed them, and obtained 4,222 BC₃F₃ seeds.

Of the BC₃F₃ seeds, we selected 53 that had recombination between SPD03 and SPD04. Phenotyping and genotyping on these 53 revealed that pod shattering factor was still completely linked with CEDG064 and was in between SPD08 and SPD09 (**Figure 2** and **Supplementary Table 3**), which is about 19 kb long and contained three genes (**Figure 2**).

Since this region seemed to have a relatively higher recombination rate, we expected it would be possible to further locate the pod shattering factor. Thus, of the 53 BC₃F₃ plants, we selected 24 plants (**Supplementary Table 3**) and selfed them to obtain BC₃F₄ seeds. We then cultivated 1–7 plants/line (81 in total; **Supplementary Table 4**) for further genotyping and phenotyping (**Figure 2**).

As a result, we again found CEDG064 was completely linked with the phenotype, but we also obtained 29 recombinants between SPD10 and CED064 and five between CEDG064 and SPD11 (**Supplementary Table 4**). Thus, the pod shattering factor was located in a 9-kb region between SPD10 and SPD11, where there is only one gene Vigan.07g34400 (**Figure 2**), which encoded VaMYB26, an ortholog of AtMYB26.

To see if there were any loss-of-function mutations in VaMYB26 in azuki bean, we searched the polymorphism data between wild and domesticated azuki bean provided by VigGS. The database showed that there were six SNPs (SNP01–SNP06) and one INDEL (INDEL01) within its open reading frame (ORF). Of these, all the SNPs were in the introns or in the 3'-untranslated region (UTR), but the INDEL was within the coding DNA sequence (CDS) (a T insertion 4 bp before the stop codon in the domesticated azuki bean). Since the wild azuki bean did not have this T insertion, the CDS could be 405 bp (125 aa) longer (**Supplementary Figure 3**). Interestingly,

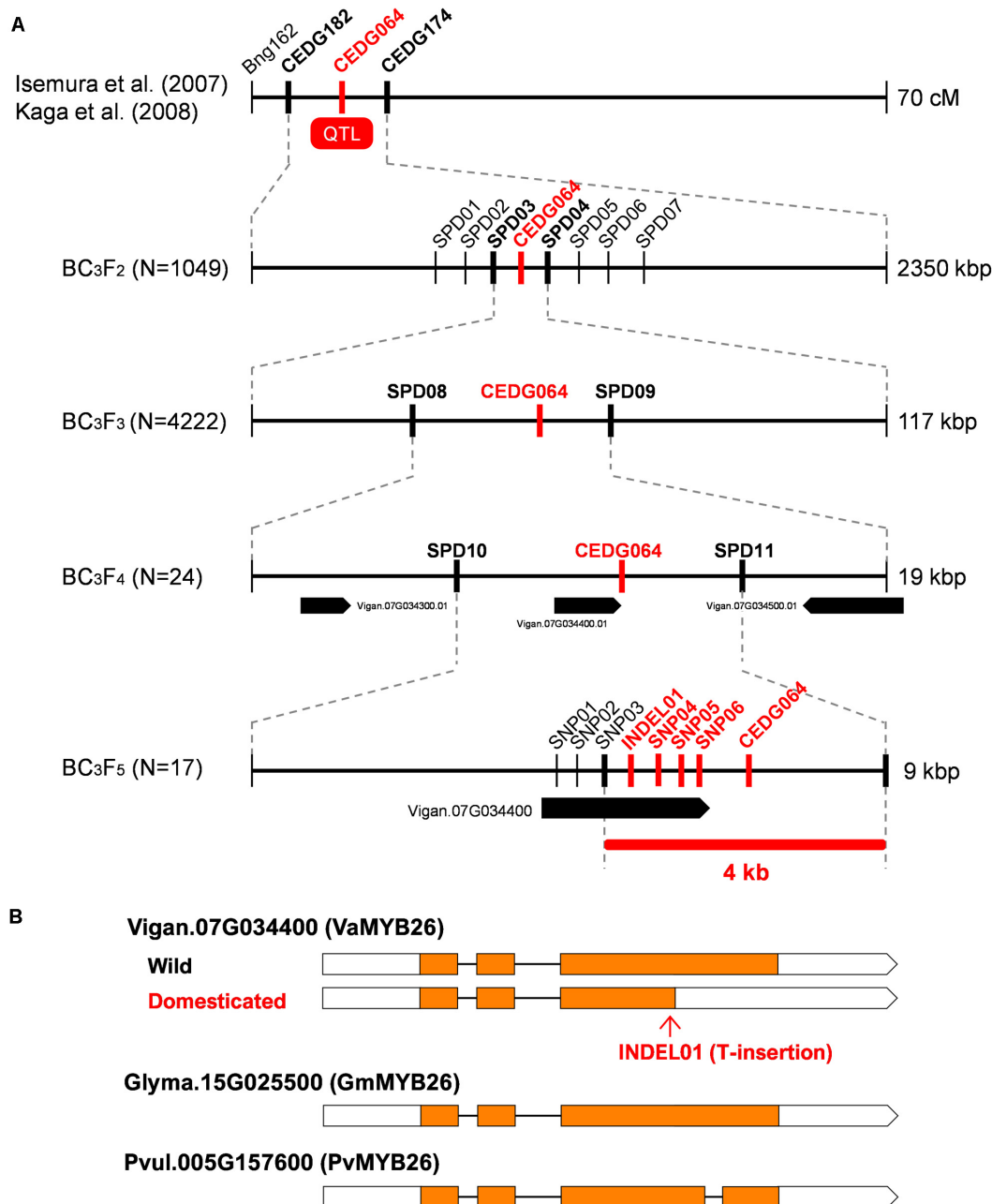


FIGURE 2 | Fine-mapping of pod shattering factor in azuki bean. **(A)** Process of fine mapping. Vertical lines indicate DNA marker sites. Red markers indicate those with complete linkage with the phenotype in the mapping population. Bold black markers indicate those neighboring the candidate region. **(B)** A schematic of the wild-type and domesticated-type alleles of Vigan.07g034400. The MYB26 orthologues of soybean (*Glyma.15G025500*) and common bean (*Pvul.005G157600*) are also shown. Boxes indicate exons, whereas lines indicate introns. Coding DNA sequences (CDS) are filled with yellow.

a BLAST search of MYB26 gene revealed that the longer version is widely conserved across plant taxa (Figure 2). Thus, the VaMYB26 in the domesticated azuki bean seemed to have an immature stop codon.

To confirm the polymorphisms identified in the database, we determined the genomic and the transcribed sequences of VaMYB26 locus of both parents and found all the polymorphisms in the ORF were true (Supplementary Figure 3). We also

sequenced the 17 lines of BC₃F₅, which we obtained from the selected BC₃F₄ plants (those with fixed genotypes between SPD03 and SPD09; Supplementary Table 5), to further confirm the relationship between the genotypes and the phenotype. Of the 17, one had recombination between SNP03 and INDEL01, but none had recombination between INDEL01 and CEDG064 (Supplementary Table 5). Thus, the genotypes at INDEL01, SNP04–SNP06, and CEDG064 were completely linked with the

pod shattering phenotype, locating the pod shattering factor within the 4-kb region between SNP03 and SPD11 (Figure 2 and Supplementary Table 5).

Assembly and Annotation of Wild Cowpea Genome

Though the reference genome of cowpea had been released (Lonardi et al., 2019), it was yet not easy to obtain polymorphic markers because the genome sequence of the wild cowpea was not available. Thus, to facilitate our fine-mapping the factor of pod tenderness, we sequenced and assembled the genome of *V. unguiculata* subsp. *dekindtiana* (JP81610), which is a wild cowpea accession used to develop the mapping population of yard-long bean. We used PacBio RSII and obtained 4,413,480 raw reads with an average read length of 6.0 kbp and N50 length of 10.6 kbp, covering 44.9X of the estimated genome size (585.8 Mbp) (Lonardi et al., 2019) (Supplementary Figure 4 and Supplementary Table 6). The assembly produced 4,285 contigs covering 488.5 Mbp (83.4%) of the cowpea genome, with the contig NG50 of 438.6 kbp and the maximum contig length of 2.6 Mbp, respectively (Supplementary Table 7). Of the assembled sequences, 45.7% composed of transposable elements, of which LTR retrotransposons share the largest content (20.5% of the assembly) (Supplementary Table 8). Our *ab initio* gene prediction detected 36,061 protein-coding genes, and 27,345 of those were non-repeat related (Supplementary Table 9). The annotated genes of the wild cowpea contained 93.2% (85.6% complete and 7.6% partial) of the 1,440 plant genes in BUSCO v3 (Waterhouse et al., 2017).

We aligned the genome sequences of the wild cowpea to those of the reference cowpea (Lonardi et al., 2019) and identified 5,661,319 SNPs and 1,626,169 INDELs. As expected, we detected fewer SNPs and INDELs around centromeric and pericentromeric regions but many in chromosome arms, which were presumably gene-rich regions (Supplementary Figure 5).

Fine-Mapping Pod Tenderness Factor in Yard-Long Bean

The QTL for pod tenderness was previously located between cp06388 and VR294 on LG7 (Kongjaimun et al., 2013) (Figure 3), which corresponds to about 47.5–45.5 Mbp region in Chr5 of the cowpea genome (*Vigna unguiculata* v1.0, NSF, UCR, USAID, DOE-JGI)⁷. To more finely locate the pod tenderness factor, we designed more markers based on the SNPs and INDELs we identified above, genotyped 2,304 BC₃F₄ seeds, and selected 195 for phenotyping (Supplementary Table 10).

The obtained data of genotype and phenotype located the pod tenderness factor between VuPT03 and VuPT04, which was about 47 kbp (Figure 3 and Supplementary Table 10). Although no marker was completely linked with the phenotype, eight plants had recombination between this region (Figure 3 and Supplementary Table 10).

Of the 195 BC₃F₄ plants we tested, we selfed the eight plants and obtained BC₃F₅ for further mapping. As a result,

we located the candidate regions in between VuPT08 and VuPT11, which was 13 kbp long and contained three genes, Vigun05g27300, Vigun05g273400, and Vigun05g273500. Of note, Vigun05g273500 encoded VuMYB26 (Figure 3 and Supplementary Table 11).

According to the SNPs and INDELs data, Vigun05g273500 had an A to G substitution which might disrupt the junction site of the first intron and the second exon (Supplementary Figure 6). On the other hand, Vigun05g27300 seemed to have only synonymous SNPs, and Vigun05g273400 was a non-coding gene. All the SNPs were confirmed by directly sequencing the ORFs of these genes. Interestingly, the ORF sequences of the three genes were exactly the same between the reference cowpea and yard-long bean, which means yard-long bean and the reference cowpea share the domestication-type allele in this locus (Supplementary Figure 6).

To test whether the A to G substitution disrupted splicing, we sequenced cDNAs of this locus prepared from both parents. As a result, we found, in the domestication-type allele, the junction site of the first intron and the second exon had shifted downstream by 8 bp, which could result in a frameshift and an immature stop codon in the middle of the second exon (Supplementary Figure 6). This product would encode a protein of 60 aa, which would be 305 aa shorter than that of the wild-type allele.

Seed Size Increase by Loss of Sclerenchymal Tissue

Because Murgia et al. (2017) suggested the pod shattering phenotypes load an energy cost on plants and limit seed size, we measured 100 seed weight of the mapping population (BC₃F₂ of azuki bean and BC₃F₄ of yard-long bean).

As expected, those with domestication-type trait produced larger seeds compared to those with the wild-type trait (Figure 4). In the azuki bean population, the seeds of non-shattering plants were 15.4 ± 2.4 g/100 seeds, whereas the seeds of shattering plants were 13.3 ± 1.9 g/100 seeds. In the yard-long bean population, the seeds of tender-pod plants were 13.6 ± 1.9 g/100 seeds, whereas the seeds of hard-pod plants were 12.9 ± 1.7 g/100 seeds. The following *t*-test revealed that in both cases, the seeds of the plants with the domestication-type phenotypes produced significantly larger seeds than the others ($p = 4.6 \times 10^{-7}$ for azuki bean and $p = 0.019$ for yard-long bean).

We also measured 100 seed weight of BC₃F₅ plants and observed the same trend, though the differences were not significant (Supplementary Figure 7). In azuki bean, the seeds of non-shattering plants were 9.7 ± 1.4 g/100 seeds, whereas the seeds of shattering plants were 8.7 ± 1.2 g/100 seeds. In yard-long bean, the seeds of tender-pod plants were 14.2 ± 1.5 g/100 seeds, whereas the seeds of hard-pod plants were 12.2 ± 1.2 g/100 seeds.

DISCUSSION

In this study, we identified Vigan.07g034400 (Figure 2) and Vigun05g273500 (Figure 3) are the best candidate genes for sclerenchyma formation in the seed pods of azuki bean and yard-long bean, respectively (Figure 1). The domestication-type alleles

⁷<http://phytozome.jgi.doe.gov/>

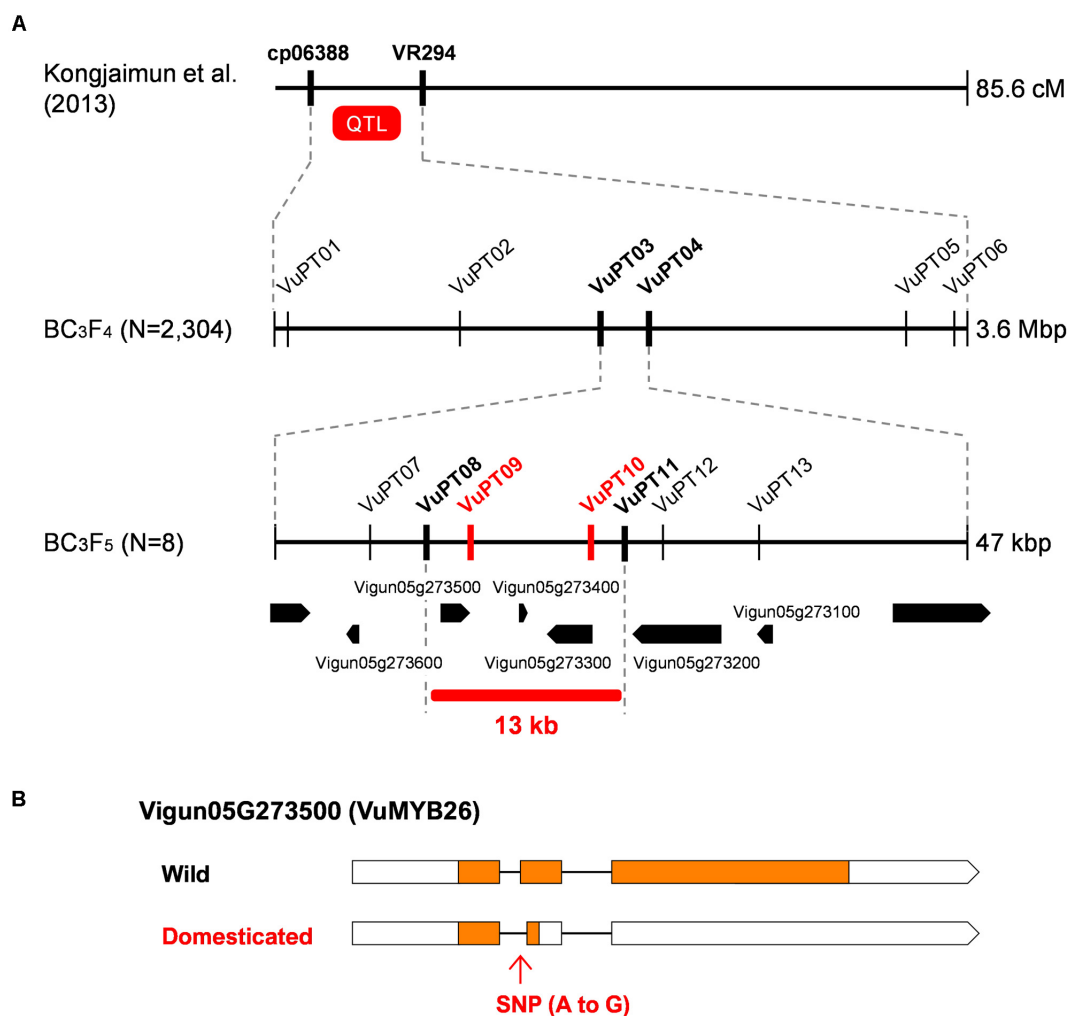


FIGURE 3 | Fine-mapping of pod tenderness factor in yard-long bean. **(A)** Process of fine-mapping. Vertical lines indicate DNA marker sites. Red markers indicate those with complete linkage with the phenotype in the mapping population. Bold black markers indicate those neighboring the candidate region. **(B)** A schematic of the wild-type and domesticated-type alleles of Vigun05g273500. Boxes indicate exons, whereas lines indicate introns. Coding DNA sequences (CDS) are filled with yellow.

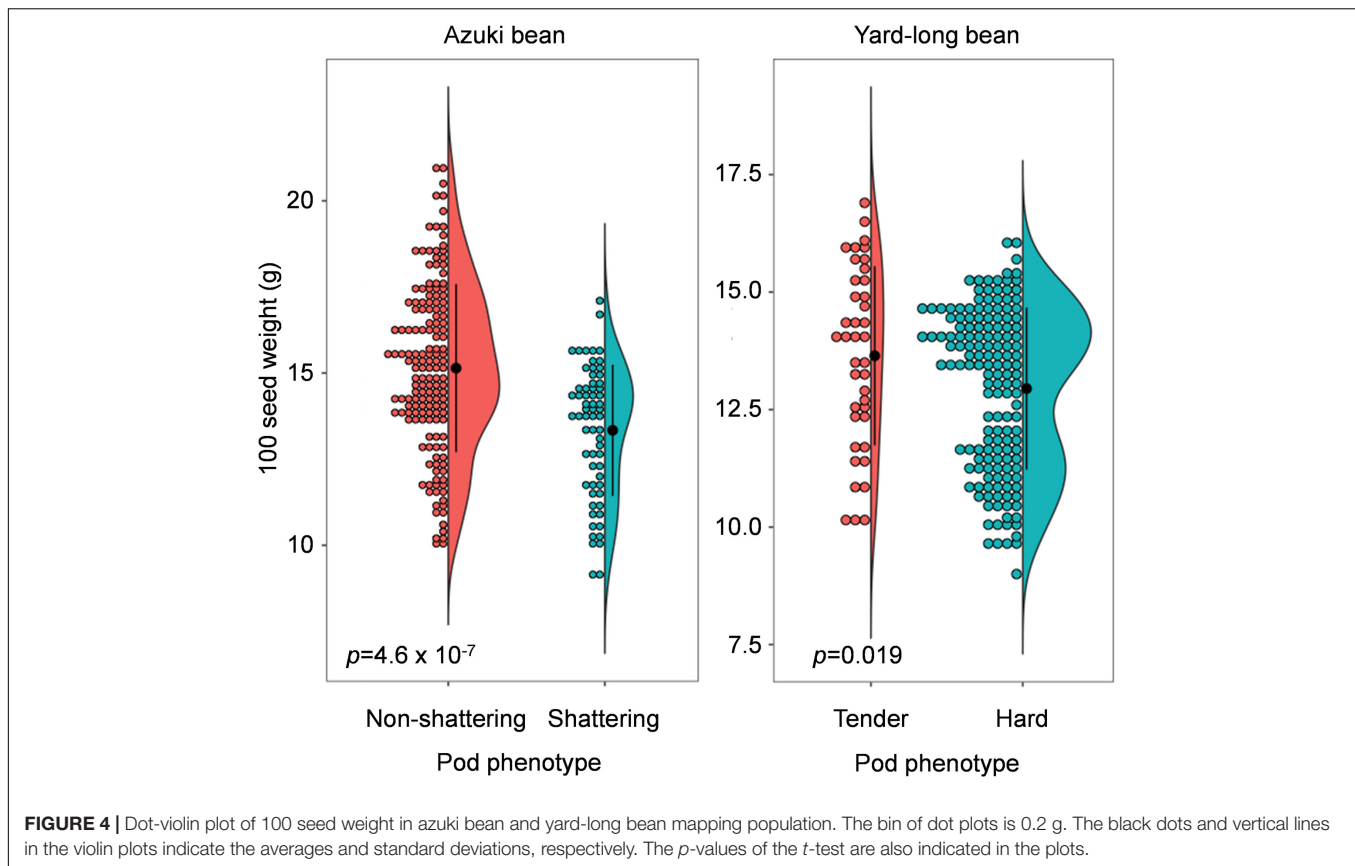
of both loci encode truncated proteins due to immature stop codons (**Supplementary Figures 3, 6**), leading to a reduction or loss of sclerenchyma, which is required to generate the helical tension for splitting open legume pods (**Figure 1** and **Supplementary Figure 1**). Although our primary goal of the fine-mapping in yard-long bean was to identify a gene for the extreme tenderness of its seed pod, the SNP we identified was in common between yard-long bean and the reference cowpea (**Supplementary Figure 6**). Thus, the domestication-type allele of Vigun05g273500 is not the specific allele to yard-long bean but is the one which has originally been selected for resistance to pod shattering during domestication of cowpea (Lush and Evans, 1981).

Vigan.07g034400 and Vigun05g273500 are orthologous to each other and to *Arabidopsis* transcription factor MYB26, which is strongly expressed in the anther and promotes secondary wall thickening in the endothelium (Yang et al.,

2007). Interestingly, this secondary wall thickening is critical for anther dehiscence and thus for releasing the matured pollen (Yang et al., 2007). However, this gene is hardly expressed in siliques and is not involved in pod dehiscence (Yang et al., 2007). On the other hand, MYB26 is expressed in seed pods in many legume species according to the expression database [VigGS, the *Phaseolus vulgaris* Gene Expression Atlas (PvGEA)⁸ and Soybean eFP Browser⁹]. In addition, Di Vittori et al. (2020) has recently revealed in common bean that reduced expression of MYB26 is a key for non-shattering genotypes. Thus, in legumes, MYB26 plays a crucial role in the development of pod sclerenchyma and has been a common target of human selection among the three independently domesticated legumes, azuki bean, cowpea, and

⁸<http://plantgrn.noble.org/PvGEA/index.jsp>

⁹<http://bar.utoronto.ca/efpsoybean/cgi-bin/efpWeb.cgi>



common bean. This fact suggests the importance of MYB26 in pod shattering, despite there are usually multiple options to achieve a domestication-related trait (Doubley et al., 2006; Takahashi et al., 2019).

Our results also indicate that the accumulation of some more mutations is necessary to achieve the extreme tenderness of seed pod in yard-long bean. Indeed, the QTL analysis on pod tenderness (Kongjaimun et al., 2013) and pod fiber content (Suanum et al., 2016) identified a few other QTLs than the locus we identified here. Thus, such mutations had been selected for the complete loss of pod sclerenchyma during the secondary domestication in Asia (Faris, 1965).

In addition, we have to admit that availability of whole-genome sequences greatly facilitated the study. Although the lower coverage and shorter read length of our assembly resulted in ~10 times lower contiguity than the reference cowpea genome (Lonardi et al., 2019), its accuracy was good enough as all the SNPs between the two assemblies, which we tried to use as markers, were confirmed to be true (Supplementary Tables 5, 11 and Supplementary Figures 3, 6). We also reemphasize the power of map-base cloning, which is a classic, time-consuming, but reliable approach to isolate responsible genes.

We also consider the knowledge obtained in this study is applicable to improve soybean or other legume crops. As described before, soybean has reduced shattering ability

by disturbing dehiscence zone formation between the valves (Dong et al., 2014) and by reducing helical tension of the sclerenchyma (Funatsuki et al., 2014). However, it still forms thick sclerenchyma in the seed pods (Supplementary Figure 2), which causes shattering problems especially under drought conditions (Philbrook and Oplinger, 1989). Thus, loss-of-function mutations or genome-editing in MYB26 genes may reduce sclerenchyma and enforce resistance to shattering. In addition, our findings indicate that loss or reduction of sclerenchyma may result in a slight (~10%) increase in seed size (Figure 4 and Supplementary Figure 7). Murgia et al. (2017) also reported that common bean plants with non-shattering genotypes produce slightly larger seeds than those with shattering genotypes. This might be because developing sclerenchyma loads an energy cost for plants that could be allocated for seed production (Murgia et al., 2017), or pod hardness limits space for seed enlargement. In any case, MYB26 may serve not only for reducing harvest loss but also for directly increasing seed size.

CONCLUSION

Our map-based cloning approach with support of whole-genome sequences identified MYB26 as an only candidate for the development of pod sclerenchyma, which generates helical tension of legumes' shattering pods. Our findings

suggest that MYB26 might be a good target to improve resistance to pod shattering in soybean and other legumes.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YT, PS, AKo, and KN planned the study. YT, AKo, CM, YK, TI, HSai, AKa, NT, and KN performed the experiments and collected the data. MK, HSak, KS, KT, AS, MS, and TH performed the genome sequencing, *de novo* assembly, and gene annotation. YT, AKa, HSak, MK, and KN analyzed the data. KN wrote

the manuscript. All the authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by JSPS KAKENHI Grant Numbers 23688002, 13J09808, and 26850006. It was also partially supported by Research Supporting Program of the Advanced Analysis Center, National Agriculture and Food Research Organization (NARO), and the Genebank Project, NARO. The manuscript has been released as a preprint at bioRxiv (Takahashi et al., 2019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00748/full#supplementary-material>

REFERENCES

- Armon, S., Efrati, E., Kupferman, R., and Sharon, E. (2011). Geometry and mechanics in the opening of chiral seed pods. *Science* 333, 1726–1730. doi: 10.1126/science.1203874
- Di Vittori, V., Bitocchi, E., Rodriguez, M., Alseekh, S., Bellucci, E., Nanni, L., et al. (2020). Pod indehiscence in common bean is associated to the fine regulation of PvMYB26 and a non-functional abscission layer. *bioRxiv [Preprint]* doi: 10.1101/2020.04.02.021972
- 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.
- Doubly, F. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321. doi: 10.1016/j.cell.2006.12.006
- Erb, R. M., Sander, J. S., Grisch, R., and Studart, A. R. (2013). Self-shaping composites with programmable microstructures. *Nat. Commun.* 4:1712.
- Fahn, A., and Zohary, M. (1955). On the pericardial structure of the legumen, its evolution and relation to dehiscence. *Phytomorphology* 5, 99–111.
- Faris, D. G. (1965). The origin and evolution of the cultivated forms of *Vigna sinensis*. *Can. J. Genet. Cytol.* 7, 433–452. doi: 10.1139/g65-058
- 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
- Isemura, T., Kaga, A., Konishi, S., Ando, T., Tomooka, N., Han, O. K., et al. (2007). Genome dissection of traits related to domestication in azuki bean (*Vigna angularis*) and comparison with other warm-season legumes. *Ann. Bot.* 100, 1053–1071. doi: 10.1093/aob/mcm155
- Kaga, A., Isemura, T., Tomooka, N., and Vaughan, D. A. (2008). The genetics of domestication of the azuki bean (*Vigna angularis*). *Genetics* 178, 1013–1036. doi: 10.1534/genetics.107.078451
- Kamiya, M., and Kiguchi, T. (2003). Rapid DNA extraction method from soybean seeds. *Breed. Sci.* 53, 277–279. doi: 10.1270/jsbbs.53.277
- Kongjaimun, A., Kaga, A., Tomooka, N., Somta, P., Vaughan, D. A., and Srinives, P. (2012). The genetics of domestication of yardlong bean, *Vigna unguiculata* (L.) Walp. ssp. *unguiculata* cv.-gr. *Sesquipedalis*. *Ann. Bot.* 109, 1185–1200. doi: 10.1093/aob/mcs048
- Kongjaimun, A., Somta, P., Tomooka, N., Kaga, A., Vaughan, D. A., and Srinives, P. (2013). QTL mapping of pod tenderness and total soluble solid in yardlong bean [*Vigna unguiculata* (L.) Walp. subsp. *unguiculata* cv.-gr. *sesquipedalis*]. *Euphytica* 189:217. doi: 10.1007/s10681-012-0781-2
- Kurtz, S., Pilliply, A., Delcher, A. L., Smoot, M., Shumway, M., Sntonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12.
- Lo, S., Muñoz-Amatriain, M., Boukar, O., Herniter, I., Cisse, N., Guo, Y. N., et al. (2018). Identification of QTL controlling domestication-related traits in cowpea (*Vigna unguiculata* L. Walp.). *Sci. Rep.* 8:6261.
- Lonardi, S., Muñoz-Amatriain, M., Liang, Q., Shu, S., Wanamaker, S. I., Lo, S., et al. (2019). The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant J.* 98, 767–782.
- Lush, W. M., and Evans, L. T. (1981). The domestication and improvement of cowpeas (*Vigna unguiculata* (L.) Walp.). *Euphytica* 30, 579–587. doi: 10.1007/bf00038783
- Murgia, M. L., Attene, G., Rodriguez, M., Bitocchi, E., Bellucci, E., Fois, D., et al. (2017). A Comprehensive Phenotypic Investigation of the “Pod-Shattering Syndrome” in Common Bean. *Front. Plant Sci.* 8:251. doi: 10.3389/fpls.2017.00251
- Parker, T. A., Miery Teran, J. C. B., Palkovic, A., Jernstedt, J., and Gepts, P. (2019). Genetic control of pod dehiscence in domesticated common bean: associations with range expansion and local aridity conditions. *bioRxiv [Preprint]* doi: 10.1101/517516
- Philbrook, B., and Oplinger, E. S. (1989). Soybean field losses as influenced by harvest delays. *Agron. J.* 81, 251–258. doi: 10.2134/agronj1989.00021962008100020023x
- Pormar, F., Merino, F., and Barceló, R. (2002). O-4-linked coniferyl and sinapyl aldehydes in lignifying cell walls are the main targets of the Wiesner (phloroglucinol-HCl) reaction. *Protoplasma* 220, 17–28. doi: 10.1007/s00709-002-0030-y
- Rau, D., Murgia, M. L., Rodriguez, M., Bitocchi, E., Bellucci, E., Fois, D., et al. (2019). Genomic dissection of pod shattering in common bean: mutations at non-orthologous loci at the basis of convergent phenotypic evolution under domestication of leguminous species. *Plant J.* 97, 693–714. doi: 10.1111/tj.14155
- Sakai, H., Naito, K., Ogiso-Tanaka, E., Takahashi, Y., Iseki, K., Muto, C., et al. (2015). The power of single molecule real-time sequencing technology in the *de novo* assembly of a eukaryotic genome. *Sci. Rep.* 5:16780.
- Sakai, H., Naito, K., Takahashi, Y., Sato, T., Yamamoto, T., Muto, I., et al. (2016). The *Vigna* Genome Server, “VigGS”: a Genomic Knowledge Base of the Genus *Vigna* Based on High-Quality, Annotated Genome Sequence of the Azuki Bean, *Vigna angularis* (Willd.) Ohwi & Ohashi. *Plant Cell Physiol.* 57:e2. doi: 10.1093/pcp/pcv189

- Smit, A. F. A., and Hubley, R. (2008). *RepeatModeler Open-1.0*. Available online at: <http://www.repeatmasker.org>
- Smit, A. F. A., Hubley, R., and Green, P. (2013). *RepeatMasker Open-4.0*. Available online at: <http://www.repeatmasker.org>
- Spence, J., Vercher, Y., Gates, P., and Harris, N. (1996). 'Pod shatter' in *Arabidopsis thaliana*, *Brassica napus* and *B. juncea*. *J. Micros.* 181, 195–203. doi: 10.1046/j.1365-2818.1996.111391.x
- Stanke, M., Diekhans, M., Baetsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24, 637–664.
- 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.* 36:80.
- Takahashi, Y., Kongjaimun, A., Muto, C., Kobayashi, Y., Kumagai, M., Sakai, H., et al. (2019). Genetic factor for twisting legume pods identified by fine-mapping of shattering-related traits in azuki bean and yard-long bean. *bioRxiv [Preprint]* doi: 10.1101/774844
- Untergasser, A., Cutcutache, I., Koressaar, T., Faircloth, B. C., Remm, M., and Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Bioinformatics* 40:e115. doi: 10.1093/nar/gks596
- Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2017). BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi: 10.1093/molbev/msx319
- Yang, C., Xu, Z., Song, J., Conner, K., Vizcay Barrena, G., and Wilson, Z. A. (2007). Arabidopsis MYB26/MALE STERILE35 regulates secondary thickening in the endothecium and is essential for anther dehiscence. *Plant Cell* 19, 534–548. doi: 10.1105/tpc.106.046391

Conflict of Interest: TH is currently employed by the company Sentan Pharma Inc., his contribution to this study was made while he was working for his former affiliation, OIAS.

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

Copyright © 2020 Takahashi, Kongjaimun, Muto, Kobayashi, Kumagai, Sakai, Satou, Teruya, Shiroma, Shimoji, Hirano, Isemura, Saito, Baba-Kasai, Kaga, Somta, Tomooka and Naito. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Fine Mapping, Candidate Gene Identification and Co-segregating Marker Development for the Phytophthora Root Rot Resistance Gene *RpsYD25*

Chao Zhong^{1,2}, Suli Sun¹, Xuecui Zhang¹, Canxing Duan¹ and Zhendong Zhu^{1*}

¹ Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, ² College of Agronomy, Shenyang Agricultural University, Shenyang, China

OPEN ACCESS

Edited by:

Jiayin Pang,
The University of Western Australia,
Australia

Reviewed by:

Hon-Ming Lam,
The Chinese University of Hong Kong,
China
Yongle Li,
The University of Adelaide, Australia

*Correspondence:

Zhendong Zhu
zhuzhendong@caas.cn

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 09 February 2020

Accepted: 03 July 2020

Published: 28 July 2020

Citation:

Zhong C, Sun S, Zhang X, Duan C
and Zhu Z (2020) Fine Mapping,
Candidate Gene Identification
and Co-segregating Marker
Development for the Phytophthora
Root Rot Resistance Gene *RpsYD25*.
Front. Genet. 11:799.
doi: 10.3389/fgene.2020.00799

Phytophthora root rot (PRR) caused by *Phytophthora sojae* is a serious disease of soybean. The most effective disease-control strategy is to deploy resistant cultivars carrying *Rps* genes. Soybean cultivar Yudou25 can effectively resist pathotypes of *P. sojae* in China. Previous studies have mapped the *Rps* gene in Yudou25, *RpsYD25*, on chromosome 3. In this study, at first *RpsYD25* was located between SSR markers Satt1k3 (2.2 cM) and BARCSOYSSR_03_0253 (4.5 cM) by using an F_{2:3} population containing 165 families derived from Zaoshu18 and Yudou25. Then the recombination sites were identified in 1127 F_{3:4} families derived from Zaoshu18 and Yudou25 using the developed PCR-based SNP, InDel and SSR markers, and *RpsYD25* was finely mapped in the a 101.3 kb genomic region. In this region, a zinc ion binding and nucleic acid binding gene Glyma.03g034700 and two NBS-LRR genes Glyma.03g034800 and Glyma.03g034900 were predicted as candidate genes of *RpsYD25*, and five co-segregated SSR markers with *RpsYD25* were identified and validated to be diagnostic markers. Combined with the resistance reaction to multiple *P. sojae* isolates, seven of 178 soybean genotypes were detected to contain *RpsYD25* using the five co-segregated SSR markers. The soybean genotypes carrying *RpsYD25* and the developed co-segregated markers can be effectively applied in the breeding for *P. sojae* resistance in China.

Keywords: Phytophthora root rot, fine mapping, resistance gene, *RpsYD25*, soybean

INTRODUCTION

Phytophthora root rot (PRR), caused by the soil-borne oomycete pathogen *Phytophthora sojae*, is one of the most devastating diseases in soybean-growing regions worldwide, since it was firstly reported in Indiana, United States (Kaufmann and Gerdemann, 1958; Schmitthenner, 1985; Kamoun et al., 2015). According to the latest reports, from 2006 to 2014, approximately 340 million bushels of yield losses were attributed to PRR in major soybean production areas of United States, and PRR causes economic losses of \$1 – 2 billion annually worldwide (Tyler, 2007; Koenning and Wrather, 2010). The disease was first reported in Northeast China and has now spread to all major

soybean producing areas in China (Zhu et al., 2003; Tian et al., 2016; Wu et al., 2016). Under saturated soil conditions, *P. sojae* infect soybean plants throughout the growing season, resulting in pre- and post-emergence damping-off, root and stem rot, yellowing and wilting of leaves, and the death of soybean plants (Schmitthenner, 1999; Chen and Wang, 2017). At present, the most economical and eco-friendly method to control PRR is deployment of resistant soybean cultivars (Kamoun et al., 2015; Dorrance, 2018).

Two different types of PRR resistance have been identified in soybean, complete resistance and partial resistance (Sugimoto et al., 2012). Complete resistance is race-specific and monogenic, and is the result of a single dominant resistance *Rps* gene that confers immunity to *P. sojae*. In contrast, partial resistance appears as quantitative trait locus (QTL) loci controlled by multiple genes (Tooley and Grau, 1984; Dorrance et al., 2003b, 2004; Sugimoto et al., 2012). Since qualitative resistance is controlled by a single gene, it is easier to introduce into susceptible cultivars, and facilitates the selection of phenotypes during breeding, discovery of utilization of the *Rps* genes has always been the focus of research. Up to now, nearly 30 *Rps* genes have been identified and mapped to nine chromosomes, which are distributed on chromosomes 2, 3, 7, 10, 13, and 16–19 (Sugimoto et al., 2012; Lin et al., 2013; Zhang et al., 2013a,b; Sun et al., 2014; Li et al., 2016, 2017; Ping et al., 2016; Cheng et al., 2017; Niu et al., 2017; Sahoo et al., 2017; Zhong et al., 2017, 2018). Among the nine chromosomes containing *Rps* genes, chromosome 3 has much more *Rps* genes identified than other chromosomes. So far, 15 identified *Rps* genes are known to be distributed on the short arm of chromosome 3 in soybean, they were *Rps1* alleles, *Rps7*, *RpsYD25*, *RpsYD29*, *Rps9*, *RpsQ*, *RpsUN1*, the *Rps* gene in Waseshiroge, *RpsWY*, *RpsHN*, and *RpsHC18* (Demirbas et al., 2001; Weng et al., 2001; Gao et al., 2005; Fan et al., 2009; Sugimoto et al., 2011, 2012; Sun et al., 2011; Wu et al., 2011; Lin et al., 2013; Zhang et al., 2013b; Li et al., 2016; Cheng et al., 2017; Niu et al., 2017; Zhong et al., 2017). The *Rps1* locus includes five alleles *Rps1a*, *Rps1b*, *Rps1c*, *Rps1d*, and *Rps1k* (Sugimoto et al., 2012). However, since the mapping intervals of these *Rps* genes on the short arm of chromosome 3 are close to or overlap with each other, it is difficult to confirm whether these genes are alleles or adjacent to each other. The individual *Rps* gene is non-durable due to the virulence complexity and rapid changes of *P. sojae* population. The emergence of new *P. sojae* pathotypes can quickly overcome the *Rps* genes (Dorrance et al., 2003a; Grau et al., 2004). Most of identified *Rps* genes would be effective for only 8–15 years under the pressures of emerging *P. sojae* pathotypes (Dorrance et al., 2004; Sugimoto et al., 2012; Stewart et al., 2014, 2016; Xue et al., 2015). No *Rps* gene that confers resistance to all *P. sojae* races or pathotypes in China has been identified (Zhu et al., 2003; Cui et al., 2010; Zhang et al., 2010, 2014). Among the *Rps* genes identified and mapped in China, some of them can effectively resist Chinese *P. sojae* pathotypes, like *RpsYD29*, *RpsYD25*, and *RpsQ* (Zhang et al., 2014; Li et al., 2017). Compared with *Rps1k*, *RpsYD25*, which was identified in the soybean cultivar Yudou25 in Henan province of China, showed a more broad-spectral resistance *P. sojae* pathotypes in China (Chen et al., 2008; Zhang et al., 2014;

Li et al., 2017). Two independent studies have been carried out to map the *Rps* gene in Yudou25 (Fan et al., 2009; Sun et al., 2011). Although both studies have mapped *RpsYD25* on the short arm of chromosome 3, however, due to the different set of polymorphic molecular markers used in each mapping population and the relatively small size of mapping population, the linkage markers and their order in genetic maps in both studies are inconsistent. Thus, it is necessary for the larger-size mapping population and more polymorphic molecular markers to construct the high-resolution genetic and accurate physical maps for *RpsYD25*. This helps to confirm genomic intervals containing *RpsYD25*, identify candidate genes and further develop functional markers, which can be applied to marker-assisted breeding.

As the origin of soybean, there are abundant soybean germplasm resources, and a lot of germplasm resistant to *P. sojae* have been identified in recent decades in China (Zhu et al., 2003; Chen et al., 2008; Zhang et al., 2014). Although some studies have carried out the *Rps* gene postulation and prediction of Chinese soybean germplasm or cultivars based on resistance reaction types to different *P. sojae* isolates, due to the complex *P. sojae* pathotypes and genetic background of soybean germplasm, it is yet unclear about the distribution of identified *Rps* genes in the soybean cultivars in China (Chen and Wang, 2017). Fine mapping *Rps* genes and further development of co-segregated markers can detect *Rps* genes in soybean cultivars and germplasm to make more effective use of *P. sojae*-resistance sources. In our previous study, we used *P. sojae* isolate PsMC1 to conduct phenotypic screening on two mapping populations of 82 and 98 F_{2:3} families, respectively, and initially mapped *RpsYD25* on chromosome 3 (Fan et al., 2009). Hence, we finely mapped *RpsYD25* using a larger mapping population to determine the genomic region that contains *RpsYD25*. Then candidate genes were analyzed, and corresponding co-segregated markers were developed. In addition, the detection of *RpsYD25* in soybean cultivars and landraces from soybean producing regions in China was carried out using co-segregated markers.

MATERIALS AND METHODS

Plant Materials and Mapping Population Development

A set of cultivars/lines each containing one known *Rps* gene was used as differential hosts for PRR phenotyping (Table 1). Two soybean cultivars Williams and Yudou21 were used as susceptible control (Zhang et al., 2014). All of these cultivars/lines were obtained from Institute of Crop Sciences, Chinese Academy of Agriculture Sciences.

A population of 165 F_{2:3} families derived from Zaoshu18 × Yudou25 was used as initial mapping for *RpsYD25*. The F₁ hybrids were self-pollinated to produce F₂ individuals. Each F₂ individual was self-pollinated and the seeds were obtained as F_{2:3} Families. Fine mapping population consisting of 1127 families was obtained from segregating families based on phenotypic result of the F_{2:3} population crossed by Yudou25 and Zaoshu18 described by Fan et al. (2009) (Supplementary Figure S1). All the remaining seeds of segregating families

TABLE 1 | Phenotyping of 25 different soybean cultivars' resistance reactions to 14 *Phytophthora sojae* isolates.

Cultivar/line (<i>Rps</i> gene)	<i>Phytophthora sojae</i> isolate ^a														No. of resistance to isolates
	PsRace1	PsRace3	PsRace4	PsRace5	PsUSAR2	Ps41-1	PsAH4	PsMC1	PsNKI	PsFJ2	PsFJ3	PsJS2	Ps6497	Ps7063	
Harlon (<i>Rps1a</i>)	S	S	S	S	R	S	R	S	R	S	S	S	R	S	4
Harosoy13XX (<i>Rps1b</i>)	R	R	R	R	S	S	S	S	S	S	S	S	S	R	5
Williams79 (<i>Rps1c</i>)	R	R	R	R	R	R	S	R	R	R	R	S	R	R	12
PI103091 (<i>Rps1d</i>)	R	S	S	R	R	S	S	S	S	S	S	S	S	S	3
Williams 82 (<i>Rps1k</i>)	R	R	R	R	R	R	S	S	R	R	S	S	R	R	10
L76-988 (<i>Rps2</i>)	R	R	R	R	S	S	S	S	S	S	S	S	S	S	4
L83-570 (<i>Rps3a</i>)	R	R	R	R	R	S	S	S	S	S	S	S	R	S	6
PRX146-36 (<i>Rps3b</i>)	R	R	S	R	R	S	S	S	S	S	S	S	R	R	6
PRX145-48 (<i>Rps3c</i>)	R	R	R	R	S	S	S	S	S	S	S	S	S	R	5
L85-2352 (<i>Rps4</i>)	R	R	R	R	R	S	S	S	S	S	S	S	R	S	6
L85-3059 (<i>Rps5</i>)	R	R	R	R	S	S	S	S	S	S	S	S	R	S	5
Harosoy62XX (<i>Rps6</i>)	R	R	R	R	R	S	R	S	S	S	S	S	R	S	7
Harosoy (<i>Rps7</i>)	R	R	R	S	R	S	S	S	S	S	S	S	S	S	4
PI399073 (<i>Rps8</i>)	R	R	R	R	R	S	R	S	S	S	S	S	R	S	7
Ludou4 (<i>Rps9</i>)	R	R	R	R	R	R	R	R	R	R	S	R	R	R	13
Wandou15 (<i>Rps10</i>)	R	R	R	R	R	R	R	R	R	S	R	S	R	S	11
Youbian30 (<i>RpsYB30</i>)	R	R	S	R	R	R	S	S	R	R	S	S	S	S	7
Yudou29 (<i>RpsYD29</i>)	R	R	R	R	R	R	S	R	R	R	R	S	R	R	12
Qichadou1 (<i>RpsQ</i>)	R	R	R	R	R	R	R	R	R	R	S	R	R	R	13
Huachun18(<i>RpsHC18</i>)	R	R	R	R	R	R	R	R	R	R	R	R	R	S	13
Yudou25 (<i>RpsYD25</i>)	R	R	R	R	R	R	S	R	R	S	R	S	R	R	11
Zaoshu18 (<i>RpsZS18</i>)	R	R	R	R	R	R	S	S	R	R	R	S	R	S	10
Zheng92116	R	R	R	R	R	R	S	R	R	S	R	S	R	R	11
Meng8206 (<i>RpsHN</i>)	R	S	S	R	S	S	S	S	S	S	S	S	S	S	2
Yudou21	S	S	S	S	S	S	S	S	S	S	S	S	S	S	0
Williams (<i>rps</i>)	S	S	S	S	S	S	S	S	S	S	S	S	S	S	0

^aPsRace1, PsRace3, PsRace4, PsRace5, Ps41-1, and PsNKI were isolated from Heilongjiang Province, Northeast of China; PsAH4 and PsMC1 were isolated from Anhui province, Huang-Huai Region of China; PsFJ2 and PsFJ3 were isolated from Fujian province in China; PsJS2 was isolated from Jiangsu province of China; PsUSAR2 was the strain of race2 from the United States with pathotype changed (Zhang et al., 2014). Ps6497 and Ps7063 were the standard strains from United States (Dou et al., 2010; Song et al., 2013). The bolded terms are the cultivars containing *RpsYD25*.

identified for phenotype and genotype were planted in the field to produce the large F_{3:4} population. To determine positional relationship between *RpsYD25* and *Rps1*, allelic test was deployed to an F₂ population consisting of 402 individuals derived from a cross Williams 82 (containing *Rps1k*) × Yudou25.

PRR Phenotyping of the Soybean Cultivars and Mapping Populations

Fourteen *P. sojae* isolates were used for characterizing phenotype of each cultivars containing *Rps* genes. The set of *P. sojae* isolates was provided by Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Northeast Agricultural University, and Nanjing Agricultural University. All *P. sojae* isolates were isolated from infected soybean plants of different soybean production regions in China except for Ps7063. All *P. sojae* isolates were activated and transferred to the V₈ juice agar medium for use. 15–20 seeds of each cultivars were planted in 1000 mL paper cups. Each of the soybean cultivars sowed for 14 cups in order to inoculate 14 different isolates, respectively. Inoculum preparation and the applied hypocotyl-inoculation technique were operated using the protocol described by Zhang et al. (2014).

Two *P. sojae* isolates PsMC1 (virulence formula: *Rps1a*, *1b*, *1d*, *1k*, *2*, *3a*, *3b*, *3c*, *4*, *5*, *6*, *7*, *8*, *YB30*, and *RpsZS18*) and Ps7063 (virulence formula: *Rps1a*, *1d*, *2*, *3a*, *4*, *5*, *6*, *7*, *8*, *10*, *YB30*, and *ZS18*) were used to evaluate resistance reaction of parental cultivars and populations. 20–25 seeds of each family in the mapping populations were planted in paper cups. For the 165 F_{2:3} families crossed by Zaoshu18 and Yudou25, PsMC1 and Ps7063 were used to identify the phenotypes. PsMC1 were used to evaluate resistance response of the 1127 derived F_{3:4} families from heterozygous families crossed by Zaoshu18 and Yudou25. 402 F₂ individuals crossed by Williams 82 × Yudou25 were inoculated by the *P. sojae* isolate Ps7063 for the allelic test. The corresponding parents of all populations were also inoculated as control.

After 6 days of inoculation, the number of dead seedlings was recorded. Families with 0–20%, 80–100%, and 21–79% dead seedlings were considered homozygous resistant (R), susceptible (S), and segregating (Rs), respectively (Gordon et al., 2006; Lin et al., 2013; Zhang et al., 2013a,b; Ping et al., 2016).

Simple Sequence Repeat (SSR) Analyses and Genetic Mapping

Equivalent amounts of leaf tissues from every seedlings of each family were mixed and stored at -80°C . Genomics DNA was extracted using the Plant Genomic DNA Kit (Tiangen, Beijing, China). Based on the previous mapping results of Fan et al. (2009) and Sun et al. (2011), SSR on the short arm of soybean chromosome 3 were selected to screen for polymorphism between parental cultivars (Song et al., 2010; Sugimoto et al., 2012; Zhang et al., 2013b). Screening and identification of SSR markers was based on previous studies (Zhong et al., 2017). The screened polymorphic SSR markers were genotyped for corresponding mapping population.

Linkage analysis was conducted with MAPMAKER/EXP version 3.0 (Lincoln et al., 1993). Kosambi mapping function was employed to calculate the Genetic distances (Kosambi, 1943). A log-likelihood threshold of 3.0 was used to determine the linkage groups. The genetic linkage map of the molecular markers linked to *RpsYD25* was prepared using MapDraw version 2.1 (Liu and Meng, 2003).

Identifying and Genotyping Recombination Events and Candidate Gene Discovery

Whole genome re-sequencing was applied to the two parental cultivars Zaoshu18 and Yudou25. Genomic DNA of the two cultivars was isolated, respectively, to construct Illumina libraries and sequenced on an Illumina HiSeqTM 4000 by Biomarker Technologies (Beijing, China) (Zhong et al., 2017). SNPs and InDels between the two parental cultivars were detected using the SNP analyses software GATK (McKenna et al., 2010). The main detection process: (1) for the results obtained by BWA (Li and Durbin, 2009), Picard's mark duplicate tool was used to remove duplicates. (2) InDel realignment was deployed to correct the error of the alignment result caused by the insertion or deletion. (3) Base recalibration was conducted using GATK to correct the base quality value. (4) GATK was used for variant calling. (5) Strict filtering was performed on SNP and InDel: 2 SNP within 5 bp were filter out; SNPs located 5 bp upstream or downstream from InDel were filtered out; Two InDels with a distance of less than 10 bp were filtered out. Of all the obtained SNP and InDel, only homozygous sites were selected for further development of molecular marker.

According to the results of SSR genetic mapping, InDels in the *RpsYD25* genomic interval were identified for development of the PCR-based makers (Li et al., 2017), and SNPs in the mapping interval were obtained and used to develop PCR-based makers for the Tetra-Primer ARMS-PCR assay (Ye et al., 2001; Collins and Ke, 2012). Genomic sequence of Williams 82 corresponding to the mapping interval of *RpsYD25* was downloaded from SoyBase¹, and then new simple sequence repeat (SSR) motifs were searched and developed to SSR makers using the website primer design tool BatchPrimer3² (You et al., 2008). Recombinant break

points in the derived 1127 $F_{3:4}$ families were identified using the developed makers in the genomic region containing *RpsYD25*. Gene models in the final genomic interval that no recombinant events occurred among the 1127 $F_{2:3}$ families were preferred as candidate genes for *RpsYD25*.

Co-segregating Marker Genotyping Among Soybean Cultivars

Soybean cultivars containing single *Rps* genes were genotyped by the identified markers co-segregated with *RpsYD25* to detect whether they can effectively distinguish other published *Rps* genes. Cluster analysis was deployed according to the polymorphic co-segregated markers in different soybean genotypes. Each marker polymorphic information content (PIC) was calculated among these genotypes using the online tool PICcalc (Nagy et al., 2012). A Neighbor-joining tree was constructed using the PowerMarker V3.0 and MEGA 6.0 software among the soybean genotypes, and further distinguish whether these markers can effectively identify the haplotype of *RpsYD25* (Liu and Muse, 2004; Tamura et al., 2013). To further detect *RpsYD25* in Chinese soybeans, 178 cultivars and landraces were selected to identify the reaction type to eight *P. sojae* isolates (PsRace1, PsRace3, PsRace4, PsRace5, PsUSAR2, Ps41-1, PsMC1, and PsJS2). The combination of co-segregated markers was also used to detect *RpsYD25* haplotype among these soybean genotypes.

RESULTS

Phenotyping for Phytophthora Resistance

To identify the resistance reaction of *RpsYD25* and other known *Rps* genes, 25 soybean cultivars containing single known *Rps* genes along with Yudou25 were selected and inoculated with 14 *P. sojae* isolates to identify reaction types. The results revealed that the 25 soybean cultivars produced a total of 18 reaction types when inoculated the 14 isolates (Table 1). Among them, Williams and Yudou21 showed susceptibility to 14 isolates, confirming that the two cultivars do not contain *Rps* genes (Zhang et al., 2014). Yudou25 was resistant to 11 of 14 isolates, and the reaction type was different from all the other soybean genotypes containing known *Rps* genes except Zheng92116, suggesting that *RpsYD25* is a distinct *Rps* gene (Fan et al., 2009). Zheng92116 is a soybean cultivar derived from Yudou25, and Sun et al. (2011) showed that Zheng92116 contained the *RpsYD25* (*RpsYu25*).

Further analysis the reaction types between resistant cultivars Zaoshu18 and Yudou25, there is a difference in the response to three *P. sojae* isolates, namely PsMC1, PsFJ2 and Ps7063. Yudou25 showed resistance to PsMC1 and Ps7063, while Zaoshu18 were susceptible to both of them; Zaoshu18 was resistant to PsFJ2, while Yudou25 showed susceptibility to PsFJ2. Therefore, in order to identify the resistance gene *RpsYD25* in Yudou25, isolates PsMC1 and Ps7063 were selected for phenotypic evaluation of 165 $F_{2:3}$ families derived from Yudou25 and Zaoshu18 (Table 1). The phenotyping results of the 165

¹<https://www.soybase.org/>

²<https://wheat.pw.usda.gov/demos/BatchPrimer3/>

TABLE 2 | Reactions of F_2 individuals and $F_{2:3}$ families derived from crosses between parental soybean cultivars to *Phytophthora sojae* isolates.

<i>P. sojae</i> isolates	Parent and the cross	Generation	Amount	Observed number ^a			Chi squared tests		
				R ^a	Rs	S	Expected ratio	χ^2	P
PsMC1, Ps7063	Zaoshu18 × Yudou25	$F_{2:3}$	165	37	82	46	1:2:1	0.99	0.61
PsMC1	Zaoshu18 × Yudou25	$F_{3:4}$	1127	281	544	302	1:2:1	2.13	0.34
Ps7063	Williams 82 × Yudou25	F_2	402	402		0			

^aR, homozygous resistant; Rs, segregating; S, susceptible.

$F_{2:3}$ families inoculated with isolates PsMC1 and Ps7063 were consistent. Among $F_{2:3}$ families, 37 were homozygous resistant families, 82 were segregating families, and 46 were susceptible families. The χ^2 test revealed that phenotyping results fit well with 1: 2: 1 ratio ($\chi^2 = 0.99 < \chi^2_{0.05} = 5.99$) (Table 2 and Supplementary Figures S2A,B). Among the 1127 families of the $F_{3:4}$ sub-population derived from the cross between Zaoshu18 and Yudou25, there were 281 homozygous resistant, 544 segregated, and 302 susceptible. The χ^2 test met the 1:2:1 ratio ($\chi^2 = 2.13 < \chi^2_{0.05} = 5.99$) (Table 2 and Supplementary Figure S2C). Therefore, it indicates that *Phytophthora* resistance in Yudou25 is controlled by a single dominant gene, which is consistent with previous studies (Fan et al., 2009; Sun et al., 2011).

402 F_2 plants from the cross between Williams 82 and Yudou25 were identified for phenotypes using isolate Ps7063. All 402 individuals showed resistance and no plants died. Allele test indicated that *RpsYD25* on chromosome 3 may be an allele of or closely linked with *Rps1k* (Table 2).

Genetic Map Construction for *RpsYD25*

Based on the mapping results of *RpsYD25* by Fan et al. (2009) and Sun et al. (2011), 120 SSR markers which had been published on the short arm of chromosome 3 were selected to screen polymorphism between Zaoshu18 and Yudou25. 15 SSR markers showed polymorphism between the two parental cultivars. The SSR markers were further used to identify the genotypes of 165 $F_{2:3}$ families, and the linkage map of *RpsYD25* was constructed based on the phenotype and genotype results. *RpsYD25* was mapped between SSR markers Satt1k3 and BARCSoySSR_03_0253 with genetic distances of 2.2 cM (LOD = 71.68, 86.5% variation explained) and 4.5 cM (LOD = 60.37, 81.5% variation explained), respectively, while BARCSoySSR_03_0244 and BARCSoySSR_03_0247 were co-segregated with *RpsYD25* (Figure 1).

Fine Mapping of *RpsYD25*

Based on the preliminary mapping results of *RpsYD25*, *RpsYD25* was located between SSR marker Satt1k3 and BARCSoySSR_03_0253. The primer sequences of Satt1k3 were searched on Williams 82 (*Glycine max* version 2.0) reference genome, and Satt1k3 was located at 4022513 bp ~ 4022856 bp on chromosome 3, while BARCSoySSR_03_0253 is located at 4132642 bp ~ 4132683 bp on chromosome 3. The physical distance between the two SSR markers is approximately 290 kb. Identification of recombinant families was carried out in 1127 $F_{3:4}$ families derived from Zaoshu18 and Yudou25 using SSR

markers Satt1k3 and BARCSoySSR_03_0253. On the Satt1k3 side, a total of 6 recombination events were identified; On the BARCSoySSR_03_0253 side, 12 recombination events were occurred in 12 different families (Figure 2A).

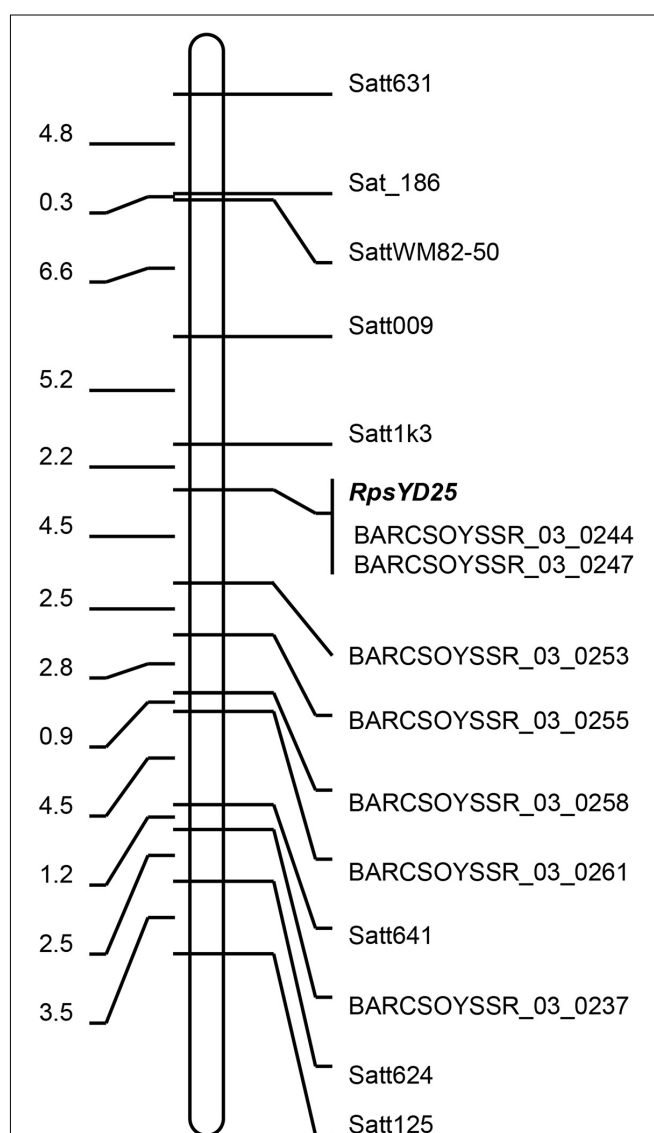


FIGURE 1 | Genetic linkage map of *RpsYD25* on chromosome 3. The genetic distances (cM) are on the left and the molecular markers are on the right.

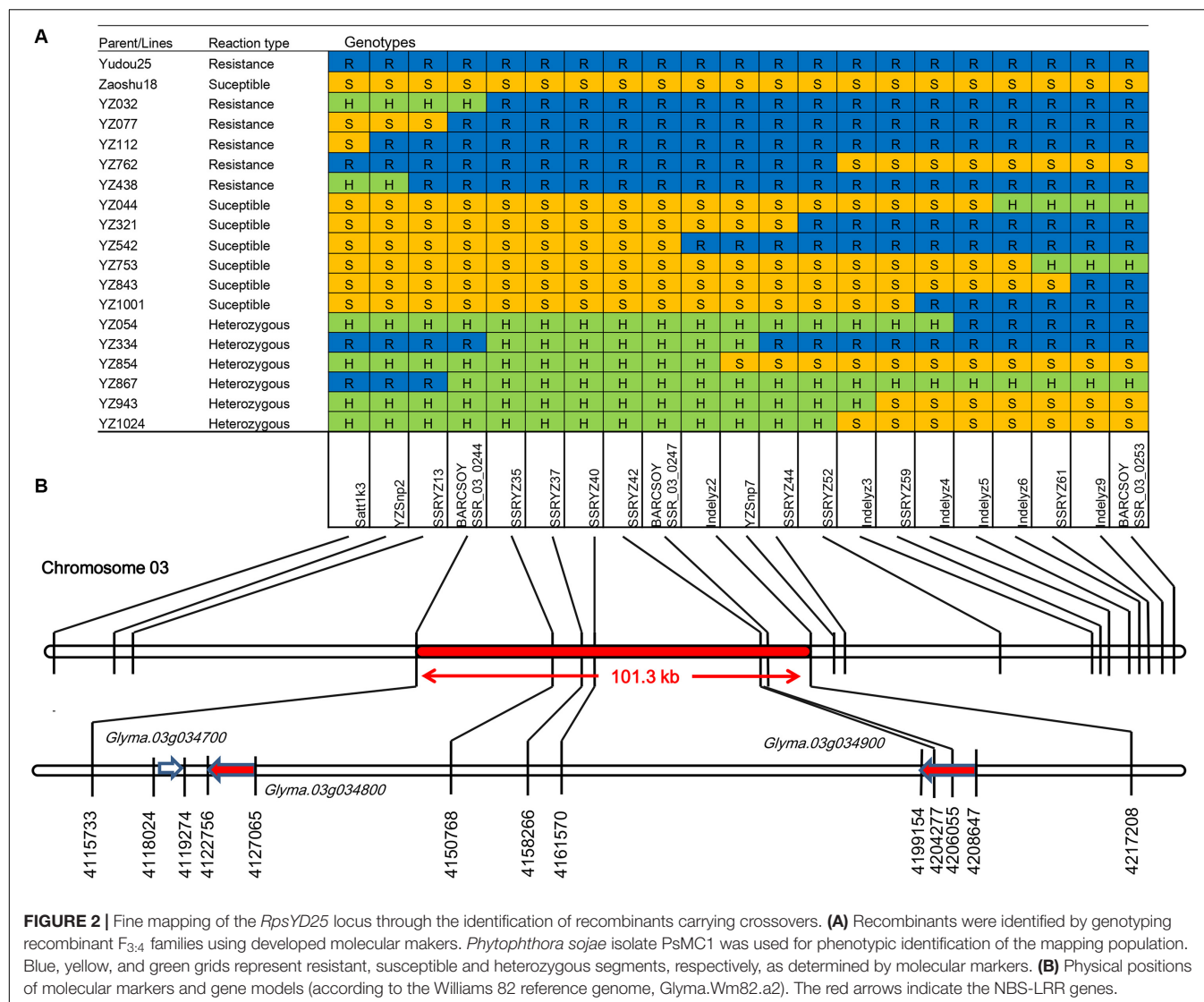


TABLE 3 | Co-segregating SSR markers in the *RpsYD25* mapping interval and the polymorphic information contents (PICs) among 25 different soybean cultivars.

Marker name		Reverse primer	SSR motif	Start site	Stop site	Product (bp)	Polymorphic information content (PIC)
SSRYZ35	Forward primer	ACGGTCATCTGATTATAAATTG	(AT)24	4150768	4150894	127	0.80
	Reverse primer	AGTGTGAAATAGTGTGCGTGT					
SSRYZ37	Forward primer	CATTATTTTGTCCGCTATAA	(AG)11	4158266	4158412	147	0.62
	Reverse primer	TATATCAAGGTTTGACGTGT					
SSRYZ40	Forward primer	ATTCCGTCACATAAAGTGCATA	(TA)9	4161570	4161692	123	0.84
	Reverse primer	TAAACATAAAGCGTGACAACA					
SSRYZ42	Forward primer	ATGACACATGCTAATTGATCC	(TA)6	4204277	4204469	192	0.78
	Reverse primer	CGCCATTTCAAAAGAATTAC					
BARCSOY	Forward primer	ATTATTATGGTGGGGCGTGA	(TAA)17	4205985	4206148	163	0.83
	Reverse primer	TGACCACCATTTCAAAGGAA					

In order to further narrow the mapping interval of *RpsYD25*, we carried out the detection of further recombination sites in 17 recombinant families by developing different types of molecular markers in the *RpsYD25*-mapped region (**Figure 2A**).

PCR markers were developed based on SNP and InDel sites identified within the *RpsYD25* genomic interval identified by whole genome re-sequencing. After detecting amplification efficiency and polymorphism between Zaoshu18 and Yudou25,

two Tetra-AMS PCR-based polymorphic SNP markers YZSn2 and YZSn7 (**Supplementary Table S1**) and six polymorphic InDel markers Indelyz2, Indelyz3, Indelyz4, Indelyz5, Indelyz6, and Indelyz9 were identified (**Supplementary Table S2**). Based on the physical interval of the Willimas82 genome sequence (*Glycine max* V2.0), new SSR loci were searched and developed as SSR markers. Nine new polymorphic SSR markers showed polymorphism between parental cultivars (**Supplementary Table S3**).

The newly developed two SNP markers, six InDel markers, and nine SSR markers and two co-segregated SSR markers from the genetic mapping were used to further identify the recombination sites among 17 recombinant families. The results show that BARCSOYSSR_03_0244 has the closest recombination event from *RpsYD25* in families YZ334 and YZ032, and on the other side, the closest recombination event happened at Indelyz2 in family YZ542 (**Figure 2A**). The distance between the BARCSOYSSR_03_0244 and Indelyz2 is 101.3 kb, and between them there are five co-segregated SSR markers, namely SSRYZ35, SSRYZ37, SSRYZ40, SSRYZ42, and BARCSOYSSR_03_0247. There are three gene models in physical region of *RpsYD25*, Glyma.03g034700 is a gene model with zinc ion binding and nucleic acid binding protein structure (Zinc ion binding; nucleic; acid binding), Glyma.03g034800 and Glyma.03g034900 are typical disease-resistant structure of the NBS-LRR gene. These three genes may be candidate gene models of *RpYD25* (**Figure 2B**).

Validation of Co-segregated Markers for Specific Detection of *RpsYD25*

To detect whether the co-segregated markers can effectively distinguish *RpsYD25* haplotype from other identified *Rps* genes, these markers were used to detect soybean genotypes containing known *Rps* genes and susceptible control genotypes. Among the 25 soybean genotypes, the PIC values of the five co-segregated markers ranged from 0.62 to 0.84, which means that these markers are relatively rich in polymorphism among different genotypes (**Table 3**). Cluster analysis showed that the five markers formed 21 haplotypes in 25 soybean genotypes (**Figure 3**). Yudou25 and Zheng92116 belonged to the same haplotype, indicating that both Zheng92116 and Yudou25 contained *RpsYD25*.

Molecular marker detection using the five co-segregated markers was performed on 178 soybean genotypes. Six genotypes with the same haplotype as Yudou25 and Zheng92116 were identified, namely Yudou23, Wansu2156, Yudou13, Yudou15, Zhoudou11, and Zhoudou12 (**Figure 4A**). The resistance reaction of the seven soybean cultivars to eight *P. sojae* isolates was consistent with that of Yudou25 except Yudou23, which showed resistant all the eight *P. sojae* isolates (**Figure 4B** and **Supplementary Table S4**), this may be due to an allelic mutation or other resistance genes existed in Yudou23. Pedigree analysis showed that these cultivars are genetically related to Yudou25 (**Figure 4C**). The seven cultivars all have the same ancestral cultivar Zheng77249, but Zheng77249 is different in

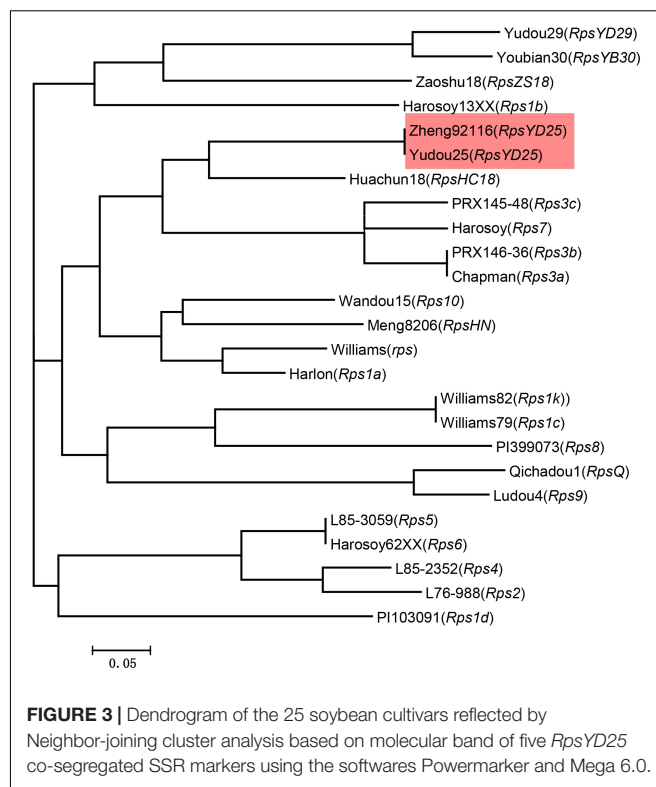


FIGURE 3 | Dendrogram of the 25 soybean cultivars reflected by Neighbor-joining cluster analysis based on molecular band of five *RpsYD25* co-segregated SSR markers using the softwares Powermarker and Mega 6.0.

molecular band and resistance reaction type from the derivatives. For phenotypic identification, Zheng77249 has a phenotypic difference with Yudou25 on three isolates (PsRace5, Ps41-1 and PsMC1), and for molecular identification, Zheng77249 is the same as Yudou25 only at the SSR marker YZSSR42 locus (**Figures 4A,B** and **Supplementary Table S4**), suggesting that Zheng77249 does not contain *RpsYD25* haplotype. These results show that the combination of the five co-segregated markers can effectively identify soybean genotypes containing *RpsYD25* haplotype.

DISCUSSION

RpsYD25 can effectively resist the *P. sojae* pathotypes in China. Zhang et al. (2014) analyzed the resistance spectrum of soybean cultivars in Henan Province, and found that Yudou25 were resistant to 21 of the 26 *P. sojae* isolates; Li et al. (2017) inoculated Yudou25 with 36 different *P. sojae* isolates in different areas of China, and found that Yudou25 showed resistance to 33 of 36 isolates. In this study, Yudou25 was inoculated with 14 different virulence isolates of *P. sojae*, and it showed resistant to 11 of them. In the comparison of the resistance reactions of 25 soybean genotypes to 14 *P. sojae* isolates, the reaction type of Yudou25 was only the same as that of Zheng92116, and it is different from other soybean genotypes containing *Rps* genes, which indicates that Zheng92116 may contain *RpsYD25*. Zheng92116 is a soybean cultivar from the hybrid of Yudou25 as the male parent and Zheng506 as the female parent. Sun et al. (2011) found that the *Rps* gene in Zheng92116 is likely to be the

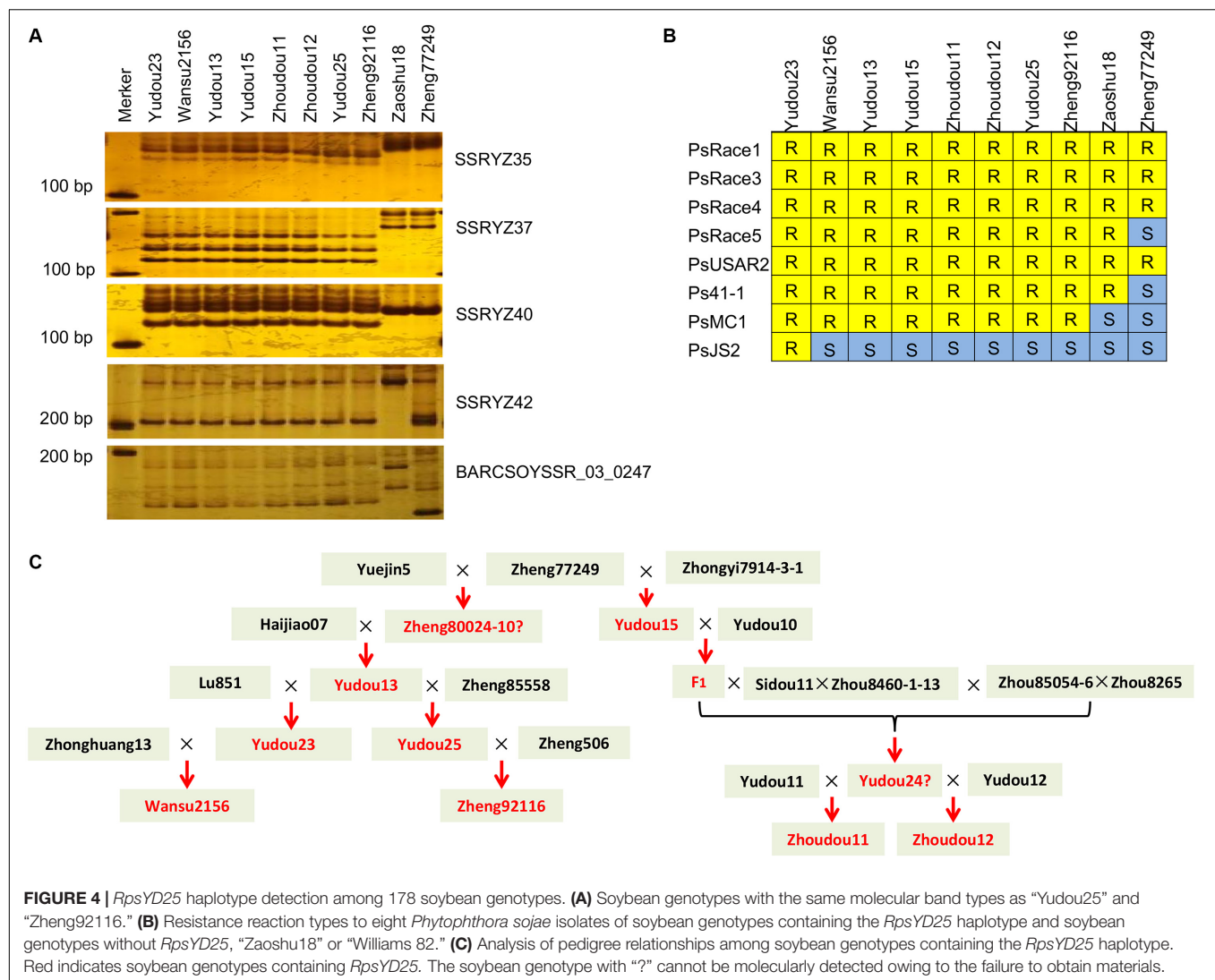


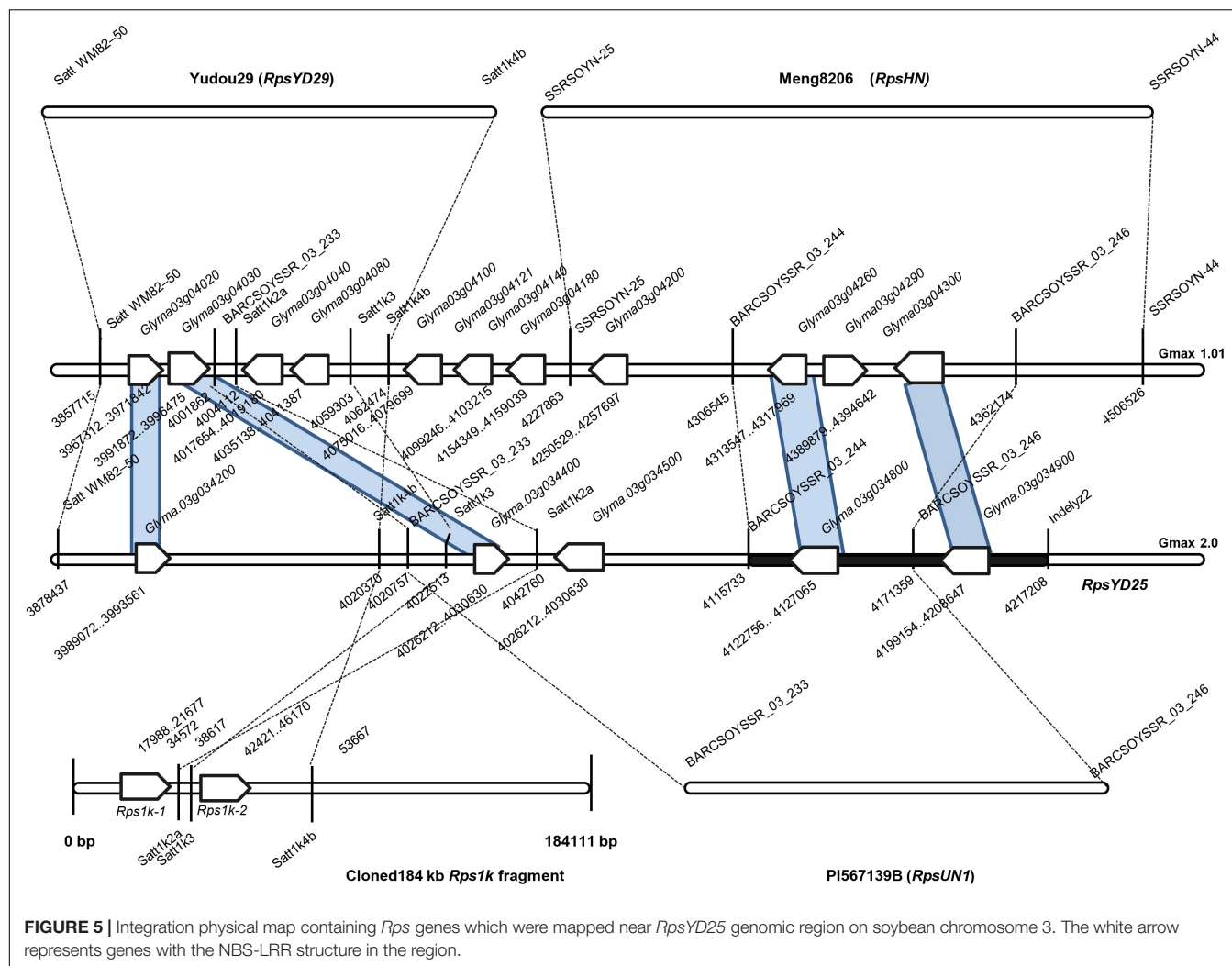
FIGURE 4 | *RpsYD25* haplotype detection among 178 soybean genotypes. **(A)** Soybean genotypes with the same molecular band types as “Yudou25” and “Zheng92116.” **(B)** Resistance reaction types to eight *Phytophthora sojae* isolates of soybean genotypes containing the *RpsYD25* haplotype and soybean genotypes without *RpsYD25*, “Zaoshu18” or “Williams 82.” **(C)** Analysis of pedigree relationships among soybean genotypes containing the *RpsYD25* haplotype. Red indicates soybean genotypes containing *RpsYD25*. The soybean genotype with “?” cannot be molecularly detected owing to the failure to obtain materials.

same as Yudou25 through genetic mapping and pedigree analysis, and this result was also verified by *RpsYD25* co-segregated SSR markers detection in this study.

At present, there are 15 *Rps* genes which are located on the short arm of chromosome 3, however, many *Rps* genes cannot be determined the positional relationship between each other, because the mapping intervals were usually too large or the order of markers cannot correspond to the reference genome (Demirbas et al., 2001; Weng et al., 2001; Gao et al., 2005; Fan et al., 2009; Schmutz et al., 2010; Sugimoto et al., 2011, 2012; Sun et al., 2011; Wu et al., 2011; Lin et al., 2013; Zhang et al., 2013b; Li et al., 2016; Cheng et al., 2017; Niu et al., 2017; Zhong et al., 2017). Previous studies have mapped *RpsYD25* in a large region of chromosome 3 (Fan et al., 2009; Sun et al., 2011). Therefore, in this study, we first reconstructed the genetic linkage map of *RpsYD25* with the published SSR markers on chromosome 3, confirming the previous mapping results, and narrowed the mapping interval to a region of 290 kb. To further finely map *RpsYD25*, we reconstructed a sub-population containing 1127

$F_{3:4}$ families using the resistant segregated families selected from the $F_{2:3}$ populations derived by the cross of Zaoshu18 and Yudou25 constructed by Fan et al. (2009). The nearly flanking markers of *RpsYD25* were used to identify the recombinant families and recombination sites in the population, and the mapping interval of *RpsYD25* was further reduced to 101.3 kb.

In the mapping interval of *RpsYD25*, there are two resistance gene models with NBS-LRR structure, Glyma.03g034800 and Glyma.03g034900, and a gene Glyma.03g034700 with zinc ion and nucleic acid binding domain structure. The mapping interval of *RpsYD25* is partially overlapped with the *Rps* genes *RpsUN1* and *RpsHN* (Figure 5). The candidate gene Glyma.03g034800 is also predicted to be a candidate for *RpsUN1* (Li et al., 2016), while Glyma.03g034800 and Glyma.03g034900 are located within the *RpsHN* mapping region and are predicted to be candidate genes for *RpsHN* (Niu et al., 2017). In this study, we performed a comparative analysis of the resistance patterns of soybean cultivar Meng8206 containing *RpsHN* and Yudou25, and found that Meng8206 was only resistant to 2 isolates PsRace1 and



PsRace5, while Yudou25 is resistant to 11 isolates. Niu et al. (2017) used 8 *P. sojae* isolates to identify the resistance spectrum of Meng8206, and the results showed that Meng8206 was only resistant to one isolate. Because *RpsYD25* and *RpsHN* showed significant difference in the resistance types to *P. sojae*, these two genes are unlikely to be the same *Rps* gene, presumably alleles or tightly linked genes. Haplotype analysis also indicates that the two cultivars contain different *Rps* genes (Figure 3). *RpsUN1* was identified in US soybean germplasm PI 567139B. PI 567139B is from Indonesia. Due to lack of this resource, we were unable to perform resistance reaction type analysis and molecular marker identification in this study. The relationship between *RpsYD25* and *RpsUN1* is currently uncertain. Although *Rps1k* has been cloned, no identical sequence to the cloned *Rps1k* gene sequence was found in the Williams 82 genome sequence, making it difficult to determine the exact locus of *Rps1k* in the reference genome. In this study, we performed an allelic test on the F₂ population derived from the hybridization of Williams 82 and Yudou25. It was found that *RpsYD25* and *Rps1k* did not separate in 402 individuals, which proved that the two genes are alleles

or tightly linked. The SSR marker Satt1k3, which is closest to the genetic distance of *RpsYD25*, is from the *Rps1k* clone sequence (Zhang et al., 2013b), so it is speculated that *RpsYD25* is closely linked to *Rps1k* (Figure 5).

Although zinc ion binding gene was not used as an *Rps* candidate gene in previous studies (Li et al., 2016; Niu et al., 2017), barley powdery mildew resistance gene *Rar1* containing zinc ion binding domain was able to increase the accumulation of H₂O₂ through signal transduction, triggered by race-specific resistance (R) genes (Shirasu et al., 1999). Therefore, we speculate that the gene *Glyma.03g034700* may be triggered by the race-specific NBS-LRR genes *Glyma.03g034800* and *Glyma.03g034900* in soybean to resist *P. sojae*. *RpsYD25* is mapped in a genomic region rich in NBS-LRR gene clusters on chromosome 3. Among the identified resistance genes, *RpsYD29*, *RpsUN1*, and *RpsHN* were mapped in the region containing the NBS-LRR gene cluster, and all located in the 200 kb region upstream and downstream of *RpsYD25* (Figure 5; Zhang et al., 2013b; Li et al., 2016; Niu et al., 2017). *Rps1k* has been cloned to prove that consist of two adjacent NBS-LRR genes. This study also confirmed that *Rps1k* should be

located near the *RpsYD25* region by allelic test and molecular mapping analysis (Gao et al., 2005; Gao and Bhattacharyya, 2008). The NBS-LRR genes are typical resistance genes expressing conserved domains of the nucleotide-binding site and leucine-rich repeats. Among the resistance genes in plants that have been cloned and characterized, 61% of them belong to NBS-LRR gene (Shao et al., 2016; Kourelis and van der Hoorn, 2018; Xing et al., 2018). The accumulation of mutations in the NBS-LRR gene can increase the complexity of the NBS-LRR gene clusters and can generate new resistance types by recombining or rearranging with each other, such as gene duplication, unequal exchange, chromosomal abnormal recombination, gene conversion (Michelmore and Meyers, 1998; Meyers et al., 2003; Nagy and Bennetzen, 2008; Marone et al., 2013; Lestari et al., 2017; Lv et al., 2017). The resistance mechanism in the NBS-LRR gene cluster region is complicated. In the cloning and functional study of the wheat powdery mildew resistance gene *Pm60*, it was found that the NBS-LRR powdery mildew resistance gene can interact with its adjacent NBS-LRR structural gene to jointly provide resistance against powdery mildew (Zou et al., 2018). The diversity of *Rps* genes on soybean chromosome 3 may be caused by the translocation, recombination or rearrangement of functional NBS-LRR genes in the NBS-LRR gene cluster to lead the different resistance types to *P. sojae* (Kang et al., 2012; Lestari et al., 2017). It requires further experimental support to explore whether there is a large sequence change in the mapping interval of *RpsYD25* compared with the reference genome Williams82 such as the presence of other genes.

Due to the large number of high similarity NBS-LRR sequences around *RpsYD25* region, it is difficult to effectively detect specific mutation sites in candidate genes based on whole genome re-sequencing, and to develop specific primers for amplifying target candidate genes. However, in the mapping interval of *RpsYD25*, five SSR markers with high PIC were found to effectively distinguish the haplotype of *RpsYD25* from other known *Rps* genes. Therefore, these SSR markers can be used to jointly detect *RpsYD25* haplotype in different soybean genotypes. The *RpsYD25* haplotype was detected in 8 related soybean genotypes, which all have broad-spectrum resistance to *P. sojae* in China (Zhang et al., 2014). However, the resistance types of Yudou23 was not the same as that of other genotypes in this study, and 7 other genotypes were consistent to the resistance reaction to 8 *P. sojae* isolates, but they differ in their reactions to individual *P. sojae* isolates in the study of Zhang et al. (2014), which may be due to virulence changes of the *P. sojae* isolates or other alleles existed in *RpsYD25* locus. The *RpsYD25* absence in the shared ancestor genotypes Zheng77249 may be due to the unequal recombination in the *RpsYD25* locus in the process of deriving other genotypes, resulting in stronger resistance to *P. sojae*, which may also explain why the Henan province in China is not the area where PRR severely occurs, but there is evidence for abundance resources containing resistance diversity to *P. sojae* (Zhang et al., 2014). Hence, Zheng77249 in this study may not be the original material used to produce Yudou25 and Yudou15, and have a novel *RpsYD25* allele generated after multiple generations.

In the present study, the *Phytophthora* resistance gene *RpsYD25* was finely mapped, and three candidate genes for *RpsYD25* were identified in the mapping interval. *RpsYD25* was found to be located in a genomic region rich in NBS-LRR genes. The co-segregated markers with *RpsYD25* was developed and identified, and can be further converted into high-throughput markers such as SNP or KASP to improve the efficiency of marker-assisted selection to detect the *RpsYD25* haplotype.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI: [SAMN15376405] [<https://www.ncbi.nlm.nih.gov/biosample/15376405>], [SAMN15376406] [<https://www.ncbi.nlm.nih.gov/biosample/15376406>], [SAMN15376407] [<https://www.ncbi.nlm.nih.gov/biosample/15376407>], and [SAMN15376408] [<https://www.ncbi.nlm.nih.gov/biosample/15376408>].

AUTHOR CONTRIBUTIONS

ZZ conceived and designed the experiments and revised the manuscript. CZ, SS, XZ, and CD performed the experiments. CZ analyzed the data and wrote the manuscript. All authors read and approved the manuscript.

FUNDING

The work was supported by the National Natural Science Foundation of China (31771822), the Program of Protection of Crop Germplasm Resources (2018NWB036-12) from the Ministry of Agriculture of the People's Republic of China, the Scientific Innovation Program of Chinese Academy of Agricultural Sciences, and the Special Fund for Agroscientific Research in the Public Interest (201303018).

ACKNOWLEDGMENTS

We thank Prof. Lijuan Qiu of the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences for supplying the soybean cultivars tested and Prof. Yuanchao Wang of Nanjing Agricultural University and Prof. Shuzhen Zhang of Northeast Agricultural University for providing the *Phytophthora sojae* isolates used in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00799/full#supplementary-material>

REFERENCES

- Chen, X., and Wang, Y. (2017). "Phytophthora sojae," in *Biological invasions and its management in China*, eds F. Wan, M. Jiang, and A. Zhan, (Singapore: Springer), 199–223.
- Chen, X., Zhu, Z., Wang, X., Xiao, Y., and Wu, X. (2008). Postulation of Phytophthora resistance genes in soybean cultivars or lines. *Sci. Agric. Sin.* 41, 1227–1234.
- Cheng, Y., Ma, Q., Ren, H., Xia, Q., Song, E., Tan, Z., et al. (2017). Fine mapping of a Phytophthora-resistance gene *RpsWY* in soybean (*Glycine max* L.) by high-throughput genome-wide sequencing. *Theor. Appl. Genet.* 130, 1041–1051. doi: 10.1007/s00122-017-2869-5
- Collins, A., and Ke, X. (2012). Primer1: primer design web service for tetra-primer ARMS-PCR. *Open Bioinform. J.* 6, 55–58. doi: 10.2174/1875036201206010055
- Cui, L., Yin, W., Tang, Q., Dong, S., Zheng, X., Zhang, Z., et al. (2010). Distribution, pathotypes, and metalaxyl sensitivity of *Phytophthora sojae* from Heilongjiang and Fujian provinces in China. *Plant Dis.* 94, 881–884. doi: 10.1094/PDIS-94-7-0881
- Demirbas, A., Rector, B. G., Lohnes, D. G., Fioritto, R. J., Graef, G. L., Cregand, P. B., et al. (2001). Simple sequence repeat markers linked to the soybean genes for Phytophthora resistance. *Crop Sci.* 41, 1220–1227. doi: 10.2135/cropsci2001.4141220x
- Dorrance, A. E. (2018). Management of *Phytophthora sojae* of soybean: a review and future perspectives. *Can. J. Plant Pathol.* 40, 210–219. doi: 10.1080/07060661.2018.1445127
- Dorrance, A. E., Jia, H., and Abney, T. S. (2004). Evaluation of soybean differentials for their interaction with *Phytophthora sojae*. *Plant Health Prog.* 5:9. doi: 10.1094/PHP-2004-0309-01-RS
- Dorrance, A. E., McClure, S. A., and DeSilva, A. (2003a). Pathogenic diversity of *Phytophthora sojae* in Ohio soybean fields. *Plant Dis.* 87, 139–146. doi: 10.1094/PDIS.2003.87.2.139
- Dorrance, A. E., McClure, S. A., and St. Martin, S. K. (2003b). Effect of partial resistance on *Phytophthora* stem rot incidence and yield of soybean in Ohio. *Plant Dis.* 87, 308–312. doi: 10.1094/PDIS.2003.87.3.308
- Dou, D., Kale, S. D., Liu, T., Tang, Q., Wang, X., Arredondo, F. D., et al. (2010). Different domains of *Phytophthora sojae* effector Avr4/6 are recognized by soybean resistance genes Rps4 and Rps6. *Mol. Plant Microbe* 23, 425–435. doi: 10.1094/MPMI-23-4-0425
- Fan, A., Wang, X., Fang, X., Wu, X., and Zhu, Z. (2009). Molecular identification of *Phytophthora* resistance gene in soybean cultivar Yudou 25. *Acta Agron. Sin.* 35, 1844–1850. doi: 10.3724/sp.j.1006.2009.01844
- Gao, H., and Bhattacharyya, M. K. (2008). The soybean-*Phytophthora* resistance locus *Rps1-k* encompasses coiled coil-nucleotide binding leucine rich repeat-like genes and repetitive sequences. *BMC Plant Biol.* 8:29. doi: 10.1186/1471-2229-8-29
- Gao, H., Narayanan, N. N., Ellison, L., and Bhattacharyya, M. K. (2005). Two classes of highly similar coiled coil-nucleotide binding-leucine rich repeat genes isolated from the Rps1-k locus encode *Phytophthora* resistance in soybean. *Mol. Plant Microbe* 18, 1035–1045. doi: 10.1094/MPMI-18-1035
- Gordon, S. G., St. Martin, S. K., and Dorrance, A. E. (2006). *Rps8* maps to a resistance gene rich region on soybean molecular linkage group F. *Crop Sci.* 46, 168–173. doi: 10.2135/cropsci2004.04-0024
- Grau, C. R., Dorrance, A. E., Bond, J., and Russin, J. S. (2004). "Fungal diseases," in *Soybeans: Improvement, Production, and Uses*, (Soybeans Improve), ed. J.R. Wilcox, (Madison, WI: American Society of Agronomy), 679–763.
- Kamoun, S., Furzer, O., Jones, J. D., Judelson, H. S., Ali, G. S., Dalio, R. J., et al. (2015). The Top 10 oomycete pathogens in molecular plant pathology. *Mol. Plant Pathol.* 16, 413–434. doi: 10.1111/mpp.12190
- Kang, Y. J., Kim, K. H., Shim, S., et al. (2012). Genome-wide mapping of *NBS-LRR* genes and their association with disease resistance in soybean. *BMC Plant Biol.* 12:139. doi: 10.1186/1471-2229-12-139
- Kaufmann, M., and Gerdemann, J. (1958). Root and stem rot of soybean caused by *Phytophthora sojae* n. sp. *Phytopathology* 48, 201–208.
- Koenning, S. R., and Wrather, J. A. (2010). Suppression of soybean yield potential in the continental United States by plant diseases from 2006 to 2009. *Plant Health Prog.* 11:5. doi: 10.1094/PHP-2010-1122-01-RS
- Kosambi, D. D. (1943). The estimation of map distances from recombination values. *Ann. Eugen.* 12, 172–175. doi: 10.1007/978-81-322-3676-4_16
- Kourelis, J., and van der Hoorn, R. A. L. (2018). Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *The Plant Cell* 30, 285–299. doi: 10.1105/tpc.17.00579
- Lestari, P., Lee, S. H., and Tasma, I. M. (2017). Gene Duplication to reveal adaptation clue of plant to environmental stress: a case study of *NBS-LRR* genes in soybean. *J. Agro. Biogen.* 12, 119–130.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, L., Lin, F., Wang, W., Ping, J., Fitzgerald, J., Zhao, M., et al. (2016). Fine mapping and candidate gene analysis of two loci conferring resistance to *Phytophthora sojae* in soybean. *Theor. Appl. Genet.* 129, 2379–2386. doi: 10.1007/s00122-016-2777-0
- Li, Y., Sun, S., Zhong, C., Wang, X., Wu, X., and Zhu, Z. (2017). Genetic mapping and development of co-segregating markers of *RpsQ*, which provides resistance to *Phytophthora sojae* in soybean. *Theor. Appl. Genet.* 130, 1223–1233. doi: 10.1007/s00122-017-2883-7
- Lin, F., Zhao, M., Ping, J., Johnson, A., Zhang, B., Abney, T. S., et al. (2013). Molecular mapping of two genes conferring resistance to *Phytophthora sojae* in a soybean landrace PI 567139B. *Theor. Appl. Genet.* 126, 2177–2185. doi: 10.1007/s00122-013-2127-4
- Lincoln, S. E., Daly, M. J., and Lander, E. S. (1993). *Constructing Genetic Maps with MAPMAKER/EXP Version 3.0: A Tutorial and Reference Manual*, 3rd Edn. Cambridge: Whitehead Institute for Biomedical Research.
- Liu, K., and Muse, S. (2004). *PowerMarker: A New Genetic Data Analysis Software. Version 3.0*.
- Liu, R. H., and Meng, J. L. (2003). MapDraw: a microsoft excel macro for drawing genetic linkage maps based on given genetic linkage data. *Hereditas* 25, 317–321.
- Lv, Q., Huang, Z., Xu, X., Tang, L., Liu, H., Wang, C., et al. (2017). Allelic variation of the rice blast resistance gene *Pid3* in cultivated rice worldwide. *Sci. Rep.* 7:10362. doi: 10.1038/s41598-017-10617-2
- Marone, D., Russo, M. A., Laidò, G., De Leonardi, A. M., and Mastrangelo, A. M. (2013). Plant nucleotide binding site-leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. *Int. J. Mol. Sci.* 14, 7302–7326. doi: 10.3390/ijms14047302
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Meyers, B. C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R. W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15, 809–834. doi: 10.1105/tpc.009308
- Michelmore, R. W., and Meyers, B. C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* 8, 1113–1130. doi: 10.1101/gr.8.11.1113
- Nagy, E. D., and Bennetzen, J. L. (2008). Pathogen corruption and site-directed recombination at a plant disease resistance gene cluster. *Genome Res.* 18, 1918–1923. doi: 10.1101/gr.078766.108
- Nagy, S., Pocai, P., Cernák, I., Gorji, A. M., Hegedűs, G., and Taller, J. (2012). PICcalc: an online program to calculate polymorphic information content for molecular genetic studies. *Biochemical. Genet.* 50, 670–672. doi: 10.1007/s10528-012-9509-1
- Niu, J., Guo, N., Sun, J., Li, L., Cao, Y., Li, S., et al. (2017). Fine mapping of a resistance gene *RpsHN* that controls *Phytophthora sojae* using recombinant inbred lines and secondary populations. *Front. Plant Sci.* 8:538. doi: 10.3389/fpls.2017.00538
- Ping, J., Fitzgerald, J. C., Zhang, C., Lin, F., Bai, Y., Wang, D., et al. (2016). Identification and molecular mapping of *Rps11*, a novel gene conferring resistance to *Phytophthora sojae* in soybean. *Theor. Appl. Genet.* 129, 445–451. doi: 10.1007/s00122-015-2638-2
- Sahoo, D. K., Abeysekara, N. S., Cianzio, S. R., Robertson, A. E., and Bhattacharyya, M. K. (2017). A Novel *Phytophthora sojae* resistance *Rps12* gene mapped to a genomic region that contains several Rps genes. *PLoS One* 12:e0169950. doi: 10.1371/journal.pone.0169950
- Schmittner, A. F. (1985). Problems and progress in control of *Phytophthora* root rot of soybean. *Plant Dis.* 69, 362–368.

- Schmitthenner, A. F. (1999). "Phytophthora rot of soybean," in *Compendium of Soybean Diseases*, 4th Edn, eds G. L. Hartman, J. C. Rupe, E. J. Sikora, L. L. Domier, J. A. Davis, and K. L. Steffey, (St. Paul, Minnesota: The American Phytopathological Society Press), 39–42.
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670
- Shao, Z. Q., Xue, J. Y., Wu, P., Zhang, Y. M., Wu, Y., Hang, Y. Y., et al. (2016). Large-scale analyses of angiosperm nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes reveal three anciently diverged classes with distinct evolutionary patterns. *Plant Physiol.* 170, 2095–2109. doi: 10.1104/pp.15.01487
- Shirasu, K., Lahaye, T., Tan, M. W., Zhou, F., Azevedo, C., and Schulze-Lefert, P. (1999). A novel class of eukaryotic zinc-binding proteins is required for disease resistance signaling in barley and development in *C. elegans*. *Cell* 99, 355–366. doi: 10.1016/S0092-8674(00)81522-6
- Song, Q., Jia, G., Zhu, Y., Grant, D., Nelson, R. T., Hwang, E. Y., et al. (2010). Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. *Crop Sci.* 50, 1950–1960. doi: 10.2135/cropsci2009.10.0607
- Song, T., Kale, S. D., Arredondo, F. D., Shen, D., Su, L., Liu, L., et al. (2013). Two RxLR avirulence genes in *Phytophthora sojae* determine soybean *Rps* 1k-mediated disease resistance. *Mol. Plant Microbe Interact.* 26, 711–720. doi: 10.1094/MPMI-12-12-0289-R
- Stewart, S., Abeysekara, N., and Robertson, A. E. (2014). Pathotype and genetic shifts in a population of *Phytophthora sojae* under soybean cultivar rotation. *Plant Dis.* 98, 614–624. doi: 10.1094/PDIS-05-13-0575-RE
- Stewart, S., Robertson, A. E., Wickramasinghe, D., Draper, M. A., Michel, A., and Dorrance, A. E. (2016). Population structure among and within Iowa, Missouri, Ohio, and South Dakota populations of *Phytophthora sojae*. *Plant Dis.* 100, 367–379. doi: 10.1094/PDIS-04-15-0437-RE
- Sugimoto, T., Kato, M., Yoshida, S., Matsumoto, I., Kobayashi, T., Kaga, A., et al. (2012). Pathogenic diversity of *Phytophthora sojae* and breeding strategies to develop *Phytophthora*-resistant soybeans. *Breed. Sci.* 61, 511–522. doi: 10.1270/jsbbs.61.511
- Sugimoto, T., Yoshida, S., Kaga, A., Hajika, M., Watanabe, K., Aino, M., et al. (2011). Genetic analysis and identification of DNA markers linked to a novel *Phytophthora sojae* resistance gene in the Japanese soybean cultivar Waseshiro. *Euphytica* 182, 133–145. doi: 10.1007/s10681-011-0525-8
- Sun, J., Li, L., Zhao, J., Huang, J., Yan, Q., Xing, H., et al. (2014). Genetic analysis and fine mapping of *RpsJ*S, a novel resistance gene to *Phytophthora sojae* in soybean [*Glycine max* (L.) Merr.]. *Theor. Appl. Genet.* 127, 913–919. doi: 10.1007/s00122-014-2266-2
- Sun, S., Wu, X., Zhao, J., Wang, Y., Tang, Q., Yu, D., et al. (2011). Characterization and mapping of *RpsYu25*, a novel resistance gene to *Phytophthora sojae*. *Plant Breed.* 130, 139–143. doi: 10.1111/j.1439-0523.2010.01794.x
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tian, M., Zhao, L., Li, S., Huang, J., Sui, Z., Wen, J., et al. (2016). Pathotypes and metalaxyl sensitivity of *Phytophthora sojae* and their distribution in Heilongjiang, China 2011–2015. *J. Gen. Plant Pathol.* 82, 132–141. doi: 10.1007/s1032
- Tooley, P. W., and Grau, C. R. (1984). The relationship between rate-reducing resistance to *Phytophthora megasperma* f. sp. *glycinea* and yield of soybean. *Phytopathology* 74, 1209–1216.
- Tyler, B. M. (2007). *Phytophthora sojae*: root rot pathogen of soybean and model oomycete. *Mol. Plant Pathol.* 8, 1–8. doi: 10.1111/j.1364-3703.2006.00373.x
- Weng, C., Yu, K., Anderson, T. R., and Poysa, V. (2001). Mapping genes conferring resistance to *Phytophthora* root rot of soybean, *Rps1a* and *Rps7*. *J. Hered.* 92, 442–446. doi: 10.1093/jhered/92.5.442
- Wu, M., Li, B., Liu, P., Weng, Q., Zhan, J., and Chen, Q. (2016). Population genetic analyses of *Phytophthora sojae* in Fujian, China. *Plant Pathol.* 66, 1182–1190. doi: 10.1111/ppa.12666
- Wu, X., Zhang, B., Sun, S., Zhao, J., Yang, F., Guo, N., et al. (2011). Identification, genetic analysis and mapping of resistance to *Phytophthora sojae* of *Pm28* in soybean. *Agric. Sci. China* 10, 1506–1511. doi: 10.1016/S1671-2927(11)60145-4
- Xing, L., Hu, P., Liu, J., Witek, K., Zhou, S., Xu, J., et al. (2018). *Pm21* from *Haynaldia villosa* encodes a CC-NBS-LRR that confers powdery mildew resistance in wheat. *Mol. Plant* 11, 874–878. doi: 10.1016/j.molp.2018.02.013
- Xue, A. G., Marchand, G., Chen, Y., Zhang, S., Cober, E. R., and Tenuta, A. (2015). Races of *Phytophthora sojae* in Ontario, Canada, 2010–2012. *Can. J. Plant Pathol.* 37, 376–383. doi: 10.1080/07060661.2015.1052562
- Ye, S., Dhillon, S., Ke, X., Collins, A. R., and Day, I. N. (2001). An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res.* 29:e88. doi: 10.1093/nar/29.17.e88
- You, F. M., Huo, N., Gu, Y. Q., Luo, M. C., Ma, Y., Hane, D., et al. (2008). BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* 9:253. doi: 10.1186/1471-2105-9-253
- Zhang, J., Sun, S., Wang, G., Duan, C., Wang, X., Wu, X., et al. (2014). Characterization of *Phytophthora* resistance in soybean cultivars/lines bred in Henan province. *Euphytica* 196, 375–384. doi: 10.1007/s10681-013-1040-x
- Zhang, J., Xia, C., Duan, C., Sun, S., Wang, X., Wu, X., et al. (2013a). Identification and candidate gene analysis of a novel *Phytophthora* resistance gene *Rps10* in a Chinese soybean cultivar. *PLoS One* 8:e69799. doi: 10.1371/journal.pone.0069799
- Zhang, J., Xia, C., Wang, X., Duan, C., Sun, S., Wu, X., et al. (2013b). Genetic characterization and fine mapping of the novel *Phytophthora* resistance gene in a Chinese soybean cultivar. *Theor. Appl. Genet.* 126, 1555–1561. doi: 10.1007/s00122-013-2073-1
- Zhang, S., Xu, P., Wu, J., Zhang, J., Li, W., Chen, C., et al. (2010). Races of *Phytophthora sojae* and their Virulences on soybean cultivars in Heilongjiang, China. *Plant Dis.* 94, 87–91. doi: 10.1094/PDIS-94-1-0087
- Zhong, C., Sun, S., Li, Y., Duan, C., and Zhu, Z. (2017). Next-generation sequencing to identify candidate genes and develop diagnostic markers for a novel *Phytophthora* resistance gene, *RpsHC18*, in soybean. *Theor. Appl. Genet.* 131, 525–538. doi: 10.1007/s00122-017-3016-z
- Zhong, C., Sun, S., Yao, L., Ding, J., Duan, C., and Zhu, Z. (2018). Fine mapping and identification of a novel *phytophthora* root rot resistance locus *RpsZS18* on Chromosome 2 in Soybean. *Front. Plant Sci.* 9:44. doi: 10.3389/fpls.2018.00044
- Zhu, Z., Wang, H., Wang, X., Chang, R., and Wu, X. (2003). Distribution and virulence diversity of *Phytophthora sojae* in China. *Agric. Sci. China* 3, 116–123.
- Zou, S., Wang, H., Li, Y., Kong, Z., and Tang, D. (2018). The NB-LRR gene *Pm60* confers powdery mildew resistance in wheat. *New Phytol.* 218, 298–309. doi: 10.1111/nph.14964

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

Copyright © 2020 Zhong, Sun, Zhang, Duan and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Impacts of Domestication and Breeding on Nitrogen Fixation Symbiosis in Legumes

Jinge Liu^{1†}, Xiaocheng Yu^{1†}, Qiulin Qin¹, Randy D. Dinkins² and Hongyan Zhu^{1*}

¹ Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY, United States, ² Forage-Animal Production Research Unit, United States Department of Agriculture-Agricultural Research Service, Lexington, KY, United States

OPEN ACCESS

Edited by:

Gaofeng Zhou,
Department of Primary Industries
and Regional Development (DPIRD),
Australia

Reviewed by:

Alessio Mengoni,
University of Florence, Italy
Maria Eugenia Zanetti,
National University of La Plata,
Argentina
Hon-Ming Lam,
The Chinese University of Hong Kong,
China

*Correspondence:

Hongyan Zhu
hzhu4@uky.edu

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 01 June 2020

Accepted: 31 July 2020

Published: 18 August 2020

Citation:

Liu J, Yu X, Qin Q, Dinkins RD and
Zhu H (2020) The Impacts
of Domestication and Breeding on
Nitrogen Fixation Symbiosis
in Legumes. *Front. Genet.* 11:00973.
doi: 10.3389/fgene.2020.00973

Legumes are the second most important family of crop plants. One defining feature of legumes is their unique ability to establish a nitrogen-fixing root nodule symbiosis with soil bacteria known as rhizobia. Since domestication from their wild relatives, crop legumes have been under intensive breeding to improve yield and other agronomic traits but with little attention paid to the belowground symbiosis traits. Theoretical models predict that domestication and breeding processes, coupled with high-input agricultural practices, might have reduced the capacity of crop legumes to achieve their full potential of nitrogen fixation symbiosis. Testing this prediction requires characterizing symbiosis traits in wild and breeding populations under both natural and cultivated environments using genetic, genomic, and ecological approaches. However, very few experimental studies have been dedicated to this area of research. Here, we review how legumes regulate their interactions with soil rhizobia and how domestication, breeding and agricultural practices might have affected nodulation capacity, nitrogen fixation efficiency, and the composition and function of rhizobial community. We also provide a perspective on how to improve legume-rhizobial symbiosis in sustainable agricultural systems.

Keywords: legume, nodulation, nitrogen fixation, symbiosis, domestication, rhizobia

INTRODUCTION

Nitrogen availability is a limiting factor for plant growth. Modern agriculture is heavily dependent on utilization of nitrogen fertilizers to sustain crop production, which is not only expensive but also contributes to severe environmental pollution and global warming. Legumes, however, are able to make their own nitrogen fertilizer through forming a root-nodule symbiosis with nitrogen-fixing soil bacteria collectively known as rhizobia. Inside nodule cells, the rhizobial symbionts find an environmental niche that enables them to convert atmospheric nitrogen to ammonia. The bacteria supply fixed nitrogen to the plant and the plant provides bacteria with carbohydrates derived from photosynthesis. This mutualistic relationship is overwhelmingly important in biosphere nitrogen cycle and plays a critical role in maintaining agricultural sustainability.

Establishment of the legume-rhizobia mutualism is a complex process (Oldroyd et al., 2011). It begins with a cross-kingdom molecular dialogue, leading to altered development of both symbiotic partners. Starting in the rhizosphere, perception of legume root exudates (flavonoids) by the bacterium activates a specific set of bacterial nodulation (Nod) genes,

leading to synthesis and secretion of a highly specific signal known as Nod factors or lipochitooligosaccharides. Perception of Nod factors by the plant LysM receptors induces a variety of host responses, including root hair curling, entry of the bacteria into the root hairs, induced expression of the host symbiotic genes, and cortical cell divisions that collectively culminate in the formation of rhizobia-infected root nodules. Different legumes vary in bacterial infection mode (root hair vs. crack/intercellular infection), nodule development (determinate vs. indeterminate), and bacterial differentiation inside the nodule cells (terminal vs. non-terminal). However, recent phylogenomic studies suggested a single evolutionary innovation of nodulation originated in a common ancestor of the so called “nitrogen-fixing clade” around 110 million years ago (Werner et al., 2014; Griesmann et al., 2018; van Velzen et al., 2018, 2019).

Conversion of nitrogen gas to ammonia is catalyzed by the rhizobial enzyme complex known as nitrogenase. This reaction is an energy-demanding process, requiring a large amount of adenosine triphosphates (ATPs) generated from aerobic metabolism of plant photosynthates. Paradoxically, however, the nitrogenase enzyme is exquisitely sensitive to, and irreversibly inactivated by, free oxygen. This conflict is perfectly resolved in the root nodule where oxygen concentration is buffered by the binding of oxygen with leghemoglobin (Terpolilli et al., 2012; Downie, 2014). To balance the trade-off and maximize the benefits of nitrogen fixation, legume hosts have evolved complex genetic mechanisms to tightly control nodule numbers or to block nodulation when sufficient nitrogen is available in the soil (Krusell et al., 2002; Searle et al., 2003; Nishida and Suzuki, 2018; Nishida et al., 2018; Ferguson et al., 2019). In particular, in autoregulation of nodulation, rhizobial infection activates the transcriptional regulator NIN that in turn triggers a systemic long-range signaling cascade between roots and shoots, resulting in inhibition of further nodule formation; under high soil nitrate conditions, a homologous NIN-Like Protein (NLP) is activated, which integrates into the autoregulation pathway to suppress nodulation and nitrogen fixation.

In contrast to wild species widely distributed in natural environments, modern varieties are developed for high input farming systems, thus decreasing their dependence on biological nitrogen fixation. It has been predicted that this dramatic transition from the wild to cultivation might have resulted in selection of genotypes that are more reliant on soil nitrogen and thus less able to take advantage of symbiotic nitrogen fixation, even though the symbiosis property is universally retained (Werner et al., 2015; Turcotte et al., 2017; Porter and Sachs, 2020). Such genetic shifts can be the consequence of relaxed selection on symbiosis traits, leading to altered host-bacteria compatibility, genotype-dependent root and soil microbial community, and the host's ability to discriminate between high-efficient and low-efficient bacterial partners (Kiers et al., 2003, 2007; Westhoek et al., 2017; Porter and Sachs, 2020). Here, we review how domestication, breeding and agricultural practices might have affected the symbiosis traits in legumes and also provide a perspective on how such knowledge can be used for development of strategies to improve symbiotic nitrogen fixation in sustainable agriculture. It is noteworthy that the potential impacts of

domestication on symbiosis discussed in this review are not independent but interrelated.

EFFECT OF DOMESTICATION AND BREEDING ON HOST-RHIZOBIAL COMPATIBILITY

The legume-rhizobial symbiosis is a highly specific interaction, such that particular legume species/genotypes form an efficient symbiosis with only a specific set of rhizobial species/strains (Wang et al., 2012, 2018). Discordance can occur at various phases of symbiosis development. Incompatibility occurring at the beginning stages of the interaction can block bacterial infection and nodule organogenesis, while incompatibility taking place at later stages of the nodule development can result in development of infected nodules incapable of fixing nitrogen (Yang et al., 2010, 2017; Liu et al., 2014; Tang et al., 2016; Wang et al., 2017). Genetic control of symbiotic specificity is complex, involving a wide range of host and bacterial genes with various modes of action. In most legumes, rhizobial infection and nodule organogenesis requires host-specific recognition of bacterially derived Nod factors, and this recognition defines the basis of host-symbiont specificity (Geurts and Bisseling, 2002). While Nod factor perception is sufficient to cause nodule primordium formation, bacterial infection is also modulated by rhizobial effectors and microbe-associated molecular patterns (Frayssé et al., 2003; Deakin and Broughton, 2009; Kawaharada et al., 2015). Therefore, symbiosis development requires the bacteria to evade or overcome effector- and pattern-triggered immunity (Benezech et al., 2020).

The effect of domestication and breeding on host-rhizobial compatibility has been evaluated only in a handful of experiments (Table 1). Mutch and Young (2004) investigated host-range evolution for a population of *Rhizobium leguminosarum* strains sampled from nodules of pea (*Pisum sativum*), broad bean (*Vicia faba*) and several closely related wild species grown in a natural soil. Nodulation assays using a diverse panel of 80 isolates demonstrated that only 34% of the isolates were able to nodulate broad bean, while 89% were able to nodulate the wild species, suggesting that the domesticated legumes are less promiscuous in their interaction with the indigenous soil strains as compared to the wild legumes. Similarly, a genomic survey of rhizobia sampled from nodules of cultivated chickpea (*Cicer arietinum*) and its wild ancestor (*Cicer reticulatum*) showed that the wild species were associated with a more diverse *Mesorhizobium* population than cultivated chickpea (Kim et al., 2014). It was also reported that a higher abundance of *Bradyrhizobium* strains was present in the rhizosphere of wild species of soybeans (*Glycine soja*) than that of cultivated soybeans (*Glycine max*) (Chang et al., 2019). However, the conclusions drawn from these small-scale studies need to be further evaluated, because significant variation could also exist between different genotypes within the same wild or cultivated species. It is noteworthy that host range alone cannot dictate a cooperative advantage. Although reduced host range might limit nitrogen fixation if compatible strains are unavailable in a given soil, a more frequent scenario is that endemic strains are able to

TABLE 1 | Significant examples of experimental studies on the effects of domestication and breeding on root nodule symbiosis in legumes.

Potential impact	Legume host	References	Conclusion
Symbiosis compatibility	Broad bean (<i>V. faba</i>) Chickpea (<i>C. arietinum</i>) Soybeans (<i>G. max</i>)	Mutch and Young (2004) Kim et al. (2014) Chang et al. (2019)	The cultivated species are less promiscuous in their interaction with indigenous symbionts than the wild legumes.
	Soybeans (<i>G. max</i>)	Devine (1985) Devine and Kuykendall (1996) Yang et al. (2010) Tang et al. (2016) Sugawara et al. (2019)	Analysis of frequencies of <i>Rj2</i> , <i>Rfg1</i> , and <i>Rj4</i> alleles in wild and cultivated soybeans did not reveal strong selection associated with domestication.
Coevolution of rhizobia with domestication	Common bean (<i>P. vulgaris</i>) Chickpea (<i>C. arietinum</i>)	Aguilar et al. (2004) Greenlon et al. (2019)	Both dispersal and novel adapted strains could have evolved with the spread of the domesticated legumes. Dispersed symbionts can transfer symbiosis genes into local strains that quickly became adapted.
Root-associated microbiome	Common bean (<i>P. vulgaris</i>) Soybeans (<i>G. max</i>)	Pérez-Jaramillo et al. (2016, 2018, 2019) Han et al. (2020)	There were significant differences in rhizosphere microbiome composition between domesticated plants and their wild relatives. These differences might have affected legume-rhizobial symbiosis in an indirect manner.
Partner choice and symbiosis capacity	Soybean (<i>G. max</i>) Pea (<i>P. sativum</i>) alfalfa (<i>Medicago sativa</i>)	Kiers et al. (2007) Provorov and Tikhonovich (2003)	The wild legumes and old varieties performed better than the advanced cultivars in terms of symbiotic nitrogen fixation, suggesting that the legume's natural capability to distinguish between less- and high-efficient strains might be disrupted during domestication bottlenecks or breeding process.
	Soybeans (<i>G. max</i>)	Muñoz et al. (2016)	Analysis of 31 cultivated and 17 wild soybeans suggested that nitrogen fixation capacity was enhanced by soybean domestication.

competitively nodulate the host legumes but with low nitrogen fixation efficiency. In this latter situation, excluding nodulation with the less efficient indigenous strains will increase the chances of the host to nodulate with more mutualistic symbiotic partners or added rhizobial inoculants.

A number of strain-specific nodulation-restrictive genes have been cloned in soybeans (e.g., *Rj2*, *Rfg1*, and *Rj4*) (Yang et al., 2010; Tang et al., 2016). These genes control rhizobial infection in the same manner as “gene-for-gene” resistance against pathogens. In particular, *Rj2* and *Rfg1* are allelic genes encoding nucleotide-binding/leucine-rich repeats (NLR) proteins that confer resistance to different groups of *Bradyrhizobium japonicum* and *Sinorhizobium fredii* strains (Yang et al., 2010). *Rj4* encodes a thaumatin-like defense-response protein that restricts infection by specific strains of *Bradyrhizobium elkanii* (Tang et al., 2016). The function of these genes is reliant on the rhizobial type III secretion system and its secreted effectors (Faruque et al., 2015; Sugawara et al., 2018). These studies revealed that effector-triggered immunity play a central role in the regulation of nodulation specificity.

The nodulation-restrictive alleles occur in both wild and cultivated populations, suggesting that these seemingly unfavorable alleles are retained during domestication and breeding processes. This could happen due to a lack of selection, or more likely because of balancing selection under the scenario that the same alleles could also confer resistance against the pathogens that secrete effectors homologous to those secreted by rhizobia (Deakin and Broughton, 2009). Alternatively, these

alleles might have been maintained during domestication and breeding because they may confer an advantage for the host to interact with high-efficient strains and to exclude nodulation with less efficient ones, which is the case of the soybean *Rj4* allele that prevents nodulation by many strains of *B. japonicum* and *B. elkanii* (Devine and Kuykendall, 1996). *B. elkanii* is a poor symbiotic partner of soybeans because of its low nitrogen-fixation efficiency and its production of rhizobitoxine, a compound that induces chlorosis in the host plant. Thus, cultivars with the *Rj4* genotype are favorable in soils where the *B. elkanii* population is dominant. A survey of over 500 accessions of *G. soja* established an *Rj4* allele frequency of 63%, which is similar to that estimated from a cultivated soybean population sampled from Southeast Asia (66%) (Devine and Kuykendall, 1996). The *Rj4* allele is less frequent for cultivars sampled from East Asia (27%) and South Asia (16%). The geographic patterns suggested coevolution of host and bacteria in determining symbiotic compatibilities, because in Southeast Asia, warm climate provides an environment conducive to strains of *B. elkanii*. The *Rj2* allele is also present in both *G. soja* and *G. max*, with an allele frequency of 7.7% and 5.4%, respectively, based on genotyping of 1,324 cultivated and 259 wild soybean accessions, showing that the *Rj2* allele has been maintained during the domestication and breeding processes (Sugawara et al., 2019). In contrast, the *Rfg1* allele occurred with a much higher frequency (0.44) in a *G. max* population consisting of 285 genotypes from Asian countries (Devine, 1985). Even though the *Rfg1* allele frequency has

not been broadly surveyed in *G. soja*, the distribution of the *Rfg1* (0.44) and *rfg1* (0.56) allele frequencies in cultivated soybeans suggests that these alleles has not been subjected to strong selection.

COEVOLUTION OF RHIZOBIA WITH LEGUME DOMESTICATION

Legume hosts play a dominant role in interaction with rhizobia, and the soil rhizobial community has a strong influence on the symbiosis capacity of the plant hosts. The mutualism is established as a result of a coadaptation process between the symbiotic partners, which is also affected by environmental and climate variables and agricultural practices (Fox et al., 2007; Weese et al., 2015; Regus et al., 2017). A survey of *Rhizobium etli* strains that nodulate common beans (*Phaseolus vulgaris*) from the three putative bean domestication centers identified polymorphisms in the common nodulation gene *nodC* that are correlated with their geographical origins (Aguilar et al., 2004). Cross-inoculating wild beans from the three centers with these strains revealed preferential nodulation of beans by geographically related *R. etli* lineages. This observation was further supported by co-inoculation assays, which demonstrated that plants were almost exclusively nodulated by strains isolated from their host region, suggesting coevolution of the symbiotic partners in the host domestication centers.

An intriguing question is how legume domestication might have driven the parallel evolution of rhizobial symbionts. Assuming that the wild relatives evolved a symbiosis with distinct bacteria, then after domestication and subsequent distribution to new environments, the domesticated crops either brought their natural symbionts with them (a process called dispersal), began to adapt to novel symbionts encountered at their new locations, or both (Bouznif et al., 2019). This question can be addressed by phylogenetic analysis of nodulating rhizobia across the crop's native and introduced ranges. Dispersal would entail clustering of bacterial strains from the native and introduced ranges, otherwise bacterial strains would cluster based on their geographic locations and show a biogeographical pattern. A recent comprehensive study in chickpea has provided significant insights into this outstanding question (Greenlon et al., 2019; **Figure 1**), which is discussed below.

Chickpea (*C. arietinum*) was domesticated from the wild species *C. reticulatum* about 11,000 years ago in modern-day southeastern Turkey (von Wettberg et al., 2018). Since then, chickpea was distributed throughout the world, first in Middle East and Mediterranean basin, later reaching the Indian subcontinent and Ethiopia, and presently widely grown in countries where high-input agricultural practices are adopted, including Canada, United States, and Australia. This well-known domestication history of chickpea along with its global distribution provided an ideal system to evaluate the effect of plant domestication on the evolution of its symbiotic partners. For this purpose, rhizobial symbionts were sampled from nodules collected from cultivated chickpea and its wild relatives grown in multiple geographic locations representative of the majority of

chickpea's agricultural and natural ranges, including the plant's center of origin in the southeastern Turkey. Phylogenomic analysis of over 1,000 draft whole-genome sequences revealed *Mesorhizobium* as the single bacterial genus consisting of a diverse panel of taxa that form symbiosis with chickpea, supporting that the wild relatives did evolve for symbiosis with specific bacterial partners.

Beyond chickpea's native range, across regions where chickpea has been cultivated without applying specific rhizobial inoculants, the majority of root nodule bacteria are phylogenetically distinct from those derived from nodules of the wild relatives. Furthermore, phylogenetic diversity of the bacterial strains correlated with their geographic origins, mirroring edaphic and environmental factors (primarily the soil type and latitude), suggesting that the bacteria that dominate each location are likely adapted to the environmental conditions in those locations. Despite being diverse at the nucleotide level and in terms of genome structure, the genes associated with symbiosis share a high-level of gene synteny and sequence similarity to those found in chickpea's natural symbionts, implying that chickpea's coevolved symbionts dispersed along with the crop. In particular, the clade 5A consists of broad geographic distribution of strains that are closely related to those natural symbionts at wild chickpea's center of origin. Taken together, both dispersal and novel adapted strains could have evolved with the spread of the domesticated chickpea. In this scenario, dispersed symbionts can transfer symbiosis genes into local strains that quickly became adapted and outcompeted the dispersal strains.

It was determined that evolutionarily divergent and geographically isolated bacterial taxa share a common set of integrative conjugative elements (ICEs) that encode the major functions of the symbiosis. The spread of the symbiosis ICEs is fixed in the microbe that originated at the crop's center of origin. The relatedness of these ICEs is associated with both geographic distance and the evolutionary history of the background genome, parallel with the chickpea crop's domestication and distribution. These ICEs could be a driving force to achieve increased diversity of compatible bacterial species by facilitating horizontal transfer of symbiosis genes across diverse *Mesorhizobium* taxa.

EFFECT OF DOMESTICATION ON ROOT-ASSOCIATED MICROBIOME

Plant microbiomes, particularly the rhizosphere microbiota, have been extensively studied in recent years using community profiling and omics analyses, but fewer studies have been carried out on legume microbiomes, especially during symbiosis development (Guttman et al., 2014; Hacquard et al., 2015; Müller et al., 2016). The root nodule symbiosis was long thought to be a highly specific binary plant-microbe interaction where the compatible nitrogen-fixing soil bacterium is recruited by the host. It has now become increasingly evident that root nodules contain not only nitrogen-fixing bacteria but many other bacterial species as well, some of which possess growth-promoting activities (Zgadzaj et al., 2015; Martínez-Hidalgo and Hirsch, 2017; Poole et al., 2018). Zgadzaj et al. (2016)

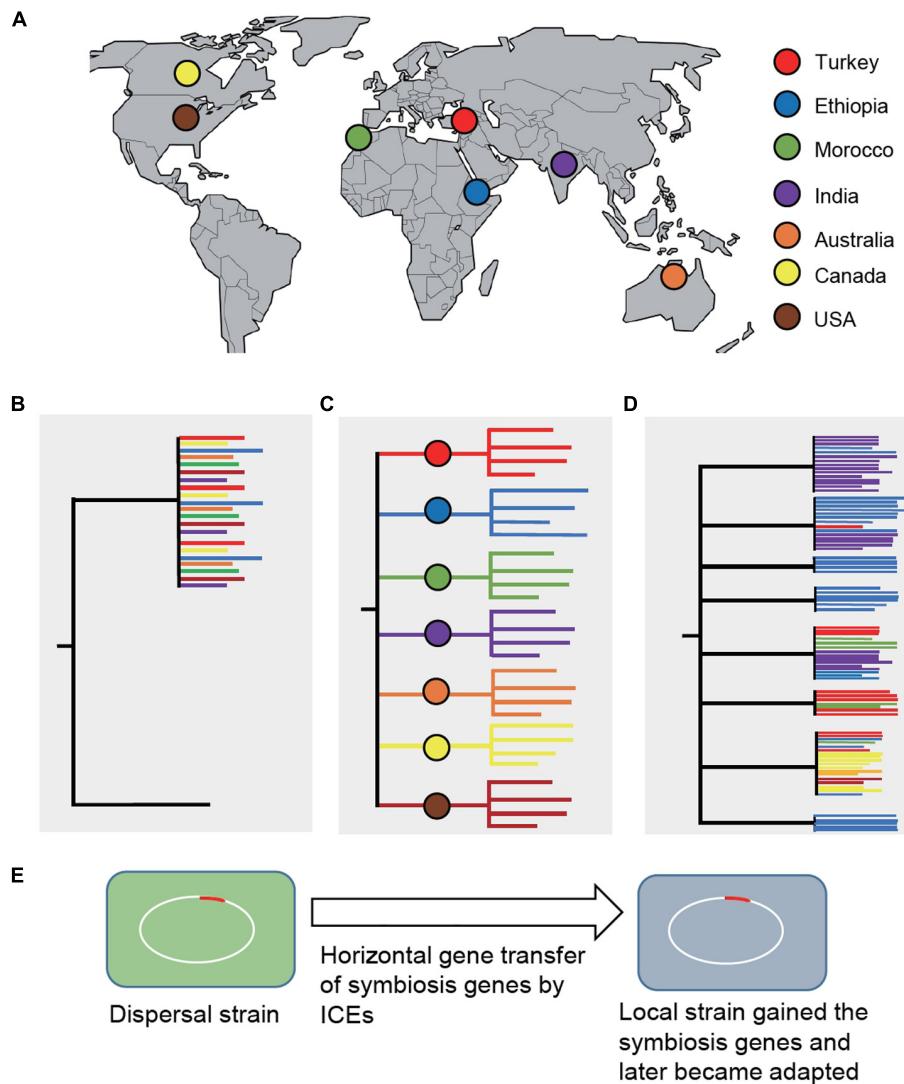


FIGURE 1 | Phylogenetic analysis of *Mesorhizobium* strains sampled across the chickpea's native and introduced ranges (Greenlon et al., 2019). **(A)** Geographic locations (countries) where the strains were sampled, including the chickpea's domestication center, Turkey. **(B)** A phylogenetic diagram showing clustering of bacterial strains from the native and introduced ranges under the dispersal (co-introduction) model. **(C)** A phylogenetic diagram showing that bacterial strains would cluster based on their geographic locations and show a biogeographical pattern if the domesticated plants acquired novel symbionts in their introduced locations. **(D)** If both co-introduction and co-adaptation happened, the tree would show mixed patterns of **(A)** and **(B)**, which is the case revealed by the study of chickpea's symbionts (Greenlon et al., 2019). Trees are not drawn in scale. **(E)** Strains in the introduced ranges might have acquired symbiosis genes from dispersed strains and subsequently became adapted and outcompeted the dispersal strains. Chromosomal symbiosis genes in *Mesorhizobium* are encoded on mobile integrative and conjugative elements (ICEs). ICEs can excise from the chromosome that can transfer and integrate into the chromosome of a recipient strain.

compared root-associated microbiomes, including those from the root and rhizosphere, between *Lotus japonicus* wild type and nodulation-impaired mutant plants. This work revealed that the root-associated microbiomes are influenced by the host's ability to accommodate symbiotic bacteria. For example, when symbiosis is lost, at least six different abundant bacterial orders become almost undetectable in the root; instead, some of these bacterial orders accumulate in the rhizosphere, suggesting that these bacteria are no longer able to enter the root. It is conceivable that these bacteria are involved in nodulation processes and symbiosis as helper bacteria, although it cannot

be excluded that some of these bacteria are opportunists that have nothing to do with the actual symbiosis but hijack the symbiosis pathways to gain entry to the root. Nonetheless, these findings strongly suggest that the legume-rhizobia symbiosis is not just mediated by the rhizobia inside legume nodules, but also engages symbiosis-linked microbial communities that together may influence plant health, productivity, and adaption to biotic and abiotic stresses. Supporting this notion, Han et al. (2020) analyzed the composition and relationship of soybean rhizosphere microbiota in three types of soils and showed that the rhizosphere community differed significantly across different

soils. Interestingly, the presence of some rhizosphere microbes was associated with the composition of *Bradyrhizobium* and *Sinorhizobium* bacteria in nodules; in particular, the *Bacillus cereus* group specifically promoted or suppressed the growth of *Sinorhizobium* and *Bradyrhizobium* bacteria, respectively, and alleviated the effects of saline-alkali conditions on the nodulation by *Sinorhizobium* strains and their colonization in nodules. This study demonstrated a crucial role of the bacterial microbiota in shaping the root nodule symbiosis in soybeans, suggesting that symbiotic capacity can be improved through the use of synthetic bacterial communities.

The impact of plant domestication and habitat expansion on root-associated microbiomes is not well understood. Several studies reported differences in rhizosphere microbiome composition between domesticated plants and their wild relatives (Zachow et al., 2014; Bulgarelli et al., 2015; Pérez-Jaramillo et al., 2016, 2018; Leff et al., 2017). In the legume common bean (*P. vulgaris*), Pérez-Jaramillo et al. (2017) reported that the rhizosphere of wild relatives was enriched in bacterial taxa from the phylum Bacteroidetes, while the rhizosphere of modern bean varieties was dominated by bacterial families belonging to the Actinobacteria and Proteobacteria, and that this compositional shift was associated with plant genotypes as well as root structure. Subsequently, the same set of common bean accessions representing a trajectory from wild to modern were grown both in a native soil and in an agricultural soil and their rhizosphere bacterial communities were profiled and compared (Pérez-Jaramillo et al., 2019). The study showed that the transition of common bean from a native soil to an agricultural soil led to a gain, rather a depletion, of rhizobacterial diversity. In particular, a small number of low-abundant bacterial genera were found only in the native soil and associated with only the rhizosphere of wild bean accessions, suggesting that these bacteria were depleted as a consequence of domestication and habitat expansion. In contrast, a higher number of bacterial taxa were found in the agricultural soil but not in the native soil, suggesting that the depleted bacterial genera in the agricultural soil were replenished with new bacterial taxa, many of which were highly abundant in the rhizosphere of all the common bean accessions. However, it is unknown whether such compositional changes in rhizosphere microbiomes have an impact on plant performance and whether such an observation is universal if more plant genotypes and soil types are tested. Unfortunately, all the above-mentioned experiments did not make a connection with symbiosis traits.

EFFECT OF DOMESTICATION AND BREEDING ON PARTNER CHOICE AND SYMBIOSIS CAPACITY

Studies of binary legume-rhizobial interactions revealed tremendous variation in nitrogen fixation efficiency between different plant-rhizobial combinations. In particular, the same bacterial strain can form a highly efficient symbiosis with a given host but a low-efficient interaction with another, and no single host genotypes or rhizobial strains are consistently associated

with the best nitrogen fixation performance. The failed or low-efficient partnership is often due to genetic incompatibility due to a lack of coadaptation between the interacting partners that negatively affects bacterial differentiation and survival inside the nodule cells. Thus, the host's ability to choose best-matching partners from a pool of compatible rhizobia is key to improve the gain of nitrogen fixation.

It was hypothesized that natural hosts (wild species) might have evolved signaling mechanisms to preferentially partner with more mutualistic symbionts and to disadvantage less-cooperative ones (Younginger and Friesen, 2019). Evolutionary models predict that such natural surveillance capability can be disrupted during domestication bottlenecks or breeding process because of: (i) trading-offs between partner quality and agricultural traits as a result of negative trait association or antagonistic gene pleiotropy, such as defense versus symbiosis (Chen et al., 2017; Wood et al., 2018) or (ii) relaxed selection against symbiosis disruption in nitrogen-rich soils owing to a lack of observable fitness costs over disrupted symbiosis (Porter and Sachs, 2020). For example, when plants are grown in high-input agricultural soils, genotypes with reduced symbiosis capability are likely to have higher yields because they can freely get sufficient nitrogen from soil without need to allocate photosynthates to rhizobia for symbiosis development. However, very few experimental studies were carried out to test the hypothesis. Kiers et al. (2007) evaluated six soybean cultivars, representing 60 years of breeding, by infection with a mixture of effective and ineffective rhizobial strains. The experiments showed that older cultivars performed better than newer cultivars in terms of seed production, and this difference mirrored the relative fitness of effective and ineffective strains in their nodules, suggesting that the newer cultivars have reduced their capability to defend against less-efficient symbiotic partners. Provorov and Tikhonovich (2003) also observed that the wild-growing populations and local varieties of pea and alfalfa performed better than the advanced cultivars in terms of symbiotic nitrogen fixation.

However, Muñoz et al. (2016) reported that nitrogen fixation capacity was enhanced by soybean domestication. For their experiments, 31 cultivated and 17 wild soybeans were assayed for their nitrogen fixation efficiency by inoculation with the slow-growing *B. japonicum* strain USDA110 or the fast-growing *S. fredii* strain CCBAU45436. The results showed that cultivated soybeans overall out-performed wild soybeans in nitrogen fixation, as indicated by total nodule fresh weight, nodule number, total nitrogen, and total ureides accumulation (in soybeans, fixed nitrogen is first converted into ureides, which are exported out of nodules and transported to shoots, where they are catabolized). The lower nitrogen fixation efficiency of wild species appeared to be associated with higher abundance of polyhydroxybutyrate (PHB) bodies in bacteroids. PHB is an energy-rich compound that increases fitness of rhizobia but represents a net cost to the host, implying a trade-off between PHB accumulation and nitrogen fixation. Furthermore, genetic mapping in a recombinant inbred population derived from a cross between a cultivated and a wild parent identified several quantitative trait loci (QTLs) for total ureides and total nodule fresh weight. Association analysis revealed single nucleotide

polymorphisms (SNPs) that are associated with these QTLs and these SNPs exhibited a low level of diversity in cultivated soybean, typical for domestication-related traits.

Despite inconsistent observations from different experiments, it is clear that there exists significant variation in symbiosis capacity between different plant genotypes in both wild and cultivated legume populations. This provides an opportunity for researchers to identify genetic loci/alleles that control the variation and integrate the genetic information into plant improvement through marker-assisted selection or gene editing.

CONCLUSION AND PERSPECTIVE

Nitrogen fixation is dependent on the availability of the legume host's bacterial partners in a given soil and the host's ability to selectively interact with the most mutualistic partners from a group of compatible indigenous strains. In intensively managed agriculture systems, legume crops are often provided specific inoculants, but inoculation can fail if the added strains are not compatible with a given variety or easily displaced by endemic strains that are highly competitive for nodulation but less efficient for nitrogen fixation. Furthermore, the inoculant can become unstable over time because of gene loss and frequent horizontal gene transfer between soil bacteria (Brambilla et al., 2018). Thus, it is important for researchers to engineer or screen for elite rhizobia that are both competitive for nodulation and capable of high rates of nitrogen fixation, while also maintaining long-term stability in the intended soil environments (Checcucci et al., 2017; Westhoek et al., 2017). Most recently, a high-throughput screening method has been developed that allows for simultaneous measurement of both competitiveness and

nitrogen fixation effectiveness of rhizobia at the single nodule level (Mendoza-Suárez et al., 2020); this technology provides the potential to revolutionize the search for super competitive and highly effective rhizobial symbionts.

It is equally important to improve host's ability to choose and cooperate with the best mutualistic symbionts. For this purpose, genetic and genomic approaches should be employed to identify genes and alleles through harnessing abundant genetic and phenotypic variation present in the wild species and cultivated varieties. Lastly, understanding the effects of rhizosphere bacterial community on symbiotic capacity will provide an additional tool to manipulate the legume-rhizobial symbiosis. These coordinated strategies will yield the optimal combinations of partners' genotypes and associated microbial helpers, ultimately leading to improvement of biological nitrogen fixation in sustainable agriculture.

AUTHOR CONTRIBUTIONS

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

FUNDING

This work was supported by the United States Department of Agriculture/National Institute of Food and Agriculture, Agriculture and Food Research Initiative Grant 2014-67013-21573, by United States National Science Foundation Grant 1758037, and by a USDA-University of Kentucky special cooperative grant.

REFERENCES

- Aguilar, O. M., Riva, O., and Peltzer, E. (2004). Analysis of *Rhizobium etli* and of its symbiosis with wild *Phaseolus vulgaris* supports coevolution in centers of host diversification. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13548–13553. doi: 10.1073/pnas.0405321101
- Benezech, C., Doudement, M., and Gourion, B. (2020). Legumes tolerance to rhizobia is not always observed and not always deserved. *Cell Microbiol.* 22:e13124. doi: 10.1111/cmi.13124
- Bouzrif, B., Guefrachi, I., Rodríguez de la Vega, R. C., Hungria, M., Mars, M., Alunni, B., et al. (2019). Phylogeography of the *Bradyrhizobium* spp. associated with peanut, *Arachis hypogaea*: fellow travelers or new associations?. *Front. Microbiol.* 10:2041. doi: 10.3389/fmicb.2019.02041
- Brambilla, S., Frare, R., Soto, G., Jozefkiewicz, C., and Ayub, N. (2018). Absence of the nitrous oxide reductase gene cluster in commercial alfalfa inoculants is probably due to the extensive loss of genes during rhizobial domestication. *Microb. Ecol.* 76, 299–302. doi: 10.1007/s00248-018-1145-9
- Bulgarelli, D., Garrido-Oter, R., Münch, P. C., Weiman, A., Dröge, J., Pan, Y., et al. (2015). Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host Microb.* 17, 392–403. doi: 10.1016/j.chom.2015.01.011
- Chang, C., Chen, W., Luo, S., Ma, L., Li, X., and Tian, C. (2019). *Rhizosphere microbiota* assemblage associated with wild and cultivated soybeans grown in three types of soil suspensions. *Arch. Agron. Soil Sci.* 65, 74–87. doi: 10.1080/03650340.2018.1485147
- Checcucci, A., DiCenzo, G. C., Bazzicalupo, M., and Mengoni, A. (2017). Trade, diplomacy, and warfare: the quest for elite rhizobia inoculant strains. *Front. Microbiol.* 8:2207. doi: 10.3389/fmicb.2017.02207
- Chen, T., Duan, L., Zhou, B., Yu, H., Zhu, H., Cao, Y., et al. (2017). Interplay of pathogen-induced defense responses and symbiotic establishment in *Medicago truncatula*. *Front. Microbiol.* 8:973. doi: 10.3389/fmicb.2017.00973
- Deakin, W. J., and Broughton, W. J. (2009). Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. *Nat. Rev. Microbiol.* 7, 312–320. doi: 10.1038/nrmicro2091
- Devine, T. E. (1985). Nodulation of soybean plant introduction lines with the fast-growing rhizobial strain USDA 205. *Crop Sci.* 25, 354–356. doi: 10.2135/cropsci1985.0011183X002500020037x
- Devine, T. E., and Kuykendall, L. D. (1996). Host genetic control of symbiosis in soybean (*Glycine max* L.). *Plant Soil* 186, 173–187. doi: 10.1007/BF00035072
- Downie, J. A. (2014). Legume nodulation. *Curr. Biol.* 24, R184–R190. doi: 10.1016/j.cub.2014.01.028
- Faruque, O. M., Miwa, H., Yasuda, M., Fujii, Y., Kaneko, T., Sato, S., et al. (2015). Identification of *Bradyrhizobium elkanii* genes involved in incompatibility with soybean plants carrying the Rj4 allele. *Appl. Environ. Microbiol.* 81, 6710–6717. doi: 10.1128/AEM.01942-15
- Ferguson, B. J., Mens, C., Hastwell, A. H., Zhang, M., Su, H., Jones, C. H., et al. (2019). Legume nodulation: the host controls the party. *Plant Cell Environ.* 42, 41–51. doi: 10.1111/pce.13348
- Fox, J. E., Gullledge, J., Engelhaupt, E., Burrow, M. E., and McLachlan, J. A. (2007). Pesticides reduce symbiotic efficiency of nitrogen-fixing rhizobia and host plants. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10282–10287. doi: 10.1073/pnas.0611710104

- Frayssé, N., Couderc, F., and Poinso, V. (2003). Surface polysaccharide involvement in establishing the rhizobium-legume symbiosis. *Eur. J. Biochem.* 270, 1365–1380. doi: 10.1046/j.1432-1033.2003.03492.x
- Geurts, R., and Bisseling, T. (2002). Rhizobium Nod factor perception and signalling. *Plant Cell* 14, S239–S249. doi: 10.1105/tpc.002451
- Greenlon, A., Chang, P. L., Damte, Z. M., Muleta, A., Carrasquilla-Garcia, N., Kim, D., et al. (2019). Global-level population genomics reveals differential effects of geography and phylogeny on horizontal gene transfer in soil bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 116, 15200–15209. doi: 10.1073/pnas.1900056116
- Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M. B., et al. (2018). Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* 361:6398. doi: 10.1126/science.aat1743
- Guttman, D. S., McHardy, A. C., and Schulze-Lefert, P. (2014). Microbial genome-enabled insights into plant-microorganism interactions. *Nat. Rev. Genet.* 12, 797–813. doi: 10.1038/nrg3748
- Hacquard, S., Garrido-Oter, R., González, A., Spaepen, S., Ackermann, G., Lebeis, S., et al. (2015). Microbiota and host nutrition across plant and animal kingdoms. *Cell Host Microbe* 17, 603–616. doi: 10.1016/j.chom.2015.04.009
- Han, Q., Ma, Q., Chen, Y., Tian, B., Xu, L., Bai, Y., et al. (2020). Variation in rhizosphere microbial communities and its association with the symbiotic efficiency of rhizobia in soybean. *ISME J.* 14, 1915–1928. doi: 10.1038/s41396-020-0648-649
- Kawaharada, Y., Kelly, S., Nielsen, M. W., Hjuler, C. T., Gysel, K., Muszyński, A., et al. (2015). Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* 523, 308–312. doi: 10.1038/nature14611
- Kiers, E. T., Hutton, M. G., and Denison, R. F. (2007). Human selection and the relaxation of legume defences against ineffective rhizobia. *Proc. Biol. Sci.* 274, 3119–3126. doi: 10.1098/rspb.2007.1187
- Kiers, E. T., Rousseau, R. A., West, S. A., and Denison, R. F. (2003). Host sanctions and the legume-rhizobium mutualism. *Nature* 425, 78–81. doi: 10.1038/nature01931
- Kim, D. H., Kaashyap, M., Rathore, A., Das, R. R., Parupalli, S., and Upadhyaya, H. D. (2014). Phylogenetic diversity of Mesorhizobium in chickpea. *J. Biosci.* 39, 513–517. doi: 10.1007/s12038-014-9429-9
- Krusell, L., Madsen, L. H., Sato, S., Aubert, G., Genua, A., Szczyglowski, K., et al. (2002). Shoot control of root development and nodulation is mediated by a receptor-like kinase. *Nature* 420, 422–426. doi: 10.1038/nature01207
- Leff, J. W., Lynch, R. C., Kane, N. C., and Fierer, N. (2017). Plant domestication and the assembly of bacterial and fungal communities associated with strains of the common sunflower. *Helianthus annuus*. *New Phytol.* 214, 412–423. doi: 10.1111/nph.14323
- Liu, J., Yang, S., Zheng, Q., and Zhu, H. (2014). Identification of a dominant gene in *Medicago truncatula* that restricts nodulation by *Sinorhizobium meliloti* strain Rm41. *BMC Plant Biol.* 14:167. doi: 10.1186/1471-2229-14-167
- Martínez-Hidalgo, P., and Hirsch, A. (2017). The nodule microbiome: N₂-fixing rhizobia do not live alone. *Phytobiomes* 1, 70–82. doi: 10.1094/PBIOMES-12-16-0019-RVW
- Mendoza-Suárez, M. A., Geddes, B. A., Sánchez-Cañizares, C., Ramírez-González, R. H., Kirchhelle, C., Jorin, B., et al. (2020). Optimizing Rhizobium-legume symbioses by simultaneous measurement of rhizobial competitiveness and N₂ fixation in nodules. *Proc. Natl. Acad. Sci. U.S.A.* 117, 9822–9831. doi: 10.1073/pnas.1921225117
- Müller, D. B., Vogel, C., Bai, Y., and Vorholt, J. A. (2016). The plant microbiota: systems-level insights and perspectives. *Annu. Rev. Genet.* 50, 211–234. doi: 10.1146/annurev-genet-120215-034952
- Muñoz, N., Qi, X., Li, M. W., Xie, M., Gao, Y., Cheung, M. Y., et al. (2016). Improvement in nitrogen fixation capacity could be part of the domestication process in soybean. *Heredity* 117, 84–93. doi: 10.1038/hdy.2016.27
- Mutch, L. A., and Young, J. P. (2004). Diversity and specificity of *Rhizobium leguminosarum* biovar viciae on wild and cultivated legumes. *Mol. Ecol.* 13, 2435–2444. doi: 10.1111/j.1365-294X.2004.02259.x
- Nishida, H., and Suzuki, T. (2018). Nitrate-mediated control of root nodule symbiosis. *Curr. Opin. Plant Biol.* 44, 129–136. doi: 10.1016/j.pbi.2018.04.006
- Nishida, H., Tanaka, S., Handa, Y., Ito, M., Sakamoto, Y., Matsunaga, S., et al. (2018). A NIN-LIKE PROTEIN mediates nitrate-induced control of root nodule symbiosis in *Lotus japonicus*. *Nat. Commun.* 9:499. doi: 10.1038/s41467-018-02831-x
- Oldroyd, G. E., Murray, J. D., Poole, P. S., and Downie, J. A. (2011). The rules of engagement in the legume-rhizobial symbiosis. *Annu. Rev. Genet.* 45, 119–144. doi: 10.1146/annurev-genet-110410-132549
- Pérez-Jaramillo, J. E., Carrión, V. J., Bosse, M., Ferrão, L., de Hollander, M., Garcia, A., et al. (2017). Linking rhizosphere microbiome composition of wild and domesticated *Phaseolus vulgaris* to genotypic and root phenotypic traits. *ISME J.* 11, 2244–2257. doi: 10.1038/ismej.2017.85
- Pérez-Jaramillo, J. E., Carrión, V. J., de Hollander, M., and Raaijmakers, J. M. (2018). The wild side of plant microbiomes. *Microbiome* 6:143. doi: 10.1186/s40168-018-0519-z
- Pérez-Jaramillo, J. E., de Hollander, M., Ramírez, C. A., Mendes, R., Raaijmakers, J. M., and Carrión, V. J. (2019). Deciphering rhizosphere microbiome assembly of wild and modern common bean (*Phaseolus vulgaris*) in native and agricultural soils from Colombia. *Microbiome* 7:114. doi: 10.1186/s40168-019-0727-1
- Pérez-Jaramillo, J. E., Mendes, R., and Raaijmakers, J. M. (2016). Impact of plant domestication on rhizosphere microbiome assembly and functions. *Plant Mol. Biol.* 90, 635–644. doi: 10.1007/s11103-015-0337-7
- Poole, P., Ramachandran, V., and Terpolilli, J. (2018). Rhizobia: from saprophytes to endosymbionts. *Nat. Rev. Microbiol.* 16, 291–303. doi: 10.1038/nrmicro.2017.171
- Porter, S. S., and Sachs, J. L. (2020). Agriculture and the disruption of plant-microbial symbiosis. *Trends Ecol. Evol.* 35, 426–439. doi: 10.1016/j.tree.2020.01.006
- Provorov, N., and Tikhonovich, I. (2003). Genetic resources for improving nitrogen fixation in legume-rhizobia symbiosis. *Genet. Resour. Crop Evol.* 50, 89–99. doi: 10.1023/A:1022957429160
- Regus, J. U., Wendlandt, C. E., Bantay, R. M., Gano-Cohen, K. A., Gleason, N. J., Hollowell, A. C., et al. (2017). Nitrogen deposition decreases the benefits of symbiosis in a native legume. *Plant Soil* 414, 159–170. doi: 10.1007/s11104-016-3114-8
- Searle, I. R., Men, A. E., Laniya, T. S., Buzas, D. M., Iturbe-Ormaetxe, I., Carroll, B. J., et al. (2003). Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. *Science* 299, 109–112. doi: 10.1126/science.1077937
- Sugawara, M., Takahashi, S., Umehara, Y., Iwano, H., Tsurumaru, H., Odake, H., et al. (2018). Variation in bradyrhizobial NopP effector determines symbiotic incompatibility with Rj2-soybeans via effector-triggered immunity. *Nat. Commun.* 9:3139. doi: 10.1038/s41467-018-05663-x
- Sugawara, M., Umehara, Y., Kaga, A., Hayashi, M., Ishimoto, M., Sato, S., et al. (2019). Symbiotic incompatibility between soybean and *Bradyrhizobium ardens* from one amino acid determinant in soybean Rj2 protein. *PLoS One* 14:e0222469. doi: 10.1371/journal.pone.0222469
- Tang, F., Yang, S., Liu, J., and Zhu, H. (2016). Rj4, a gene controlling nodulation specificity in soybeans, encodes a thaumatin-like protein but not the one previously reported. *Plant Physiol.* 170, 26–32. doi: 10.1104/pp.15.01661
- Terpolilli, J. J., Hood, G. A., and Poole, P. S. (2012). What determines the efficiency of N(2)-fixing Rhizobium-legume symbioses? *Adv. Microb. Physiol.* 60, 325–389. doi: 10.1016/B978-0-12-398264-3.00005-X
- Turcotte, M. M., Araki, H., Karp, D. S., Poveda, K., and Whitehead, S. R. (2017). The eco-evolutionary impacts of domestication and agricultural practices on wild species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372:20160033. doi: 10.1098/rstb.2016.0033
- van Velzen, R., Doyle, J. J., and Geurts, R. (2019). A resurrected scenario: single gain and massive loss of nitrogen-fixing nodulation. *Trends Plant Sci.* 24, 49–57. doi: 10.1016/j.tplants.2018.10.005
- van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., et al. (2018). Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc. Natl. Acad. Sci. U.S.A.* 115, E4700–E4709. doi: 10.1073/pnas.1721395115
- von Wettberg, E. J. B., Chang, P. L., Başdemir, F., Carrasquilla-Garcia, N., Korbu, L. B., Moenga, S. M., et al. (2018). Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nat. Commun.* 9:649. doi: 10.1038/s41467-018-02867-z
- Wang, D., Yang, S., Tang, F., and Zhu, H. (2012). Symbiosis specificity in the legume: rhizobial mutualism. *Cell. Microbiol.* 14, 334–342. doi: 10.1111/j.1462-5822.2011.01736.x

- Wang, Q., Liu, J., and Zhu, H. (2018). Genetic and molecular mechanisms underlying symbiotic specificity in legume-rhizobium interactions. *Front. Plant Sci.* 9:313. doi: 10.3389/fpls.2018.00313
- Wang, Q., Yang, S., Liu, J., Terecskei, K., Ábrahám, E., Gombár, A., et al. (2017). Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* 114, 6854–6859. doi: 10.1073/pnas.1700715114
- Weese, D. J., Heath, K. D., Dentinger, B. T., and Lau, J. A. (2015). Long-term nitrogen addition causes the evolution of less-cooperative mutualists. *Evolution* 69, 631–642. doi: 10.1111/evo.12594
- Werner, G. D., Cornwell, W. K., Cornelissen, J. H., and Kiers, E. T. (2015). Evolutionary signals of symbiotic persistence in the legume-rhizobia mutualism. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10262–10269. doi: 10.1073/pnas.1424030112
- Werner, G. D., Cornwell, W. K., Sprent, J. I., Kattge, J., and Kiers, E. T. (2014). A single evolutionary innovation drives the deep evolution of symbiotic N₂-fixation in angiosperms. *Nat. Commun.* 5:4087. doi: 10.1038/ncomms5087
- Westhoek, A., Field, E., Rehling, F., Mulley, G., Webb, I., Poole, P. S., et al. (2017). Policing the legume-*Rhizobium* symbiosis: a critical test of partner choice. *Sci. Rep.* 7:1419. doi: 10.1038/s41598-017-01634-2
- Wood, C. W., Pilkington, B. L., Vaidya, P., Biel, C., and Stinchcombe, J. R. (2018). Genetic conflict with a parasitic nematode disrupts the legume-rhizobia mutualism. *Evol. Lett.* 2, 233–245. doi: 10.1002/evl3.51
- Yang, S., Tang, F., Gao, M., Krishnan, H. B., and Zhu, H. (2010). R gene-controlled host specificity in the legume-rhizobia symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18735–18740. doi: 10.1073/pnas.1011957107
- Yang, S., Wang, Q., Fedorova, E., Liu, J., Qin, Q., Zheng, Q., et al. (2017). Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* 114, 6848–6853. doi: 10.1073/pnas.1700460114
- Younginger, B. S., and Friesen, M. L. (2019). Connecting signals and benefits through partner choice in plant-microbe interactions. *FEMS Microbiol. Lett.* 366:fnz217. doi: 10.1093/femsle/fnz217
- Zachow, C., Müller, H., Tilcher, R., and Berg, G. (2014). Differences between the rhizosphere microbiome of *Beta vulgaris* ssp. *maritima*-ancestor of all beet crops-and modern sugar beets. *Front. Microbiol.* 5:415. doi: 10.3389/fmicb.2014.00415
- Zgadaj, R., Garrido-Oter, R., Jensen, D. B., Koprivova, A., Schulze-Lefert, P., and Radutoiu, S. (2016). Root nodule symbiosis in *Lotus japonicus* drives the establishment of distinctive rhizosphere, root, and nodule bacterial communities. *Proc. Natl. Acad. Sci. U.S.A.* 113, E7996–E8005. doi: 10.1073/pnas.1616564113
- Zgadaj, R., James, E. K., Kelly, S., Kawaharada, Y., de Jonge, N., Jensen, D. B., et al. (2015). A legume genetic framework controls infection of nodules by symbiotic and endophytic bacteria. *PLoS Genet.* 11:e1005280. doi: 10.1371/journal.pgen.1005280

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

Copyright © 2020 Liu, Yu, Qin, Dinkins and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Soybean Endo-1,3-Beta-Glucanase (GmGLU) Interaction With Soybean mosaic virus-Encoded P3 Protein May Contribute to the Intercellular Movement

Feifei Shi^{1,2†}, Ying Wang^{1,3†}, Fang Zhang⁴, Xingxing Yuan¹, Huatao Chen¹, Xuehao Chen², Xin Chen^{1,5*} and Xiaoyan Cui^{1,5*}

¹ Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences/Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, Nanjing, China, ² Department of Plant Pathology, College of Plant Protection, China Agricultural University, Beijing, China, ³ Department of Horticulture, College of Horticulture and Plant Protection, Yangzhou University, Yangzhou, China, ⁴ Central Laboratory, Jiangsu Academy of Agricultural Sciences, Nanjing, China, ⁵ Institute of Life Science, Jiangsu University, Zhenjiang, China

OPEN ACCESS

Edited by:

Lifeng Zhu,
Nanjing Normal University, China

Reviewed by:

Sergey Morozov,
Lomonosov Moscow State University,
Russia
Hiroshi Nyunoya,
Waseda University, Japan

*Correspondence:

Xiaoyan Cui
cxy@jaas.ac.cn
Xin Chen
cx@jaas.ac.cn

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 21 February 2020

Accepted: 26 August 2020

Published: 15 September 2020

Citation:

Shi F, Wang Y, Zhang F, Yuan X,
Chen H, Chen X, Chen X and Cui X
(2020) Soybean
Endo-1,3-Beta-Glucanase (GmGLU)
Interaction With Soybean mosaic
virus-Encoded P3 Protein May
Contribute to the Intercellular
Movement. *Front. Genet.* 11:536771.
doi: 10.3389/fgene.2020.536771

Soybean mosaic virus (SMV), a member of the genus *Potyvirus*, is a prevalent and devastating viral pathogen in soybean-growing regions worldwide. Potyvirus-encoded P3 protein is reported to participate in virus replication, movements, and pathogenesis. This study provides evidence that the soybean (*Glycine max*) endo-1,3-beta-glucanase protein (designated as GmGLU) interacts with SMV-P3 by using a yeast two-hybrid system to screen a soybean cDNA library. A bimolecular fluorescence complementation assay further confirmed the interaction, which occurred on the cytomembrane in *Nicotiana benthamiana* cells. Subcellular localization experiment indicated that GmGLU localized in cytomembrane and could co-localized at PD with PD marker. The transient expression of GmGLU promoted the coupling of Turnip mosaic virus replication and cell-to-cell movement in *N. benthamiana*. Meanwhile, qRT-PCR experiment demonstrated that the expression of GmGLU which involved in callose regulation increased under SMV infection. Under SMV infection, callose deposition at PD was observed obviously by staining with aniline blue, which raise a physical barrier restricting cell-to-cell movement of SMV. When overexpression the GmGLU into the leaves under SMV infection, the callose induced by SMV was degraded. Coexpression the GmGLU and SMV in soybean leaves, callose was not found, whereas a large amount of callose deposition on soybean leaves which were only under SMV infection. The results show that GmGLU can degrade the callose induced by SMV infection and indicate that GmGLU may be an essential host factor involvement in potyvirus infection.

Keywords: Soybean mosaic virus, P3, GmGLU, interaction, callose

Abbreviations: BiFC, bimolecular fluorescence complementation; ER, endoplasmic reticulum; GmGLU, soybean (*Glycine max*) endo-1,3-beta-glucanase; MP, movement protein; mY2HS, membrane yeast two-hybrid system; PBS, phosphatic buffer solution; PD, plasmodesmata; PR, pathogenesis-related; PVX, *Potato virus X*; PVY, *Potato virus Y*; SEL, size exclusion limit; SMV, *Soybean mosaic virus*; TMV, *Tobacco mosaic virus*; TuMV, *Turnip mosaic virus*.

HIGHLIGHTS

- GmGLU interacted with SMV-P3 in a screening of soybean cDNAs using mY2HS.
- The interaction between GmGLU and SMV-P3 was confirmed using BiFC.
- The subcellular location of GmGLU was on the cytomembrane and form the dots at plasmodesmata sites.
- GmGLU was up-expressed in response to SMV infection.
- GmGLU overexpression of soybean decreases callose deposition.
- GmGLU overexpression promotes viral cell-to-cell movement and replication.

INTRODUCTION

Soybean mosaic virus (SMV), a member of the family *Potyviridae*, is the causative agent of one of the most well-known viral diseases affecting soybean (Kim et al., 2016). Specifically, the virions are transmitted by seeds and aphids, reducing the yield and quality of soybeans (Shukla et al., 1994; Yang and Gai, 2011). The SMV genome is a single positive-stranded RNA, about 9600 nucleotides long, containing an open reading frame (ORF) and 11 different multifunctional mature proteins (Urcuqui-Inchima et al., 2001; Chung et al., 2008; Wen and Hajimorad, 2010).

P3, a 43 kDa non-structural protein, is the most variable potyviral proteins (Eiamtanaset et al., 2007). Many previous studies have shown that P3 participates in virus replication, movement, and pathogenesis (Johansen et al., 2001; Jenner et al., 2003; Suehiro et al., 2004). Some studies further revealed that the P3 cistron may also function as a virulence determinant for the SMV elicitors of Rsv1-mediated resistance (Hajimorad et al., 2006, 2008, 2011; Chowda-Reddy et al., 2010; Wen et al., 2011, 2012; Wang et al., 2014). Previous studies have revealed that the *Tobacco etch virus* P3 protein is localized in the ER membrane and forms punctate inclusions in association with the Golgi apparatus (Cui et al., 2010). Moreover, the P3 punctate structure was found to traffic along the actin filaments and co-localize with the 6 kDa peptide-containing replication vesicles (Cui et al., 2010). The potyviral P3 cistron encodes two viral proteins, P3N-PIPO and P3, which have two distinct functions: P3N-PIPO is a dedicated MP (Wei et al., 2010) and P3 is an essential component of the viral replication complexes (Cui et al., 2017). Furthermore, P3N-PIPO interacts with P3 via the shared N-terminal domain to recruit viral replication vesicles for cell-to-cell movement (Chai et al., 2020). Therefore, we speculate that P3 may act as a bridge between virus replication and movement.

Previous studies have shown that P3 targets the host elongation factor eEF1A, inducing the associated unfolded protein response, which in turn facilitates SMV replication (Luan et al., 2016). It has been reported that P3 interacts with the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO), thus contributing to the development of viral infection symptoms in host plants (Lin et al., 2011). Furthermore,

previous reports have studied the interaction of soybean proteins with P3, including soybean actin-depolymerizing factor 2 (Lu et al., 2015), hypersensitive response-like lesion-inducing protein (Luan et al., 2019), endoplasmic reticulum homologous protein (GmRHP) (Cui et al., 2018), which might play important role in SMV infection. Taken together, these studies suggest that studying the function of P3 and its interactors can help us better understand viral infection.

Endo-1,3- β -glucanases (GLU) are grouped in the PR-2 family of PR proteins (Balasubramanian et al., 2012; Piršelová and Matušíková, 2013), which catalyze the hydrolysis of 1,3- β -glucan linkages and play important roles in plant defense and development. For instance, the previously reported that abscisic acid inhibits the transcription of 1,3- β -glucanase and thus remarkably strengthens the ability of plants to resist viruses (Mauch-Mani and Mauch, 2005; Fernández-Crespo et al., 2017). Callose, a 1,3- β -glucan which is a common substrate of GLU, is reversibly and transiently deposited in plant cell walls as part of the hypersensitive response to fungi, viruses, and abiotic stresses (Stone and Clarke, 1992; Kauss, 1996). Callose deposition occurs in the wall surrounding plasmodesmata [the intercellular connections between plant cells that allow cell-to-cell transport of sugars, amino acids, inorganic ions, proteins, and nucleic acids (reviewed by Lucas et al., 1993)] at both ends of the channel, compressing the plasma membrane inward, thus creating a narrowed neck region (Radford et al., 1998), which inhibits the transport of macromolecules through plasmodesma (PD). GLU as hydrolytic enzymes which specifically degrade callose, change the dynamic balance between callose synthesis and degradation, while promoting cell-to-cell viral movement and spread. Thus, callose is thought to act as a physical barrier limiting or preventing the spread of viruses in the host (Allison and Shalla, 1974; Iglesias and Meins, 2000). Previous study has shown that increased expression of GLU I in virus-infected cells can increase the size of local lesions and in the absence of the hypersensitive response, can promote virus spread (Gregor et al., 2001). As opposed to GLU overexpression, the deficiency in GLU results in reduced susceptibility to viral infection due to net increase in PD callose levels (Zavaliev et al., 2011).

In this study, we demonstrated that the soybean factor endo-1,3- β -glucanase (designated as GmGLU) interacts with SMV P3 in a mY2H analysis and BiFC. *GmGLU* transformed and transiently expressed into *Nicotiana benthamiana* leaf epidermal cells by agroinfiltration, formed punctate inclusions on the cytomembrane, and co-located with a PD marker at the PD in *N. benthamiana*. The expression of GmGLU in soybean leaves inoculated with SMV was higher than that in leaves mock-inoculated with PBS at 48 and 72 h post-infiltration. Furthermore, we demonstrated that overexpression of soybean *GLU* can decrease callose deposition at the PD which was induced by SMV. We also observed that the viral spread speed and genomic RNA accumulation significantly increased when coexpression GmGLU in *N. benthamiana* leaves. Our data support that GmGLU plays an important role in viral infection.

TABLE 1 | Primer sequences used for cloning GLU genes.

Name	Sequence (5' → 3') ^a
SMV-P3-F	ATTAACAAGGCCATTACGGCCGGGGAAGTGCAACAAAGGATGA
SMV-P3-R	AACTGATTGGCCGAGGCGGCCCTGTGCGGAGACATCTTCTGA
GmGLU-F	GGGGACAAGTTTGTACAAAAAGCAGGCTTCATGATTTTCTCAA GAGGCAAC
GmGLU-R	GGGGACCACTTTGTACAAAGAAAGCTGGGTCAATTGAACTGAG TTGGTATTAGGT

^a The restriction enzyme sequences or gateway sequences are underlined.

MATERIALS AND METHODS

Plants, Viral Strains, and Plasmids

Seeds of *Glycine max* 'Nannong 1138-2' were obtained from the Soybean Improvement Center of Nanjing Agricultural University (Nanjing, China). Leaves of 2-weeks-old 'Nannong 1138-2' plants were inoculated via a point inoculation approach or direct rubbing by a brush or toothpick with SMV, isolate 6067-1 (strain SC15) (GenBank Accession: JF833015.1) in controlled environment chambers at 22°C with a 16-h light/8-h dark photoperiod. The *N. benthamiana* seedlings were grown in growth chambers under a 16-h light/8-h dark photoperiod at 22 to 24°C. *Escherichia coli* strain DH5a, *Saccharomyces cerevisiae* strain NMY51 (Dualsystems Biotech, Zurich, Switzerland), yeast expression vectors (pPR3-N, and pBT3-STE) (Dualsystems Biotech), and the pPR3-gateway vector (modified from pPR3-N) were used in this study. All vectors were constructed using Gateway technology. *Agrobacterium tumefaciens* strain EHA105 and the fluorescence expression vectors pEarleyGate104, pEarleyGate201-YC, pEarleyGate202-YN and the plant expression vector pEarleyGate201 were stored at the Jiangsu Province Academy of Agricultural Sciences (Nanjing, China).

Gene Cloning and Plasmid Construction

Gateway technology (Invitrogen, Burlington, ON, Canada) was used to generate all plasmid clones used in this work. Gene sequences were amplified by PCR using PrimeSTAR HS DNA Polymerase (Takara, Dalian, China). The resulting DNA fragments were purified and transferred by recombination into the entry vector pDONR221 (Invitrogen) using BP clonase II (Invitrogen) following the supplier's recommendations. Insertions in the resulting pDONR clones were verified by DNA sequencing. The destination vectors, yeast expression vector (pPR3-gateway) or fluorescence expression vectors (pEarleyGate104, pEarleyGate201-YC, and pEarleyGate202-YN), or plant expression vector (pEarleyGate201) were produced using LR clonase II (Invitrogen) according to the conditions and procedures recommended by the supplier. The primer sequences used in this study are listed in **Table 1**.

mY2H Analysis

The mY2H tests were performed using the Matchmaker DUAL membrane system (Clontech, Palo Alto, CA, United States)

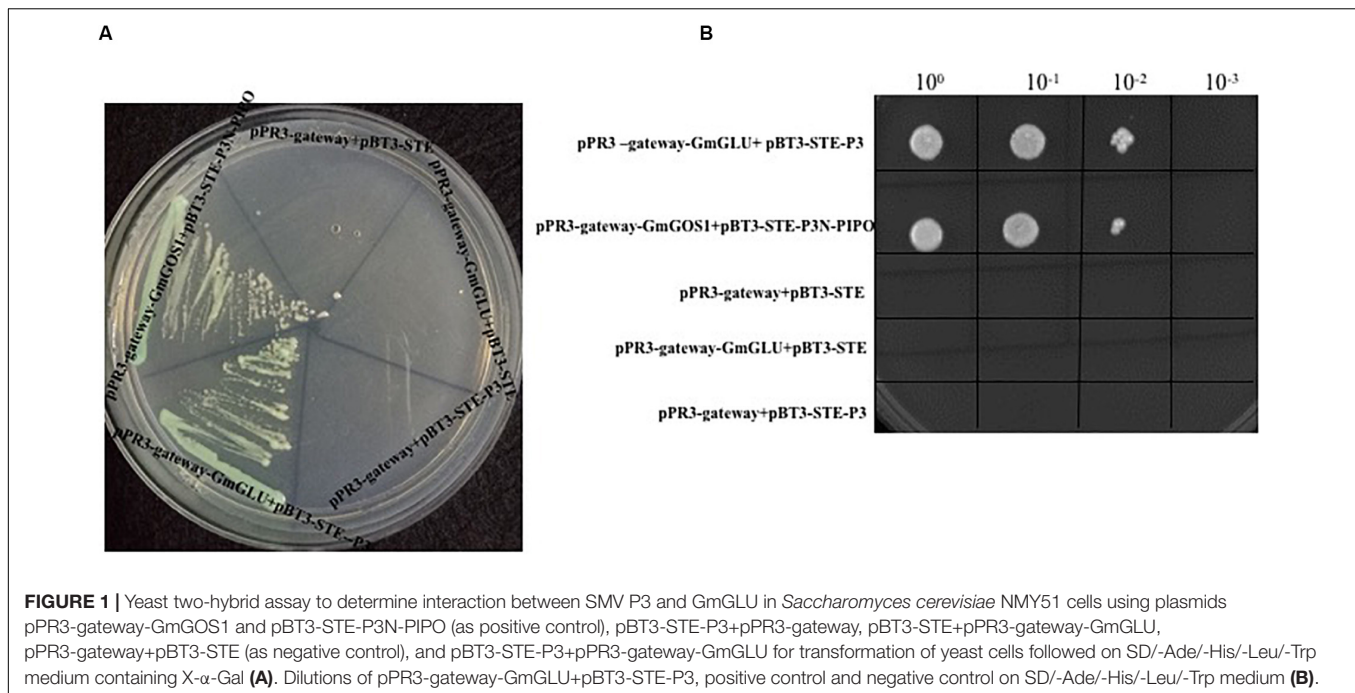
according to the manufacturer's protocol. The DUAL membrane system was developed to identify and characterize protein-protein interactions between integral membrane proteins, membrane-associated proteins, and soluble proteins in their natural condition. The fusion vector pPR3-gateway-GmGLU was constructed by inserting the GmGLU gene into the pPR3-gateway vector. pPR3-gateway-GmGLU and pBT3-STE-P3 (constructed previously) were co-transformed into yeast cells using the small-scale lithium acetate transformation method. The yeast cells were then grown on SD/-Leu/-Trp and SD/-Ade/-His/-Leu/-Trp medium for 3–7 days at 30°C. Colonies growing on SD/-Ade/-His/-Leu/-Trp medium were transferred to SD/-Ade/-His/-Leu/-Trp/X-α-Gal solid medium and grown for 3–5 days at 30°C. Blue colonies were considered as positive clones. Yeast cells co-transformed with empty pBT3-STE-P3+pPR3-gateway, pBT3-STE+pPR3-gateway-GmGLU, and pBT3-STE+pPR3-gateway were used as a negative control, and those co-transformed with pBT3-STE-P3N-PIPO+pPR3-gateway-GmGOS1 were used as a positive control (Song et al., 2016). The experiments were repeated at least three times.

Agroinfiltration in *N. benthamiana*

All experiments were performed using *N. benthamiana*, an experimental host of SMV. For the BiFC assay, the *A. tumefaciens* strain EHA105 was transformed with pEarleyGate202-YN-GmGLU and pEarleyGate201-YC-P3 via cryoapplication. For the subcellular localization assay or overexpression, *A. tumefaciens* strain EHA105 was transformed with pEarleyGate104-GmGLU or pEarleyGate201-GmGLU, then cultured overnight with appropriate antibiotics at 28°C (Sparkes et al., 2006). For intercellular movement analysis, the infectious clone TuMV-GFP was described in Cui et al. (2017). The *Agrobacterium* cells were diluted to an optical density (OD₆₀₀) of 0.5–0.6 for BiFC experiments (Tian et al., 2011) and 0.1–0.2 for subcellular expression and 0.001 for intercellular movement analyses. The diluted *A. tumefaciens* was injected into *N. benthamiana* and fluorescence was observed by confocal microscopy. Each BiFC experiment, subcellular localization assay and intercellular movement analysis was repeated at least three times, and each treatment included at least three independent *N. benthamiana* plants.

RNA Extraction and RT-PCR

RNA was isolated from 50 to 100 mg of frozen soybean leaf tissue using the total RNA isolation kit according to the manufacturer's instructions (TIANGEN, Nanjing, China). The quality and quantity of RNA were measured with a nucleic acid and protein detection instrument. Reverse transcription (RT) and first-strand cDNA synthesis were carried out using Fast-RT SuperMix (TIANGEN, Nanjing, China). The primer pair SMV-CP-F (5'-ATGGTTGAGAGGGAAAGATCTATCTC-3') and SMV-CP-R (5'-AGAGCTGCAGCCTTCATCTTG-3') was used to detect soybean genomic RNA. An SMV coat protein gene DNA fragment at the expected size of 500 bp was amplified as follows: initial denaturation at 94°C for 2 min followed by 30 cycles of denaturation at 94°C for 0.5 min, annealing at 55°C for



0.5 min and elongation at 72°C for 1 min. The final extension step was at 72°C for 10 min.

Quantitative Reverse Transcription PCR

Quantitative reverse transcription PCR was carried out in a LightCycler 480 (Roche, Penzberg, Germany) using SYBR® Green I (Takara, Dalian, China). The primer pair GmGLU-F (5'-TATTTAGGTGATTTGTCAGGCC-3') and GmGLU-R (5'-CCGCAATTTGATTAATCATGCC-3') was used to detect soybean genomic RNA, and the primer pair GmTubulin-F (5'-CGAGTTCACAGAGGCAGAG-3') and GmTubulin-R (5'-CACTTACGCATCACATAGCA-3') was used to detect the reference gene Tubulin in *G. max*, which was used for normalization. The primer pair TuCP-F (5'-GGCACTCAAGAAAGGCAAGG-3') and TuCP-R (5'-CTCCGTCAGTTCGTAATCAGC-3') was used for detection of TuMV viral genomic RNA, and the primers NbActin-F (5'-GGGATGTGAAGGAGAAGTTGGC-3') and NbActin-R (5'-ATCAGCAATGCCCGGAACA-3') for the reference gene Actin in *N. benthamiana* were used for normalization. The final reaction mixtures contained 250 nM of each primer in a volume of 20 μL. All genes were amplified under the same reaction conditions: 30 s at 95°C, followed by 40 cycles of 10 s at 95°C, 30 s at 53°C, and 34 s at 72°C. The specificity of amplification was verified using a melting curve analysis (60–95°C). qRT-PCRs with no template control were performed under the same conditions. All reactions were performed in triplicate, and the mean value was calculated. Cycle threshold (CT) values were determined using LightCycler 480 Real-Time PCR software (Roche, Penzberg, Germany) and were used for gene expression analysis. Bar charts and error lines were created using OriginLab OriginPro 8.5.

Aniline Blue Staining

Callose was stained by a modification of the aniline blue fluorochrome method (Conrath et al., 1998; Li et al., 2012). In brief, sections of about 1 cm × 1 cm were cut at different time points after inoculation from unifoliate leaves, which do not have large veins. Chlorophyll was cleared by heating the samples at 55°C in lactophenol (water: glycerol: phenol: lactic acid, 1:2:2:1 by volume) diluted with two volumes of ethanol for 30 min. The cleared leaf samples were stained in the dark with 0.1% aniline blue (Macklin, China) in 0.15M K₂HPO₄ buffer, pH 9.0, for 30–60 min. The leaves were observed under a confocal microscope. The experiment was repeated at least three times, and each treatment included at least three independent 'Nannong 1138-2' plants.

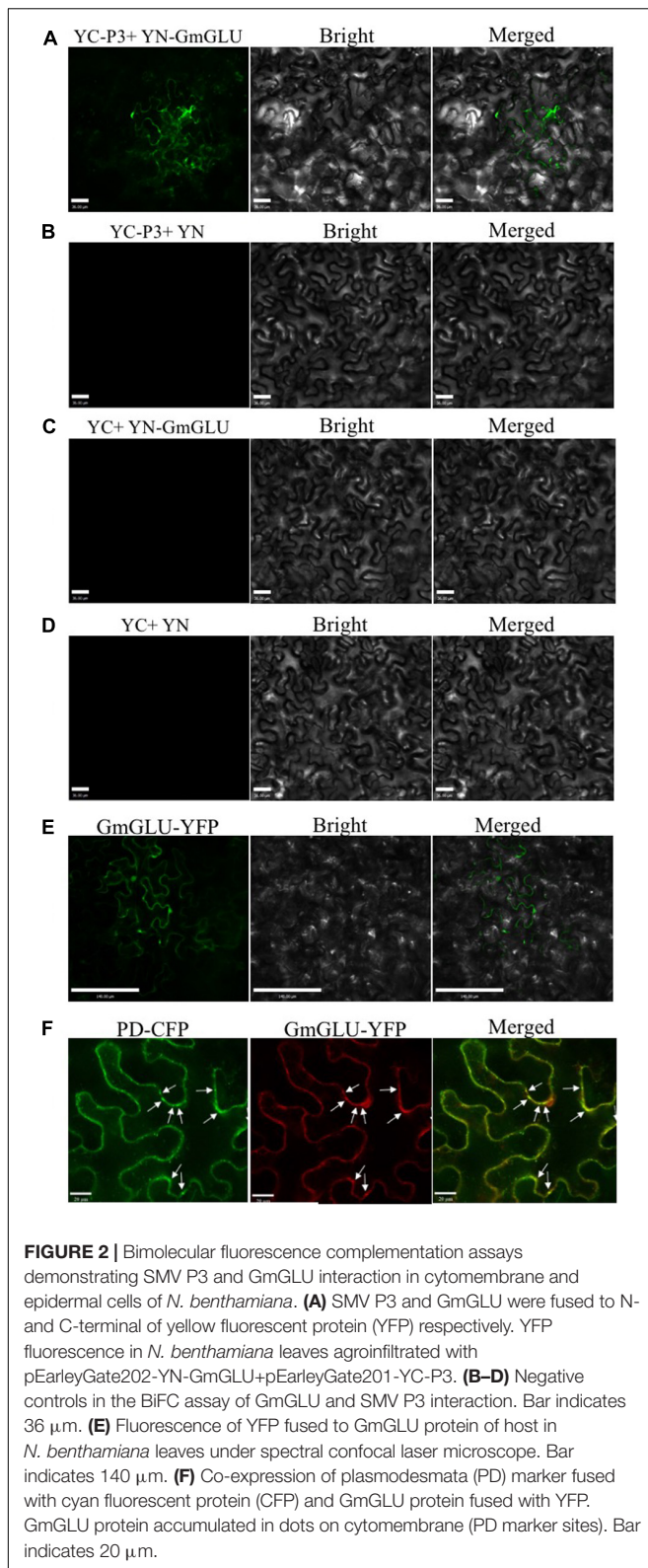
Confocal Microscopy

Epidermal cells of *N. benthamiana* were examined with Ultra View VOX inverted confocal microscope. For confocal microscopy, CFP was excited at 458 nm, and the emitted light was captured at 440 to 470 nm; YFP was excited at 514 nm, and the emitted light was captured at 525 to 650 nm; GFP was excited at 488 nm and the emitted light was captured at 505 to 555 nm. Post-acquisition images were processed by Adobe Photoshop software.

RESULTS

Interaction Between GmGLU and SMV-P3 in Y2HS

In a mY2HS analysis using the SMV-P3 protein as bait and a soybean cDNA library as prey, one candidate gene was



identified and shared a 94.65% nucleotide sequence identity with the *G. max* endo-1,3- β -glucanase (AY461847.1). Thus, this isolated soybean factor was named GmGLU. The gene

encoding SMV-P3 was inserted into pBT3-STE and the complete cDNA of soybean GmGLU (Accession Number: AY461847.1) was inserted into pPR3-gateway. The clones harboring pPR3-gateway-GmGLU+pBT3-STE-P3 grew well on SD/-Ade/-His/-Leu/-Trp/X- α -Gal medium (**Figure 1A**). Negative control P3N-PIPO cannot interact with GmGLU (**Supplementary Figure S1**). The positive control harboring pPR3-gateway-GmGOS1+pBT3-STE-P3N-PIPO also grew well on the selective medium, while the negative controls harboring pPR3-gateway-GmGLU+pBT3-STE/pPR3-gateway+pBT3-STE-P3 and pPR3-gateway+pBT3-STE did not grow on the selective medium. Different dilutions of the cells harboring pPR3-gateway-GmGLU+pBT3-STE-P3 grew as well as the positive control on SD/-Ade/-His/-Leu/-Trp medium (**Figure 1B**).

Confirmation of Interaction by BiFC

To further confirm the interaction between SMV-P3 and GmGLU *in vivo*, we recombined GmGLU into pEarlygate202-YN to create a fusion construct containing the N-terminal fragment of yellow fluorescent protein (YFP) and recombined SMV-P3 into pEarlygate201-YC to create a fusion construct containing the C-terminal fragment of YFP. These two vectors were co-transformed into *N. benthamiana* via *Agrobacterium*-mediated transformation, allowing the transient expression of fusion proteins. The green fluorescence indicative of the P3 and GmGLU combination was observed on the cytomembrane at 48 hpi (**Figure 2A**). No BiFC fluorescence was detected in the negative control samples expressing combinations with non-hybrid YC or/and YN (**Figures 2B–D**). These results confirmed that GmGLU could interact with SMV-P3 *in vivo*.

Subcellular Localization of GmGLU in *N. benthamiana* Cells

To explore the possible functions of GmGLU, we analyzed its subcellular localization in *N. benthamiana* leaf epidermal cells. The gene was fused with YFP to form pEarlyGate104-GmGLU and agroinfiltrated into *N. benthamiana*. Using fluorescence microscopy at 48 hpi, the GmGLU protein fluoresced on the cytomembrane and formed small dots (**Figure 2E**). To investigate that where is small dots, the PD marker fused with cyan fluorescent protein (Wei et al., 2010; Park et al., 2017) and GmGLU protein fused with YFP coexpress in *N. benthamiana* leaf epidermal cells. The punctate inclusions on the cytomembrane from GmGLU were observed to overlap with PD marker on the points (**Figure 2F**). These results indicated that GmGLU was located at the PD site in leaf cells of *N. benthamiana*.

SMV Infection Induced the Overexpression of GmGLU

To investigate whether GmGLU is involved in response to SMV infection, we inoculated SMV onto the soybean variety 'Nannong 1138-2' and then detected the expression level of GmGLU at time-course after infection. The expression level of GmGLU in soybean leaves inoculated with SMV were significantly higher

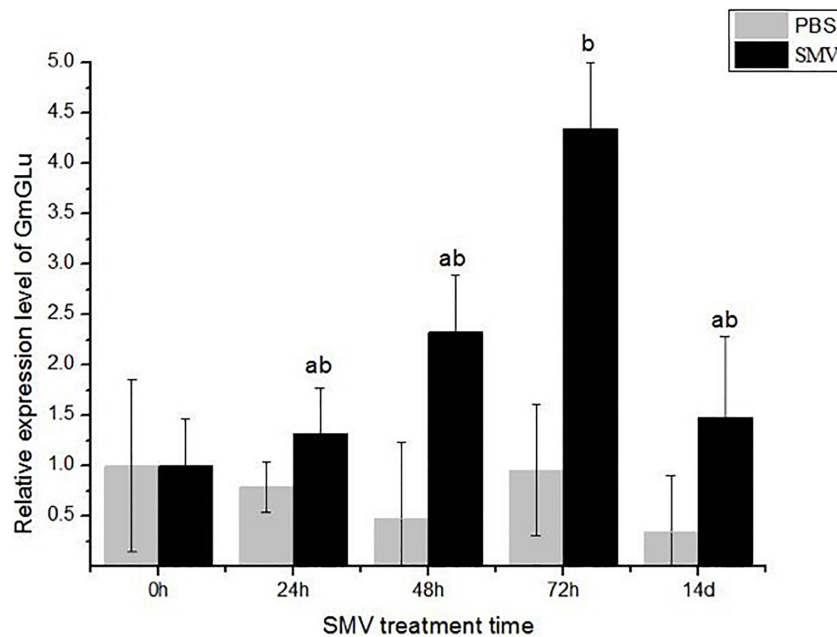


FIGURE 3 | qRT-PCR analysis of GmGLU expression levels. Y-axes show relative expression levels of genes in SMV-infected and mock (PBS)-infected samples at different time points. Each error bar represents mean $\pm$ SD. Alphabetic symbols represent statistical significance: ^{ab}0.01 < $P \leq 0.05$; ^b $P \leq 0.01$.

than those in leaves mock-inoculated with PBS at 24, 48, 72 hpi and 14 dpi (**Figure 3**). Additionally, a significant difference ($P < 0.01$) was found between the leaves treatment by SMV and PBS at 72 hpi (**Figure 3**). These findings confirmed that the infection of SMV induced the overexpression of GmGLU in soybean leaves.

SMV Infection Induces Callose Deposition

Callose can modulate the PD permeability, and the deposition of callose is one of the critical defense responses of host plants against viral infection. In order to detect the effect of SMV infection on callose deposition, we inoculated SMV into the soybean variety 'Nannong 1138-2' and then detected the deposition of callose 14 days after infection. PBS treatment was used as a negative control. Many points with blue fluorescence, unevenly distributed on the cell membrane, were observed by confocal microscopy (**Figures 4A,C**). The control group had a few points of callose deposition, caused by friction (**Figures 4B,C**). The accumulation of callose was quantified with the software ImageJ and was shown in **Figure 4C**. These results suggested that SMV infection can significantly induce callose deposition.

GmGLU Decreases Callose Deposition at the PD

To further examine whether the soybean GmGLU plays a role in callose deposition at the PD, we inoculated SMV by toothpick on the leaves of soybean, where have been agroinfiltrated with GmGLU 24 h in advance. The leaves without agroinfiltration

of GmGLU were used as control. After 24 h, we observed there was no callose deposition on soybean leaves with GmGLU overexpression (**Figures 5B,F**), whereas large callose depositions were found on soybean leaves which were only infected by SMV (**Figures 5A,F**). We therefore speculated that GmGLU inhibited the deposition of callose.

To further confirm this hypothesis, we observed callose deposition through overexpressing GmGLU into SMV-infected leaves. At 12 h post-SMV inoculation, a few fluorescence spots resulting from callose formation were found by staining with aniline blue (**Figures 5C,F**). We then transfected GmGLU into the SMV-infected leaves. At 36 h post-SMV inoculation, we found that the fluorescence formed by callose at the inoculation point disappeared following the transfection of GmGLU (**Figures 5E,F**), while there was considerable callose deposition on soybean leaves which were only infected with SMV (**Figures 5D,F**). No fluorescent spots produced by callose deposition was detected on leaves only inoculated with GmGLU and healthy plant leaves (**Supplementary Figures S2A,B**). This result suggests that GmGLU could degrade callose. Further, the viral RNA on the leaves of the inoculation was detected by RT-PCR analysis to prove that the plants were indeed infected by SMV (**Figures 5G,H**). These results indicated that the overexpression of soybean GmGLU decreases callose deposition induced by SMV infection at the PD.

Overexpression of GmGLU Promotes Viral Proliferation

Systemic viral spread was monitored with a recombinant form of TuMV-GFP (Cui et al., 2017). The infectious

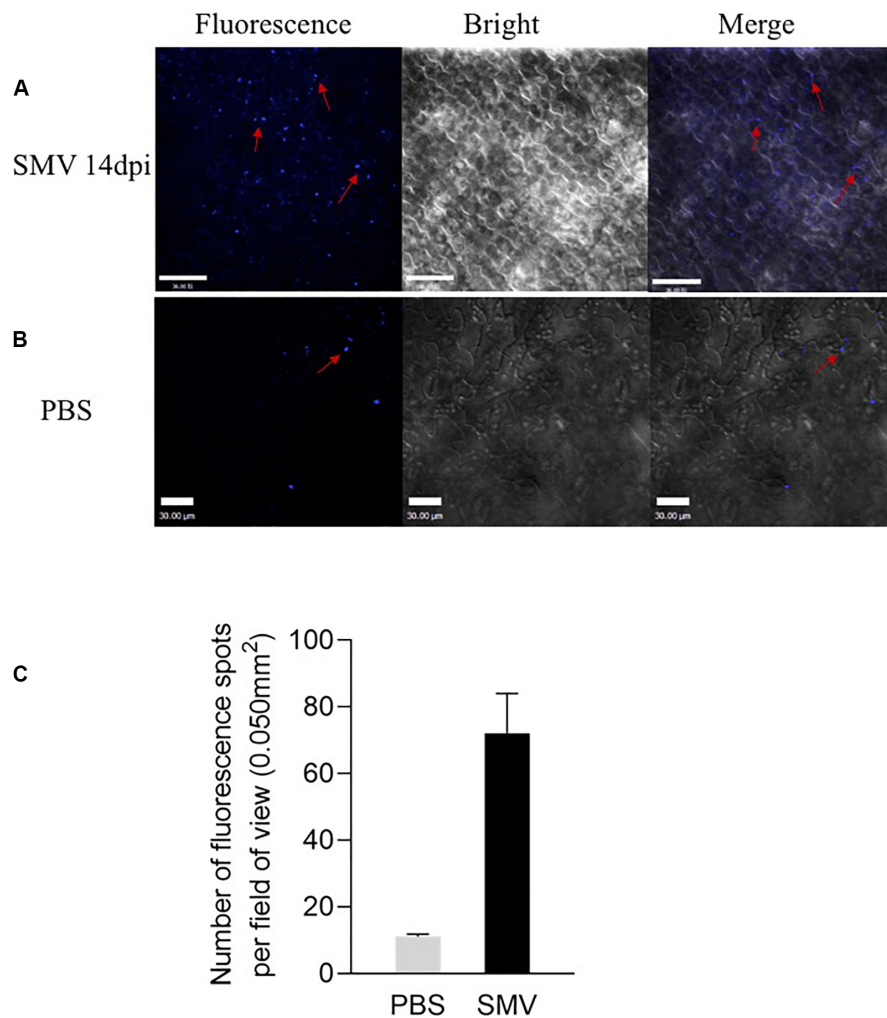


FIGURE 4 | Fluorescence spots at the inoculation sites detected from soybean cv. Nannong 1138-2 leaves inoculated with SMV strain SC-15. **(A)** Lots of fluorescence spots were detected as uneven distribution on the cell membrane 14 days post-inoculation. **(B)** A few of callose deposition, caused by friction was observed as negative control. Bar indicates 30 μm . **(C)** Number of fluorescence spots per field of view (0.050 mm^2), as indicated callose deposition in systemically infected leaves at 14 dpi. Three plants were examined in each treatment, two small areas of each plant were inoculated, and five fields of view are randomly observed in each area. Error bars represent the standard deviation from 30 fields of view and data were from three independent experiments.

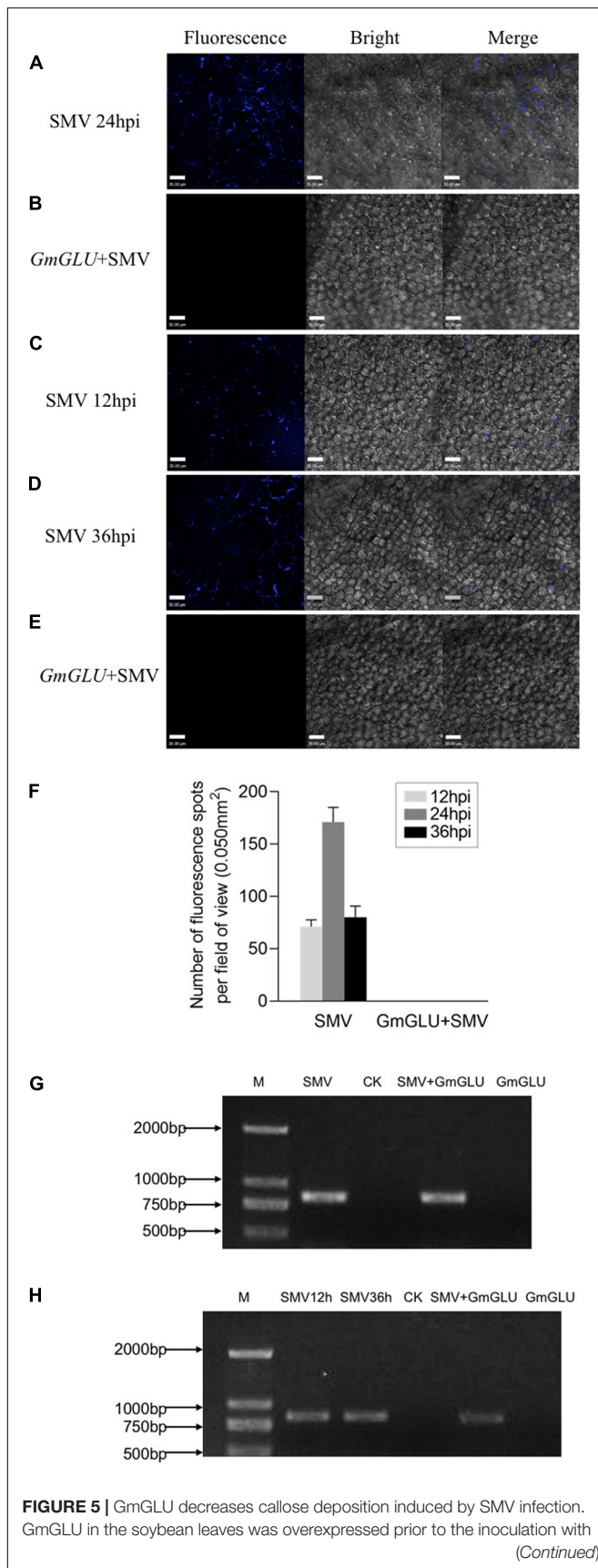
clone TuMV-GFP and pEarleyGate201-GmGLU were agroinfiltrated into *N. benthamiana* leaf cells, and the spread of fluorescence due to TuMV-GFP was monitored 3.5 days later. As **Figure 6B** shows, leaves co-infected with TuMV-GFP/GmGLU-201 showed a cluster of virus-infected cells, while sporadic luminescence occurred at the site of the initial infection in *N. benthamiana* leaves agroinfiltrated with the control TuMV-GFP (**Figure 6A**). The virus spread significantly increased of the GLU-treated plants compared with the control (**Figure 6B**). These results suggested that the movement of viruses is facilitated in the plant with GmGLU overexpression.

To test if GmGLU overexpression affects virus accumulation, we carried out qRT-PCR assay. *N. benthamiana* leaves were inoculated with the infectious clone TuMV-GFP and GmGLU-201, qRT-PCR was performed to quantify TuMV genomic

RNA 48 and 72 h post-transfection. We found that GmGLU overexpression significantly promotes TuMV genomic RNA accumulation compared with control leaves which were only agroinfiltrated by TuMV-GFP (**Figure 6C**). Taken together, these results indicated that GmGLU overexpression promotes viral cell-to-cell movement and accumulation.

DISCUSSION

Soybean mosaic virus is composed of a single-stranded, positive-sense RNA genome and coat protein (CP) that protects nucleic acid. Since viruses lack their own reproduction system, they achieve systemic infection by interacting with host proteins by taking advantage of the host's cellular mechanisms for replication and intracellular/intercellular movement (Gergerich

**FIGURE 5 |** Continued

SMV. **(A)** Confocal microscopic observation. There was a large amount of callose deposition on soybean leaves which were only infected by SMV. **(B)** There was no callose deposition on soybean leaves with overexpression GmGLU. We observed callose deposition through overexpressing GmGLU into soybean leaves which inoculated by SMV. **(C,D)** There were many the fluorescence spots formed by callose at 12 h and 36 post-SMV inoculation. **(E)** The fluorescence formed by callose at the inoculation point were disappeared following the transfection of GmGLU. Bar indicates 30 μ m. **(F)** Number of fluorescence spots per field of view (0.050 mm²), as indicated callose deposition in primarily infected leaves. Three plants were examined in each treatment, two small areas of each plant were inoculated, and five fields of view are randomly observed in each area. Error bars represent the standard deviation from 30 fields of view and data were from three independent experiments. **(G,H)** Detection of viral CP-RNA by RT-PCR analysis in the upper leaves. The viral CP-RNA was detected in the SMV-infected soybean cv. Nannong 1138-2 (G: lane 1, lane 3; H: lane 1, lane 2, lane 4). The viral CP-RNA was not detected from the negative control and GLU-treatment (G: lane 2, lane 4; H: lane 3, lane 5). Independent experiments were performed three times.

and Dolja, 2006). If the virus loses the ability to move, infection will be limited to an initial infection site or infected organ (Hong and Ju, 2017).

Previous reports indicated that the successful infection of plants by most viruses depends on cell-to-cell movement via PD and rapid long distance movement through the sieve tubes of the phloem (Lucas, 1995; Van Bel, 2003; Lucas and Lee, 2004; Oparka, 2004). For plant viruses, the ability of crossing cell walls through PD into adjacent healthy cells is necessary to establish systemic infection (Carrington et al., 1996; Seron and Haenni, 1996). However, virions are too large to move through the PD, and thus transport through PD is usually limited to small molecules less than 1 kDa in size (Lucas, 1995). To overcome this limitation, most viruses encode MPs, which interact with host proteins to increase the SEL of the PD and help virions pass through PD into adjacent cells (Lucas, 1995; Iglesias and Meins, 2000). Our results suggest that GLU interacts with P3 to reduce the SEL, which can play a role in potyviral intercellular spread in infected plants.

Callose is a linear 1,3- β -glucan playing an important role on the plant allergy defense response (Piršelová and Matušíková, 2013). The callose content at PD determines the aperture of the PD channel and thus controls cell-to-cell transport of macromolecules (Bucher et al., 2001; Levy and Epel, 2009; Zavaliev et al., 2011). Once the accumulated amount of callose is down-regulated, it will cause PD to relax and allow more macromolecules trafficking (Beffa et al., 1996; Iglesias and Meins, 2000). The net callose level at PD is determined by the balance between the activity of callose synthase and 1,3- β -glucanase, because loss of either enzyme activity affects callose accumulation and transport through PD (Guseman et al., 2010; Vaten et al., 2011). The accumulation of callose in susceptible hosts is very low or similar to uninfected plants (Krasavina et al., 2002; Zavaliev et al., 2011). In this study, we explored the mechanism by which GLU degrades callose by targeting P3, and our results provide evidence for the regulatory effect of GmGLU on callose under SMV infection.

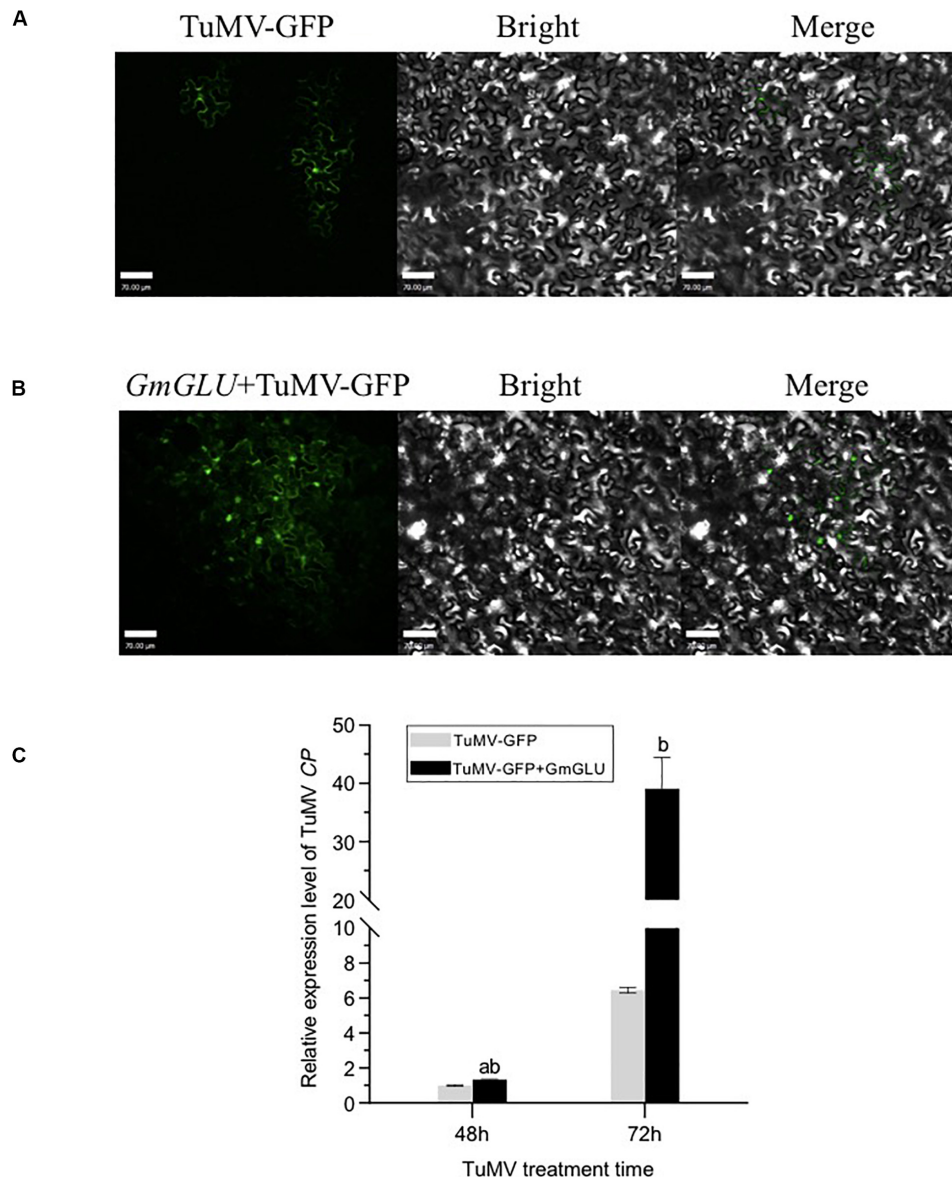


FIGURE 6 | Expression of GmGLU promote the viral proliferation. **(A,B)** Confocal microscopic observation of viral intercellular movement of TuMV-GFP, TuMV-GFP/GmGLU-201. Images were taken 3.5 days post-agroinfiltration. Independent experiments were performed three times. Typical imaging was shown. Bar indicates 70 μ m. **(C)** qRT-PCR detection of the viral accumulation level in *N. benthamiana* leaves transfected with TuMV-GFP, TuMV-GFP/GmGLU-201. Leaves were collected at 48 h post-transfection. Viral RNA was quantified by qRT-PCR analysis of the CP RNA level using the *N. benthamiana* ACTIN gene transcripts as the internal control. TuMV-GFP genomic RNA level was normalized to 1. Statistical differences from three biological replicates, determined by unpaired two-tailed Student's test, are indicated by ab and b as follows: ^{ab}0.001 < $P \leq 0.01$; ^b $P \leq 0.001$. Each error bar represents mean $\pm$ SD.

In this study, GmGLU interacted with SMV-P3 in a screening of soybean cDNAs using mY2HS. The expression of *GmGLU* in soybean leaves inoculated with SMV was increased. GLU is a PR protein, located on the cytomembrane and forms the dots at plasmodesmata sites, playing important roles in plant defense and development (Leubner-Metzger and Meins, 1999; Wan et al., 2011). GLU are cell wall enzymes that catalyze the callose and enhances viral transmission by degrading callose on PD (Deom et al., 1992; Zavaliev et al., 2011). The role of the GLU in virus spread has

been described in TMV (Bucher et al., 2001; Gregor et al., 2001). Infection of tobacco with TMV led to a 20-fold increase in class I GLU levels (Vogeli-Lange et al., 1988; Epel, 2009). Furthermore, by combining SMV infection with aniline blue staining, we demonstrate that callose deposition induced by SMV infection was decreased when soybean *GLU* is overexpressed. The viral spread speed and genomic RNA accumulation significantly increased under overexpression of *GmGLU*. This conclusion is consistent with the research of Bucher et al. (2001).

It is widely believed that the deposition of callose is a critical defense response of host plants against viral infection (Li et al., 2012). These data suggest that GmGLU regulate the callose deposition at the PD to allow viral intercellular trafficking in viral infection by interacting with SMV-P3.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at NCBI (AY461847.1).

AUTHOR CONTRIBUTIONS

XYC conceived the project. XYC and XC supervised the work. FS and YW performed most of the experiments with assistance from FZ, XY, HC, and XHC. YW, FS, and XYC wrote the manuscript with contributions from all the authors. All authors analyzed the data.

FUNDING

This work was supported by the National Natural Science Foundation of China (31101411), National Key Research and Development Project (2018YFE0112200 and 2016YFE0203800), the Key R&D Project of Jiangsu Province (BE2019376), and Jiangsu Agricultural Industry Technology System [JATS(2019)399]. The research was

carried out in the joint lab of China-AAFC for legume research in JAAS.

ACKNOWLEDGMENTS

We thank Dr. Aiming Wang (Agriculture and Agri-Food Canada) for providing the pPR3-N and pBT3-STE vectors and Dr. Yuhai Cui (Agriculture and Agri-Food Canada) for providing pEarleyGate202-YN and pEarleyGate201-YC vectors.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.536771/full#supplementary-material>

FIGURE S1 | Yeast two-hybrid assay used to determine that SMV P3N-PIPO and GmGLU cannot interact in *Saccharomyces cerevisiae* NMY51 cells using plasmids pPR3-gateway-GmGLU+pBT3-STE-P3 (1) and pPR3-gateway-GmGOS1+pBT3-STE-P3N-PIPO (2) as positive control, pPR3-gateway+pBT3-STE-P3 (4) and pPR3-gateway+pBT3-STE-P3N-PIPO (5) and pPR3-gateway-GmGLU+pBT3-STE (6) and pPR3-gateway+pBT3-STE (7) as negative control, pPR3-gateway-GmGLU+pBT3-STE-P3N-PIPO (3) for transformation of yeast cells followed on SD/-Ade/-His/-Leu/-Trp medium containing X- α -Gal. In addition, negative controls 4–7 and pPR3-gateway-GmGLU+pBT3-STE-P3N-PIPO could not grow on the selective medium.

FIGURE S2 | No fluorescent spots produced by callose deposition were detected on leaves only inoculated with GmGLU and healthy plant leaves. **(A)** No callose deposits can be detected on leaves overexpressing GmGLU. **(B)** No callose deposits can be detected on healthy plant leaves. Bar indicates 30 μ m.

REFERENCES

- Allison, A. V., and Shalla, T. A. (1974). The ultrastructure of local lesions induced by potato virus X: a sequence of cytological events in the course of infection. *Phytopathology*, 64, 784–793. doi: 10.1094/Phyto-64-784
- Balasubramanian, V., Vashisht, D., Cletus, J., and Sakthivel, N. (2012). Plant β -1,3-glucanases: their biological functions and transgenic expression against phytopathogenic fungi. *Biotechnol. Lett.* 34, 1983–1990. doi: 10.1007/s10529-012-1012-6
- Beffa, R. S., Hofer, R. M., Thomas, M., and Meins, F. Jr. (1996). Decreased susceptibility to viral disease of beta-1,3-glucanase-deficient plants generated by antisense transformation. *Plant Cell* 8, 1001–1011. doi: 10.2307/3870211
- Bucher, G. L., Tarina, C., Heinlein, M., Serio, F. D., Meins, F. Jr., and Iglesias, V. A. (2001). Local expression of β -1,3-glucanase enhances symptoms of TMV infection in tobacco. *Plant J.* 28, 361–369. doi: 10.1046/j.1365-3113x.2001.01181.x
- Carrington, J. C., Kasschau, K. D., Mahajan, S. K., and Schaad, M. C. (1996). Cell-to-cell and long-distance transport of viruses in plants. *Plant Cell* 8, 1669–1681. doi: 10.1105/tpc.8.10.1669
- Chai, M. Z., Wu, X. Y., Liu, J. H., Fang, Y., Luan, Y. M., Cui, X. Y., et al. (2020). P3N-PIPO interacts with P3 via the shared N-terminal domain to recruit viral replication vesicles for cell-to-cell movement. *J. Virol.* 94:e01898-19. doi: 10.1128/JVI.01898-19
- Chowda-Reddy, R. V., Sun, H. Y., Chen, H. Y., Poysa, V., Ling, H., Gijzen, M., et al. (2010). Mutations in the P3 protein of *Soybean mosaic virus* G2 isolates determine virulence on Rsv4-genotype *Soybean*. *Mol. Plant Microbe Interact.* 24, 37–43. doi: 10.1094/MPMI-07-10-0158
- Chung, B. Y., Miller, W. A., Atkins, J. F., and Firth, A. E. (2008). An overlapping essential gene in the Potyviridae. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5897–5902. doi: 10.1073/pnas.0800468105
- Conrath, U., Klessig, D. F., and Bachmair, A. (1998). Tobacco plants perturbed in the ubiquitin-dependent protein degradation system accumulate callose, salicylic acid, and pathogenesis-related protein. *Plant Cell Rep.* 17, 876–880. doi: 10.1007/s002990050501
- Cui, X. Y., Lu, L., Wang, Y., Yuan, X. X., and Chen, X. (2018). The interaction of *Soybean reticulon* homology domain protein (GmRHP) with *Soybean mosaic virus* encoded P3 contributes to the viral infection. *Biochem. Biophys. Res. Commun.* 495, 2105–2110. doi: 10.1016/j.bbrc.2017.12.043
- Cui, X. Y., Wei, T. Y., Chowda-Reddy, R. N., Sun, G. Y., and Wang, A. M. (2010). The *Tobacco etch virus* P3 protein forms mobile inclusions via the early secretory pathway and traffics along actin microfilaments. *Virology* 397, 56–63. doi: 10.1016/j.virol.2009.11.015
- Cui, X. Y., Yaghmaiean, H., Wu, G. W., Wu, X. Y., Chen, X., Thornc, G., et al. (2017). The C-terminal region of the *Turnip mosaic virus* P3 protein is essential for viral infection via targeting P3 to the viral replication complex. *Virology* 510, 147–155. doi: 10.1016/j.virol.2017.07.016
- Deom, C. M., Lapidot, M., and Beachy, R. N. (1992). Plant virus movement proteins. *Cell* 69, 221–224. doi: 10.1016/0092-8674(92)90403-Y
- Eiamtanassate, S., Juricek, M., and Yap, Y. K. (2007). C-terminal hydrophobic region leads PRSV P3 protein to endoplasmic reticulum. *Virus Genes* 35, 611–617. doi: 10.1007/s11262-007-0140-x
- Epel, B. L. (2009). Plant viruses spread by diffusion on ER-associated movement-protein-rafts through plasmodesmata gated by viral induced host β -1,3 glucanases. *Semin. Cell Dev. Biol.* 20, 1074–1081. doi: 10.1016/j.semcdb.2009.05.010

- Fernández-Crespo, E., Navarro, J., Serra-Soriano, M., Finiti, I., García-Agustín, P., Pallás, V., et al. (2017). Hexanoic acid treatment prevents systemic MNSV movement in *cucumis melo* plants by priming callose deposition correlating SA and OPDA accumulation. *Front. Plant Sci.* 8:1793. doi: 10.3389/fpls.2017.01793
- Gergerich, R., and Dolja, V. (2006). Introduction to plant viruses, the invisible foe. *Plant Health Instr.* doi: 10.1094/PHI-I-2006-0414-01
- Gregor, L. B., Corina, T., Manfred, H., Di Serio, F., Meins, F., and Iglesias, V. A. (2001). Local expression of enzymatically active class I β -1, 3-glucanase enhances symptoms of TMV infection in tobacco. *Plant J.* 28, 361–369. doi: 10.1046/j.1365-313X.2001.01181.x
- Guseman, J. M., Lee, J. S., Bogenschutz, N. L., Peterson, K. M., Virata, R. E., Xie, B., et al. (2010). Dysregulation of cell-to-cell connectivity and stomatal patterning by loss-of-function mutation in *Arabidopsis* chorus (glucan synthase-like 8). *Development* 137, 1731–1741. doi: 10.1242/dev.049197
- Hajimorad, M. R., Eggenberger, A. L., and Hill, J. H. (2006). Strain-specific P3 of *Soybean mosaic virus* elicits Rsv1-mediated extreme resistance, but absence of P3 elicitor function alone is insufficient for virulence on Rsv1-genotype *Soybean*. *Virology* 345, 156–166. doi: 10.1016/j.virol.2005.09.055
- Hajimorad, M. R., Eggenberger, A. L., and Hill, J. H. (2008). Adaptation of *Soybean mosaic virus* avirulent chimeras containing P3 sequences from virulent strains to Rsv1-genotype soybeans is mediated by mutations in HC-Pro. *Mol. Plant Microbe Interact.* 21, 937–946. doi: 10.1094/mpmi-21-7-0937
- Hajimorad, M. R., Wen, R. H., Eggenberger, A. L., Hill, J. H., and Saghai Maroof, M. A. (2011). Experimental adaptation of an RNA virus mimics natural evolution. *J. Virol.* 85, 2557–2564. doi: 10.1128/JVI.01935-10
- Hong, J. S., and Ju, H. J. (2017). The plant cellular systems for plant virus movement. *Plant Pathol.* 33, 213–228. doi: 10.5423/PPJ.RW.09.2016.0198
- Iglesias, V. A., and Meins, M. F. Jr. (2000). Movement of plant viruses is delayed in a β -1,3-glucanase-deficient mutant showing a reduced plasmodesmatal size exclusion limit and enhanced callose deposition. *Plant J.* 21, 157–166. doi: 10.1046/j.1365-313X.2000.00658.x
- Jenner, C. E., Wang, X., Tomimura, K., Ohshima, K., Ponz, F., and Walsh, J. A. (2003). The dual role of the Potyvirus P3 protein of *Turnip mosaic virus* as a symptom and avirulence determinant in Brassicas. *Mol. Plant Microbe Interact.* 16, 777–784. doi: 10.1094/MPMI.2003.16.9.777
- Johansen, I. E., Lund, O. S., Hjulsgaard, C. K., and Laursen, J. (2001). Recessive resistance in *Pisum sativum* and Potyvirus pathotype resolved in a gene-for-citron correspondence between host and virus. *J. Virol.* 75, 6609–6614. doi: 10.1128/JVI.75.14.6609-6614.2001
- Kauss, H. (1996). “Callose synthesis,” in *Membranes: Specialized Functions in Plants*, ed. M. Smallwood (Oxford: BIOS Press), 77–92.
- Kim, H. J., Kim, M. J., Pak, J. H., Im, H. H., Lee, D. H., Kim, K. H., et al. (2016). RNAi-mediated *Soybean mosaic virus* (SMV) resistance of a Korean *Soybean cultivar*. *Plant Biotechnol. Rep.* 10, 257–267. doi: 10.1007/s11816-016-0402-y
- Krasavina, M. S., Malysenko, S. I., Raldugina, G. N., Burmistrova, N. A., and Novos, A. V. (2002). Can salicylic acid affect the intercellular transport of the *Tobacco mosaic virus* by changing plasmodesmal permeability? *Russ. J. Plant Physiol.* 49, 71–77. doi: 10.1023/a:1013760227650
- Leubner-Metzger, G., and Meins, F. (1999). “Functions and regulation of plant β -1,3-glucanases (PR-2),” in *Pathogenesis-Related Proteins in Plants*, ed. S. K. Datta (Boca Raton, FL: CRC Press), 49–76.
- Levy, A., and Epel, B. L. (2009). “Cytology of the (1-3)-beta-glucan (callose) in plasmodesmata and sieve plate pores,” in *Chemistry, Biochemistry and Biology of (1-3)-beta-Glucans and Related Polysaccharides*, ed. A. Bacic (London: Academic Press), 439–463. doi: 10.1016/b978-0-12-373971-1.00013-3
- Li, W. L., Zhao, Y. S., Liu, C. J., Yao, G. B., Wu, S. S., Hou, C. Y., et al. (2012). Callose deposition at plasmodesmata is a critical factor in restricting the cell-to-cell movement of *Soybean mosaic virus*. *Plant Cell Rep.* 31, 905–916. doi: 10.1007/s00299-011-1211-y
- Lin, L., Luo, Z., Yan, F., Lu, Y., Zheng, H., and Chen, J. (2011). Interaction between potyvirus P3 and ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) of host plants. *Virus Genes* 43, 90–92. doi: 10.1007/s11262-011-0596-6
- Lu, L., Wu, G. W., Xu, X. M., Luan, H. X., Zhi, H. J., Cui, J., et al. (2015). *Soybean actin*-depolymerizing factor 2 interacts with *Soybean mosaic virus*-encoded P3 protein. *Virus Genes* 50, 333–339. doi: 10.1007/s11262-014-1150-0
- Luan, H. X., Liao, W. L., Niu, H. P., Cui, X. Y., Chen, X., and Zhi, H. J. (2019). Comprehensive analysis of *Soybean mosaic virus* P3 protein interactors and hypersensitive response-like lesion-inducing protein function. *Int. J. Mol. Sci.* 20:3388. doi: 10.3390/ijms20143388
- Luan, H. X., Shine, M. B., Cui, X. Y., Chen, X., Ma, N., Kachroo, P., et al. (2016). The Potyviral P3 protein targets eukaryotic elongation factor 1A to promote the unfolded protein response and viral pathogenesis. *Plant Physiol.* 172, 231–234. doi: 10.1104/pp.16.00505
- Lucas, W. J. (1995). Plasmodesmata-intercellular channels for macromolecular transport in plants. *Curr. Opin. Cell Biol.* 7, 673–680. doi: 10.1016/0955-0674(95)80109-X
- Lucas, W. J., Ding, B., and Van der Schoot, C. (1993). Plasmodesmata and the supracellular nature of plants. *New Phytol.* 125, 435–476. doi: 10.1111/j.1469-8137.1993.tb03897.x
- Lucas, W. J., and Lee, J. Y. (2004). Plasmodesmata as a supracellular control network in plants. *Nat. Rev. Mol. Cell Biol.* 5, 712–726. doi: 10.1038/nrm1470
- Mauch-Mani, B., and Mauch, F. (2005). The role of abscisic acid in plant–pathogen interactions. *Curr. Opin. Plant Biol.* 8, 409–414. doi: 10.1016/j.pbi.2005.05.015
- Oparka, K. J. (2004). Getting the message across: how do plant cells exchange macromolecular complexes? *Trends Plant Sci.* 9, 33–41. doi: 10.1016/j.tplants.2003.11.001
- Park, S. H., Li, F. F., Renaud, J., Shen, W. T., Li, Y. Z., Guo, L. H., et al. (2017). NbEXPA1, an α -expansin, is plasmodesmata-specific and a novel host factor for potyviral infection. *Plant J.* 92, 846–861. doi: 10.1111/tpj.13723
- Piršelová, B., and Matusíková, I. (2013). Callose: the plant cell wall polysaccharide with multiple biological functions. *Acta Physiol. Plant* 35, 635–644. doi: 10.1007/s11738-012-1103-y
- Radford, J. E., Vesk, M., and Overall, R. L. (1998). Callose deposition at plasmodesmata. *Protoplasma* 201, 30–37. doi: 10.1007/bf01280708
- Seron, K., and Haenni, A. L. (1996). Vascular movement of plant viruses. *Mol. Plant Microbe Interact.* 9, 435–442. doi: 10.1094/MPMI-9-0435
- Shukla, D. D., Ward, C. W., and Bunt, A. A. (1994). *The Potyviridae*. Wallingford: C.A.B. International.
- Song, P. W., Chen, X., Wu, B. Y., Gao, L., Zhi, H. J., and Cui, X. Y. (2016). Identification for soybean host factors interacting with P3N-PIPO protein of *Soybean mosaic virus*. *Acta Physiol. Plant* 38:131. doi: 10.1007/s11738-016-2126-6
- Sparkes, A., Kearns, R. J., and Hawes, A. C. (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat. Protoc.* 1, 2019–2025. doi: 10.1038/nprot.2006.286
- Stone, B. A., and Clarke, A. C. (1992). *Chemistry and Biology of 1,3- β -Glucans*. Bundoora, VI: La Trobe University Press.
- Suehiro, N., Natsuaki, T., Watanabe, T., and Okuda, S. (2004). An important determinant of the ability of *Turnip mosaic virus* to infect *Brassica* spp. and/or *Raphanus sativus* is in its P3 protein. *J. Gen. Virol.* 85, 2087–2098. doi: 10.1099/vir.0.79825-0
- Tian, G., Lu, Q., Zhang, L., Kohalmi, S. E., and Cui, Y. H. (2011). Detection of protein interactions in plant using a gateway compatible bimolecular fluorescence complementation (BiFC) system. *J. Vis. Exp.* 55, e3473. doi: 10.3791/3473
- Urcuqui-Inchima, S., Haenni, A. L., and Bernardi, F. (2001). Potyvirus proteins: a wealth of functions. *Virus Res.* 74, 157–175. doi: 10.1016/S0168-1702(01)00220-9
- Van Bel, A. J. E. (2003). The phloem, a miracle of ingenuity. *Plant Cell Environ.* 26, 125–149. doi: 10.1046/j.1365-3040.2003.00963.x
- Vaten, A., Dettmer, J., Wu, S., Stierhof, Y. D., Miyashima, S., Yadav, S. R., et al. (2011). Callose biosynthesis regulates symplastic trafficking during root development. *Dev. Cell* 21, 1144–1155. doi: 10.1016/j.devcel.2011.10.006
- Vogeli-Lange, R., Hansen-Gehri, A., Boller, T., and Meins, F. Jr. (1988). Induction of the defense-related glucanohydrolases, β -1,3-glucanase and chitinase, by *Tobacco mosaic virus* infection of tobacco leaves. *Plant Sci.* 54, 171–176. doi: 10.1016/0168-9452(88)90110-0
- Wan, L. L., Zha, W. J., Cheng, X. Y., Liu, C., Lv, L., Liu, C. X., et al. (2011). A rice-1,3-glucanase gene Osg1 is required for callose degradation in pollen development. *Planta* 233, 309–323. doi: 10.1007/s00425-010-1301-z
- Wang, Y., Khatibi, B., and Hajimorad, M. R. (2014). Amino acid substitution in P3 of *Soybean mosaic virus* to convert avirulence to virulence on Rsv4-genotype

- soybean is influenced by the genetic composition of P3. *Mol. Plant Pathol.* 16, 301–307. doi: 10.1111/mpp.12175
- Wei, T. Y., Zhang, C. W., Hong, J., Xiong, R. Y., Kasschau, K. D., Zhou, X. P., et al. (2010). Formation of complexes at plasmodesmata for potyvirus intercellular movement is mediated by the viral protein P3N-PIPO. *PLoS Pathog.* 6:e1000962. doi: 10.1371/journal.ppat.1000962
- Wen, R. H., and Hajimorad, M. R. (2010). Mutational analysis of the putative pipo of *Soybean mosaic virus* suggests disruption of PIPO protein impedes movement. *Virology* 400, 1–7. doi: 10.1016/j.virol.2010.01.022
- Wen, R. H., Khatabi, B., Ashfield, T., Saghai Maroof, M. A., and Hajimorad, M. R. (2012). The HC-Pro and P3 cistrons of an avirulent *Soybean mosaic virus* are recognized by different resistance genes at the complex Rsv1 locus. *Mol. Plant Microbe Interact.* 26, 203–215. doi: 10.1094/MPMI-06-12-0156-R
- Wen, R. H., Saghai Maroof, M. A., and Hajimorad, M. R. (2011). Amino acid changes in P3, and not the overlapping pipo-encoded protein, determine virulence of *Soybean mosaic virus* on functionally immune Rsv1-genotype soybean. *Mol. Plant Pathol.* 12, 799–807. doi: 10.1111/j.1364-3703.2011.00714.x
- Yang, Q. H., and Gai, J. Y. (2011). Identification, inheritance and gene mapping of resistance to a virulent *Soybean mosaic virus* strain SC15 in soybean. *Plant Breed.* 130, 128–132. doi: 10.1111/j.1439-0523.2010.01797.x
- Zavaliev, R., Ueki, S., Epel, B. L., and Citovsky, V. (2011). Biology of callose (β -1,3-glucan) turnover at plasmodesmata. *Protoplasma* 248, 117–130. doi: 10.1007/s00709-010-0247-0
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Shi, Wang, Zhang, Yuan, Chen, Chen, Chen and Cui. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genome- and Transcriptome-Wide Identification of C3Hs in Common Bean (*Phaseolus vulgaris* L.) and Structural and Expression-Based Analyses of Their Functions During the Sprout Stage Under Salt-Stress Conditions

Qi Zhang¹, Wen-jing Zhang¹, Zhen-gong Yin², Wei-jia Li¹, Hao-hao Zhao¹, Shuo Zhang¹, Lin Zhuang¹, Yu-xin Wang¹, Wen-Hui Zhang¹ and Ji-Dao Du^{1,3*}

¹ Laboratory Crop Genetics and Breeding, College of Agriculture, Heilongjiang Bayi Agricultural University, Daqing, China,

² Crop Resources Institute of Heilongjiang Academy of Agricultural Sciences, Heilongjiang, China, ³ Laboratory Crop Genetics and Breeding, National Coarse Cereals Engineering Research Center, Daqing, China

OPEN ACCESS

Edited by:

Xin Chen,
Jiangsu Academy of Agricultural
Sciences, China

Reviewed by:

Pengtao Ma,
Yantai University, China
Longbiao Guo,
Chinese Academy of Agricultural
Sciences, China

*Correspondence:

Ji-Dao Du
djdbynd@163.com

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 22 May 2020

Accepted: 19 August 2020

Published: 15 September 2020

Citation:

Zhang Q, Zhang W, Yin Z, Li W,
Zhao H, Zhang S, Zhuang L, Wang Y,
Zhang W-H and Du J-D (2020)
Genome- and Transcriptome-Wide
Identification of C3Hs in Common
Bean (*Phaseolus vulgaris* L.)
and Structural and Expression-Based
Analyses of Their Functions During
the Sprout Stage Under Salt-Stress
Conditions. *Front. Genet.* 11:564607.
doi: 10.3389/fgene.2020.564607

CCCH (C3H) zinc-finger proteins are involved in plant biotic and abiotic stress responses, growth and development, and disease resistance. However, studies on C3H genes in *Phaseolus vulgaris* L. (common bean) are limited. Here, 29 protein-encoding C3H genes, located on 11 different chromosomes, were identified in *P. vulgaris*. A phylogenetic analysis categorized the *PvC3Hs* into seven subfamilies on the basis of distinct features, such as exon-intron structure, *cis*-regulatory elements, and MEME motifs. A collinearity analysis revealed connections among the *PvC3Hs* in the same and different species. The *PvC3H* genes showed tissue-specific expression patterns during the sprout stage, as assessed by real-time quantitative PCR (RT-qPCR). Using RNA-sequencing and RT-qPCR data, *PvC3Hs* were identified as being enriched through Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses in binding, channel activity, and the spliceosome pathway. These results provide useful information and a rich resource that can be exploited to functionally characterize and understand *PvC3Hs*. These *PvC3Hs*, especially those enriched in binding, channel activity, and the spliceosome pathway will further facilitate the molecular breeding of common bean and provide insights into the correlations between *PvC3Hs* and salt-stress responses during the sprout stage.

Keywords: C3H gene family, common bean, sprout stage, salt stress, gene expression analysis, transcriptome

INTRODUCTION

The zinc transcription factor family is among the largest in eukaryotes (Li et al., 2001). Zinc finger proteins have various functions, such as DNA and RNA binding and transcriptional activation, as well as being involved in protein-protein interactions (Gao et al., 2002). The zinc-finger motif is composed of one or more cysteine residues and one histidine residue. This superfamily is divided

into nine categories on the basis of the numbers and distances between conserved cysteine and histidine residues (Wolfe et al., 2000). However, there are limited studies on CCCH (C3H) zinc-finger proteins. C3H type zinc-finger proteins typically possess one or multiple zinc-finger motifs, characterized by three cysteine residues followed by a histidine residue, and they account for approximately 0.8% of all zinc-finger proteins (Hall, 2005).

In plants, at present, the functions of only some C3H zinc-finger proteins have been revealed. The C3H gene families are involved in various plant developmental and adaptation-related processes and play important roles in hormone regulation. In the jasmonic acid (JA) pathway of rice (*Oryza sativa*), the overexpression of *OsDOS* results in negative regulation that significantly delays leaf senescence (Kong et al., 2006). In *Arabidopsis thaliana*, *ATML1* is a specific C3H gene that is indispensable for the embryogenesis of seeds (Li, 1998). *AtC3H14* and *AtC3H15* regulate secondary wall thickening, anther development, and male fertility (Kim et al., 2014; Lu et al., 2014; Chai et al., 2015). Additionally, some C3H zinc-finger proteins participate in abiotic and biotic stress responses. In *A. thaliana*, *TZF1* is involved in sugar signal transduction, and its overexpression delayed an increase in salt tolerance (Lin et al., 2014). *AtZFP1* improves plant salt tolerance (Han et al., 2014) by maintaining the ion balance and alleviating oxidation and osmotic stress. *AtSZF1* and *AtSZF2* modulate the expression of many salt-sensitive genes, resulting in enhanced salt tolerance in *A. thaliana* (Sun et al., 2007). Rice *C3H12* regulates rice bacterial blight resistance (Deng et al., 2012) through the JA signaling pathway, and after the over-expression of *C3H18* in sweet potato, plant resistance to drought and high salt conditions is enhanced (Zhang et al., 2019) through the regulation of many stress-response genes. In cotton (*Gossypium hirsutum*), *ZFP1* controls salt tolerance, as well as disease resistance, by interacting with dehydrin and disease resistance-related proteins, respectively (Guo et al., 2009).

Phaseolus vulgaris L. is a valuable legume crop and a livelihood source for many people (Broughton et al., 2003; Wu et al., 2020). Additionally, the crop is rich in proteins and micro-nutrients that are essential for human health. The crop yield has slowly improved, particularly in tropical regions, and favorable yields have been realized in temperate countries because of the development of new cultivars (Rosenzweig and Liverman, 1992). The C3H genes participate in many responses related to various abiotic stresses in several plant species. However, the C3H family in *P. vulgaris* has not been characterized. To screen for potential salt-tolerance genes expressed during the sprout stage of *P. vulgaris*, members of C3H gene family were identified at the genome level. Here, 29 PvC3H members were detected in the *P. vulgaris* genome. Furthermore, their distribution among the chromosomes was determined and their conserved domains and collinearity investigated. PvC3H gene expression levels were evaluated in multiple *P. vulgaris* varieties. Additionally, the expression patterns of the PvC3H genes were analyzed in various tissues at the sprout stage, and the possible regulation of specifically expressed PvC3H genes under salt stress and other stress conditions was investigated. These findings offer

crucial information for studying the evolution and functional differentiation of the C3H gene family in common bean.

MATERIALS AND METHODS

Identification and Phylogenetic Analysis of C3H Genes in *P. vulgaris*

The HMMER software¹ and the Pfam protein family database² (Finn et al., 2016) were employed to identify candidate C3H proteins containing the C3H domain (PF00642). The protein annotation file was retrieved from the website of Ensembl plants³. Subsequently, InterPro⁴ (Finn et al., 2017) and SMART⁵ (Letunic et al., 2015; Han et al., 2019) software were applied to verify the reliability of the C3H domain prediction. Lastly, Interpro⁶, WoLF PSORT⁷, P3DB⁸, and the ExPASy Proteomics Server⁹ were employed to verify the integrity of the C3H domains in the candidate genes. Each C3H gene was named on the basis of their precise position on the chromosome.

The C3H protein sequences of the other species, like *Arabidopsis*, were obtained from Ensembl plants using the C3H domain. C3Hs were introduced into MEGA X to perform multiple sequence alignments. The Maximum-likelihood method with the JTT + G model, which MEGA X predicted as the optimal model, and 1,000 replicates were used to produce bootstrap values. MEGA X was used to revise and construct a phylogenetic tree.

Analyses of Gene Structure and Conserved Motifs, and Promoter Predictions

The GSDS platform¹⁰ (Guo et al., 2007) was used to display and analyze the exon-intron structures of the PvC3H genes. GeneWise (Birney et al., 2004) was used to examine the correspondence to coordinates in the DNA (comprising both exons and introns) versus protein sequences. Subsequently, in-house perl scripts were used to transform the C3H domain coordinates in the protein sequence to those in the nucleotide sequence. The intron-splicing phase in the hinge and the basic regions of the C3H domains from all the PvC3H genes were characterized and separated into various types. The MEME tool¹¹ (Bailey et al., 2009) was employed to detect additional motifs outside the C3H domain in the protein sequences. Motifs containing 10–50 amino acids and *E*-value < 1e⁻²⁰ were characterized. All of the motifs found among the PvC3Hs

¹<http://hmmer.janelia.org/>

²<http://pfam.sanger.ac.uk/>

³<http://plants.ensembl.org/index>

⁴<http://www.ebi.ac.uk/interpro/>

⁵<http://smart.embl-heidelberg.de/>

⁶<http://prosite.expasy.org/>

⁷<https://wolfsort.hgc.jp/>

⁸<http://www.p3db.org/>

⁹<http://prosite.expasy.org/>

¹⁰<http://gsds.cbi.pku.edu.cn/>

¹¹<http://meme.nbcr.net/meme/>

were compared to identify the group-specific or group-conserved signatures. The motifs were numbered based on the order in which they occurred in the protein sequences.

Collinearity Analysis of *PvC3Hs*

The mapping of the *PvC3Hs* to the chromosomes was performed using the chromosomal locations provided by Esembl plants. The Multiple Collinearity Scan toolkit (MCScanX) was used to examine gene duplication events based on default parameters (Wang et al., 2012). Additionally, circos (version 0.69¹²) was used to generate the visualization as described previously (Krzywinski et al., 2009).

Salt-Tolerant and Salt-Sensitive Common Beans Exposed to Different Treatments

Two kinds of common beans, which are salt-tolerant (R) and salt-sensitive (N), respectively, were selected as the test materials. They were treated with water (W) and 0.4% NaCl (S). The plants were then placed in a 28°C incubator without light for 7 days. The numbers of germination events were recorded consecutively for 7 days.

Transcriptome Analysis

After 5 days, the bean-free hypocotyls of WN, SN, WR, and SR were selected, with three repeated treatments. All the materials were rapidly frozen in liquid nitrogen at a low temperature and then rapidly ground independently into powder. The RNAs were extracted using TRIzol (Invitrogen, Carlsbad, CA, United States), and the RNA quality was determined using 1% agarose gel electrophoresis and NanoDrop (Thermo, Carlsbad, CA, United States) instrumentation. The concentration and purity of the RNAs were determined using a NanoDrop spectrophotometer. The OD_{260/280} was required to be between 1.8 and 2.2, and that of the 28S/18S was required to be 1.0 or greater. An Agilent Bioanalyzer 2100 (Agilent Technologies, Carlsbad, California) was used to determine the RNA integrity. All the sequencing samples were rated as A-level, meeting the requirements of library construction. After qualified RNA detection, sequencing libraries were constructed (Supplementary Table 1) on an Illumina HiSeq 2500 PE 150 and referring to the NCBI Common bean (*P. vulgaris*) reference genome. The heat map was constructed using Tbttools software (Chen et al., 2020). The Kyoto Encyclopedia of Genes and Genomes (KEGG) database¹³ and the phytozome database¹⁴ were used for gene annotation. Gene ontology (GO) annotations were predicted using Blast2GO software (Conesa et al., 2005), and the functional classification of genes was performed using WEGO software (Ye et al., 2006).

Quantitative Real-Time PCR (RT-qPCR) Analysis of *PvC3Hs*

Different tissues (cotyledon, hypocotyl, and radicle) of WR were selected on the fifth day during the sprout stage, and

hypocotyls of the four samples (WR, WN, SR, and SN) were selected on the fourth, fifth and sixth days. Each sample had three biological repeats. Total RNA was extracted using RNA isolater Total RNA Extraction Reagent (Vazyme Biotech, China). After the extraction, the RNA samples were electrophoresed in 1% agarose gel for detection and quality control. Additionally, a NanoDrop spectrophotometer was used to check the concentration and purity of the RNAs. Samples of poor quality were discarded. Premier 5.0 (Pi et al., 2018) was used to design the RT-qPCR primers (Supplementary Table 2). *PvC3Hs* were selected from the GO enrichment and KEGG pathways, and some *PvC3Hs* were selected as candidate genes for verification. *ACTIN-11* of common bean served as an internal control gene (Borges et al., 2012). The RNA was transcribed to cDNA, and then, RT-qPCR was performed using 2 × ChamQ Universal SYBR qPCR Master Mix Kit (Vazyme Biotech, China). The RT-qPCR was performed in triplicate. The relative gene expression was determined as follows (Livak and Schmittgen, 2001):

Relative Expression = $2^{\Delta \Delta Ct}$, Where $\Delta \Delta Ct = [Ct(Pv \text{ target genes}) - Ct(PvACTIN-11)]$.

Data Analyses

Office 2013 and SPSS 19.0 were used for data analyses (Li et al., 2018).

RESULTS

Identification of C3H Family Members in *P. vulgaris* L.

In total, 65 protein sequences with C3H domains were identified using a HMM profile analysis from the common bean genome. Furthermore, InterPro and SMART analyses assigned 29 putative members, *PvC3H01*–29, to the common bean C3H family (Table 1). The 29 *PvC3H* members were located on the 11 chromosomes (Chrs) of common bean.

Specifically, most of the *PvC3Hs* were located on Chrs 7 and 3, while few genes were distributed on Chrs 2 and 10 (Figure 1). At least one *PvC3H* member was mapped to each chromosome.

Phylogenetic Analysis of C3H Proteins

A comprehensive phylogenetic tree comprising 147 C3H protein sequences from three plants, 29 sequences from common bean (*P. vulgaris*), 50 from rice (*O. sativa*), and 68 from the *A. thaliana*, was constructed (Figure 2). The *PvC3H*, rice protein sequences and *A. thaliana* C3H sequences were obtained from a public database retrieved using the pfam method. The phylogenetic tree of the C3H sequences was generated with the Maximum-likelihood method using a JTT + G model, which was the optimal model as determined by MEGA X software. Using the phylogenetic tree and the classification method applied to the C3H genes in *A. thaliana* (Wang et al., 2008), which were separated into 11 subfamilies, I–X I, the C3H genes in *P.* were classified into only seven subfamilies (I, III, V, VI, VII, VIII, and XI).

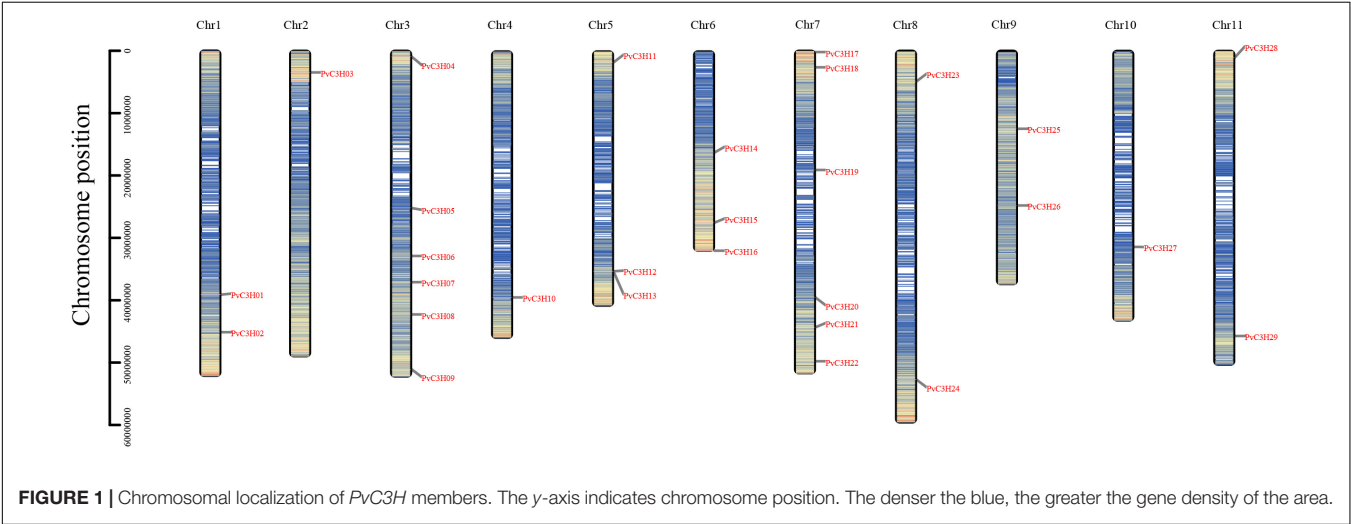
¹²<http://circos.ca/>

¹³<http://www.genome.jp/keg>

¹⁴<https://phytozome.jgi.doe.gov/pz/portal.html>

TABLE 1 | Information regarding *Phaseolus vulgaris* L. C3H family members.

Gene name	Gene Bank	Gene ID	chr	Location	Protein length	CDS length	Lsoelectric point	Molecular weight	Instability index	Aliphatic index
<i>PvC3H01</i>	ESW34310.1	<i>Phvul.001G141900</i>	1	39106940:39110116	295	888	9.31	31402.61	38.53	57.66
<i>PvC3H02</i>	ESW34844.1	<i>Phvul.001G186200</i>	1	45111242:45114277	315	948	9.74	36528.44	71.26	47.05
<i>PvC3H03</i>	ESW28984.1	<i>Phvul.002G034300</i>	2	3447337:3452995	699	2100	6.19	77929.46	49.69	72.42
<i>PvC3H04</i>	ESW25121.1	<i>Phvul.003G009100</i>	3	904447:908755	258	777	9.79	30270.78	81.33	53.33
<i>PvC3H05</i>	ESW26238.1	<i>Phvul.003G102600</i>	3	25189363:25194111	483	1452	8.62	50645.69	65.52	54.53
<i>PvC3H06</i>	ESW26651.1	<i>Phvul.003G136700</i>	3	32863086:32868451	601	1806	6.30	66954.69	47.72	72.38
<i>PvC3H07</i>	ESW26973.1	<i>Phvul.003G162600</i>	3	37078874:37081522	344	1035	8.26	38457.22	47.77	55.32
<i>PvC3H08</i>	ESW27506.1	<i>Phvul.003G207900</i>	3	42243339:42248077	485	1458	4.89	52877.12	58.05	48.02
<i>PvC3H09</i>	ESW28409.1	<i>Phvul.003G284200</i>	3	51035876:51040126	376	1131	6.56	39672.21	50.24	42.85
<i>PvC3H10</i>	ESW24345.1	<i>Phvul.004G122600</i>	4	39407374:39416501	424	1275	7.86	46083.01	22.55	75.57
<i>PvC3H11</i>	ESW20849.1	<i>Phvul.005G019900</i>	5	1724921:1727678	363	873	6.94	40635.13	55.13	70.69
<i>PvC3H12</i>	ESW22104.1	<i>Phvul.005G127700</i>	5	35217300:35220277	308	1146	9.24	36262.07	81.14	48.47
<i>PvC3H13</i>	ESW22107.1	<i>Phvul.005G127800</i>	5	35227175:35228656	290	873	9.06	34119.87	76.75	53.10
<i>PvC3H14</i>	ESW18515.1	<i>Phvul.006G047900</i>	6	16181121:16187473	496	1491	5.31	55805.75	51.79	74.15
<i>PvC3H15</i>	ESW19882.1	<i>Phvul.006G163300</i>	6	27430746:27431063	483	1452	8.49	50620.70	64.74	55.55
<i>PvC3H16</i>	ESW20584.1	<i>Phvul.006G221300</i>	6	31916147:31924712	565	1698	6.40	66815.55	56.23	44.69
<i>PvC3H17</i>	ESW14610.1	<i>Phvul.007G002700</i>	7	186479:191473	298	897	9.34	31581.67	38.29	56.71
<i>PvC3H18</i>	ESW14969.1	<i>Phvul.007G033200</i>	7	2618882:2622003	365	1098	8.00	40310.00	49.76	62.77
<i>PvC3H19</i>	ESW15970.1	<i>Phvul.007G118500</i>	7	19087659:19092941	444	1335	8.86	48721.00	55.66	59.37
<i>PvC3H20</i>	ESW16542.1	<i>Phvul.007G165200</i>	7	39536306:39540749	466	1401	8.97	49805.21	57.55	64.23
<i>PvC3H21</i>	ESW17023.1	<i>Phvul.007G203900</i>	7	44306014:44306584	125	378	8.63	14241.05	53.39	53.84
<i>PvC3H22</i>	ESW17675.1	<i>Phvul.007G259300</i>	7	49744488:49748916	293	882	9.37	30682.19	43.58	64.37
<i>PvC3H23</i>	ESW11726.1	<i>Phvul.008G054800</i>	8	4888914:4897088	426	1281	8.41	46803.62	65.21	53.40
<i>PvC3H24</i>	ESW13654.1	<i>Phvul.008G214600</i>	8	52694068:52694121	427	1284	7.86	46764.29	57.51	53.51
<i>PvC3H25</i>	ESW08797.1	<i>Phvul.009G075500</i>	9	12478833:12479986	251	756	8.30	27883.32	64.45	59.12
<i>PvC3H26</i>	ESW09958.1	<i>Phvul.009G169600</i>	9	24777928:24782632	492	1479	4.81	53385.37	58.93	50.30
<i>PvC3H27</i>	ESW06876.1	<i>Phvul.010G084000</i>	10	31414254:31417666	329	990	6.58	37373.60	49.50	50.21
<i>PvC3H28</i>	ESW03434.1	<i>Phvul.011G013700</i>	11	1051759:1053851	700	2103	5.96	76237.53	54.75	68.60
<i>PvC3H29</i>	ESW05458.1	<i>Phvul.011G180700</i>	11	45672467:45688469	484	1455	7.51	52859.50	54.37	56.92

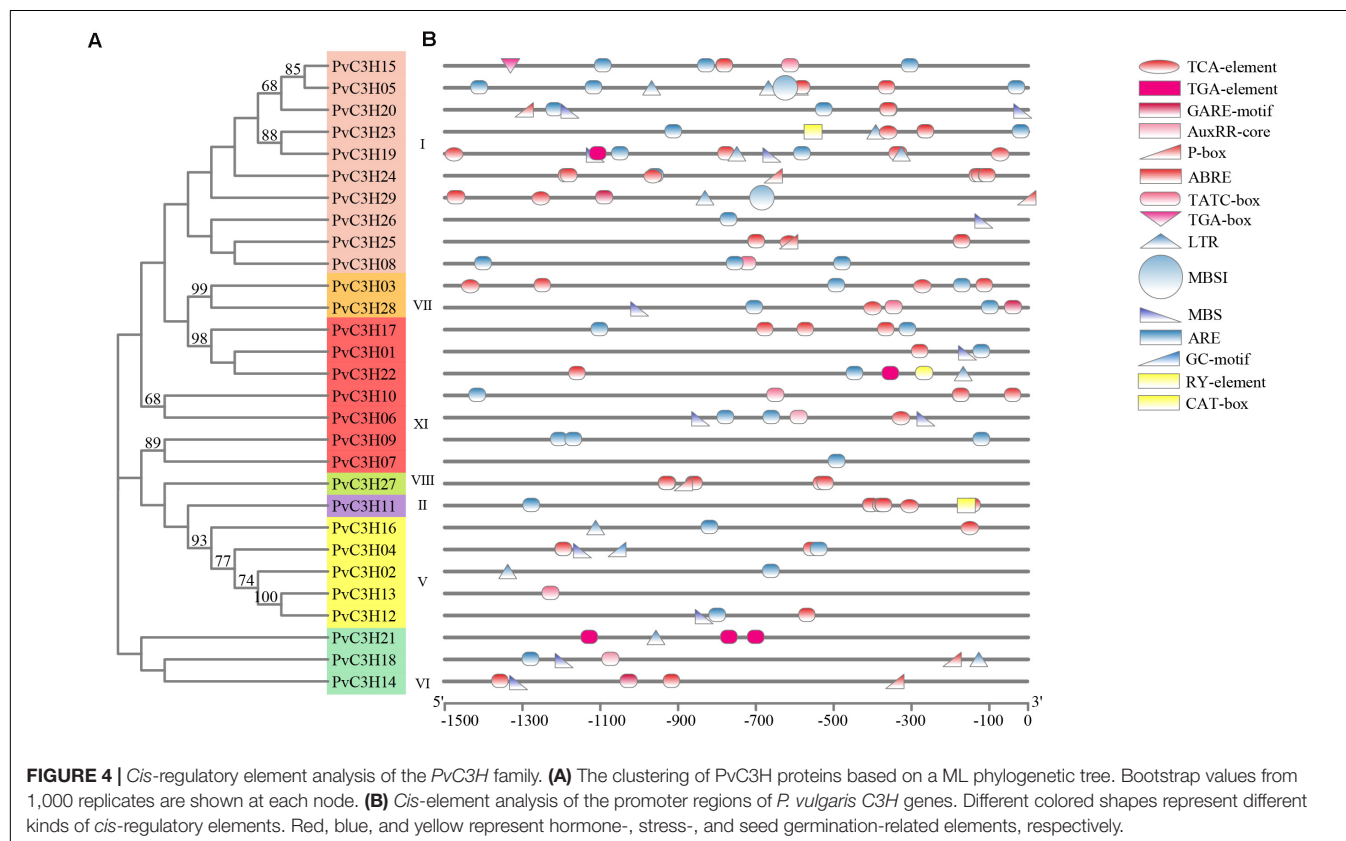
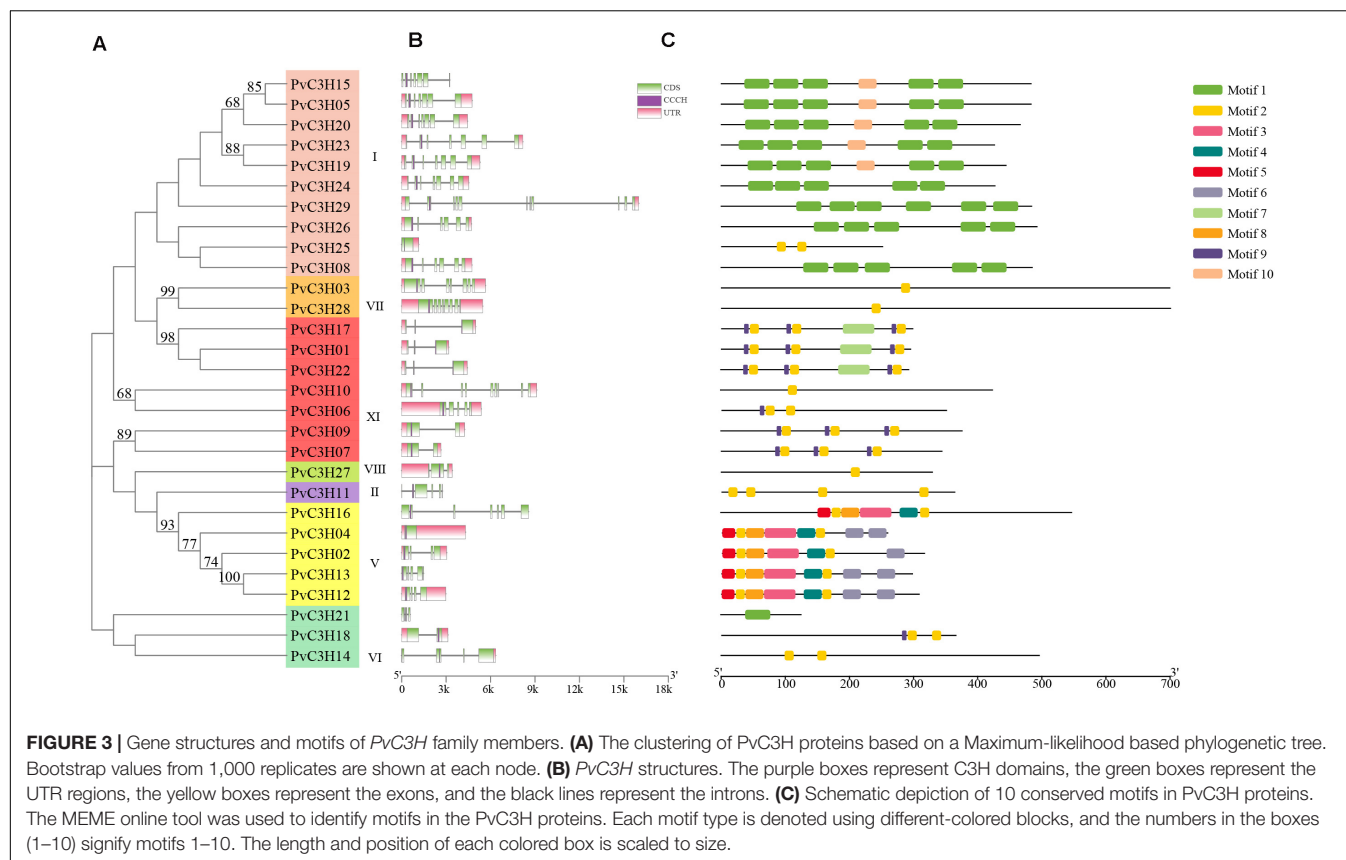


Gene Structures and Conserved Motif Compositions of the *PvC3Hs*

The exon–intron structures of the 29 *PvC3Hs* were examined to reveal the gene structures and the evolutionary

trajectory (Figure 3A). Each subfamily had a conserved C3H domain, and the exon and intron numbers, ranging from 0 to 11, were similar in each subfamily. Generally, the *PvC3Hs* that clustered together showed similar structures (Figure 3B).





are shown in **Figure 5**. Only four pairs of *C3H* family genes in *P. vulgaris* were collinear. In leguminous species, there was collinearity between the *C3H* family genes and 43 genes in soybean (*Glycine max*), 23 genes in adzuki bean (*Vigna angularis*), and 11 genes in mung bean (*Vigna radiata*), which indicated that the *P. vulgaris* *C3H* family genes were closely related to genes in the legume family. Additionally, *PvC3H05* and *PvC3H26* were collinear with *AT5G18550*, and *PvC3H08* and *PvC3H15* were collinear with *AT5G16540*, indicating that the four *C3H* family genes may have similar functions in *P. vulgaris*. Selection pressure analyses of collinear genes have revealed that a ratio of the non-synonymous mutation rate (K_a) to the synonymous mutation rate (K_s) = 1 usually represents neutral selection, whereas $K_a/K_s < 1$ represents negative or purified selection. **Supplementary Table 4** shows that the K_a/K_s values of most pairs of replicated *CH3* genes in *P. vulgaris* were less than 1, indicating that they underwent purification and selection after replication, and we speculated that their gene functions did not undergo differentiation,

thereby largely maintaining the functional similarities of the *PvC3H* family members.

Variation in the Germination of R and N Common Beans

To clarify the sampling time, the germination numbers of N and R were investigated. The number of germination events for R after each treatment was similar. There was no difference in the number of germination events under S conditions. For N, a significant difference in the number of germination events was found between the two treatments. The difference appeared on the fifth day. The number of germination events between WN and SN differed after 5 day. There was no difference in the germination number of R after 5 days of the W treatment compared with the S treatment, but the germination number of N had decreased (**Figure 6**). Therefore, the fifth day was used as the sampling time.

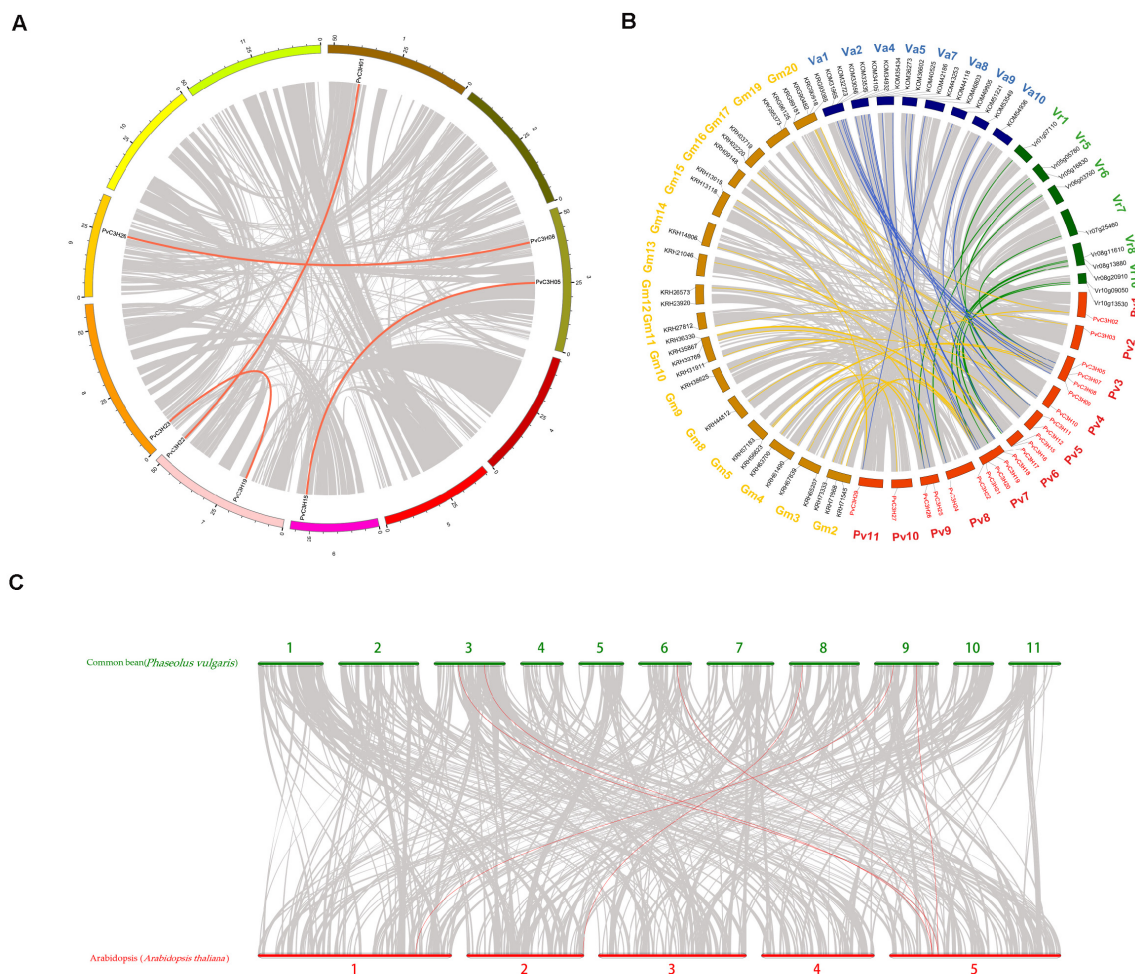


FIGURE 5 | Collinearity analysis of the *C3H* gene family in *P. vulgaris*. The bright lines represent collinearity, and the gray lines represent all isomorphic blocks in the genome. The collinear relationship **(A)** within the common bean species, **(B)** between common bean and three other species of legumes, and **(C)** between common bean and *A. thaliana*.

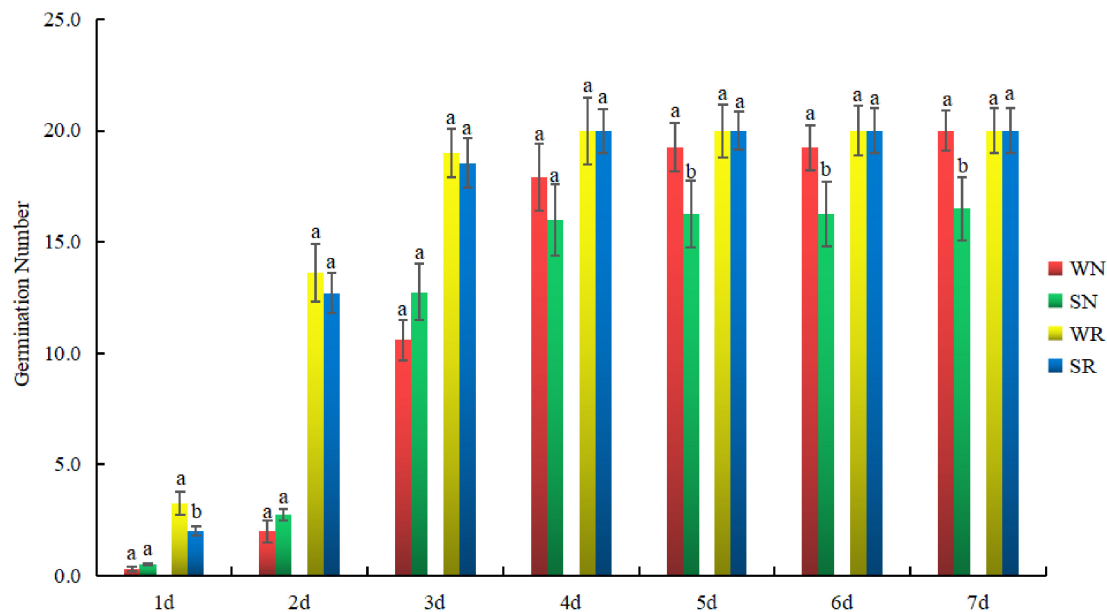


FIGURE 6 | Dynamic variation in germination numbers of salt-tolerant (R) and salt-sensitive (N) common bean varieties. They were treated with water (W) and 0.4% NaCl (S). Green and red bars represent WN and SN, respectively, while yellow and blue bars represent WR and SR, respectively. The x-axis indicates time, and the y-axis indicates the number of germination events. Lower case letter(s) above bars indicate significant differences ($\alpha = 0.05$, LSD) among the treatments. The x-axis indicates germination time, and the y-axis indicates germination number.

Expression Analysis of PvC3Hs in Different Tissues of Common Bean During the Sprout Stage

The transcriptional patterns of *PvC3Hs* in many tissues, including flower buds, flowers, green mature pods, leaves, nodules, roots, trifoliates, stems, pods, and seeds, were examined using high-throughput sequencing data from the Phytozome database. The expression levels of *PvC3Hs* in various tissues were high, and the expression differences were relatively great (Figure 7A). To elucidate the functions of the *C3Hs* in *P. vulgaris* at the sprout stage, six *PvC3H* genes were chosen randomly for analysis. RT-qPCR analyses of three tissues (cotyledon, hypocotyl, and radicle; Figure 7B) from R during the sprout stage was performed to assess the expression levels of these *PvC3H* genes on the fifth day. As shown in Figure 7C, the relative expression levels of some *PvC3Hs* in the radicle and hypocotyl were relatively high, while the expression levels of other *PvC3Hs* showed no differences among the three tissues. Therefore, the hypocotyl and radicle can be used as sprout tissues for *PvC3Hs* research.

Expression Analysis of *PvC3Hs* in R and N Under Salt-Stress Conditions

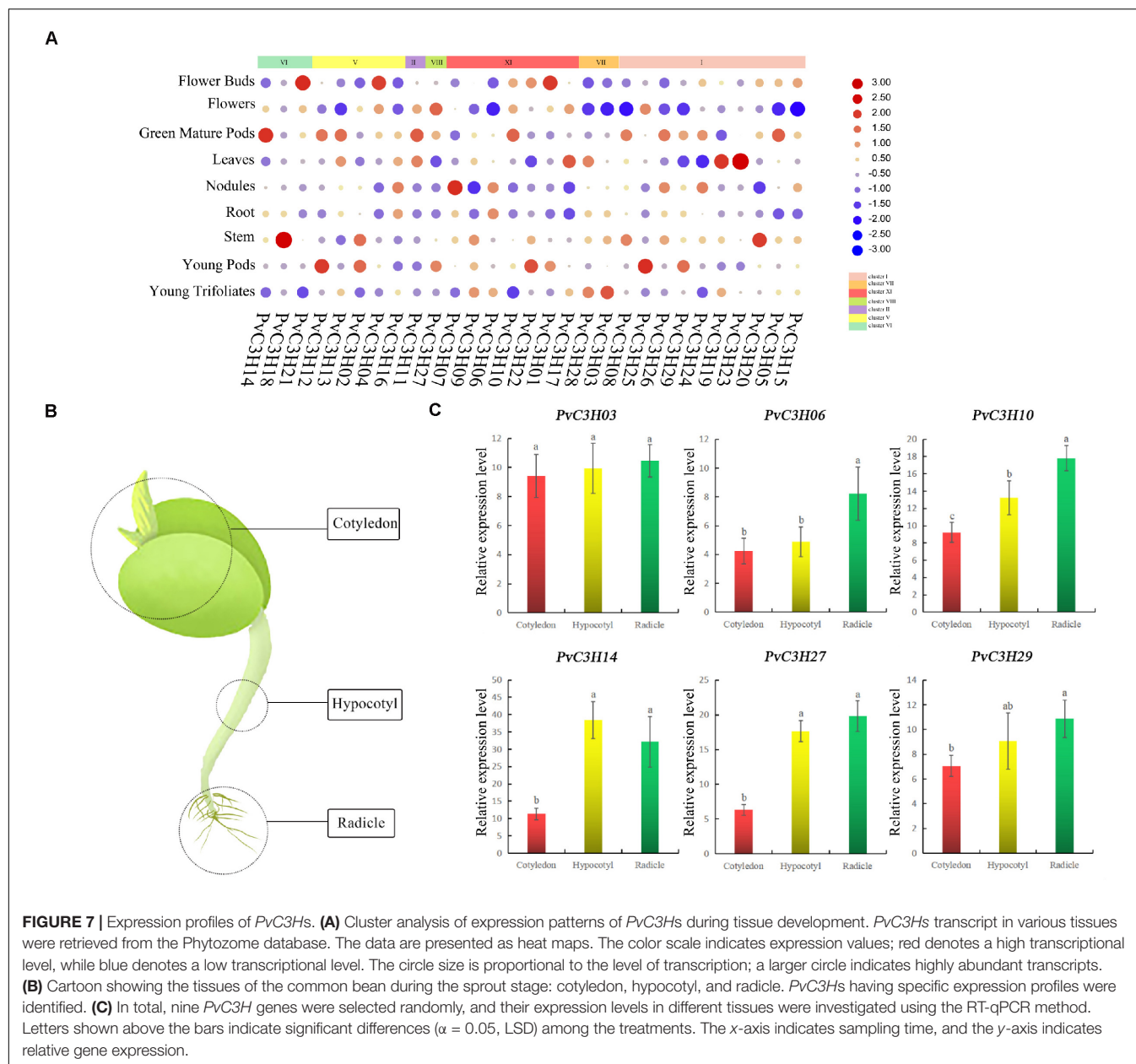
Two common bean varieties were tested to response to salt stress. R and N, were subjected to two treatments, W and S, and were analyzed after 5 days. The assembled gene dataset, deposited at the National Center for Biotechnology Information under the accession number PRJNA558376¹⁵, was used as a

reference for further analyses (Supplementary Table 5). Using the constructed bean-free hypocotyl transcriptome, differences in *PvC3H* gene expression levels were identified (Figure 8A). Six *PvC3Hs* were randomly selected for RT-qPCR analyses, and the results under W and S conditions were consistent with the transcriptome results (Figure 8B). Compared with N, the relative expression levels of the *PvC3Hs* showed more differences between WR and SR. Two *PvC3Hs* were selected randomly and RT-qPCR was performed using samples taken over 3 days. Although the change trend of the curve was different, the *PvC3Hs* in R showed greater changes in expression under the W and S treatments compared with those in N (Figure 8C). Thus, under salt-stress conditions, the differential expression of *PvC3Hs* may affect the salt tolerance of plants.

GO Enrichment of *PvC3H* Genes' Expression Analysis

The GO enrichment analysis of *PvC3Hs* showed that they were enriched in many terms, such as ion binding, cation binding, metal ion binding, amide transport, sodium channel regulator, inhibitor activities, binding, and ion channel inhibitor activity (Figure 9A). Two GO categories were enriched after screened: ion binding and channel activity. Three genes enriched in the two terms were randomly selected for RT-qPCR verification. For the three *PvC3Hs* enriched in channel activity (*PvC3H01*, *PvC3H17*, and *PvC3H22*; Figure 9B), and the three genes enriched in ion binding (*PvC3H07*, *PvC3H18*, and *PvC3H19*; Figure 9C), the differences in R between the W and S treatments were greater those in N. Thus, these

¹⁵<http://www.ncbi.nlm.nih.gov/bioproject/>



genes may affect the expression of salt-tolerance traits during the sprout stage.

The KEGG analysis of *PvC3Hs* showed that they were enriched in the “Spliceosome” (pvu03040) pathway in common bean, having a corrected P -value of $7.03e^{-8}$. The *PvC3Hs* belonged to the U2-related module in the spliceosome (Figure 10). The RT-qPCR analysis of radicles from the two varieties showed that, compared with the S treatment, the changes in some genes’ expression levels (*PvC3H02*, *PvC3H04*, *PvC3H12*, and *PvC3H13*) in R were greater than in N, which indicated that the change in the abscisic acid content affected the expression levels of the *PvC3H* genes under salt-stress conditions, which may further affect the salt tolerance of the common bean during the sprout stage.

DISCUSSION

Characterization of C3H Gene Family Members in Common Bean

Here, 29 *PvC3Hs* were recognized from the genome of *P. vulgaris* L, suggesting that the C3H gene family has a wider distribution in common bean than in *A. thaliana* or rice. The *PvC3H* family was separated into seven subfamilies on the basis of the phylogenetic analysis. A similar method was used to identify subfamilies in maize (Peng et al., 2012), citrus (Liu et al., 2014), and *Brassica rapa* (Pi et al., 2018). Furthermore, the phylogenetic relationships of the *PvC3Hs* were confirmed using their conserved motifs and gene structures. A gene structure analysis revealed that the *PvC3Hs* contain from 0 to 11 introns.

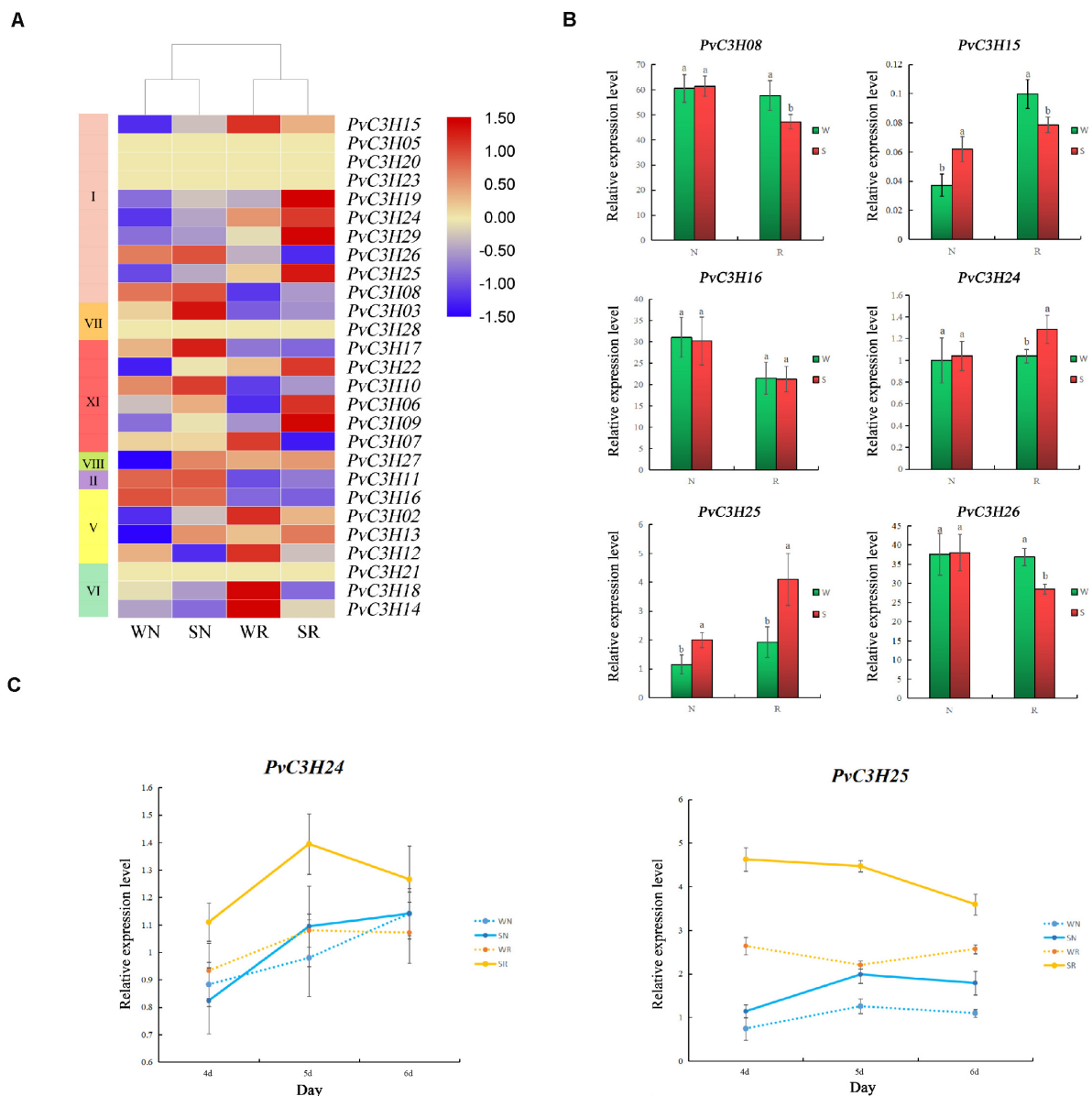
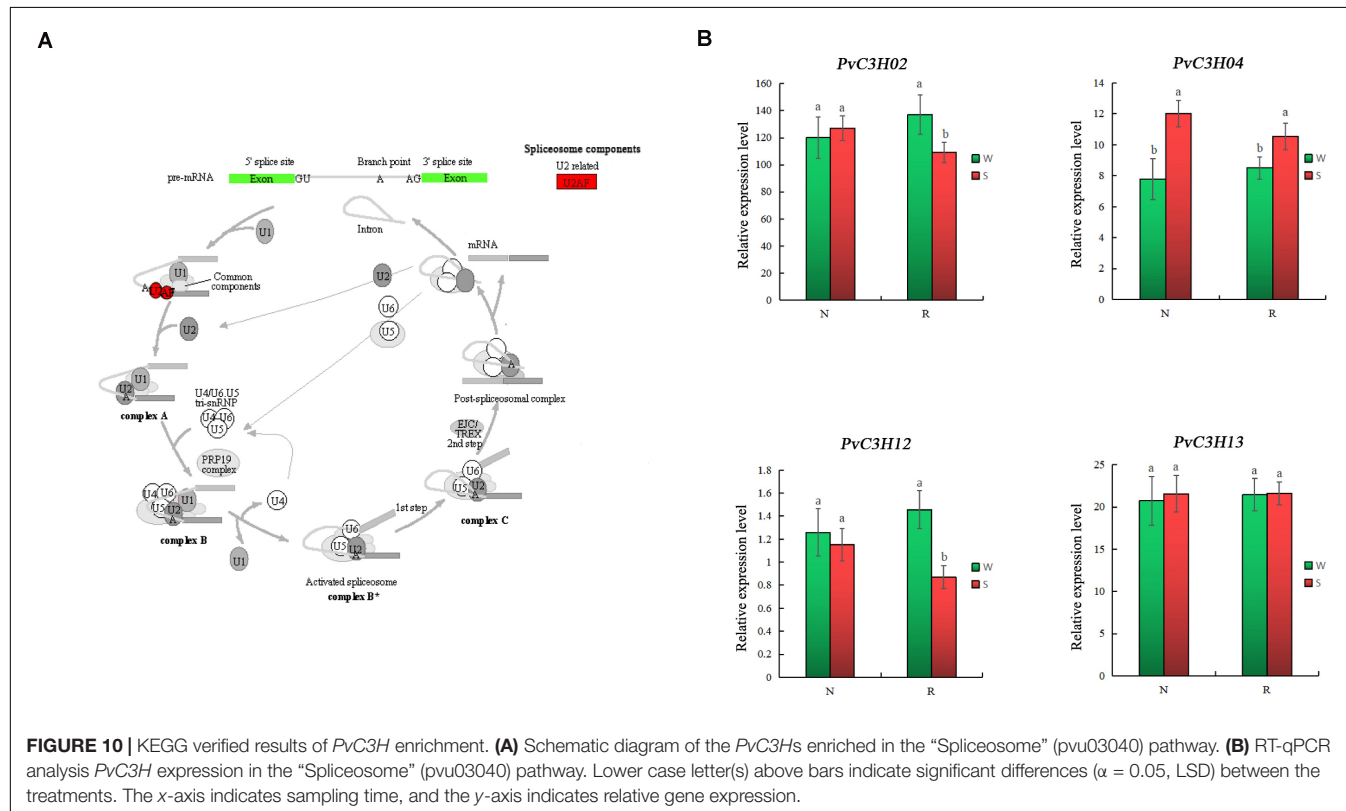
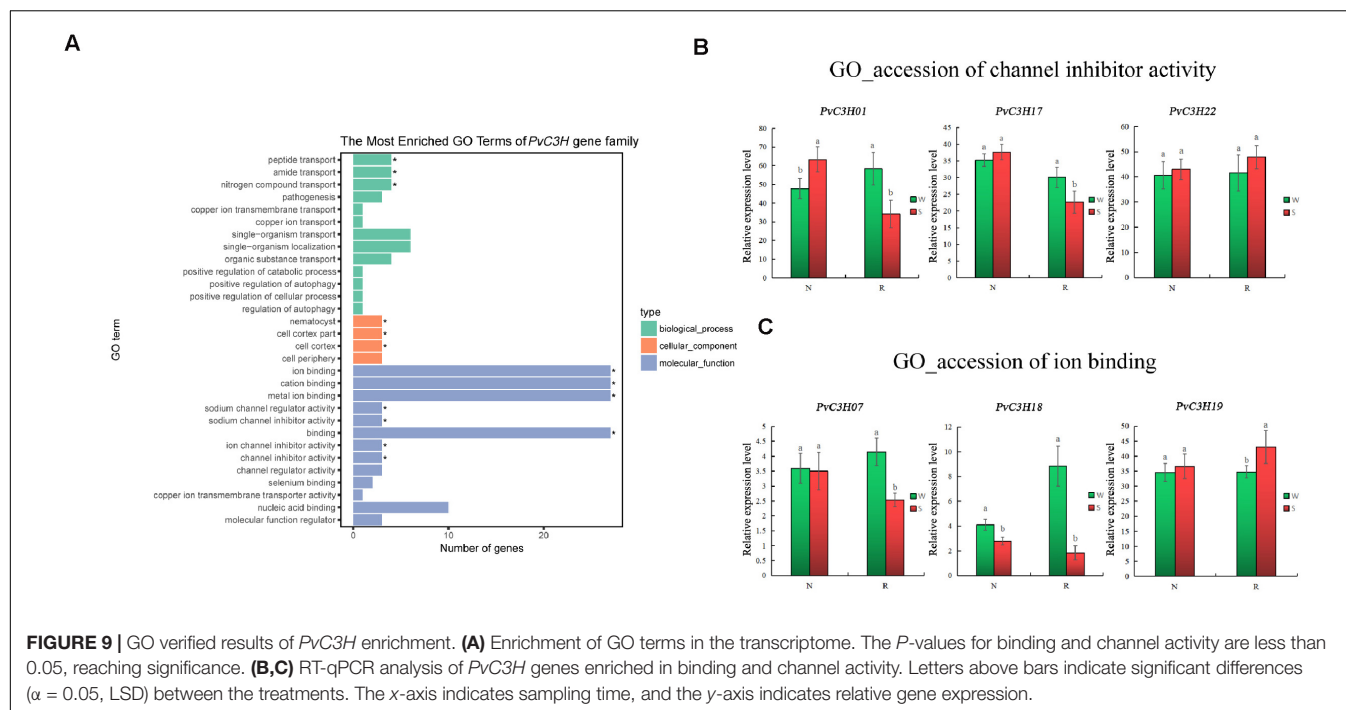


FIGURE 8 | Transcriptome analysis of salt-tolerant (R) and salt-sensitive (N) common bean varieties under salt-stress conditions. The x-axis indicates sampling time, and the y-axis indicates relative gene expression. **(A)** A heat map with clustering was constructed using *PvC3H*s in the different varieties under salt-stress conditions. The color scale varies from blue to red, representing relatively low to high expression, respectively. **(B)** Six *PvC3H*s were selected and their expression assessed by RT-qPCR to determine the reliability of the transcriptome, the y-axis indicates relative gene expression. Lower case letter(s) above bars indicate significant differences ($\alpha = 0.05$, LSD) between the treatments. **(C)** The expression changes of the two *PvC3H*s on the fourth, fifth, and sixth day. The yellow line represents R, and the blue line represents N; The x-axis indicates sampling time, and the y-axis indicates relative gene expression.

Additionally, every subfamily exhibited a similar exon-intron organization and contained *cis*-regulatory elements. These results may help reveal the functions of various gene families. According to the conserved motif analysis, the C3H domain (motif 1) was common to all the *PvC3H* proteins. Each subfamily harbored similar motifs, and the motif sequences in each subfamily were relatively similar, which indicated that *PvC3H* subfamily members may be conserved. Generally, the same subfamily shared an analogous exon-intron organization, as well as motif

composition, and this has also been observed in rapeseed, poplar (Chai et al., 2012), rice (Wang et al., 2008), chickpea (Pradhan et al., 2017), *Solanum lycopersicum* (Xu, 2014), and *Medicago truncatula* (Zhang et al., 2013).

A comparative genomic analysis was performed to assess the structure of the genome at the syntenic block level, which reveals features that are conserved in multiple genomes (Ghiurcuta and Moret, 2014). This enables functional information to be transferred from a known taxon to a less-studied taxon.



In this study, collinearity revealed that some *PvC3H* genes may have been generated through gene duplication. Moreover, segmental duplication events were vital in *PvC3H* evolution. The *Arabidopsis* homologous gene *HSFC1* responds to stress

and also has a binding effect (Gaudet et al., 2011). *AtC3H14*, which contains a C3H zinc-finger protein, also binds to its target mRNAs during *A. thaliana* growth (Kim et al., 2014). The *PvC3H* homologous genes in *Arabidopsis* produce JA to induce

the biosynthesis of the metabolite N δ -acetylornithine, which enhances the defense functions (Adewale et al., 2011).

Therefore, the collinear PvC3Hs of common bean revealed in this study may also be involved in plant resistance. Thus comparative genomic analyses between common bean and other species form an important basis for future studies on the roles of PvC3H genes.

PvC3H Expression During the Sprout Stage

Salt, which acts as important abiotic stress in agricultural production (Liang et al., 2018), is different from other biological stresses. Salt stress may exist from the plant's germination stage, through its growth, and into maturity. The sprout stage is the initial period of crop growth, and stress during this period negatively affects the entire growth period. Additionally, seed germination is a critical stage for crop establishment (Hubbard et al., 2012), and it is the least tolerant to abiotic stress (Patade et al., 2011). Various abiotic stresses, especially salt stress, reduce the germination rate and delay the germination of high-quality but salt-sensitive seeds (Khan and Gul, 2006; Ansari and Sharif-Zadeh, 2012; Fazlali et al., 2013; Thiam et al., 2013). Therefore, it is very important to analyze the expression levels of PvC3Hs during the sprout stage.

In this study, through the germination of R and N during W and S treatments, day five was selected as the sampling time. The tissue-specific expression results showed that the hypocotyl and radicle may be used as the sampling materials when studying PvC3Hs, and hypocotyl growth has also been reported to be affected by salt stress (Rossi et al., 2016).

Enrichment of PvC3Hs

Recent evidence showed that C3H genes play crucial roles in the adaptation to various abiotic stresses, especially salt stress, by plants, such as *A. thaliana* (Seok et al., 2018), rice (Jan et al., 2013), tobacco (Guo et al., 2009), and sweet potato (Zhang et al., 2019), but there is limited information on common bean. Transcriptome data from R and N indicated that the expression levels of PvC3H genes may cause salt tolerance in common bean sprouts. PvC3Hs were enriched in the binding and channel activity GO terms and in the "Spliceosome" pathway. A RT-qPCR analysis of the PvC3Hs enriched in GO and KEGG analyses was conducted. The differences in the expression levels of PvC3Hs in R were more significant than those in N, suggesting that PvC3Hs participate in the enrichment of the pathway and traits. The pathways for inhibitor activities, binding, and ion channel inhibitor, which were enriched in the GO analysis have been reported in previous studies on salt tolerance in plants, such as *A. thaliana* (Li et al., 2014; Kim et al., 2019; Wang et al., 2019), rice (Mansuri et al., 2019), cotton (Mu et al., 2019), poplar (Gaudet et al., 2011), and alfalfa (Lai et al., 2014). The spliceosome pathway has anti-stress effects in *A. thaliana*

(Feng et al., 2015; Narasimha et al., 2018), *O. sativa* (Ashwini et al., 2018), and *Cannabis sativa* (Liu et al., 2016). Additionally, C3H has been reported to enhance stress resistance through spliceosome driven RNA splicing (Bogamuwa and Jang, 2014). Thus, the enrichment results of these analyses provide new ideas and valuable information for studying the relationship between PvC3Hs and salt stress.

CONCLUSION

Despite the importance of C3H genes in responses to salt stress, the precise roles of C3H gene family members in common bean have not been elucidated. Here, a reference genome of common bean was used to comprehensively analyze 29 PvC3Hs, including the identification of sequences, construction of phylogenetic trees, characterization of motif composition, and analyses of cis-regulatory elements, gene structure, collinearity, and expression levels in different tissues during the sprout stage. RNA-seq and RT-qPCR analyses confirmed that the expression levels of the PvC3Hs underwent greater changes in R than in N. Through a GO analysis and KEGG enrichment, the PvC3Hs were found to be enriched in ion binding, channel regulator activity, and the spliceosome pathway. These findings will aid further studies on the C3H and other gene families in common bean.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

QZ and J-DD conceived the study and designed and managed the experiments. W-JZ, WL, and ZY provided the plant lines. SZ, HZ, LZ, YW, and QZ performed the trials and collected the data. ZY and QZ completed the statistical analysis of the phenotypic data and wrote the manuscript. All authors contributed to writing the manuscript.

FUNDING

This study was supported by the Research Project of Heilongjiang Bayi Agricultural University (XDB2011-02).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.564607/full#supplementary-material>

REFERENCES

- Adewale, M. A., Clare, L. C., Martin, D. V., Jae, H. K., Vijay, J., Baohua, L., et al. (2011). Biosynthesis and defensive function of Nδ-acetylornithine, a jasmonate-induced *Arabidopsis* metabolite. *Plant Cell* 23, 3303–3318. doi: 10.1105/tpc.111.088989
- Ansari, O., and Sharif-Zadeh, F. (2012). Osmo and hydro priming improvement germination characteristics and enzyme activity of Mountain Rye (*Secale montanum*) seeds under drought stress. *J. Stress Physiol. Biochem.* 8, 253–261.
- Ashwini, N., Sajeevan, R. S., Udayakumar, M., and Nataraja, K. N. (2018). Identification of splice variant of OsGBF1 in *Oryza sativa* ssp. indica genotypes under salinity stress. *3 Biotech* 8:345. doi: 10.1007/s13205-018-1370-4
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335
- Birney, E., Clamp, M., and Durbin, R. (2004). Genewise and genomewise. *Genome Res.* 14, 988–995. doi: 10.1101/gr.1865504
- Bogamuwa, S. P., and Jang, J. C. (2014). Tandem CCCH zinc finger proteins in plant growth, development and stress response. *Plant Cell Physiol.* 55, 1367–1375. doi: 10.1093/pcp/pcu074
- Borges, A., Tsai, S. M., and Caldas, D. G. (2012). Validation of reference genes for RT-qPCR normalization in common bean during biotic and abiotic stresses. *Plant Cell* 31, 827–838. doi: 10.1007/s00299-011-1204-x
- Broughton, W. J., Hernandez, G., Blair, M. W., Beebe, S., and Vanderleyden, J. (2003). Beans (*Phaseolus* spp.): model food legumes. *Plant Soil* 252:74. doi: 10.1023/a:1024146710611
- Chai, G., Hu, R., Zhang, D., Qi, G., Zuo, R., Cao, Y., et al. (2012). Comprehensive analysis of CCCH zinc finger family in poplar (*Populus trichocarpa*). *BMC Genomics* 13:253. doi: 10.1186/1471-2164-13-253
- Chai, G., Kong, Y., Zhu, M., Yu, L., Qi, G., Tang, X., et al. (2015). Arabidopsis C3H14 and C3H15 have overlapping roles in the regulation of secondary wall thickening and anther development. *J. Exp. Bot.* 66, 2595–2609. doi: 10.1093/jxb/erv060
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools - an integrative toolkit developed for interactive analyses of big biological data [published online ahead of print, 2020 Jun 22]. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Deng, H. Q., Liu, H. B., Li, X. H., Xiao, J. L., and Wang, S. P. (2012). A CCCH-type zinc finger nucleic acid-binding protein quantitatively confers resistance against rice bacterial blight disease. *Plant Physiol.* 158, 876–889. doi: 10.1104/pp.111.191379
- Fazlali, R., Asli, D. E., and Moradi, P. (2013). The effect of seed priming by ascorbic acid on bioactive compounds of naked seed pumpkin (*Cucurbita pepo* var. striata) under salinity stress. *Int. J. Farm. Alli. Sci.* 2, 587–590.
- Feng, J., Li, J., Gao, Z., Lu, Y., Yu, J., Zheng, Q., et al. (2015). SKIP confers osmotic tolerance during salt stress by controlling alternative gene splicing in *Arabidopsis*. *Mol. Plant* 8, 1038–1052. doi: 10.1016/j.molp.2015.01.011
- Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., et al. (2017). InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* 45, D190–D199. doi: 10.1093/nar/gkw1107
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344
- Gao, G., Guo, X., and Goff, S. P. (2002). Inhibition of retroviral RNA production by ZAPa CCCH-type zinc finger protein. *Science* 297, 1703–1706. doi: 10.1126/science.1074276
- Gaudet, P., Livstone, M. S., Lewis, S. E., and Thomas, P. D. (2011). Phylogenetic-based propagation of functional annotations within the gene ontology consortium. *Brief Bioinform.* 12, 449–462. doi: 10.1093/bib/bbr042
- Ghiurcuta, C. G., and Moret, B. M. (2014). Evaluating synteny for improved comparative studies. *Bioinformatics* 30, i9–i18. doi: 10.1093/bioinformatics/btu259
- Guo, A. Y., Zhu, Q. H., Chen, X., and Luo, J. C. (2007). GSDS: a gene structure display server. *Yi Chuan* 29, 1023–1026. doi: 10.1360/yc-007-1023
- Guo, Y.-H., Yu, Y.-P., Wang, D., Wu, C.-A., Yang, G.-D., Huang, J.-G., et al. (2009). GhZFP1, a novel CCCH-type zinc finger protein from cotton, enhances salt stress tolerance and fungal disease resistance in transgenic tobacco by interacting With GZIRD21A and GZIPR5. *New Phytol.* 183, 62–75. doi: 10.1111/j.1469-8137.2009.02838.x
- Hall, T. M. (2005). Multiple models of RNA recognition by zinc finger proteins. *Curr. Opin. Struct. Biol.* 15, 367–373. doi: 10.1016/j.sbi.2005.04.004
- Han, G. L., Wang, M. J., Yuan, F., Sui, N., Song, J., and Wang, B. S. (2014). The CCCH zinc finger protein gene AtZFP1 improves salt resistance in *Arabidopsis thaliana*. *Plant Mol. Biol.* 86, 237–253. doi: 10.1007/s11103-014-0226-5
- Han, Z., Liu, Y., Deng, X., Liu, D., Liu, Y., Hu, Y., et al. (2019). Genome-wide identification and expression analysis of expansin gene family in common wheat (*Triticum aestivum* L.). *BMC Genom.* 20:101. doi: 10.1186/s12864-019-5455-1
- Hubbard, M., Germida, J., and Vujanovic, V. (2012). Fungal endophytes improve wheat seed germination under heat and drought stress. *Botany* 90, 137–149. doi: 10.1139/b11-091
- Jan, A., Maruyama, K., Todaka, D., Kidokoro, S., Abo, M., Yoshimura, E., et al. (2013). OsTZF1, a CCCH-tandem zinc finger protein, confers delayed senescence and stress tolerance in rice by regulating stress-related genes. *Plant Physiol.* 161, 1202–1216. doi: 10.1104/pp.112.205385
- Khan, M., and Gul, B. (2006). “Halophyte seed germination,” in *Ecophysiology of High Salinity Tolerant Plants. Tasks for Vegetation Science*, vol 40., eds M. Khan and D. Weber (Dordrecht: Springer).
- Kim, W. C., Kim, J. Y., Ko, J. H., Kang, H., Kim, J., and Han, K. H. (2014). AtC3H14, a plant-specific tandem CCCH zinc-finger protein, binds to its target mRNAs in a sequence-specific manner and affects cell elongation in *Arabidopsis thaliana*. *Plant J.* 80, 772–784. doi: 10.1111/tpj.12667
- Kim, Y.-O., Kang, H., and Ahn, S.-J. (2019). Overexpression of phytochelatin synthase AtPCS2 enhances salt tolerance in *Arabidopsis thaliana*. *J. Plant Physiol.* 240:153011. doi: 10.1016/j.jplph.2019.153011
- Kong, Z., Li, M., Yang, W., Xu, W., and Xue, B. (2006). A novel nuclear-localized CCCH-type zinc finger protein, OsDOS, is involved in delaying leaf senescence in rice. *Plant Physiol.* 141, 1376–1388. doi: 10.1104/pp.106.082941
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., and Horsman, D. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Lai, D., Mao, Y., Zhou, H., Li, F., and Xie, Y. (2014). Endogenous hydrogen sulfide enhances salt tolerance by coupling the reestablishment of redox homeostasis and preventing salt-induced K⁺ loss in seedlings of medicago sativa. *Plant Sci.* 225, 117–129. doi: 10.1016/j.plantsci.2014.06.006
- Letunic, I., Doerks, T., and Bork, P. (2015). Smart: recent updates, new developments and status in 2015. *Nucleic Acids Res.* 43, D257–D260. doi: 10.1093/nar/gku949
- Li, J., Jia, D. X., and Chen, J. X. (2001). HUA1, a regulator of stamen and carpel identities in *Arabidopsis*, codes for a nuclear RNA binding protein. *Plant Cell* 13, 2269–2281. doi: 10.2307/3871507
- Li, J., Jia, H., Wang, J., Cao, Q., and Wen, Z. (2014). Hydrogen sulfide is involved in maintaining ion homeostasis via regulating plasma membrane Na⁺/H⁺ antiporter system in the hydrogen peroxide-dependent manner in salt-stress *Arabidopsis thaliana* root. *Protoplasma* 251, 899–912. doi: 10.1007/s00709-013-0592-x
- Li, M., Zhang, K., Sun, Y., Cui, H., Cao, S., Yan, L., et al. (2018). Growth, physiology, and transcriptional analysis of Two contrasting *Carex rigescens* genotypes under Salt stress reveals salt-tolerance mechanisms. *J. Plant Physiol.* 229, 77–88. doi: 10.1016/j.jplph.2018.07.005
- Li, Z. (1998). PEI1, an embryo-specific zinc finger protein gene required for heart-stage embryo formation in *Arabidopsis*. *Plant Cell* 10, 383–398. doi: 10.1105/tpc.10.3.383
- Liang, W., Ma, X., Wan, P., and Liu, L. (2018). Plant salt-tolerance mechanism: a review. *Biochem. Biophys. Res. Commun.* 495, 286–291. doi: 10.1016/j.bbrc.2017.11.043
- Lin, P. C., Pomeranz, M. C., Jikumaru, Y., Kang, S. G., Hah, C., Fujioka, S., et al. (2014). The *Arabidopsis* tandem zinc finger protein AtTZF1 affects ABA and GA mediated growth, stress and gene expression responses. *Plant J.* 65, 253–268. doi: 10.1111/j.1365-313X.2010.04419.x

- Liu, J., Qiao, Q., Cheng, X., Du, G., Deng, G., Zhao, M., et al. (2016). Transcriptome differences between fiber-type and seed-type *Cannabis sativa* variety exposed to salinity. *Physiol. Mol. Biol. Plants* 22, 429–443. doi: 10.1007/s12298-016-0381-z
- Liu, S., Rehman, M., Khan, G., Li, Y., Zhang, J., and Hu, C. (2014). Comprehensive analysis of CCCH-type zinc finger gene family in citrus (*Clementine mandarin*) by genome-wide characterization. *Mol. Genet. Genomics* 289, 855–872. doi: 10.1007/s00438-014-0858-9
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT. *Method* 25, 402–408. doi: 10.1006/meth.2001.1262
- Lu, P. L., Chai, M. F., Yang, J. E., Ning, G., Wang, G. L., and Ma, H. (2014). The *Arabidopsis* callose defective microfibril gene is required for male fertility through regulating callose metabolism during microsporogenesis. *Plant Physiol.* 164, 1893–1904. doi: 10.1104/pp.113.233387
- Mansuri, R. M., Shobbar, Z. S., Jelodar, N. B., Ghaffari, M. R., and Asari, S. (2019). Dissecting molecular mechanisms underlying salt tolerance in rice: a comparative transcriptional profiling of the contrasting genotypes. *Rice* 12:13. doi: 10.1186/s12284-019-0273-2
- Mu, C., Zhou, L., Shan, L., Li, F., and Li, Z. (2019). Phosphatase GhDPTP3a interacts with annexin protein GhANN8b to reversely regulate salt tolerance in cotton (*Gossypium* Spp.). *New Phytol.* 223, 1856–1872. doi: 10.1111/nph.15850
- Narasimha, A., Sivarajan, S. R., Makarala, U., and Nataraja, K. N. (2018). Identification of splice variant of OsGBF1 in *Oryza sativa* ssp. indica genotypes under salinity stress. *Biotech* 8:345. doi: 10.1007/s13205-018-1370-4
- Patade, V. Y., Maya, K., and Zakwan, A. (2011). A. Seed priming mediated germination improvement and tolerance to subsequent exposure to cold and salt stress in capsicum. *Res. J. Seed Sci.* 4, 125–136. doi: 10.3923/rjss.2011.125.136
- Peng, X., Zhao, Y., Cao, J., Zhang, W., Jiang, H., Li, X., et al. (2012). CCCH-type zinc finger family in maize: genome-wide identification, classification and expression profiling under abscisic acid and drought treatments. *PLoS One* 7:e40120. doi: 10.1371/journal.pone.0040120
- Pi, B., He, X., Ruan, Y., Jang, J. C., and Huang, Y. (2018). Genome-wide analysis and stress-responsive expression of CCCH zinc finger family genes in *Brassica rapa*. *BMC Plant Biol.* 18:373. doi: 10.1186/s12870-018-1608-7
- Pradhan, S., Kant, C., Verma, S., and Bhatia, S. (2017). Genome-wide analysis of the CCCH zinc finger family identifies tissue specific and stress responsive candidates in chickpea (*Cicer arietinum* L.). *PLoS One* 12:e0180469. doi: 10.1371/journal.pone.0180469
- Rosenzweig, C., and Liverman, D. (1992). *Global Climate Change: Implications, Challenges and Mitigation Measures*. Philadelphia, PA: Pennsylvania Academy of Sciences, 342–361.
- Rossi, L., Borghi, M., Francini, A., Lin, X., Xie, D. Y., and Sebastiani, L. (2016). Salt stress induces differential regulation of the phenylpropanoid pathway in *Olea europaea* cultivars Frantoio (salt-tolerant) and Leccino (salt-sensitive). *J. Plant Physiol.* 204, 8–15. doi: 10.1016/j.jplph.2016.07.014
- Seok, H. Y., Nguyen, L. V., Park, H. Y., Tarte, V. N., Ha, J., Lee, S. Y., et al. (2018). *Arabidopsis* non-TZF gene AtC3H17 functions as a positive regulator in salt stress response. *Biochem. Biophys. Res. Commun.* 15, 954–959. doi: 10.1016/j.bbrc.2018.03.088
- Sun, J., Hongling, J., Yingxiu, X., Hongmei, L., Xiaoyan, W., Qi, X., et al. (2007). The CCCH-type zinc finger proteins AtSZF1 and AtSZF2 regulate salt stress responses in *Arabidopsis*. *Plant Cell Physiol.* 178, 2622–2631. doi: 10.1093/pcp/pcm088
- Thiam, M., Champion, A., Diouf, D., and Mame Ourèye, S. Y. (2013). NaCl effects on in vitro germination and growth of some senegalese cowpea (*Vigna unguiculata* (L.) Walp.) cultivars. *ISRN Biotech.* 2013:382417.
- Wang, D., Guo, Y., Wu, C., Yang, G., Li, Y., and Zheng, C. (2008). Genome-wide analysis of CCCH zinc finger family in *Arabidopsis* and rice. *BMC Genomics* 9:44. doi: 10.1186/1471-2164-9-44
- Wang, P. H., Lee, C. E., Lin, Y. S., Lee, M. H., Chen, P. Y., Chang, H. C., et al. (2019). The glutamate receptor-like protein GLR3.7 interacts with 14-3-3 ω and participates in salt stress response in *Arabidopsis thaliana*. *Front. Plant Sci.* 10:1169. doi: 10.3389/fpls.2019.01169
- Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293
- Wolfe, S. A., Neklodova, L., and Pabo, C. O. (2000). DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* 29, 183–212. doi: 10.1146/annurev.biophys.29.1.183
- Wu, J., Wang, L., Fu, J., Chen, J., Wei, S., Zhang, S., et al. (2020). Resequencing of 683 common bean genotypes identifies yield component trait associations across a north–south cline. *Nat. Genet.* 52, 118–125. doi: 10.1038/s41588-019-0546-0
- Xu, R. (2014). Genome-wide analysis and identification of stress-responsive genes of the CCCH zinc finger family in *Solanum lycopersicum*. *Mol. Genet. Genomics* 289, 965–979. doi: 10.1007/s00438-014-0861-1
- Ye, J., Fang, L., Zheng, H. K., Zhang, Y., Chen, J., Zhang, J., et al. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34, W293–W297. doi: 10.1093/nar/gkl031
- Zhang, C., Zhang, H., Zhao, Y., Jiang, H., Zhu, S., Cheng, B., et al. (2013). Genome-wide analysis of the CCCH zinc finger gene family in *Medicago truncatula*. *Plant Cell Rep.* 32, 1543–1555. doi: 10.1007/s00299-013-1466-1466
- Zhang, H., Gao, X., Zhi, Y., Li, X., Zhang, Q., Niu, J., et al. (2019). A non-tandem CCCH-type zinc-finger protein, IbC3H18, functions as a nuclear transcriptional activator and enhances abiotic stress tolerance in sweet potato. *New Phytol.* 223, 1918–1936. doi: 10.1111/nph.15925

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

Copyright © 2020 Zhang, Zhang, Yin, Li, Zhao, Zhang, Zhuang, Wang, Zhang and Du. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Effects of Domestication on Secondary Metabolite Composition in Legumes

Yee-Shan Ku¹, Carolina A. Contador¹, Ming-Sin Ng¹, Jeongjun Yu², Gyuhwa Chung^{2*} and Hon-Ming Lam^{1*}

¹Centre for Soybean Research of the State Key Laboratory of Agrobiotechnology and School of Life Sciences, The Chinese University of Hong Kong, Shatin, China, ²Department of Biotechnology, Chonnam National University, Yeosu, South Korea

OPEN ACCESS

Edited by:

Jiayin Pang,
University of Western Australia,
Australia

Reviewed by:

Nicolas Rispaïl,
Spanish National Research Council,
Spain
Eric Von Wettberg,
University of Vermont, United States

*Correspondence:

Gyuhwa Chung
chung@chonnam.ac.kr
Hon-Ming Lam
honming@cuhk.edu.hk

Specialty section:

This article was submitted to
Evolutionary and Population
Genetics,
a section of the journal
Frontiers in Genetics

Received: 08 July 2020

Accepted: 31 August 2020

Published: 18 September 2020

Citation:

Ku Y-S, Contador CA, Ng M-S, Yu J,
Chung G and Lam H-M (2020)
The Effects of Domestication on
Secondary Metabolite
Composition in Legumes.
Front. Genet. 11:581357.
doi: 10.3389/fgene.2020.581357

Legumes are rich in secondary metabolites, such as polyphenols, alkaloids, and saponins, which are important defense compounds to protect the plant against herbivores and pathogens, and act as signaling molecules between the plant and its biotic environment. Legume-sourced secondary metabolites are well known for their potential benefits to human health as pharmaceuticals and nutraceuticals. During domestication, the color, smell, and taste of crop plants have been the focus of artificial selection by breeders. Since these agronomic traits are regulated by secondary metabolites, the basis behind the genomic evolution was the selection of the secondary metabolite composition. In this review, we will discuss the classification, occurrence, and health benefits of secondary metabolites in legumes. The differences in their profiles between wild legumes and their cultivated counterparts will be investigated to trace the possible effects of domestication on secondary metabolite compositions, and the advantages and drawbacks of such modifications. The changes in secondary metabolite contents will also be discussed at the genetic level to examine the genes responsible for determining the secondary metabolite composition that might have been lost due to domestication. Understanding these genes would enable breeding programs and metabolic engineering to produce legume varieties with favorable secondary metabolite profiles for facilitating adaptations to a changing climate, promoting beneficial interactions with biotic factors, and enhancing health-beneficial secondary metabolite contents for human consumption.

Keywords: legume, domestication, secondary metabolite, defense, health benefit

INTRODUCTION

Climate change, farmland deterioration, and the resulting food insecurity are major challenges facing the world. An increase in food supply is required to feed the expanding human population. The cultivation of high-yield crops has been used as a strategy to improve food supply. Grain legumes have been suggested as the potential solution to maintaining food and protein security (Considine et al., 2017). Legumes are also beneficial for sustainable agriculture due to the reduced release of greenhouse gases compared to other crops (Stagnari et al., 2017). Besides the beneficial effects on the improvement of soil fertility, legumes could enhance the resistance of soil to ecosystem disturbance, possibly due to the enhanced soil food web complexity (Gao et al., 2020).

In agriculture, legumes are common candidates for crop rotation for promoting the growth of other crops such as cereals (Bagayoko et al., 2000; Uzoh et al., 2019). In addition, legumes produce unique secondary metabolites such as isoflavones, which are beneficial to human health (Gepts et al., 2005; Ku et al., 2020). Legumes are known to protect humans from chronic diseases, including cardiovascular diseases, diabetes, obesity, osteoporosis, or even cancer (Kushi et al., 1999; Al-Anazi et al., 2011). Based on the mode of consumption, legumes can be classified into four groups: oil seeds, pulses, vegetable crops, and feed crops (McCrary et al., 2010). Examples of oil seeds are soybean and peanut (McCrary et al., 2010). Pulses are legumes, which are exclusively harvested as dry seeds, such as chickpea, lentils, and peas. Green bean and garden pea are examples of vegetable crops while clover and alfalfa are examples of feed crops (McCrary et al., 2010). Human selection of legumes during domestication has resulted in the alteration, and even loss of diversity, of secondary metabolite contents in these crops, directly and indirectly through the selection pressure on the genes that control the production of secondary metabolites. Understanding the differences in secondary metabolites, and the underlying genetic differences, between the domesticated legume cultivars and their wild progenitors would promote the preservation of legume accessions, which possess the genes for the biosynthesis of beneficial secondary metabolites. This knowledge will facilitate breeding programs and metabolite engineering to produce legume crops with favorable traits for adapting to the changing climate and for human pharmaceutical/nutraceutical use.

SEVERAL DOMESTICATION-RELATED TRAITS ALTERED THE SECONDARY METABOLITE CONTENTS

Domestication traits refer to morphological, biochemical, developmental, or physiological traits that are different between domesticated plants and their immediate wild progenitors (Abbo et al., 2014). A key part of domestication is the improvement of crop yield and harvestability compared to the wild progenitors (Dehaan et al., 2016). Several crop traits, including pod shattering, peduncle length, floral color, days to flowering, 100-seed weight, pod length, leaf length, leaf width, and seed number per pod, have been regarded as domestication-related traits (Lo et al., 2018).

Besides yield and harvestability related traits, other agronomic traits, such as seed size, appearance, and taste, are also subject to selection by breeders. These traits could be regarded as improvements due to post-domestication selection (Abbo et al., 2014). It has been suggested that the selection for larger seeds is related to facilitating single-seed planting (Kaplan, 1981). Breeders have also selected seeds of light colors. The ease of sowing and religious reasons have been proposed to be behind such conscious selections (Heiser, 1988). Therefore, seeds of modern legumes tend to have larger sizes and lighter colors compared to their wild counterparts. Moreover, the bitter taste of seeds has been intentionally

eliminated through breeding (Muzquiz et al., 1994). Behind the loss of bitter taste is the loss of the corresponding bitter-tasting secondary metabolites such as alkaloids (Muzquiz et al., 1994).

During domestication, secondary metabolite compositions which facilitate cultivation and improve the appearance and taste of food grains were intentionally selected for by breeders. In some cases, the secondary metabolite composition may be unintentionally selected due to the close proximity of the genes or quantitative trait loci (QTLs) for secondary metabolite biosynthesis to those regulating other traits such as major nutrients and yield. The selection of favorable cultivation areas and the protection by breeders during crop growth limit natural selection pressures due to abiotic and biotic stresses. Domestication brings forth better yield, better taste, and better appearance but also reduces the availability of secondary metabolites in legumes. As a result, domesticated legumes are usually less resistant to biotic stresses compared to their wild counterparts (Muzquiz et al., 1994; Pavan et al., 2016; Bazghaleh et al., 2018; Abraham et al., 2019). The reduced availability of health-beneficial secondary metabolites (Muzquiz et al., 1994; Wang et al., 2010; Fernández-marín et al., 2014; Kaur et al., 2019) also limits the potential of legumes as sources of bioactive compounds for pharmaceutical use. For the growth of the legume plants, the loss of the secondary metabolites in modern cultivars possibly renders the plants more susceptible to abiotic stress and biotic stress. The importance of the secondary metabolites to combating these stresses will be introduced in section “The Roles of Secondary Metabolites in Combating Abiotic and Biotic Interactions.”

INTRODUCTION TO SECONDARY METABOLITES IN LEGUMES

Definition of Plant Secondary Metabolites

Secondary metabolites are organic compounds derived from primary metabolism that serves key roles in defense and signaling in plants. They contribute to adaptive traits and ecological fitness, including defense mechanisms, tolerance to abiotic/biotic stresses, and interactions with insect pollinators, root-associated microbes, and herbivores. In contrast, primary metabolites are essential for cellular functions, such as growth, development, and reproduction. For example, secondary metabolites can attract insects for pollination or symbiotic rhizobia for nitrogen-fixing nodule formation. They can also be part of the defense mechanisms against herbivores, disease-causing bacteria, fungi, viruses, and parasites. There are also a wide range of secondary metabolites with pharmaceutical, nutraceutical, and toxicological values for humans (Wink, 2013). The contents of secondary metabolites vary among different plant species (Böttger et al., 2018). Legumes are rich in secondary metabolites, such as polyphenols, alkaloids, and saponins (Gupta, 1987).

The Health Benefits of Secondary Metabolites From Legumes

In the recent past, many secondary metabolites in legumes were considered non-nutritive. For example, tannins,

glycosides, alkaloids, and saponins affect the digestibility of beans (Gupta, 1987). However, more and more evidence suggests there are health benefits from the secondary metabolites of legumes (Dixon and Sumner, 2003). The health benefits of carotenoids, polyphenols, alkaloids, and saponins, all abundant in legumes, are discussed below and summarized in **Table 1**.

Carotenoids

Carotenoids are a type of tetraterpenoids, ranging from bright yellow and orange to red, found in algae, photosynthetic bacteria, and plants, including carrot, pumpkin, tomato, sweet potato, and papaya. They can be classified into two groups, carotenes and xanthophylls (Roberts et al., 2009). Xanthophylls differ from carotenes by having oxygenated substituents in their molecules (Roberts et al., 2009). Carotenoids with unsubstituted

β -rings, including α -carotene, β -carotene, and β -cryptoxanthin, act as provitamin A (Roberts et al., 2009). The carotenoid compositions in various legume seeds have been previously summarized (Tee et al., 1995). Lutein and zeaxanthin constitute the macular pigments in the retina of the mammalian eye. The oxygenated nature of the lutein and zeaxanthin molecules provides antioxidative protection for the eye from damage by free radicals (Roberts et al., 2009). The prevention of age-related macular degeneration by the consumption of carotenoid-rich foods has been recommended (Bernstein et al., 2016).

Polyphenols

Polyphenols are the major determinants of tissue colors, and generally possess antioxidative activities (Abbas et al., 2017). Polyphenols in plants can be classified into two groups: phenolic acids and flavonoids (Abbas et al., 2017). The occurrence and

TABLE 1 | Classification of secondary metabolites in legumes and their benefits to human health.

Groups	Sub-groups	Examples in legumes	Occurrence in legumes	Benefit(s) to human health	References
Polyphenols	Flavonoids	Quercetin, kaempferol	Widely distributed	Reduction in ischemic heart disease, reduction in body weight	(Knekt et al., 2002)
	Isoflavones	Genistein, daidzin	Soybean seeds	Phytoestrogen, antioxidant, antimicrobial and anti-inflammatory properties, reduction of risk in cardiovascular diseases, diabetes, obesity, and osteoporosis	(Křížová et al., 2019)
	Catechin	Catechin, epicatechin, gallo-catechin	Broad bean, chickpea, cowpea, kidney-bean, lentil, peanut	Reduction in heart disease, improvement of sperm motility and viability	(Arts et al., 2000; Hollman and Arts, 2000; Ojwang et al., 2013; Dias et al., 2016; López-cortez et al., 2016; Quintero-soto et al., 2018)
	Anthocyanins	Pelargonidin, cyanidin, malvidin, petunidin	Widely distributed	Antioxidant and anti-inflammatory properties, lipid peroxidation, DNA cleavage protection	(Acquaviva et al., 2003; Pietta et al., 2003; Rossi et al., 2012)
Terpenoids and steroid	Triterpenoid saponins	Saponins	Chickpea, soybean, lentils, peanut, common bean, and alfalfa sprouts	Reduction of cholesterol content, antimicrobial and anti-cancer properties	(Shi et al., 2004, 2014; Hassan et al., 2010; Man et al., 2010; Marrelli et al., 2016)
	Tetraterpenes	Carotenoids	Widely distributed	Antioxidant, better visual function, reduction of cardiovascular diseases	(Voutilainen et al., 2006; Roberts et al., 2009)
Alkaloids	Quinolizidine alkaloids (QA)	Sparteine	<i>Lupinus</i> spp.	Antimicrobial properties	(Romeo et al., 2018)
	Pyrrroloindole alkaloids	Physostigmine	Ordeal bean	Treatment of Alzheimer's disease and Parkinson's disease	(Zhu et al., 2014; Kumar et al., 2015)
Peptides	Polypeptide	Lunasin	Soybean	anti-inflammatory properties, reduction of cholesterol content, antioxidant, anticancer and anti-atherosclerotic activities	(Jeong et al., 2002, 2003, 2007, 2009; Hsieh et al., 2017; Fernández-tomé and Hernández-ledesma, 2019)
	Protease inhibitors	Angiotensin-I converting enzyme inhibitors	Pea, chickpea, mung bean, soybean, lentil	Lowering blood pressure and risk of heart failure	(Zhang et al., 2018)
Amines	Polyamine	spermine, spermidine	Common bean, white clover, mung bean	Antioxidant activities, reduction of cardiovascular diseases	(Soda, 2010; Menéndez et al., 2019; Muñoz-Esparza et al., 2019)

health benefits of phenolic acids in grain legumes have been previously summarized (Singh et al., 2017). Flavonoids are classified into several sub-classes: flavones, flavonols, flavanones, flavanonols, anthocyanins, flavanols, and isoflavones (Ku et al., 2020). Among flavonoids, isoflavones are only found in legumes. Flavonoids have multiple functions in plants, for example, mediating the responses to biotic and abiotic stresses, controlling the transport of auxins, acting as UV radiation-absorbing pigments to protect the plant against UV damage, attracting pollinating insects, interacting with rhizobia to initiate nodulation for symbiotic nitrogen fixation, and regulating defense against pathogens and herbivores through phytoalexin activities (Kumar and Pandey, 2013). For human health, it has been reported that flavonoids can act as protectants against cellular oxidation, inflammation, viral infections, and cancer (Kleemann et al., 2011). The molecular mechanisms of the health benefits of flavonoids have been recently reviewed (Ku et al., 2020).

Alkaloids

Alkaloids are nitrogen-containing organic heterocyclic compounds that are biologically active. Many alkaloids have pharmaceutical properties. For example, some alkaloids were found to have anti-malarial activities (Onguéné et al., 2013), anticancer activities (Gupta et al., 2015), and abilities to facilitate blood circulation in the brain and to prevent stroke (Kumar and Khanum, 2012). Moreover, several studies reported that alkaloids have potential therapeutic effects on neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, and Huntington disease (Amirkia and Heinrich, 2014).

Saponins

Saponins are a group of terpenoids found in plants, including onion, ginger, garlic, ginseng, fenugreek, and legumes (Oakenfull, 1981; Sauvaire et al., 1996). These crops are important sources of saponins in the human diet (Oakenfull, 1981; Sauvaire et al., 1996). Chickpea, soybean, lentils, peanut, garden pea, broad bean, and alfalfa are rich in saponins (Oakenfull, 1981). The antibacterial and foaming properties of saponins led to the use of saponins as vaccine adjuvants (Marciani, 2018). In the human body, saponins can bind to bile salts to reduce cholesterol absorption (Marrelli et al., 2016). Moreover, in rats, it was shown that a saponin-rich diet resulted in the reduction of body weight, total cholesterol, triglycerides, very-low-density lipoproteins (VLDL), and low-density lipoproteins (LDL) in serum (Latha et al., 2011; Reddy et al., 2012). Alfalfa saponin extract (ASE) was found to have cholesterol-lowering effects (Wang et al., 2011; Marrelli et al., 2016). The treatment of rats with ASE led to the enhanced expression of cholesterol 7- α -hydroxylase (Cyp7a1), an enzyme involved in the bile acid biosynthetic pathway in the livers of hyperlipidemic rats (Marrelli et al., 2016). Besides, ASE treatment also enhanced the expression of low-density lipoprotein receptor (Ldlr), which promotes the uptake and clearance of LDL cholesterol in plasma (Marrelli et al., 2016). Moreover, saponins also have anti-microbial and antioxidant properties, and exhibit cancer-related immunomodulatory effects (Avato et al., 2006).

The Roles of Secondary Metabolites in Combating Abiotic and Biotic Interactions Polyphenols

Plant roots communicate actively with the soil microbes for mutualistic cycles. Flavonoids are important signaling molecules for the legume-microbe interactions. The ability to form nitrogen fixing nodules with rhizobia is a unique characteristic of legumes (Hirsch et al., 2001). Such mutualism between legume and rhizobium is initiated by flavonoids. Flavonoids released from roots attract rhizobia to migrate toward the roots and stimulate the *nod* genes, which are essential genes to synthesize Nod factors for infecting the plants (Spaink, 1995). Flavonoids in the root exudates of various legumes for attracting rhizobia have been summarized in a previous review (Haldar and Sengupta, 2015). Moreover, flavonoids stimulate the germination of mycorrhizal fungus spores and enhance hyphal growth (Abdel-lateif et al., 2012). Mycorrhizal fungi form hyphae which penetrate plant roots for the transport of nutrients in rhizosphere to the host plant (Harrison, 2005).

The importance of polyphenols to combating abiotic stress has been discussed in recent reviews (Di Ferdinando et al., 2014; Isah, 2019; Sharma et al., 2019). The antioxidating characteristics of polyphenols help alleviate the oxidative stress brought forth by abiotic stress (Di Ferdinando et al., 2014; Isah, 2019; Sharma et al., 2019). A recent method for screening legume crops for abiotic stress tolerance suggested the accumulation of anthocyanin, which is also an osmolyte, as one of the indicators of abiotic stress tolerance of legume crops (Sinha et al., 2020).

Strigolactones

Based on the molecular structure, strigolactones belong to a group of lactone, which is derived from carotenoid (Jia et al., 2018). Functionally, strigolactones are plant hormones that are released by roots to attract symbiotic arbuscular mycorrhizal fungi and induce the germination of parasitic weed seeds (Jia et al., 2018). Strigolactones have been identified from a broad range of legumes, including *Arachis hypogaea*, *Astragalus sinicus*, *Cicer arietinum*, *Glycine max*, *Lupinus albus*, *Medicago sativa*, *Phaseolus vulgaris*, *Pisum sativum*, *Psophocarpus tetragonolobus*, *Trifolium incarnatum*, *Vicia faba*, and *Vigna angularis* (Yoneyama et al., 2008).

A study showed that the expression of several secretory proteins of *Rhizophagus irregularis*, an arbuscular mycorrhizal fungus, was induced by strigolactone treatment (Tsuzuki et al., 2016). Among these proteins, Strigolactone-Induced Putative Secreted Protein 1 (SIS1) showed the highest induction fold by both strigolactone treatment and *Medicago truncatula* root symbiosis. SIS1 is important for colonization and the formation of stunted arbuscules (Tsuzuki et al., 2016). Therefore, the strigolactone-induced is an essential protein for the symbiosis (Tsuzuki et al., 2016).

Broomrapes, especially *Orobancha crenata*, are believed to be the major parasitic weeds of legumes. The effects of the parasitic weeds on legumes include local damage of the plants and yield loss (Rubiales and Fernández-Aparicio, 2012). The germination of *Orobancha* seeds is induced by strigolactones (Yoneyama et al., 2008).

Alkaloids and Saponins

Alkaloids and saponins are known for contributing to the bitter taste of plants (Drewnowski and Gomez-Carneros, 2000). The toxicity of alkaloids has been reported (Wink, 2013). Several studies report that alkaloids and saponins are related to the resistance to herbivores. For example, yellow lupin cultivar with higher level of alkaloids in the leaves is more resistant to aphid than the cultivar with lower level of alkaloids (Adhikari et al., 2012). The removal of the bitter taste from modern lupin cultivars has enabled them to be a protein source in animal feed to reduce the dependence on soybean (Abraham et al., 2019). However, “sweet” lupins are more susceptible to predators (Muzquiz et al., 1994). Saponins have been thought to be responsible for the resistance to insect attacks, as the saponin preparation garden pea (*Pisum sativum* L.) resistant to Azuki bean beetle (*Callosobruchus chinensis* L.) inhibited the development of the beetle (Applebaum et al., 1969).

ALTERATIONS OF SECONDARY METABOLITE PROFILES IN LEGUMES DURING DOMESTICATION

Polyphenols and Carotenoids Determine the Colors of Seeds and Flowers

The seed coat color is mainly determined by polyphenols such as tannins (Heiser, 1988; Espinosa-Alonso et al., 2006). It is common for the pigmentation patterns of domesticated crops to be altered compared to their wild relatives. The loss of pigment in the seed coat of cultivated *P. vulgaris* is an obvious example of the effects of domestication (McClean et al., 2018). In a survey of 18 *Lablab purpureus* (*L. purpureus*) germplasms, including wild, semi-domesticated, and cultivated accessions, it was found that all the wild accessions have gray-brown and mottled seed coat (Maass, 2006). However, cultivated accessions display a spectrum of seed coat colors, including cream-white, cream, tan, and black (Maass, 2006). Unlike the wild accessions, some cultivated accessions do not have mottled seed coats (Maass, 2006). Among 11 landraces and two cultivated accessions of peanut (*Arachis hypogaea* L.), it was found that all the cultivated accessions have a single seed coat color: tan (Husain and Mallikarjuna, 2012), while the landraces are either red or tan (Husain and Mallikarjuna, 2012). Some landraces even have variegated seed coats (Husain and Mallikarjuna, 2012). In another study, it was shown that cultivated peanut (*A. hypogaea*) could have purple, brown, red, or white seed coats and some have variegated seed coats (Bertioli et al., 2011). In a survey of a soybean population consisting of 1,957 domesticated and 1,079 wild accessions, it was found that almost all wild accessions have purple flowers and black seed coats (Jeong et al., 2019), whereas the domesticated soybean accessions have more diverse seed coat colors, including colorless (yellow or green seeds), brown, or black, and more diverse floral colors, including white or purple (Jeong et al., 2019). In another study on 110 cultivated, 130 landrace, and 62 wild soybean accessions, it was reported that all cultivated accessions

have yellow seeds, and landrace accessions have yellow, green, brown, or black seeds, while all the wild accessions have black seeds (Wang et al., 2018). Similarly, the modern cultivated pea cv. Cameor (*P. sativum*) has transparent seed coat while the wild accession (*P. sativum* subsp. *elatius* JI64) has pigmented seed coat (Smýkal et al., 2014). In another study on cultivated (*Lens culinaris* ssp. *Culinaris*) and wild lentils (*Lens culinaris* ssp. *orientalis*, *L. culinaris* ssp. *odemensis*, *L. culinaris* ssp. *tomentosus*, *Lens nigricans*, and *Lens ervoides*, *Lens lamottei*), although wild accessions do not necessarily have darker seed coats, wild accessions have more complexed patterns on the seed coats (Singh et al., 2014). The seed coats of the cultivated accessions have either no or dotted patterns (Singh et al., 2014). However, many of the wild accessions have marbled pattern on seed coats (Singh et al., 2014). For chickpea (*C. arietinum*), the light color of the cultivated seeds is thought to be non-existing in wild accessions (Penmetsa et al., 2016). The seed coat color is related to the defense against herbivore. It has been suggested that a black seed coat protects the seed from night-time foragers (Porter, 2013).

Polyphenols also give rise to the colors of flowers (Wiesner et al., 2017). In cowpea, cultivated accessions have a wide range of floral colors while most of the wild accessions have only purple flowers (Lo et al., 2018). Similarly, cultivated soybean accessions have purple, white, or other colors of flowers (Sundaramoorthy et al., 2015; Jeong et al., 2019), whereas most of the wild soybean accessions have only purple flowers (Sundaramoorthy et al., 2015). On the contrary, in common bean, most of the cultivated accessions have only white flowers while the wild accessions have white, pink, or purple flowers (García et al., 1997). Cultivated lentils (*L. culinaris* ssp. *Culinaris*) have white or purple flowers but some wild lentils, *L. culinaris* ssp. *odemensis*, *L. culinaris* ssp. *tomentosus* and *Lentil ervoides*, have only purple flowers (Singh et al., 2014). The white flowers of cultivated chickpea (*C. arietinum*) is thought to be non-existing in wild accessions (Penmetsa et al., 2016). For pea, cultivated peas (*P. sativum*) usually have white flowers while purple flowers are found in wild peas (Hellens et al., 2010). The contrasting flower colors contributed to the establishment of the Mendel's Laws.

The Co-Evolution of Seed and Floral Colors With Foragers and Pollinators

As discussed above, cultivated legumes usually have lighter seed coat colors compared to the wild counterparts. During domestication, light seed coat colors have been preferred by farmers. The loss of color is associated with the loss of secondary metabolites, such as tannins (Heiser, 1988). As mentioned before, a dark seed coat may protect the seeds from night-time foragers in the wild (Porter, 2013). However, the potential increase in loss of sown seeds to wild animals may not be significant as farmers usually have measures to keep foragers away from crops. Another example is the loss of bitter compounds, such as alkaloids and saponins in domesticated legumes. The loss of such compounds would have enhanced the loss of seeds due to foraging by animals and is usually not advantageous for the survival of the crops without the protection provided by farmers. Therefore, the loss of bitter compounds in domesticated legumes is also

known as a conscious selection by breeders during domestication. In soybean, most of the elite cultivated soybean seeds are yellow. It was found that the stay-green *G* gene is associated with green seeds and it controls seed dormancy, but is lost in elite cultivated soybean seeds (Wang et al., 2018). In the survey of 110 cultivated, 130 landrace, and 62 wild soybean accessions, it was found that the *G* genotype is present in only 4% of the cultivated accessions, 21% of the landraces while it is found in 100% of the wild accessions (Wang et al., 2018).

It has been suggested that floral color has co-evolved with pollinators such as birds and bees. Bees tend to be attracted to yellow flowers while birds tend to prefer red flowers due to their different visual sensitivities (Toon et al., 2014). Bee-pollinated plants usually have yellow, white, or blue flowers while bird-pollinated plants usually have red flowers (Toon et al., 2014). The transition from bee-pollination to bird-pollination of Australian egg-and-bacon pea is related to the number of bird species in the geographical region where the plants grow (Toon et al., 2014). The yellow color of the *Lotus* flower, together with the orientation, size, petal morphology, sucrose-dominant nectar composition, and scent of the flower, was reported as a factor contributing to the transition to pollination by birds (Cronk and Ojeda, 2008).

Carotenoid Level Is Related to Seed Dispersal by Animals

Besides polyphenols, carotenoids also play a role in determining tissue colors. During domestication, the profitability of seeds is a major concern for farmers. Therefore, genotypes with reduced seed dispersal, including through pod shattering and seed dispersal by animals, were actively selected for by breeders and farmers. In a study on the seeds of 10 legume genera: *Arachis* (peanut), *Cicer* (chickpea), *Glycine* (soybean), *Lathyrus* (vetch), *Lens* (lentil), *Lupinus* (lupin), *Phaseolus* (bean), *Pisum* (pea), *Vicia* (fava bean), and *Vigna* (cowpea), drastic changes in the levels and compositions of carotenoids in seeds were found in domesticated cultivars compared to their wild counterparts (Fernández-marín et al., 2014). An average of 48% reduction in carotenoids was found in the seeds of these 10 legumes. Besides, the compositions of carotenoids were more complex in the wild species of *Cicer*, *Glycine*, *Lathyrus*, *Lens*, *Lupinus*, and *Vigna*. In the study, neoxanthin, violaxanthin, lutein epoxide, and antheraxanthin were only found in the wild species but not the domesticated varieties. It was suggested that seeds with lower carotenoid levels are less attractive to seed dispersers (Fernández-marín et al., 2014). In contrast, attracting seed dispersers has been suggested to be an adaptation of wild legumes (Brónnviik and von Wettberg, 2019). It was suggested that seed dispersal by birds is an important factor contributing to the widespread of *P. vulgaris* from Mexico to South America (Brónnviik and von Wettberg, 2019).

Isoflavones Are Unique to Legumes

Isoflavones are a sub-class of flavonoid uniquely found in legumes. Soybean is a rich and common source of isoflavones for human consumption (Ku et al., 2020). In a study of seed isoflavone contents using 209 wild, 580 landrace, and 106

cultivated soybean accessions, it was found that landraces had the highest average level of total seed isoflavone, followed by wild accessions and then cultivated accessions (Wang et al., 2010). The higher average total seed isoflavone content in landraces compared to cultivated accessions was also reported in another study using 927 landraces and 241 cultivars (Azam et al., 2020). For individual isoflavone contents, it was suggested that high genistin and glycitin contents, with low daidzin levels, were artificially selected for. The significantly lower daidzin contents lead to the lower average total seed isoflavone levels in cultivated accessions compared to wild accessions (Wang et al., 2010). There are debates over the reasons behind the artificial selection of such seed isoflavone traits in domesticated legumes. Regarding seed nutrient content, a negative correlation between the total isoflavone level and the protein level has been reported in seeds (Primomo et al., 2005; Morrison et al., 2008; Liang et al., 2010; Smallwood et al., 2014), and a positive correlation between total seed isoflavone level and seed oil level has also been reported (Morrison et al., 2008; Liang et al., 2010). However, there has also been a report on the negative correlation between total seed isoflavone level and seed oil level (Smallwood et al., 2014). The total seed isoflavone has also been correlated to yield (Primomo et al., 2005; Smallwood et al., 2014; Zhang et al., 2014), as well as the resistance against pathogens (Carter et al., 2018). When two soybean cultivars, RCAT1004 and DH4202, which are resistant and sensitive to cyst nematodes respectively, were grown in a cyst nematode-infested environment, the resistant cultivar had a higher seed isoflavone level (Carter et al., 2018). A putative QTL related to cyst nematode susceptibility was found close to that related to total seed isoflavone content (Carter et al., 2018). During domestication, besides the deliberate selection for reduced seed isoflavone level to reduce the bitterness of the seed, the isoflavone level may also be unintentionally selected together with other desirable traits, such as nutrient composition, yield, and resistance to biotic stress.

Alkaloid and Saponin Contents Are Related to Taste-Focused Breeding

The bitter taste of legume seeds tends to be eliminated during domestication. For example, domesticated lupin cultivars are less bitter than the wild relatives, which have significantly higher levels of alkaloids in their seeds. Modern lupin cultivars are referred to as “sweet” lupins. In a survey of 20 sweet lupins and 29 bitter lupins, the bitter taste of lupins was found to be positively correlated to the seed alkaloid content, with lupanine being the main alkaloid (Muzquiz et al., 1994). Although seed saponin level has been correlated to the bitterness of seeds in general (Mohan et al., 2016), it may not be related to the bitterness of lupin seeds. In a survey of the seed saponin contents in sweet vs. bitter lupins, the level of saponin was undetectable in the seeds of both sweet and bitter varieties of *L. albus* (Shim et al., 2003).

Saponin is also a contributing factor to the bitterness of seeds (Okubo et al., 1992). It was found that many of the wild ancestors of *Vigna* spp. are more resistant than their cultivated counterparts to *Callosobruchus chinensis* or *Callosobruchus maculatus* (Tomooka et al., 2000). It is possible

that the drop in saponin contents when legumes became domesticated is related to the loss of insect resistance capability in cultivated species. Several wild chickpea accessions have higher seed saponin levels than cultivated chickpea accessions (Kaur et al., 2019). Several wild pigeonpea (*Cajanus scarabaeoides*) accessions have higher seed saponin contents than the cultivated pigeonpea accessions (*Cajanus cajan*; Sekhon et al., 2017). However, cultivated pigeonpea accessions do not necessarily have lower seed saponin contents than the wild accessions (Sekhon et al., 2017). Seed saponin content is not the sole factor leading to the insect resistance of legumes.

Besides total seed saponin content, individual saponin components in legumes are also studied. Saponins can be classified into four groups: group A saponins, group B saponins, group E saponins, and 2,3-dihydro-2,5-dihydroxy-6-methyl-4H-pyran-4-one (DDMP) saponins (Sawai and Saito, 2011; Krishnamurthy et al., 2013). The aglycone form of group A saponins is named soyasapogenol A, while that of DDMP saponins is named soyasapogenol B. The basic structure of soyasapogenol A and B is β -amyrin. Soyasapogenol A is a β -amyrin with a hydroxyl group at C-21, C-22, and C-24, while soyasapogenol B has a hydroxyl group at C-22 and C-24 only (Sawai and Saito, 2011). DDMP saponins are relatively unstable and are often degraded into group B and group E saponins during food processing (Sundaramoorthy et al., 2019). Among the various groups of saponins, group A saponins, which have an acetylated oligosaccharide chain attached to C-22 of soyasapogenol A, are thought to be mostly responsible for the undesirable taste of soybean seeds (Shiraiwa et al., 1991).

In a survey of saponin compositions among 800 cultivated soybean accessions and 329 wild soybean accessions, it was found that the saponin type Aa was predominant in cultivated soybean accessions, while the saponin type AaBc was predominant in wild soybean accessions (Tsukamoto et al., 1993). In another survey of the total seed saponin levels in 17 wild and one cultivated legumes, it was found that the total saponin level was highest in *Glycine soja* (*G. soja*; wild soybean; Shim et al., 2003). In a study of seed saponin composition of 3,025 *G. soja* accessions, diverse compositions of seed saponins were found among the accessions (Krishnamurthy et al., 2013). Moreover, naturally occurring wild soybean mutants that lack group A saponins were found (Krishnamurthy et al., 2013; Takahashi et al., 2016; Rehman et al., 2018). Wild legumes do not necessarily have higher seed saponin contents. Instead, the diverse genetic backgrounds among wild legumes allow the discovery of novel allelic forms for desirable seed saponin compositions.

Polyphenols and Strigolactones Are Related to Biotic Interactions

Flavonoids are signaling molecules for legume-microbe interaction (Abdel-lateif et al., 2012). In a test of nodulating capability of *Rhizobium japonicum* (*R. japonicum*), it was found that all the strains of *R. japonicum* in the test could nodulate cowpea, sirato, and wild soybean (Heront and Pueppket, 1984). However, nine out of the 11 strains could not form infection threads with two of the three commercial soybean cultivars in the test (Heront and Pueppket, 1984). In another study, after inoculating

36 *G. soja* (wild soybean) accessions with *R. japonicum*, 20 formed normal nodules while 16 could not form nodules or formed abnormal nodule-like structures (Pueppke et al., 1998). It was hypothesized that the different nodulating phenotypes were due to the different flavonoid profiles in the root exudates (Pueppke et al., 1998). However, the flavonoid profiles of root exudates are similar between the nodulating group and the non-nodulating group (Pueppke et al., 1998). The flavonoid profiles of root exudates were also compared between wild soybean accessions and the cultivated soybean Peking (Pueppke et al., 1998). Although many of the wild soybeans showed a more complexed root exudate profile, a strong correlation between the different root exudates and the nodulating phenotypes was not found (Pueppke et al., 1998). The effects of domestication on the flavonoid profiles in legume root exudates remain unclear. On the other hand, the root polyphenol compositions of wild lentil (*Lens ervoides*) and cultivated lentil (*Lens culinaris*) were compared after the infection of *Aphanomyces euteiches*, which is a legume pathogen (Bazghaleh et al., 2018). The wild lentil was more tolerant to *A. euteiches* than the cultivated lentil pathogen (Bazghaleh et al., 2018). The wild lentil generally had higher levels of polyphenols compared to the cultivated lentil (Bazghaleh et al., 2018). Although the amount of legume species and accessions is not enough to conclude the effect of domestication on the root polyphenol compositions after pathogen infection, genotypic difference exists between wild and cultivated legumes and is associated with polyphenol accumulation in roots under biotic stress.

Strigolactones are stimulants of seed germination (Brun et al., 2018). The yield of faba bean (*V. faba*) is also limited by parasitic weeds. Faba bean (*V. faba*) germplasms resistant to parasitic weeds, broomrape (*Orobancha* and *Phelipanche* spp.) were found (Fernández-Aparicio et al., 2014). The resistant germplasms have low or undetectable levels of strigolactones in the root exudates at all plant ages (Fernández-Aparicio et al., 2014). It was suggested that the screening of germplasms with low strigolactone levels in root exudates is a strategy to breed for weed resistant germplasms. Like faba bean (*V. faba*), most of the commercial pea (*Pisum sativum* L.) cultivars are susceptible to the attack by crenate broomrape (*O. crenata* Forsk.), which is a parasitic weed of legumes (Pavan et al., 2016). In a screen of *O. crenata* resistant pea germplasms, a landrace pea germplasm was selected. Repeated self-pollination of the landrace germplasm resulted in the *O. crenata* resistant line ROR12 (Pavan et al., 2016), which exhibited several unique characters: (1) compared to a *O. crenata* susceptible cultivar, the root exudates of ROR12, which had a lower strigolactone level, had a lower capability to stimulate the germination of *O. crenata* seeds; (2) in the field, the number of *O. crenata* shoots per host plant of ROR12 was lower; and (3) the emergence of *O. crenata* on ROR12 was delayed. It was proposed that the resistance to *O. crenata* was related to the reduced strigolactone level in the root exudates (Pavan et al., 2016).

Strigolactones are also involved in legume-microbe interaction. The treatment of synthetic strigolactone (GR24) to pea (*Pisum sativum* L.) roots enhanced the nodule number on the roots due to *Rhizobium leguminosarum* bv. *viciae*

(RLV248) inoculation (Foo and Davies, 2011). Mutant *rms1* of pea (*Pisum sativum* L.) had undetectable levels of orobanchol and orobanchyl acetate and a low level of fabacyl acetate in the root exudates (Foo and Davies, 2011). Compared to the wild type, *rms1* mutant had less nodules on the roots after being inoculated with *R. leguminosarum* bv. *viciae* (RLV248; Foo and Davies, 2011). Commercial legume germplasms usually have lower levels of strigolactones in the root exudate (Fernández-Aparicio et al., 2014; Pavan et al., 2016).

The low levels of strigolactones may result in the reduced number of nodules on the roots. However, the nodulating phenotype may not be of consideration during domestication as the application of nitrogen fertilizer is a common practice during domestication.

GENES THAT REGULATE THE DIFFERENT SECONDARY METABOLITE-RELATED TRAITS

Several methods are currently being employed to identify the genes and mutations underlying legume domestication phenotypes (Olsen and Wendel, 2013). In general, secondary metabolites are present at higher levels in wild progenitors than in the domesticated counterparts as a result of artificial selection (Nagl et al., 1997; Lindig-Cisneros et al., 2002; Gepts, 2014). The selection of cultivars based on ease of farming and other commercial attributes may have occurred at the expense of potentially beneficial secondary metabolites. The reduction in genetic diversity is one of the main impacts of domestication. However, the genetic richness of wild populations can be used to improve cultivated legumes. Traditional plant breeding is a millenary process for the improvement and development of new crop varieties. According to breeding objectives, new legume varieties are produced by crossing parents with desired traits and selecting among segregating progenies those individuals with both high yield and the target trait. In this way, pest-resistant varieties have been developed with genetic resistance to pathogens (Lavaud et al., 2015). Traits related to pigmentation and defense against pathogens or herbivores are characteristically domestication-related traits governed by secondary metabolites. Besides biosynthesis-related genes, transport-related genes are also important. The roles of transporters, including ATP-binding cassette (ABC) transporters and multidrug and toxic compound extrusion (MATE) transporters in secondary metabolite secretion and accumulation have been summarized in previous reviews (Yazaki, 2005; Ku et al., 2020). In this section, examples of genes and loci controlling secondary metabolite biosynthesis and transport in legumes will be discussed.

Biosynthesis-Related Genes Pigmentation-Related Traits

Polyphenols are the major determinants of tissue colors, including the colors of seed coats of legumes, both by their presence and their quantities (Espinosa-Alonso et al., 2006). The major polyphenols responsible for seed coat color in

legumes are flavonoids, such as anthocyanins, flavonol glycosides, and proanthocyanidins (condensed tannins). Flavonoid quantities vary according to the seed developmental stages, genotypes, and species. The biosynthetic pathway leading to the biosynthesis of flavonoids has been elucidated and is conserved among seed-producing plants. Flavonoids and isoflavonoids are derived from the phenylpropanoid pathway (Dastmalchi and Dhaubhadel, 2014). Many genes in this pathway, including enzymes, transporters, and regulatory factors, have been characterized. The first committed step is the formation of a bicyclic tetrahydroxy chalcone (naringenin chalcone) catalyzed by a chalcone synthase (CHS). Legumes produce an additional trihydroxy chalcone (THC), isoliquiritigenin chalcone (Dastmalchi and Dhaubhadel, 2014). This THC is the end product of the coupled activities of CHS and the legume-specific chalcone reductase (CHR). Compounds such as daidzein, medicarpin, and glyceollin are derived from isoliquiritigenin. Flavonoid production follows the conversion of naringenin chalcone to (2S)-naringenin by chalcone isomerase (CHI). Flavone 3-hydroxylase (F3H) catalyzes the hydroxylation of (2S)-naringenin, eryodictyol, and pentahydroxyl flavanones to yield (2R,3R)-dihydrokaempferol, dihydroquercetin, and dihydromyricetin, respectively (Tanaka et al., 2008). Flavonoid 3'-hydroxylase (F3'H) and flavonoid 3',5'-hydroxylase (F3'5'H) catalyze the hydroxylation of flavanones, flavanols, and flavones, and determine the structures of flavonoids and anthocyanins (Tanaka, 2006). Other enzymes in the pathway include dihydroflavonol 4-reductase (DFR) and anthocyanidin synthase (ANS). The biosynthesis pathway of flavonoids is illustrated in **Figure 1**.

Pigmentation mechanisms have been studied in different legumes. A transcriptomic analysis was performed to identify the genes associated with seed coat color in peanut (*A. hypogaea*; Wan et al., 2016). Lower proanthocyanidin and anthocyanin contents were detected in a peanut mutant with a brown cracking seed coat (*pscb*). Transcriptomic analyses revealed that the structural genes of the phenylpropanoid biosynthetic pathway were downregulated in the *pscb* mutant, while the genes related to melanin production were upregulated at the late developmental stages. This expression pattern was consistent with the higher melanin content in the *pscb* mutant compared to the wild type. Differential expression analyses of RNA-seq data between the wild type and *pscb* mutant revealed three candidate genes (*c36498_g1*, *c40902_g2*, and *c33560_g1*) as being responsible for the seed coat color trait. *C33560_g1* encodes a R2R3-MYB transcription factor. Its homologs in *Arabidopsis* and apple are associated with the regulation of the phenylpropanoid biosynthesis pathway (Rowan et al., 2009; Vimolmangkang et al., 2013). *C36498_g1* and *c40902_g2* encode a caffeoyl-CoA O-methyltransferase and a kinesin-4-like protein, respectively. Putative functions of the encoded proteins were associated with cell wall organization.

Soybeans cultivated for the commercial market are either completely yellow or have pigmentation restricted to the hilum (Palmer et al., 2004). Wild soybeans accumulate flavonoids and anthocyanins within the entire epidermal layer of the seed coat, giving them a black or brown color (Todd and Vodkin, 1993; Song et al., 2016). Quantitative trait loci (QTL) governing

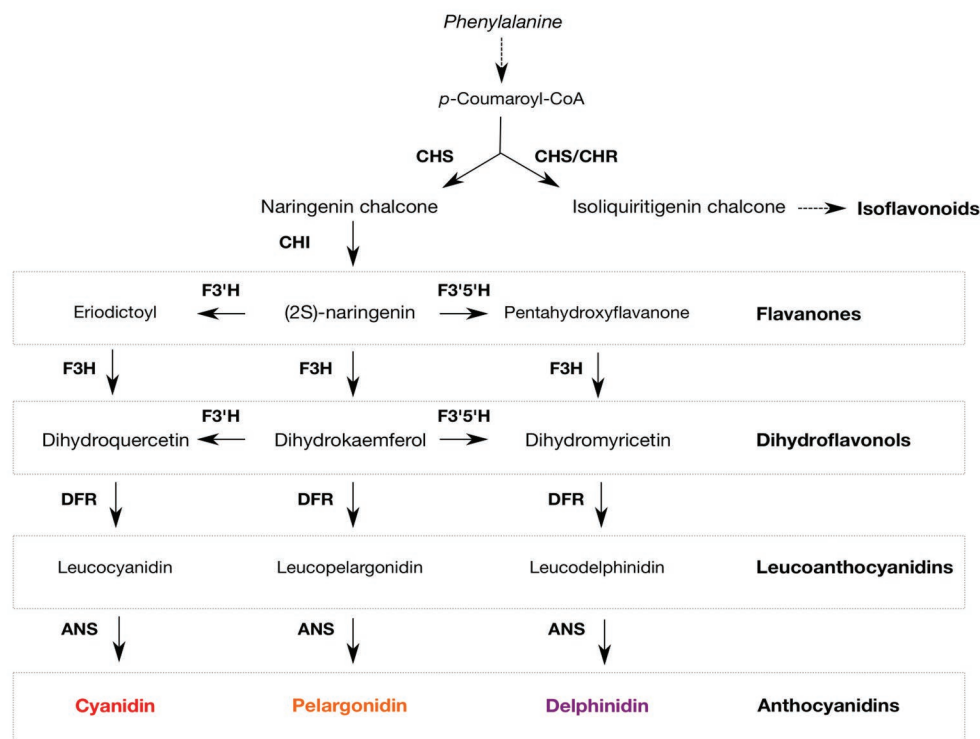


FIGURE 1 | Schematic representation of the flavonoid biosynthetic pathway. Enzymes involved in the pathway are indicated in bold: chalcone synthase (CHS), chalcone reductase (CHR), flavone 3-hydroxylase (F3H), flavonoid 3'-hydroxylase (F3'H), flavonoid 3',5'-hydroxylase (F3'5'H), dihydroflavonol 4-reductase (DFR), and anthocyanidin synthase (ANS).

seed coat color in soybean have been identified using genetic and genomic analyses to elucidate the genetic changes that resulted in this domestication trait (Todd and Vodkin, 1996; Tuteja et al., 2004, 2009; Song et al., 2016). The *I*, *R*, and *T* loci were found to be involved in the flavonoid biosynthesis pathway (Palmer et al., 2004; Yang et al., 2010). The *I* locus on chromosome eight inhibits pigmentation of the seed coat. There are four alleles (*I*, *i^h*, *i^k*, and *i*) at the *I* locus where *I* and *i^h* are the two dominant forms (Song et al., 2016). The presence of the *I* allele results in the absence of pigmentation and a yellow seed coat at maturity. This allele contains an inverted repeat of the *CHS* gene cluster. This structure triggers posttranscriptional gene silencing (PTGS), which inhibits the expression of *CHS* gene family members and their functions in the flavonoid biosynthesis pathway (Tuteja et al., 2004). The *i^h* allele inhibits pigmentation, resulting in a yellow seed coat with a pigmented hilum (Palmer et al., 2004). Meanwhile, the recessive *i^k* and *i* alleles allow pigment production, with the *i^k* allele restricting pigments to the saddle and hilum regions of the seed coat (Palmer et al., 2004). The *R* and *T* loci determine the type and accumulation of pigments in the seed coat (Buzzett et al., 1987; Todd and Vodkin, 1993). Higher flavonoid and anthocyanin contents of seeds are currently of great interest due to the antioxidant properties and flavors of these compounds. Recently, the wild soybean reference genome of *G. soja* W05 was used to identify additional alleles of the

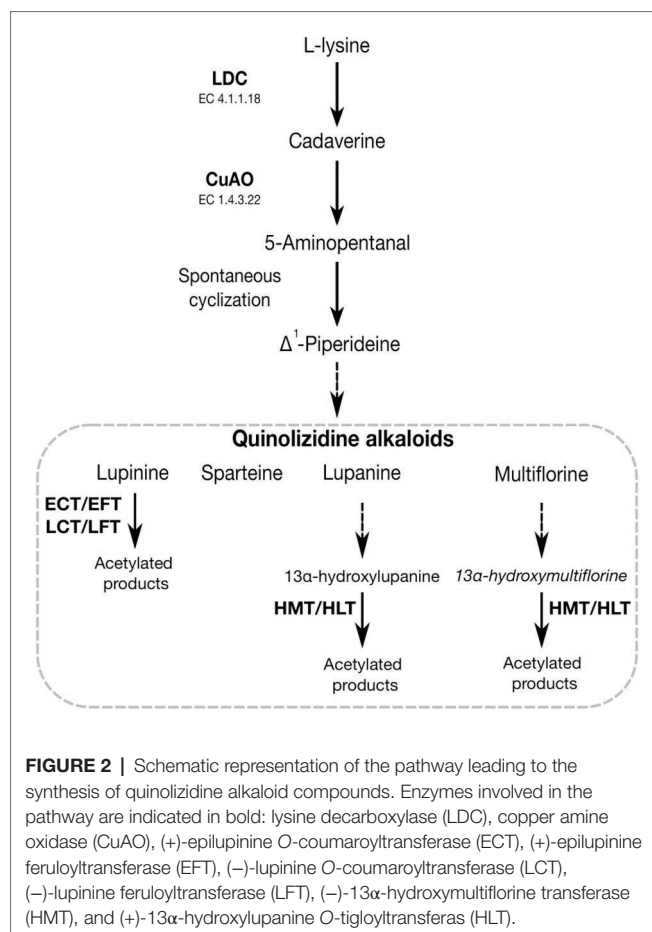
causal structural gene variation that controls soybean seed coat pigmentation (Xie et al., 2019). The analysis of a seed coat color QTL that overlaps with the known *I* locus showed that the W05 reference genome possesses the same inverted repeat of the *CHS* gene cluster as the domesticated soybean reference genome, *G. max* (Williams 82). This indicates that additional factors also played a role in causing the seed color changes during domestication. A comparative genomic analysis of W05 against two domesticated soybeans (Wm82 and ZH13) revealed the generation of a small interfering RNA (siRNA) from a large structural rearrangement next to the *CHS* gene cluster in Wm82 and ZH13. Through experimental validation, a subtilisin promoter was shown to drive the expression of a chimeric transcript that reads through a subtilisin gene fragment and an anti-*CHS1* gene region, resulting in PTGS and inhibits the expression of *CHS* genes.

Flavonoids also contribute to floral pigmentation (Tanaka, 2006; Tanaka et al., 2008). Domesticated cowpea (*Vigna unguiculata* L. Walp) shows phenotypic variation compared to its wild relatives. Among the domestication traits, a wide range of floral and seed coat colors can be found in the cultivated cowpea. The wild variety shows purple flowers and dark seed pigmentation. Purple flowers are the results of diacylated delphinidin-based anthocyanins (Tanaka et al., 2008). A QTL analysis of the determinants of floral color in cowpea was performed in a biparental mapping population (wild × cultivated crosses;

Lo et al., 2018). A single major QTL for floral color, CFcol7, was mapped in a 64-cM region on chromosome Vu07 containing 254 annotated genes, among which a transcription factor, *Vigun07g110700*, was identified as a homolog of *Arabidopsis* AT4G09820.1 and *Medicago truncatula* (Mt) TT8, involved in the regulation of flavonoid biosynthesis (Nesi et al., 2000; Li et al., 2016). In soybean, one QTL for floral pigmentation was identified on linkage group G (Josie et al., 2007).

Defense-Related Traits

Toxic secondary metabolites in legumes confer resistance against pathogens and herbivores. However, the accumulation of these compounds in legume crops is not desirable for human consumption or as animal feed since they impart a bitter taste and could present acute toxicity if ingested in sufficient quantities (Daverio et al., 2014). Alkaloids are some of the main secondary metabolites produced and stored by legumes with characteristic toxicity (Wink, 2013). Examples are quinolizidine alkaloids (QAs) produced by the genera *Lupinus*, *Baptisia*, *Thermopsis*, *Genista*, *Cytisus*, *Echinosophora*, and *Sophora* (Ohmiya et al., 1995). The breeding of low-alkaloid (sweet) varieties changed the agronomic roles of lupins from green manure and forage crops to grain legumes with high protein and fiber contents and health benefits (Sweetingham and Kingwell, 2008; Arnoldi et al., 2015). Four species within the genus *Lupinus* have been domesticated and are important legume crops: *Lupinus angustifolius* (narrow-leaved lupin or blue lupin), *L. albus* (white lupin), *Lupinus luteus* (yellow lupin), and *Lupinus mutabilis* (Andean lupin) (Gustafsson and Gadd, 1965; Reinhard et al., 2006). QAs produced by lupins include lupanine, angustifoline, lupinine, sparteine, multiflorine, and aphylline (Frick et al., 2017). The use of lupins for food purposes depends on their QA levels and each species has a characteristic alkaloid composition. Domestication has reduced the amount of alkaloids in lupins, but it has also increased their susceptibility to several aphid species (Philippi et al., 2015). QAs are derived from the decarboxylation of L-lysine (Lys) by a lysine decarboxylase (LDC, EC 4.1.1.18) to form cadaverine (Bunsupa et al., 2012a,b), which is then converted to 5-aminopentanal via oxidative deamination by a copper amine oxidase (CuAO, EC 1.4.3.22). 5-aminopentanal is spontaneously cyclized to a Δ^1 piperidine Schiff base formation (Leistner and Spenser, 1973; Gupta et al., 1979; Golebiewski and Spenser, 1988; Bunsupa et al., 2012b), which then undergoes further modifications (e.g., oxygenation, dehydrogenation, hydroxylation, or esterification) to produce a range of Lys-derived alkaloids, including lupinine, sparteine, lupanine, and multiflorane (Ohmiya et al., 1995; Bunsupa et al., 2012b; Frick et al., 2017). QAs are stored in the form of QA esters. In lupins, QA esters are converted from lupinine/epilupinine and 13 α -tigloyloxymultiflorine/13 α -tigloyloxylupanine by two types of acetyltransferases (ATs): (+)-epilupinine/(−)-lupinine O-coumaroyl/feruloyltransferase (ECT/EFT-LCT/LFT) and (−)-13 α -hydroxymultiflorine/(+)-13 α -hydroxylupanine O-tigloyltransferase (HMT/HLT; EC 2.3.1.93), respectively (Saito et al., 1992; Okada et al., 2005; Bunsupa et al., 2012a). The biosynthesis pathway of QAs is illustrated in Figure 2. An HMT/HLT-type acetyltransferase was isolated and characterized at the molecular level in *L. albus* (Okada et al., 2005) while an



acyltransferase-like gene, *Lupinus angustifolius* acyltransferase (*LaAT*), has been proposed to be involved in the QA biosynthetic pathway (Bunsupa et al., 2011).

Domestication-related genetic modifications resulting in low-alkaloid phenotypes are generally results of naturally occurring (spontaneous) mutations (Gustafsson and Gadd, 1965). The domestication of lupins led to the active selection by farmers/breeders for sweet varieties which were low in alkaloids. In the late 1920s, the first low-alkaloid lines were obtained from wild germplasms of *L. luteus* and *L. angustifolius* (von Sengbusch, 1942). Subsequently, sweet types were also obtained for *L. albus* and *L. mutabilis* in the 1930s (Taylor et al., 2020). Several recessive low-alkaloid mutations have been discovered in *L. angustifolius*: *iucundus* (*iuc*), *esculentus* (*es*), *depressus* (*depr*), and *tantalus* (Swiecicki and Swiecicki, 1995; Kurlovich, 2002; Taylor et al., 2020), among which, the *iucundus* locus is the most prevalent allele in cultivars (Taylor et al., 2020). Molecular mapping efforts have allowed researchers to map the *iucundus* locus to a 746-kb region on chromosome NLL-07 (Nelson et al., 2006, 2010; Hane et al., 2017). The reference *L. angustifolius* genome also facilitated the identification of markers linked to *iucundus* that are suitable for marker-assisted selection (MAS). Specifically, an allele marker, *IucLi*, has been identified for the *iucundus* locus, and could be used for MAS in wild × domesticated crosses in lupin breeding

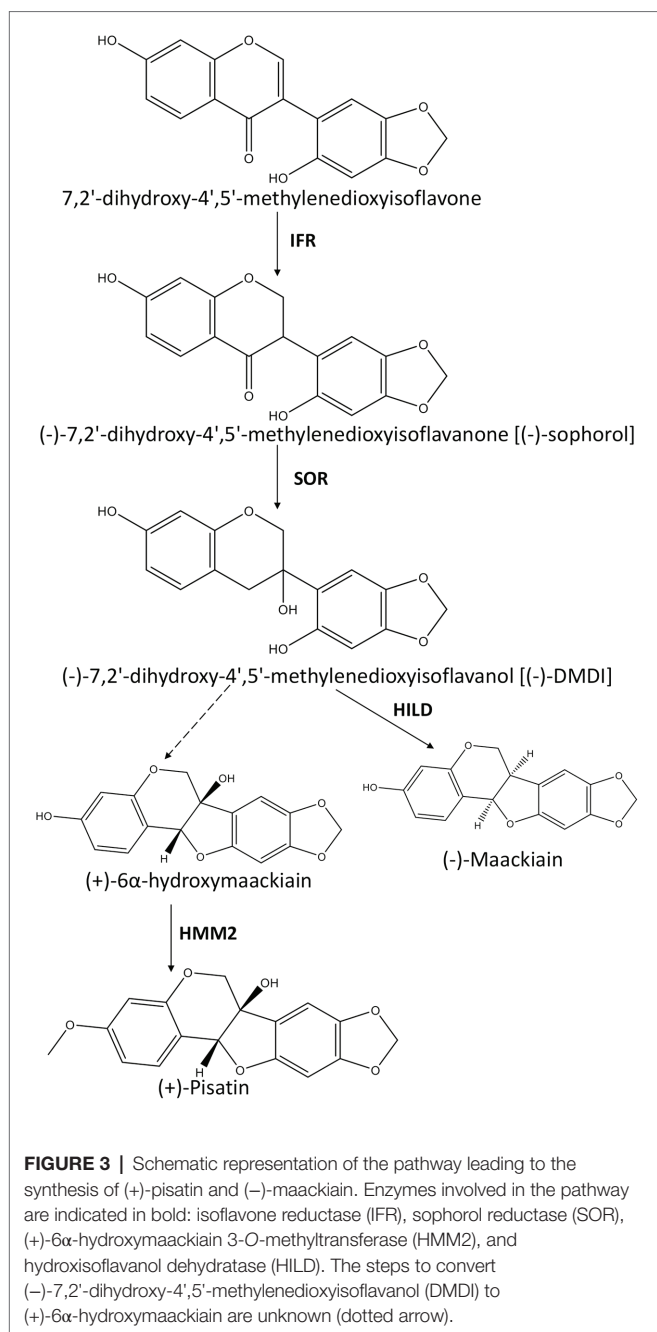
programs (Li et al., 2011). Recently, 12 candidate genes for the alkaloid locus *iucundus* and the major QTLs associated with total QA contents were identified using a transcriptomic approach (Kroc et al., 2019b). The most promising candidate, *RAP2-7*, encodes an ethylene-responsive transcription factor (ERF) that co-segregated with the *iucundus* locus and is likely to be involved in the regulation of QA biosynthesis in *L. angustifolius* (Kroc et al., 2019a). Other candidate genes include a 4-hydroxy-tetrahydronicotinic acid synthase (*DHDPS*) involved in Lys biosynthesis as well as genes involved in plant secondary metabolism (Kroc et al., 2019b).

Recessive low-alkaloid mutations in *L. albus* have also been identified: *pauper*, *mitis*, *reductus*, *exiguus*, and *nutricius* (Hackbarth, 1957; Troll, 1958; Porsche, 1964). As in the case of *L. angustifolius*, one locus, *pauper*, is the most studied in *L. albus* (Rychel and Książkiewicz, 2019). The Kiev mutant × P27174 recombinant inbred lines (RILs) population was used for the first genetic map of *L. albus* where the *pauper* locus was located on linkage group 11 (Phan et al., 2007). Recently, a high-resolution map was developed to provide a high-resolution QTL assay of the agronomic traits of *L. albus* (Michał et al., 2017). The *pauper* locus was localized in the linkage group ALB18 (Michał et al., 2017). The *Lup021586* gene was identified in the region and showed 100% nucleotide identity to *LaAT*, the acyltransferase gene previously identified in *L. angustifolius* (Bunsupa et al., 2011). *LAG101_35805*, an *L. albus* homolog of *LaAT* that is highly similar to *L. angustifolius* *Lup021586* gene, has been proposed as a molecular marker for the *pauper* locus (Rychel and Książkiewicz, 2019). Meanwhile, four low-alkaloid alleles have been identified in *L. luteus*, including *dulcis*, *amoenus*, *liber*, and *v* (von Sengbusch, 1942; Gustafsson and Gadd, 1965). However, there is limited information on the genetic basis for the low-alkaloid trait in this species. Efforts to improve the genomic resources of *L. luteus* are underway. The first genetic map for *L. luteus* has been recently released (Iqbal et al., 2019). A high-quality reference genome will help to implement MAS and identify loci responsible for the low-alkaloid content in *L. luteus*.

On the other hand, phytoalexins are a class of secondary metabolites with antimicrobial activities that are synthesized *de novo* after biotic and abiotic stresses (Walton, 1997). Phytoalexin biosynthesis can be induced by pathogens or a type of stress-mimicking compounds called elicitors (Angelova et al., 2006), and are produced by a range of crops including those in the *Fabaceae* family (Ahuja et al., 2012). Phytoalexins produced by the family *Leguminosae* comprise a variety of chemical compounds, including flavonoid phytoalexins derived from the shikimic acid pathway. In species such as soybean, prenylated pterocarpan, i.e., glyceollins, are synthesized in response to fungal pathogens such as *Phytophthora sojae* and *Macrophomina phaseolina* (Lygin et al., 2013). Soybean produces six forms of the isoflavonoid phytoalexin, glyceollin, where glyceollin I, glyceollin II, and glyceollin III are the predominant isomers (Banks and Dewick, 1983), derived from the addition of a dimethylallyl chain to (6*S*,11*aS*)-3,9,6*a*-trihydroxypterocarpan (glycinol) at either C-4 or C-2 by prenyltransferases (PTs). Two isoflavonoid PTs have been identified in soybean: 4-dimethylallyltransferase (*G4DT*)

and glycinol 2-dimethylallyltransferase (*G2DT*; Akashi et al., 2009; Yoneyama et al., 2016). Molecular characterization of *PT* genes revealed that *G4DT* and *G2DT* are paralogs resulting from a whole-genome duplication (Yoneyama et al., 2016). A genome-wide analysis of *PT* genes in *G. max* Wm82 identified 77 *PT*-encoding genes with 11 putative isoflavonoid-specific *PTs* (Sukumaran et al., 2018). One of the candidate genes, *GmPT01* (*G2DT-2*) was induced by *P. sojae* infection and AgNO₃, which mimics pathogen attack and lies in the QTL linked to *P. sojae* resistance. It was suggested that *GmPT01* is one of the genes involved in the partial resistance and could be used in breeding for increased fungal resistance. Other genes related to *P. sojae* resistance include a CHS gene, *GmCHR2A*, located near a QTL linked to *P. sojae* resistance (Sepiol et al., 2017). Additionally, studies have shown that resistant and susceptible genotypes differ in their timing of activating glyceollin biosynthesis (Yoshikawa et al., 1978; Hahn et al., 1985). A rapid activation of the biosynthetic pathway allows a high level of accumulation of these low-molecular weight compounds and confers resistance to pathogens. Soybean genotypes encoding the *P. sojae* resistance gene, *Rps1k*, have shown a rapid activation of glyceollin biosynthesis and higher resistance to the pathogen (Yoshikawa et al., 1978; Hahn et al., 1985). Recently, a member of the NAC (NAM/ATAF1/2/CUC2)-family of transcription factor (TF) genes, *GmNAC42-1*, was identified using comparative transcriptomics (Jahan et al., 2020). *GmNAC42-1* binds the promoter of *G4DT* and plays a role in the accumulation of glyceollin I. However, additional TFs are expected to participate in the regulation of glyceollin biosynthesis.

The phytoalexins in pea (*P. sativum*) are pisatin and maackiain (Ahuja et al., 2012). Pisatin, a 6*α*-hydroxyl-pterocarpan phytoalexin, is the major phytoalexin in pea produced in response to fungal infections by *Nectria haematococca*, *Botrytis cinerea*, and *Mycosphaerella pinodes* (Van den Heuvel and Glazener, 1975; Shiraishi et al., 1978; Etebu and Osborn, 2010). Its biosynthetic pathway has been partially characterized (Paiva et al., 1994; DiCenzo and Vanetten, 2006; Kaimoyo and VanEtten, 2008; Celoy and VanEtten, 2014). Pisatin and maackiain are synthesized via two chiral intermediates, (–)-7,2'-dihydroxy-4',5'-methylenedioxyisoflavanone [(–)-sophorol] and (–)-7,2'-dihydroxy-4',5'-methylenedioxyisoflavanol [(–)-DMDI; Preisig et al., 1989; Akashi et al., 2006; DiCenzo and Vanetten, 2006]. Sophorol reductase (SOR) is responsible for the production of sophorol, and it can be inactivated by RNA-mediated genetic interference (RNAi) which inhibits the production of pisatin in transgenic pea hairy roots (Kaimoyo and VanEtten, 2008). The pathway diverges for pisatin production after (–)-DMDI formation. In the last step of the pathway, the methylation of (+)-6*α*-hydroxymaackiain (6*α*-HMK) at the C-3 position by 6*α*-hydroxymaackiain methyltransferase (HMM2) results in the formation of (+)-pisatin. (–)-DMDI is converted to (–)-maackiain by hydroxyisoflavanol dehydratase (HILD). The biosynthesis pathway of (+)-pisatin and (–)-maackiain is illustrated in **Figure 3**. *M. pinodes* causes ascochyta blight, the most important foliar disease of field pea, which responds by accumulating pisatin (Shiraishi et al., 1978). Efforts have been made to elucidate the QTLs associated with the disease resistance and



to facilitate the introgression of resistance into pea cultivars (Wroth, 1999; Timmerman-Vaughan et al., 2004; Prioul-Gervais et al., 2007; Fondevilla et al., 2011). However, only moderate resistance has been reported with such efforts in pea cultivars (Kraft et al., 1998). Wild relatives present a high phytoalexin diversity (Lindig-Cisneros et al., 2002). The genetic controls for the resistance to *M. pinodes* were studied in a wild *P. sativum* subsp. *syriacum* accession P665 (Fondevilla et al., 2008). Six QTLs associated with *M. pinodes* resistance were detected in linkage groups II, III, IV, and V (Timmerman-Vaughan et al., 2004; Prioul-Gervais et al., 2007). Quantitative trait loci *MpV.1* and *MpII.1* were specific for seedlings under growth chamber

conditions and *MpIII.3* and *MpIV.1* for adult plant resistance in field conditions (Timmerman-Vaughan et al., 2004; Prioul-Gervais et al., 2007). In contrast, QTLs *MpIII.1* and *MpIII.2* were detected both for seedling and field resistance (Timmerman-Vaughan et al., 2004; Prioul-Gervais et al., 2007). *MpIII.2* overlaps with a QTL previously reported to be related to the resistance against ascochyta blight complex *Asc3.1* (Timmerman-Vaughan et al., 2004; Prioul-Gervais et al., 2007). A resistance-gene analog (*RGA1.1*) was identified in the vicinity of this QTL using *P. sativum* populations (Timmerman-Vaughan et al., 2004; Prioul-Gervais et al., 2007). QTLs associated with partial resistance to the root rot-causing *A. euteiches* have also been identified in pea, and would be useful for improving and facilitating the existing recurrent selection-based breeding program (Kraft, 1988; Lewis and Gritton, 1992). Pilet-Nayel et al. (2002) crossed susceptible lines with partially resistant ones in order to map the QTLs associated with resistance against *A. euteiches*. The genetic map revealed seven such genomic regions and *Aph1* located on the linkage group IVb was the most promising. Other minor QTLs were also identified in the 13 linkage groups obtained in the genetic mapping (Pilet-Nayel et al., 2002). Meta-analyses of four RILs of pea revealed 27 meta-Aphanomyces resistance QTLs, including 11 with high consistency across populations, locations, years, and isolates (Hamon et al., 2013). Seven highly consistent genomic regions were identified with the potential for use in MAS for pea improvement. Resistance QTLs located in these seven regions were further validated (Lavaud et al., 2015). Backcross-assisted selection programs were used to generate near-isogenic lines (NILs) carrying the resistance alleles of individual or combined resistance QTLs. The effects of two major QTLs, *Ae-Ps4.5*, and *Ae-Ps7.6*, were validated. The NILs carrying the resistance alleles of these two QTLs showed the highest resistance to *A. euteiches* strains. Several minor-effect QTLs were also validated, including *Ae-Ps2.2* and *Ae-Ps5.1*. Genome-wide analyses further validated most of these resistance QTLs and detected additional novel resistance loci (Desgroux et al., 2016). Putative candidate genes in these loci were related to biotic stress responses.

Transporters

ATP-binding cassette transporters and multidrug and toxic compound extrusion transporters play important roles in the secretion and accumulation of secondary metabolites (Yazaki, 2005; Ku et al., 2020). These transporters are associated with microbe interaction and nutrient accumulation of legumes (Sugiyama et al., 2007; Zhang et al., 2010; Fondevilla et al., 2011; Li et al., 2016).

ABC Transporter

In soybean (*G. max*), an ABC transporter was reported to be involved in the root secretion of genistein, which is an important signaling molecule for mediating the symbiosis with rhizobia (Sugiyama et al., 2007). In *M. truncatula*, two half-ABC transporters, STR and STR2, are essential for arbuscule development in arbuscular mycorrhizal symbiosis (Zhang et al., 2010). The expression of the STR and STR2 genes was induced in cortical

cells containing arbuscules (Zhang et al., 2010). STR and STR2 dimerize to form a transporter, which is located in the periarbuscular membrane and is important for the arbuscule development and therefore the symbiosis (Zhang et al., 2010). The *str* mutant and STR2-silenced transgenic roots exhibited stunted arbuscules after inoculating with *Glomus versiforme* (Zhang et al., 2010). In pea, using microarray technology, an ABC transporter was found to have a higher expression in *P. sativum* ssp. *syriacum* accession P665, which is resistant to *Mycosphaerella pinodes*, than the sensitive accession Messire (Fondevilla et al., 2011).

MATE Transporter

Seeds of wild soybeans (*G. soja*) generally have higher antioxidant contents than cultivated soybeans (*G. max*; Li et al., 2016). Statistical analysis showed the high correlation among the levels of seed total antioxidants, phenolics, and flavonoids (Li et al., 2016). Using RILs resulted from the cross between the wild soybean W05 (*G. soja*) and the cultivated soybean C08 (*G. max*), QTLs regulating the contents of antioxidants, phenolics, and flavonoids in soybean seeds were identified, which share a common genomic region (Li et al., 2016). In the target genomic region, three genes, *GmMATE1*, *GmMATE2*, and *GmMATE4*, were predicted to encode MATE transporters (Li et al., 2016). These MATE genes are possible candidates for investigating the basis behind the different seed antioxidant contents between wild soybeans (*G. soja*) and cultivated soybeans (*G. max*).

MOLECULAR BREEDING AND SECONDARY METABOLITE CONTENT

As covered in this review, legumes produce a diverse array of secondary metabolites including a large subset of compounds with biopharmaceutical/nutraceutical properties. The production of these phytochemicals can be increased through crop improvement using classical breeding to genetic approaches (Jacob et al., 2016). Legumes with increased health-beneficial secondary metabolites are potential raw materials for producing pharmaceutical products.

The genetic variability of legume species is fundamental to identify parental lines to be used in breeding programs and exploit legume secondary metabolites. Modern targeted breeding programs use tools, such as quantitative trait loci, marked-assisted selection, and genomics applications (Collard and Mackill, 2008; Jacob et al., 2016). DNA-based molecular markers are used to characterize genomic regions (insertions, deletions, mutations) controlling a particular trait or gene to differentiate individuals for germplasm identification and characterization (Nadeem et al., 2018). Molecular markers provide breeders with a valuable resource to accelerate selection programs and mark complex traits, which are influenced by environmental factors or not observable at early stages of plant development. Flavonoids have pharmacological effects, such as antioxidants for human nutrition or anti-inflammatory effects among others. Also, nutritional value of legumes can be enhanced by increasing flavonoid content through breeding selection (D'Amelia et al., 2018). In this case, molecular

markers have been used to study genetic variability in legumes to obtain varieties with high total flavonoid content. Genetic heritability of flavonoids is high and germplasms with different flavonoid content can lead to the identification of potential markers to use in breeding (Caseys et al., 2015). Flavonoid content was determined in 57 peanut accessions to evaluate the association between molecular markers and flavonoid content (Hou et al., 2017). Four expressed sequence tag-simple sequence repeat (EST-SSRs) markers were identified related to high flavonoid content in Chinese peanut germplasm. Functions of these markers were analyzed and related to outer membrane protein porin, heat-shock transcription factor, and lectins (Hou et al., 2017). Further studies are required to confirm the functions of these ESTs in flavonoid synthesis in peanuts. In soybean, three novel alleles were identified associated to flavonoid hydroxylase genes, *F3'H* and *F3'5'H*, related to pigmentation traits (Guo and Qiu, 2013). These molecular markers were identified using a set of gene-tagged markers based on the sequence variation of *GmF3'H* and *GmF3'5'H* in different soybean accessions, including cultivars, landraces, and wild soybeans (Guo and Qiu, 2013). Domestication process does not appear to erode diversity since four *GmF3'H* alleles were identified among cultivated soybeans, while *G. soja* contained only the *GmF3'H* allele. In the case of *GmF3'5'H*, 92.2% of wild soybean contained the *GmF3'5'H-a* allele, while three *GmF3'5'H* alleles occurred among cultivated soybeans (Guo and Qiu, 2013). In white clover (*Trifolium repens*), diversity array technology (DART) and microsatellite markers were used to discover marker-trait associations for flavonoid accumulation and biomass (Ballizany et al., 2016). Significant associations to concentrations of flavonols quercetin, kaempferol, and Quercetin:Kaempferol ratio were found to markers on linkage group 1–2. Additionally, the study revealed deleterious alleles in an elite cultivar indicating that genetic variability from wild germplasm could be used for white clover improvement (Ballizany et al., 2016).

ENGINEERING SECONDARY METABOLITE CONTENTS IN LEGUMES

In addition to breeding programs to improve domesticated varieties and broaden the gene pool of cultivars, secondary metabolite contents can also be modified through plant metabolic engineering (DellaPenna, 2001). The identification of genes involved in the biosynthesis pathways of diverse secondary metabolites has helped to drive strategies to optimize the production of target compounds. Increased production of target metabolites can be achieved by altering the primary or secondary metabolism of an organism, for example, through the overexpression of genes in biosynthetic pathways or by knocking out gene expressions and hence the enzymatic activities of competing pathways. In soybean, the manipulation of the (iso) flavonoid pathway and its effect on the resistance to *P. sojae* has been studied (Cheng et al., 2015; Chen et al., 2018; Zhou et al., 2018). *GmIFR*, encoding an isoflavone reductase (IFR), was identified and overexpressed in soybean (Cheng et al., 2015). IFR catalyzes an intermediate step in the biosynthesis of glyceollins (Graham et al., 1990) and its constitutive expression

in transgenic soybean plants enhances the resistance to *P. sojae*, along with higher glyceollin contents. Similar effects on pathogen resistance were obtained by the overexpression of *coenzyme A ligase (GmPI4L)* in transgenic soybean plants (Chen et al., 2018). Further attempts to elucidate enzymes/genes responsible for the resistance to pathogens include the overexpression of a chalcone isomerase, *GmCHI1A*, in soybean hairy roots, which enhanced daidzein accumulation and resistance to *P. sojae* strain P6497R compared to the control (Zhou et al., 2018). In alfalfa, nutritional value was increased by engineering genistein glucoside production (Deavours and Dixon, 2005). A transgenic alfalfa was developed by constitutively expressing an isoflavone synthase, *MtIFS1*, from *M. truncatula*. However, in the *MtIFS1*-expressing transgenic alfalfa, isoflavonoid production and accumulation was tissue-specific and affected by environmental factors such as UV-B and the disease-causing pathogen, *Phoma medicaginis* (Deavours and Dixon, 2005). RNAi-mediated gene silencing of isoflavone reductase, SOR, and hydroxymaackiain-3-O-methyltransferase in pea (*P. sativum*) allowed the identification of DMDI, an intermediary in the production of pisatin and maackiain (Kaimoyo and VanEtten, 2008). Other viable strategies of engineering secondary metabolite pathways include biosynthesis in microorganisms and modulation of gene expressions through manipulating the expressions of transcription factors (Du et al., 2010). Recently, genome-scale models have been used to represent the metabolic capabilities of legumes, including alfalfa and soybean (Pfau et al., 2018; Moreira et al., 2019). This approach allows the integration of different kinds of omics data to get new insights into plant-microbe interactions (diCenzo et al., 2016; Pfau et al., 2018; Contador et al., 2020). Models of plant metabolic pathways could also be used in the design of optimal-use biosynthesis pathways of secondary metabolites.

CONCLUSION

Domestication generally results in the reduction in secondary metabolites, which are often related to the bitter taste of seeds

and the resistance of plants to biotic stresses. The phenomenon is consistent with the reported decrease in crop biodiversity due to domestication (Food and Agricultural Organization of the United Nations, 2010). Having numerous health-beneficial secondary metabolites, legumes have the great potential to be employed as the sources of bioactive compounds for pharmaceutical use. On the other hand, besides abiotic stresses, the changing climate may also bring forth unpredictable biotic stresses such as insect infestations. From these perspectives, it is important to retain the biodiversity of legumes in order to maintain a healthy gene pool to produce new cultivars that can respond to future changes in their environments. Understanding the genes that govern the beneficial secondary metabolite compositions in legumes will facilitate the use of wild legumes in breeding programs or metabolic engineering to promote crop diversity, as well as to produce legumes with favorable secondary metabolite profiles.

AUTHOR CONTRIBUTIONS

H-ML planned and coordinated the writing. Y-SK put together the first complete draft and prepared the table. CC drew **Figures 1, 2**. M-SN prepared the table and drew **Figure 3**. All authors contributed to the search of literature and writings. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the Hong Kong Research Grants Council General Research Fund (14143916) and Area of Excellence Scheme (AoE/M-403/16).

ACKNOWLEDGMENTS

J. Chu copy-edited this manuscript.

REFERENCES

- Abbas, M., Saeed, F., Anjum, F. M., Afzaal, M., Tufail, T., Bashir, M. S., et al. (2017). Natural polyphenols: an overview. *Int. J. Food Prop.* 20, 1689–1699. doi: 10.1080/10942912.2016.1220393
- Abbo, S., van-Oss, P. R., Gopher, A., Saranga, Y., Ofner, I., and Peleg, Z. (2014). Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends Plant Sci.* 19, 351–360. doi: 10.1016/j.tplants.2013.12.002
- Abdel-lateif, K., Bogusz, D., and Hoher, V. (2012). The role of flavonoids in the establishment of plant roots endosymbioses with arbuscular mycorrhiza fungi, rhizobia and *Frankia* bacteria. *Plant Signal. Behav.* 7, 636–641. doi: 10.4161/psb.20039
- Abraham, E. M., Ganopoulos, I., Madesis, P., Mavromatis, A., Mylona, P., Nianiou-Obeidat, I., et al. (2019). The use of lupin as a source of protein in animal feeding: genomic tools and breeding approaches. *Int. J. Mol. Sci.* 20:851. doi: 10.3390/ijms20040851
- Acquaviva, R., Russo, A., Galvano, F., Galvano, G., Barcellona, M. L., Volti, G. L., et al. (2003). Cyanidin and cyanidin 3-O- β -D-glucoside as DNA cleavage protectors and antioxidants. *Cell Biol. Toxicol.* 19, 243–252. doi: 10.1023/b:cbto.0000003974.27349.4e
- Adhikari, K. N., Edwards, O. R., Wang, S., Ridsdill-Smith, T. J., and Buirchell, B. (2012). The role of alkaloids in conferring aphid resistance in yellow lupin (*Lupinus luteus* L.). *Crop Pasture Sci.* 63, 444–451. doi: 10.1071/CP12189
- Ahuja, I., Kissen, R., and Bones, A. M. (2012). Phytoalexins in defense against pathogens. *Trends Plant Sci.* 17, 73–90. doi: 10.1016/j.tplants.2011.11.002
- Akashi, T., Sasaki, R., Aoki, T., Ayabe, S., and Yazaki, K. (2009). Molecular cloning and characterization of a cDNA for pterocarpan 4-dimethylallyltransferase catalyzing the key prenylation step in the biosynthesis of glyceollin, a soybean phytoalexin. *Plant Physiol.* 149, 683–693. doi: 10.1104/pp.108.123679
- Akashi, T., VanEtten, H. D., Sawada, Y., Wasmann, C. C., Uchiyama, H., and Ayabe, S. (2006). Catalytic specificity of pea O-methyltransferases suggests gene duplication for (+)-pisatin biosynthesis. *Phytochemistry* 67, 2525–2530. doi: 10.1016/j.phytochem.2006.09.010
- Al-Anazi, A. F., Qureshi, V. F., Javaid, K., and Qureshi, S. (2011). Preventive effects of phytoestrogens against postmenopausal osteoporosis as compared

- to the available therapeutic choices: an overview. *J. Nat. Sci. Biol. Med.* 2, 154–163. doi: 10.4103/0976-9668.92322
- Amirkia, V., and Heinrich, M. (2014). Alkaloids as drug leads – a predictive structural and biodiversity-based analysis. *Phytochem. Lett.* 10, xlviii–liii. doi: 10.1016/j.phytol.2014.06.015
- Angelova, Z., Georgiev, S., and Roos, W. (2006). Elicitation of plants. *Biotechnol. Biotechnol. Equip.* 20, 72–83. doi: 10.1080/13102818.2006.10817345
- Applebaum, S. W., Marco, S., and Birk, Y. (1969). Saponins as possible factors of resistance of legume seeds to the attack of insects. *J. Agric. Food Chem.* 17, 618–622. doi: 10.1021/jf60163a020
- Arnoldi, A., Boschin, G., Zanoni, C., and Lammi, C. (2015). The health benefits of sweet lupin seed flours and isolated proteins. *J. Funct. Foods* 18, 550–563. doi: 10.1016/j.jff.2015.08.012
- Arts, I. C. W., van de Putte, B., and Hollman, P. C. H. (2000). Catechin contents of foods commonly consumed in the Netherlands. 1. Fruits, vegetables, staple foods, and processed foods. *J. Agric. Food Chem.* 48, 1746–1751. doi: 10.1021/jf000025h
- Avato, P., Bucci, R., Tava, A., Vitali, C., Rosato, A., Bialy, Z., et al. (2006). Antimicrobial activity of saponins from *Medicago* sp.: structure-activity relationship. *Phytother. Res.* 20, 454–457. doi: 10.1002/ptr.1876
- Azam, M., Zhang, S., Abdelghany, A. M., Shaibu, A. S., Feng, Y., Li, Y., et al. (2020). Seed isoflavone profiling of 1168 soybean accessions from major growing ecoregions in China. *Food Res. Int.* 130:108957. doi: 10.1016/j.foodres.2019.108957
- Bagayoko, M., Buerkert, A., Lung, G., Bationo, A., and Römhild, V. (2000). Cereal/legume rotation effects on cereal growth in Sudano-Sahelian West Africa: soil mineral nitrogen, mycorrhizae and nematodes. *Plant Soil* 218, 103–116. doi: 10.1023/A:1014957605852
- Ballizany, W. L., Griffiths, A. G., Franzmayr, B. K., Jahufer, M. Z. Z., Hofmann, R. W., and Barrett, B. A. (2016). “Marker-trait associations for flavonoids and biomass in white clover (*Trifolium repens* L.)” in *Breeding in a world of scarcity*. eds. I. Roldán-Ruiz, J. Baert and D. Reheul (Cham: Springer International Publishing), 225–229.
- Banks, S. W., and Dewick, P. M. (1983). Biosynthesis of glyceollins I, II and III in soybean. *Phytochemistry* 22, 2729–2733. doi: 10.1016/S0031-9422(00)97682-9
- Bazghaleh, N., Prashar, P., Purves, R. W., and Vandenberg, A. (2018). Polyphenolic composition of lentil roots in response to infection by *Aphanomyces euteiches*. *Front. Plant Sci.* 9:1131. doi: 10.3389/fpls.2018.01131
- Bernstein, P. S., Li, B., Vachali, P. P., Gorusupudi, A., Shyma, R., Henriksen, B. S., et al. (2016). Lutein, zeaxanthin, and meso-zeaxanthin: the basic and clinical science underlying carotenoid-based nutritional interventions against ocular disease. *Prog. Retin. Eye Res.* 50, 34–66. doi: 10.1016/j.preteyeres.2015.10.003
- Bertioli, D. J., Seijo, G., Freitas, F. O., Valls, J. F. M., Leal-Bertioli, S. C. M., and Moretzsohn, M. C. (2011). An overview of peanut and its wild relatives. *Plant Genet. Resour.* 9, 134–149. doi: 10.1017/S1479262110000444
- Böttger, A., Vothknecht, U., Bolle, C., and Wolf, A. (eds.) (2018). “Plant secondary metabolites and their general function in plants” in *Lessons on caffeine, cannabidiol & co: Plant derived drugs and their interaction with human receptors* (Switzerland: Springer Nature), 3–17.
- Brönnvik, H., and von Wettberg, E. J. (2019). Bird dispersal as a pre-adaptation for domestication in legumes: insights for neo-domestication. *Front. Plant Sci.* 10:1293. doi: 10.3389/fpls.2019.01293
- Brun, G., Braem, L., Thoiron, S., Gevaert, K., Goormachtig, S., and Delavault, P. (2018). Seed germination in parasitic plants what insights can we expect from strigolactone research? *J. Exp. Bot.* 69, 2265–2280. doi: 10.1093/jxb/erx472
- Bunsupa, S., Katayama, K., Ikeura, E., Oikawa, A., Toyooka, K., Saito, K., et al. (2012a). Lysine decarboxylase catalyzes the first step of quinolizidine alkaloid biosynthesis and coevolved with alkaloid production in *leguminosae*. *Plant Cell* 24, 1202–1216. doi: 10.1105/tpc.112.095885
- Bunsupa, S., Okada, T., Saito, K., and Yamazaki, M. (2011). An acyltransferase-like gene obtained by differential gene expression profiles of quinolizidine alkaloid-producing and nonproducing cultivars of *Lupinus angustifolius*. *Plant Biotechnol. J.* 28, 89–94. doi: 10.5511/plantbiotechnology.10.1109b
- Bunsupa, S., Yamazaki, M., and Saito, K. (2012b). Quinolizidine alkaloid biosynthesis: recent advances and future prospects. *Front. Plant Sci.* 3:239. doi: 10.3389/fpls.2012.00239
- Buzzetti, R. I., Buttery, B. R., and MacTavish, D. C. (1987). Biochemical genetics of black pigmentation of soybean seed. *J. Hered.* 78, 53–54. doi: 10.1093/oxfordjournals.jhered.a110309
- Carter, A., Tenuta, A., Rajcan, I., Welacky, T., Woodrow, L., and Eskandari, M. (2018). Identification of quantitative trait loci for seed isoflavone concentration in soybean (*Glycine max*) against soybean cyst nematode stress. *Plant Breed.* 137, 721–729. doi: 10.1111/pbr.12627
- Caseys, C., Stritt, C., Glauser, G., Blanchard, T., and Lexer, C. (2015). Effects of hybridization and evolutionary constraints on secondary metabolites: the genetic architecture of phenylpropanoids in European *Populus* species. *PLoS One* 10:e0128200. doi: 10.1371/journal.pone.0128200
- Celoy, R. M., and VanEtten, H. D. (2014). (+)-Pisatin biosynthesis: from (–) enantiomeric intermediates via an achiral 7,2-dihydroxy-4',5'-methylenedioxyisoflav-3-ene. *Phytochemistry* 98, 120–127. doi: 10.1016/j.phytochem.2013.10.017
- Chen, X., Fang, X., Zhang, Y., Wang, X., Zhang, C., Yan, X., et al. (2018). Overexpression of a soybean 4-coumaric acid: coenzyme a ligase (*GmPIL4L*) enhances resistance to *Phytophthora sojae* in soybean. *Funct. Plant Biol.* 46, 304–313. doi: 10.1071/FP18111
- Cheng, Q., Li, N., Dong, L., Zhang, D., Fan, S., Jiang, L., et al. (2015). Overexpression of soybean isoflavone reductase (*GmIFR*) enhances resistance to *Phytophthora sojae* in soybean. *Front. Plant Sci.* 6:1024. doi: 10.3389/fpls.2015.01024
- Collard, B. C. Y., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 363, 557–572. doi: 10.1098/rstb.2007.2170
- Considine, M. J., Siddique, K. H. M., and Foyer, C. H. (2017). Nature's pulse power: legumes, food security and climate change. *J. Exp. Bot.* 68, 1815–1818. doi: 10.1093/jxb/erx099
- Contador, C. A., Lo, S. -K., Chan, S. H. J., and Lam, H. -M. (2020). Metabolic analyses of nitrogen fixation in the soybean microsymbiont *Sinorhizobium fredii* using constraint-based modeling. *mSystems* 5, e00516–e00519. doi: 10.1128/mSystems.00516-19
- Cronk, Q., and Ojeda, I. (2008). Bird-pollinated flowers in an evolutionary and molecular context. *J. Exp. Bot.* 59, 715–727. doi: 10.1093/jxb/ern009
- D'Amelia, V., Aversano, R., Chiaiese, P., and Carputo, D. (2018). The antioxidant properties of plant flavonoids: their exploitation by molecular plant breeding. *Phytochem. Rev.* 17, 611–625. doi: 10.1007/s1101-018-9568-y
- Dastmalchi, M., and Dhaubhadel, S. (2014). “Soybean seed isoflavonoids: biosynthesis and regulation” in *Phytochemicals – Biosynthesis, function and application*. ed. R. Jetter (Cham, Switzerland: Springer), 1–21.
- Daverio, M., Cavicchiolo, M. E., Grotto, P., Lonati, D., Cananzi, M., and Da Dalt, L. (2014). Bitter lupine beans ingestion in a child: a disregarded cause of acute anticholinergic toxicity. *Eur. J. Pediatr.* 173, 1549–1551. doi: 10.1007/s00431-013-2088-2
- Deavours, B. E., and Dixon, R. A. (2005). Metabolic engineering of isoflavonoid biosynthesis in alfalfa. *Plant Physiol.* 138, 2245–2259. doi: 10.1104/pp.105.062539
- Dehaan, L. R., Van Tassel, D. L., Anderson, J. A., Asselin, S. R., Barnes, R., Baute, G. J., et al. (2016). A pipeline strategy for grain crop domestication. *Crop Sci.* 56, 917–930. doi: 10.2135/cropsci2015.06.0356
- DellaPenna, D. (2001). Plant metabolic engineering. *Plant Physiol.* 125, 160–163. doi: 10.1104/pp.125.1.160
- Desgroux, A., Anthoëne, V. L., Roux-Duparque, M., Rivière, J. -P., Aubert, G., Tayeh, N., et al. (2016). Genome-wide association mapping of partial resistance to *Aphanomyces euteiches* in pea. *BMC Genomics* 17:124. doi: 10.1186/s12864-016-2429-4
- Di Ferdinando, M., Brunetti, C., Fini, A., and Tattini, M. (2014). “Flavonoids as antioxidants in plants under abiotic stresses” in *Abiotic stress responses in plants: Metabolism, productivity and sustainability*. eds. P. Ahmad and M. N. V. Prasad (New York, NY: Springer), 159–179.
- Dias, T. R., Alves, M. G., Casal, S., Silva, B. M., and Oliveira, P. F. (2016). The single and synergistic effects of the major tea components caffeine, epigallocatechin-3-gallate and L-theanine on rat sperm viability. *Food Funct.* 7, 1301–1305. doi: 10.1039/c5fo01611h
- diCenzo, G. C., Checucci, A., Bazzicalupo, M., Mengoni, A., Viti, C., Dziewit, L., et al. (2016). Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat. Commun.* 7:12219. doi: 10.1038/ncomms12219
- DiCenzo, G. L., and Vanetten, H. D. (2006). Studies on the late steps of (+) pisatin biosynthesis: evidence for (–) enantiomeric intermediates. *Phytochemistry* 67, 675–683. doi: 10.1016/j.phytochem.2005.12.027
- Dixon, R. A., and Sumner, L. W. (2003). Legume natural products: understanding and manipulating complex pathways for human and animal health. *Plant Physiol.* 131, 878–885. doi: 10.1104/pp.102.017319

- Drewnowski, A., and Gomez-Carneros, C. (2000). Bitter taste, phytonutrients, and the consumer: a review. *Am. J. Clin. Nutr.* 72, 1424–1435. doi: 10.1093/ajcn/72.6.1424
- Du, H., Huang, Y., and Tang, Y. (2010). Genetic and metabolic engineering of isoflavonoid biosynthesis. *Appl. Microbiol. Biotechnol.* 86, 1293–1312. doi: 10.1007/s00253-010-2512-8
- Espinosa-Alonso, L. G., Lygin, A., Widholm, J. M., Valverde, M. E., and Paredes-Lopez, O. (2006). Polyphenols in wild and weedy Mexican common beans (*Phaseolus vulgaris* L.). *J. Agric. Food Chem.* 54, 4436–4444. doi: 10.1021/jf060185e
- Etebu, E., and Osborn, A. M. (2010). Molecular quantification of the pea footrot disease pathogen (*Nectria haematococca*) in agricultural soils. *Phytoparasitica* 38, 447–454. doi: 10.1007/s12600-010-0122-8
- Fernández-Aparicio, M., Kisugi, T., Xie, X., Rubiales, D., and Yoneyama, K. (2014). Low strigolactone root exudation: a novel mechanism of broomrape (*Orobancha* and *Phelipanche* spp.) resistance available for faba bean breeding. *J. Agric. Food Chem.* 62, 7063–7071. doi: 10.1021/jf5027235
- Fernández-marín, B., Milla, R., Martín-robles, N., Arc, E., Kranner, I., Becerril, J. M., et al. (2014). Side-effects of domestication: cultivated legume seeds contain similar tocopherols and fatty acids but less carotenoids than their wild counterparts. *BMC Plant Biol.* 14:1599. doi: 10.1186/s12870-014-0385-1
- Fernández-tomé, S., and Hernández-ledesma, B. (2019). Current state of art after twenty years of the discovery of bioactive peptide lunasin. *Food Res. Int.* 116, 71–78. doi: 10.1016/j.foodres.2018.12.029
- Fondevilla, S., Küster, H., Krajinski, F., Cubero, J. I., and Rubiales, D. (2011). Identification of genes differentially expressed in a resistant reaction to *Mycosphaerella pinodes* in pea using microarray technology. *BMC Genomics* 12:28. doi: 10.1186/1471-2164-12-28
- Fondevilla, S., Satovic, Z., Rubiales, D., Moreno, M. T., and Torres, A. M. (2008). Mapping of quantitative trait loci for resistance to *Mycosphaerella pinodes* in *Pisum sativum* subsp. *syriacum*. *Mol. Breed.* 21, 439–454. doi: 10.1007/s11032-007-9144-4
- Foo, E., and Davies, N. W. (2011). Strigolactones promote nodulation in pea. *Planta* 234, 1073–1081. doi: 10.1007/s00425-011-1516-7
- Food and Agricultural Organization of the United Nations (2010). Crop biodiversity: use it or lose it. Available at: <http://www.fao.org/news/story/en/item/46803/icode/> [Accessed July 2, 2020].
- Frick, K. M., Kamphuis, L. G., Kadambot, S. H. M., Singh, K. B., and Foley, R. C. (2017). Quinolizidine alkaloid biosynthesis in lupins and prospects for grain quality improvement. *Front. Plant Sci.* 8:87. doi: 10.3389/fpls.2017.00087
- Gao, D., Wang, X., Fu, S., and Zhao, J. (2020). Legume plants enhance the resistance of soil to ecosystem disturbance site description and experimental. *Front. Plant Sci.* 8:1295. doi: 10.3389/fpls.2017.01295
- García, E. H., Peña-valdivia, C. B., Aguirre, J. R. R., and Muruaga, J. S. M. (1997). Morphological and agronomic traits of a wild population and an improved cultivar of common bean (*Phaseolus vulgaris* L.). *Ann. Bot.* 79, 207–213. doi: 10.1006/anbo.1996.0329
- Gepts, P. (2014). “Domestication of plants” in *Encyclopedia of agriculture and food systems*. ed. N. K. Van Alfen (Cambridge, Massachusetts: Academic Press), 474–486.
- Gepts, P., Beavis, W. D., Brummer, E. C., Shoemaker, R. C., Stalker, H. T., Weeden, N. F., et al. (2005). Legumes as a model plant family.
- Genomics for food and feed report of the cross-legume advances through genomics conference. *Plant Physiol.* 137, 1228–1235. doi: 10.1104/pp.105.060871.1228
- Golebiewski, W. M., and Spenser, I. D. (1988). Biosynthesis of the lupine alkaloids. II. Sparteine and lupanine. *Can. J. Chem.* 66, 1734–1748. doi: 10.1139/v88-280
- Graham, T. L., Kim, J. E., and Graham, M. Y. (1990). Role of constitutive isoflavone conjugates in the accumulation of glyceollin in soybean infected with *Phytophthora megasperma*. *Mol. Plant-Microbe Interact.* 3, 157–166. doi: 10.1094/MPMI-3-157
- Guo, Y., and Qiu, L. -J. (2013). Allele-specific marker development and selection efficiencies for both flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase genes in soybean subgenus *soja*. *Theor. Appl. Genet.* 126, 1445–1455. doi: 10.1007/s00122-013-2063-3
- Gupta, Y. P. (1987). Anti-nutritional and toxic factors in food legumes: a review. *Plant Foods Hum. Nutr.* 37, 201–228. doi: 10.1007/BF01091786
- Gupta, R. N., Horwood, P., Koo, S. H., and Spenser, I. D. (1979). The biosynthesis of the *Lythraceae* alkaloids. I. The lysine-derived fragment. *Can. J. Chem.* 57, 1606–1614. doi: 10.1139/v79-260
- Gupta, A. P., Pandotra, P., Kushwaha, M., Khan, S., Sharma, R., and Gupta, S. (2015). “Alkaloids: a source of anticancer agents from nature” in *Studies in natural products chemistry*. ed. Atta-ur-Rahman (Amsterdam, Netherlands: Elsevier B.V.), 341–445.
- Gustafsson, A., and Gadd, I. (1965). Mutations and crop improvement. II. The genus *Lupinus* (*Leguminosae*). *Hereditas* 53, 15–39. doi: 10.1111/j.1601-5223.1965.tb01977.x
- Hackbarth, J. (1957). Die Gene der Lupinenarten. III Weiße Lupine (*Lupinus albus*). *Zeitschrift für Pflanzenzüchtung* 37, 185–191.
- Hahn, M. G., Bonhoff, A., and Grisebach, H. (1985). Quantitative lociglyceollin I in relation to fungal hyphae in soybean roots infected with *Phytophthora megasperma* f. sp. *glycinea*. *Plant Physiol.* 77, 591–601. doi: 10.1104/pp.77.3.591
- Haldar, S., and Sengupta, S. (2015). Plant-microbe cross-talk in the rhizosphere: insight and biotechnological potential. *Open Microbiol. J.* 9, 1–7. doi: 10.2174/1874285801509010001
- Hamon, C., Coyne, C. J., MacGee, R. J., Lesné, A., Esnault, R., Mangin, P., et al. (2013). QTL meta-analysis provides a comprehensive view of loci controlling partial resistance to *Aphanomyces euteiches* in four sources of resistance in pea. *BMC Plant Biol.* 13:45. doi: 10.1186/1471-2229-13-45
- Hane, J. K., Ming, Y., Kamphuis, L. G., Nelson, M. N., Garg, G., Atkins, C. A., et al. (2017). A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: insights into plant-microbe interactions and legume evolution. *Plant Biotechnol. J.* 15, 318–330. doi: 10.1111/pbi.12615
- Harrison, M. J. (2005). Signaling in the arbuscular mycorrhizal symbiosis. *Annu. Rev. Microbiol.* 59, 19–42. doi: 10.1146/annurev.micro.58.030603.123749
- Hassan, S. M., Byrd, J. A., Cartwright, A. L., and Bailey, C. A. (2010). Hemolytic and antimicrobial activities differ among saponin-rich extracts from guar, quillaja, yucca, and soybean. *Appl. Biochem. Biotechnol.* 162, 1008–1017. doi: 10.1007/s12010-009-8838-y
- Heiser, C. B. (1988). Aspects of unconscious selection and the evolution of domesticated plants. *Euphytica* 37, 77–81. doi: 10.1007/BF00037227
- Hellens, R. P., Moreau, C., Lin-Wang, K., Schwin, K. E., Thomson, S. J., Fiers, M. W. E. J., et al. (2010). Identification of Mendel's white flower character. *PLoS One* 5:e13230. doi: 10.1371/journal.pone.0013230
- Heront, D. S., and Pueppket, S. G. (1984). Mode of infection, nodulation specificity, and indigenous plasmids of 11 fast-growing *Rhizobium japonicum* strains. *J. Bacteriol.* 160, 1061–1066. doi: 10.1128/JB.160.3.1061-1066.1984
- Hirsch, A. M., Lum, M. R., and Downie, J. A. (2001). What makes the rhizobia-legume symbiosis so special? *Plant Physiol.* 127, 1484–1492. doi: 10.1104/pp.010866.1484
- Hollman, P. C. H., and Arts, I. C. W. (2000). Flavonols, flavones and flavanols – nature, occurrence and dietary burden. *J. Sci. Food Agric.* 80, 1081–1093. doi: 10.1002/(SICI)1097-0010(20000515)80:7<1081::AID-JSFA566>3.0.CO;2-G
- Hou, M., Mu, G., Zhang, Y., Cui, S., Yang, X., and Liu, L. (2017). Evaluation of total flavonoid content and analysis of related EST-SSR in Chinese peanut germplasm. *Crop Breed. Appl. Biotechnol.* 17, 221–227. doi: 10.1590/1984-70332017v17n3a34
- Hsieh, C., Martínez-Villaluenga, C., De Lumen, B. O., and Hernández-Ledesma, B. (2017). Updating the research on the chemopreventive and therapeutic role of the peptide lunasin. *J. Sci. Food Agric.* 98, 2070–2079. doi: 10.1002/jsfa.8719
- Husain, F., and Mallikarjuna, N. (2012). Genetic diversity in Bolivian landrace lines of groundnut (*Arachis hypogaea* L.). *Indian J. Genet. Plant Breed.* 72, 384–389.
- Iqbal, M. M., Huynh, M., Udall, J. A., Kilian, A., Adhikari, K. N., Berger, J. D., et al. (2019). The first genetic map for yellow lupin enables genetic dissection of adaptation traits in an orphan grain legume crop. *BMC Genet.* 20:68. doi: 10.1186/s12863-019-0767-3
- Isah, T. (2019). Stress and defense responses in plant secondary metabolites production. *Biol. Res.* 52:39. doi: 10.1186/s40659-019-0246-3
- Jacob, C., Carrasco, B., and Schwember, A. R. (2016). Advances in breeding and biotechnology of legume crops. *Plant Cell Tissue Organ Cult.* 127, 561–584. doi: 10.1007/s11240-016-1106-2
- Jahan, M. A., Harris, B., Lowery, M., Infante, A. M., Percifield, R. J., and Kovinich, N. (2020). Glyceollin transcription factor GmMYB29A2 regulates soybean resistance to *Phytophthora sojae*. *Plant Physiol.* 183, 530–546. doi: 10.1104/pp.19.01293

- Jeong, H. J., Jeong, J. B., Kin, D. S., and De Lumen, B. O. (2007). Inhibition of core histone acetylation by the cancer preventive peptide lunasin. *J. Agric. Food Chem.* 55, 632–637. doi: 10.1021/jf062405u
- Jeong, H. J., Lam, Y., and De Lumen, B. O. (2002). Barley lunasin suppresses ras-induced colony formation and inhibits core histone acetylation in mammalian cells. *J. Agric. Food Chem.* 50, 5903–5908. doi: 10.1021/jf0256945
- Jeong, H. J., Lee, J. R., Jeong, J. B., Park, J. H., and De Lumen, B. O. (2009). The cancer preventive seed peptide lunasin from rye is bioavailable and bioactive. *Nutr. Cancer* 61, 680–686. doi: 10.1080/01635580902850082
- Jeong, S. C., Moon, J. -K., Park, S. -K., Kim, M. -S., Lee, K., Lee, S. R., et al. (2019). Genetic diversity patterns and domestication origin of soybean. *Theor. Appl. Genet.* 132, 1179–1193. doi: 10.1007/s00122-018-3271-7
- Jeong, H. J., Park, J. H., Lam, Y., and De Lumen, B. O. (2003). Characterization of lunasin isolated from soybean. *J. Agric. Food Chem.* 51, 7901–7906. doi: 10.1021/jf034460y
- Jia, K., Baz, L., and Al-babili, S. (2018). From carotenoids to strigolactones. *J. Exp. Bot.* 69, 2189–2204. doi: 10.1093/jxb/erx476
- Josie, J., Alcivar, A., Rainho, J., and Kassem, M. A. (2007). Genomic regions containing QTL for plant height, internodes length, and flower color in soybean [*Glycine max* (L.) Merr.]. *BIOS* 78, 119–126. doi: 10.1893/0005-3155(2007)78[119:RAGRCQ]2.0.CO;2
- Kaimoyo, E., and VanEtten, H. D. (2008). Inactivation of pea genes by RNAi supports the involvement of two similar O-methyltransferases in the biosynthesis of (+)-pisatin and of chiral intermediates with a configuration opposite that found in (+)-pisatin. *Phytochemistry* 69, 76–87. doi: 10.1016/j.phytochem.2007.06.013
- Kaplan, L. (1981). What is the origin of the common bean? *Econ. Bot.* 35, 240–254. doi: 10.1007/BF02858692
- Kaur, K., Grewal, S. K., Gill, P. S., and Singh, S. (2019). Comparison of cultivated and wild chickpea genotypes for nutritional quality and antioxidant potential. *J. Food Sci. Technol.* 56, 1864–1876. doi: 10.1007/s13197-019-03646-4
- Kleemann, R., Verschuren, L., Morrison, M., Zadelaar, S., van Erk, M. J., Wielinga, P. Y., et al. (2011). Anti-inflammatory, anti-proliferative and anti-atherosclerotic effects of quercetin in human *in vitro* and *in vivo* models. *Atherosclerosis* 218, 44–52. doi: 10.1016/j.atherosclerosis.2011.04.023
- Knekt, P., Kumpulainen, J., Järvinen, R., Rissanen, H., Heliövaara, M., Reunanen, A., et al. (2002). Flavonoid intake and risk of chronic diseases. *Am. J. Clin. Nutr.* 76, 560–568. doi: 10.1093/ajcn/76.3.560
- Kraft, J. M. (1988). *Aphanomyces* root rot resistance in peas. *Phytopathology* 78:1545.
- Kraft, J. M., Dunne, B., Goulden, D., and Armstrong, S. (1998). A search for resistance in peas to *Mycosphaerella pinodes*. *Plant Dis.* 82, 251–253. doi: 10.1094/PDIS.1998.82.2.251
- Krishnamurthy, P., Chung, G., and Singh, R. J. (2013). Saponin polymorphism in the Korean wild soybean (*Glycine soja* Sieb. And Zucc.). *Plant Breed.* 132, 121–126. doi: 10.1111/pbr.12016
- Křížová, L., Dadáková, K., Kašparovský, J., and Kašparovský, T. (2019). Isoflavones. *Molecules* 24:1076. doi: 10.3390/molecules24061076
- Kroc, M., Czepiel, K., Wilczura, P., Mokrzycka, M., and Święcicki, W. (2019a). Development and validation of a gene-targeted dCAPS marker for marker-assisted selection of low-alkaloid content in seeds of narrow-leaved lupin (*Lupinus angustifolius* L.). *Genes* 10:428. doi: 10.3390/genes10060428
- Kroc, M., Koczyk, G., Kamel, K. A., Czepiel, K., Fedorowicz-Strońska, O., Paweł, K., et al. (2019b). Transcriptome-derived investigation of biosynthesis of quinolizidine alkaloids in narrow-leaved lupin (*Lupinus angustifolius* L.) highlights candidate genes linked to *iucundus* locus. *Sci. Rep.* 9:2231. doi: 10.1038/s41598-018-37701-5
- Ku, Y. -S., Ng, M. -S., Cheng, S. -S., Lo, A. W. -Y., Xiao, Z., Shin, T. -S., et al. (2020). Understanding the composition, biosynthesis, accumulation and transport of flavonoids in crops for the promotion of crops as healthy sources of flavonoids for human consumption. *Nutrients* 12:1717. doi: 10.3390/nut12061717
- Kumar, G. P., and Khanum, F. (2012). Neuroprotective potential of phytochemicals. *Pharmacogn. Rev.* 6, 81–90. doi: 10.4103/0973-7847.99898
- Kumar, S., and Pandey, A. K. (2013). Chemistry and biological activities of flavonoids: an overview. *Sci. World J.* 2013:162750. doi: 10.1155/2013/162750
- Kumar, A., Singh, A., and Ekavali (2015). A review on Alzheimer's disease pathophysiology and its management: an update. *Pharmacol. Rep.* 67, 195–203. doi: 10.1016/j.pharep.2014.09.004
- Kurlovich, B. S. (2002). "Genetics of lupins" in *Lupins: Geography, classification, genetic resources and breeding*. ed. B. S. Kurlovich (St Petersburg, Russia; Pellosniemi, Finland: OY International North Express), 313–350.
- Kushi, L. H., Meyer, K. A., and Jacobs, D. R. Jr. (1999). Cereals, legumes, and chronic disease risk reduction: evidence from epidemiologic studies. *Am. J. Clin. Nutr.* 70(3 suppl), 451S–458S. doi: 10.1093/ajcn/70.3.451S
- Latha, B. P., Reddy, I. R. M., Vijaya, T., Rao, S. D., Ismail, S. M., and Girish, B. P. (2011). Effect of saponin rich extract of *Achyranthes aspera* on high fat diet fed male wistar rats. *J. Pharm. Res.* 4, 3190–3193.
- Lavaud, C., Lesné, A., Piriou, C., Le Roy, G., Boutet, G., Moussart, A., et al. (2015). Validation of QTL for resistance to *Aphanomyces euteiches* in different pea genetic backgrounds using near-isogenic lines. *Theor. Appl. Genet.* 128, 2273–2288. doi: 10.1007/s00122-015-2583-0
- Leistner, E., and Spenser, I. D. (1973). Biosynthesis of the piperidine nucleus. Incorporation of chirally labeled [1-3H]cadaverine. *J. Am. Chem. Soc.* 95, 4715–4725. doi: 10.1021/ja00795a041
- Lewis, M. E., and Gritton, E. T. (1992). Use of one cycle of recurrent selection per year for increasing resistance to *Aphanomyces* root rot in peas. *J. Am. Soc. Hortic. Sci.* 117, 638–642. doi: 10.21273/JASHS.117.4.638
- Li, M. W., Muñoz, N. B., Wong, C. F., Wong, F. L., Wong, K., Wong, J. W. H., et al. (2016). QTLs regulating the contents of antioxidants, phenolics, and flavonoids in soybean seeds share a common genomic region. *Front. Plant Sci.* 7:854. doi: 10.3389/fpls.2016.00854
- Li, X., Yang, H., Buirchell, B., and Yan, G. (2011). Development of a DNA marker tightly linked to low-alkaloid gene *iucundus* in narrow-leaved lupin (*Lupinus angustifolius* L.) for marker-assisted selection. *Crop Pasture Sci.* 62, 218–224. doi: 10.1071/CP10352
- Liang, H., Yu, Y., Wang, S., Lian, Y., Wang, T., Wei, Y., et al. (2010). QTL mapping of isoflavone, oil and protein contents in soybean (*Glycine max* L. Merr.). *Agric. Sci. China* 9, 1108–1116. doi: 10.1016/S1671-2927(09)60197-8
- Lindig-Cisneros, R., Dirzo, R., and Espinosa-Garcia, F. J. (2002). Effects of domestication and agronomic selection on phytoalexin antifungal defense in *Phaseolus* beans. *Ecol. Res.* 17, 315–321. doi: 10.1046/j.1440-1703.2002.00491.x
- Lo, S., Muñoz-amatriáin, M., Boukar, O., He, I., Cisse, N., Guo, Y., et al. (2018). Identification of QTL controlling domestication-related traits in cowpea (*Vigna unguiculata* L. Walp.). *Sci. Rep.* 8:6261. doi: 10.1038/s41598-018-24349-4
- López-cortez, M. D. S., Rosales-martínez, P., Arellano-cárdenas, S., and Cornejo-Mazón, M. (2016). "Antioxidants properties and effect of processing methods on bioactive compounds of legumes" in *Grain legumes*. ed. A. K. Goyal (Rijeka: IntechOpen).
- Lygin, A. V., Zernova, O. V., Hill, C. B., Kholina, N. A., Widholm, J. M., Hartman, G. L., et al. (2013). Glyceollin is an important component of soybean plant defense against *Phytophthora sojae* and *Macrophoma phaseolina*. *Phytopathology* 103, 984–994. doi: 10.1094/PHYTO-12-12-0328-R
- Maass, B. L. (2006). Changes in seed morphology, dormancy and germination from wild to cultivated hyacinth bean germplasm (*Lablab purpureus*: Papilionoideae). *Genet. Resour. Crop. Evol.* 53, 1127–1135. doi: 10.1007/s10722-005-2782-7
- Man, S., Gao, W., Zhang, Y., Huang, L., and Liu, C. (2010). Chemical study and medical application of saponins as anti-cancer agents. *Fitoterapia* 81, 703–714. doi: 10.1016/j.fitote.2010.06.004
- Marciani, D. J. (2018). Elucidating the mechanisms of action of saponin-derived adjuvants. *Trends Pharmacol. Sci.* 39, 573–585. doi: 10.1016/j.tips.2018.03.005
- Marrelli, M., Conforti, F., Araniti, F., and Statti, G. A. (2016). Effects of saponins on lipid metabolism: a review of potential health benefits in the treatment of obesity. *Molecules* 21:1404. doi: 10.3390/molecules21101404
- McClean, P. E., Bett, K. E., Stonehouse, R., Lee, R., Pflieger, S., Moghaddam, S. M., et al. (2018). White seed color in common bean (*Phaseolus vulgaris*) results from convergent evolution in the P (pigment) gene. *New Phytol.* 219, 1112–1123. doi: 10.1111/nph.15259
- McCorry, M. A., Hamaker, B. R., Lovejoy, J. C., and Eichelsdoerfer, P. E. (2010). Pulse consumption, satiety, and weight management. *Adv. Nutr.* 1, 17–30. doi: 10.3945/an.110.1006
- Menéndez, A. B., Calzadilla, P. I., Sansberro, P. A., Espasandín, F. D., Gazquez, A., Bordenave, C. D., et al. (2019). Polyamines and legumes: joint stories of

- stress, nitrogen fixation and environment. *Front. Plant Sci.* 10:1415. doi: 10.3389/fpls.2019.01415
- Michał, K., Nazzicari, N., Yang, H., Nelson, M. N., Renshaw, D., Rychel, S., et al. (2017). A high-density consensus linkage map of white lupin highlights synteny with narrow-leaved lupin and provides markers tagging key agronomic traits. *Sci. Rep.* 7:15335. doi: 10.1038/s41598-017-15625-w
- Mohan, V. R., Tresina, P. S., and Daffodil, E. D. (2016). *Antinutritional factors in legume seeds: Characteristics and determination*. 1st Edn. Cambridge, Massachusetts: Academic Press.
- Moreira, T. B., Shaw, R., Luo, X., Ganguly, O., Kim, H. -S., Ferreira Coelho, L. G., et al. (2019). A genome-scale metabolic model of soybean (*Glycine max*) highlights metabolic fluxes in seedlings. *Plant Physiol.* 180, 1912–1929. doi: 10.1104/pp.19.00122
- Morrison, M. J., Cober, E. R., Saleem, M. F., McLaughlin, N. B., Frégeau-Reid, J., Ma, B. L., et al. (2008). Changes in isoflavone concentration with 58 years of genetic improvement of short-season soybean cultivars in Canada. *Crop Sci.* 48, 2201–2208. doi: 10.2135/cropsci2008.01.0023
- Muñoz-Esparza, N., Latorre-Moratalla, M. L., Comas-Basté, O., Toro-Funes, N., Veciana-Nogués, M. T., and Vidal-Carou, C. V. (2019). Polyamines in food. *Front. Nutr.* 6:108. doi: 10.3389/fnut.2019.00108
- Muzquiz, M., Cuadrado, C., Ayet, G., De Cuadra, C., Burbano, C., and Osagie, A. (1994). Variation of alkaloid components of lupin seeds in 49 genotypes of *Lupinus albus* L. from different countries and locations. *J. Agric. Food Chem.* 42, 1447–1450. doi: 10.1021/jf00043a011
- Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., et al. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equip.* 32, 261–285. doi: 10.1080/13102818.2017.1400401
- Nagl, W., Ignacimuthu, S., and Becker, J. (1997). Genetic engineering and regeneration of *Phaseolus* and *Vigna*. State of the art and new attempts. *J. Plant Physiol.* 150, 625–644. doi: 10.1016/S0176-1617(97)80277-5
- Nelson, M. N., Moolhuijzen, P. M., Boersma, J. G., Chudy, M., Lesniewska, K., Bellgard, M., et al. (2010). Aligning a new reference genetic map of *Lupinus angustifolius* with the genome sequence of the model legume *Lotus japonicus*. *DNA Res.* 17, 73–83. doi: 10.1093/dnares/dsq001
- Nelson, M. N., Phan, H. T. T., Ellwood, S. R., Moolhuijzen, P. M., Hane, J., Williams, A., et al. (2006). The first gene-based map of *Lupinus angustifolius* L.-location of domestication genes and conserved synteny with *Medicago truncatula*. *Theor. Appl. Genet.* 113, 225–238. doi: 10.1007/s00122-006-0288-0
- Nesi, N., Debeaujon, I., Jond, C., Pelletier, G., Caboche, M., and Lepiniec, L. (2000). The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in *Arabidopsis* siliques. *Plant Cell* 12, 1863–1878. doi: 10.1105/tpc.12.10.1863
- Oakenfull, D. (1981). Saponins in food—a review. *Food Chem.* 7, 19–40. doi: 10.1016/0308-8146(81)90019-4
- Ohmiya, S., Saito, K., and Murakoshi, I. (1995). “Lupine alkaloids” in *The alkaloids: Chemistry and pharmacology*. ed. G. A. Cordell (San Diego, CA: Academic Press), 1–114.
- Ojwang, L. O., Yang, L., Dykes, L., and Awika, J. (2013). Proanthocyanidin profile of cowpea (*Vigna unguiculata*) reveals catechin-O-glucoside as the dominant compound. *Food Chem.* 139, 35–43. doi: 10.1016/j.foodchem.2013.01.117
- Okada, T., Hirai, M. Y., Suzuki, H., Yamazaki, M., and Saito, K. (2005). Molecular characterization of a novel quinolizidine alkaloid O-tigloyltransferase: cDNA cloning, catalytic activity of recombinant protein and expression analysis in *Lupinus* plants. *Plant Cell Physiol.* 46, 233–244. doi: 10.1093/pcp/pci021
- Okubo, K., Iijima, M., Kobayashi, Y., Yoshikoshi, M., Uchida, T., and Kudou, S. (1992). Components responsible for the undesirable taste of soybean seeds. *Biosci. Biotechnol. Biochem.* 56, 99–103. doi: 10.1271/bbb.56.99
- Olsen, K. M., and Wendel, J. F. (2013). A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu. Rev. Plant Biol.* 64, 47–70. doi: 10.1146/annurev-arplant-050312-120048
- Onguéné, P. A., Ntie-kang, F., Lifongo, L. L., Ndom, J. C., Sippl, W., and Mbaze, L. M. (2013). The potential of anti-malarial compounds derived from African medicinal plants. Part I: a pharmacological evaluation of alkaloids and terpenoids. *Malar. J.* 12:449. doi: 10.1186/1475-2875-13-81
- Paiva, N. L., Sun, Y., Dixon, R. A., VanEtten, H. D., and Hrazdina, G. (1994). Molecular cloning of isoflavone reductase from pea (*Pisum sativum* L.): evidence for a 3R-isoflavone intermediate in (+)-pisatin biosynthesis. *Arch. Biochem. Biophys.* 312, 501–510. doi: 10.1006/abbi.1994.1338
- Palmer, R. G., Pfeiffer, T. W., Buss, G. R., and Kilen, T. C. (2004). “Qualitative genetics” in *Soybeans: Improvement, production, and uses*. eds. R. M. Shibles, J. E. Harper, R. F. Wilson and R. C. Shoemaker (Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America), 137–233.
- Pavan, S., Schiavulli, A., Marcotrigiano, A. R., Bardaro, N., Bracuto, V., Ricciardi, F., et al. (2016). Characterization of low-strigolactone germplasm in pea (*Pisum sativum* L.) resistant to crenate broomrape (*Orobanche crenata* Forsk.). *Mol. Plant-Microbe Interact.* 29, 743–749. doi: 10.1094/MPMI-07-16-0134-R
- Penmetsa, R. V., Carrasquilla-Garcia, N., Bergmann, E. M., Vance, L., Castro, B., Kassa, M. T., et al. (2016). Multiple post-domestication origins of *kabuli* chickpea through allelic variation in a diversification-associated transcription factor. *New Phytol.* 211, 1440–1451. doi: 10.1111/nph.14010
- Pfau, T., Christian, N., Masakapalli, S. K., Sweetlove, L. J., Poolman, M. G., and Ebenhöf, O. (2018). The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. *Sci. Rep.* 8:12504. doi: 10.1038/s41598-018-30884-x
- Phan, H. T. T., Ellwood, S. R., Adhikari, K. N., Nelson, M. N., and Oliver, R. P. (2007). The first genetic and comparative map of white lupin (*Lupinus albus* L.): identification of QTLs for anthracnose resistance and flowering time, and a locus for alkaloid content. *DNA Res.* 14, 59–70. doi: 10.1093/dnares/dsm009
- Philippi, J., Schliephake, E., Jürgens, E., Jansen, H. U., and Ordon, F. (2015). Feeding behavior of aphids on narrow-leaved lupin (*Lupinus angustifolius*) genotypes varying in the content of quinolizidine alkaloids. *Entomol. Exp. Appl.* 156, 37–51. doi: 10.1111/eea.12313
- Pietta, P., Minoggio, M., and Bramati, L. (2003). Plant polyphenols: structure, occurrence and bioactivity. *Stud. Nat. Prod. Chem.* 28, 257–312. doi: 10.1016/S1572-5995(03)80143-6
- Pilet-Nayel, M. -L., Muehlbauer, F. J., McGee, R. J., Kraft, J. M., Baranger, A., and Coyne, C. J. (2002). Quantitative trait loci for partial resistance to *Aphanomyces* root rot in pea. *Theor. Appl. Genet.* 106, 28–39. doi: 10.1007/s00122-002-0985-2
- Porsche, W. (1964). Untersuchungen über die Vererbung der Alkaloidarmut und die Variabilität des Restalkaloidgehaltes Bei *lupinus albus* L. *Der Züchter* 34, 251–256. doi: 10.1007/BF00705826
- Porter, S. S. (2013). Adaptive divergence in seed color camouflage in contrasting soil environments. *New Phytol.* 197, 1311–1320. doi: 10.1111/nph.12110
- Preisig, C. L., Matthews, D. E., and VanEtten, H. D. (1989). Purification and characterization of S-Adenosyl-L-methionine:6a-hydroxymaackain 3-O-methyltransferase from *Pisum sativum*. *Plant Physiol.* 91, 559–566. doi: 10.1104/pp.91.2.559
- Primomo, V. S., Poysa, V., Ablett, G. R., Jackson, C. -J., Gijzen, M., and Rajcan, I. (2005). Mapping QTL for individual and total isoflavone content in soybean seeds. *Crop Sci.* 45, 2454–2464. doi: 10.2135/cropsci2004.0672
- Prioul-Gervais, S., Deniot, G., Receveur, E. -M., Frankewitz, A., Fourmann, M., Rameau, C., et al. (2007). Candidate genes for quantitative resistance to *Mycosphaerella pinodes* in pea (*Pisum sativum* L.). *Theor. Appl. Genet.* 114, 971–984. doi: 10.1007/s00122-006-0492-y
- Pueppke, S. G., Bolaños-Vásquez, M. C., Werner, D., Bec-Ferté, M. -P., Promé, J. -C., and Krishnan, H. B. (1998). Release of flavonoids by the soybean cultivars McCall and Peking and their perception as signals by the nitrogen-fixing symbiont *Sinorhizobium fredii*. *Plant Physiol.* 117, 599–608. doi: 10.1104/pp.117.2.599
- Quintero-soto, M. F., Saracho-peña, A. G., Chavez-ontiveros, J., and Garzon-tiznado, J. A. (2018). Phenolic profiles and their contribution to the antioxidant activity of selected chickpea genotypes from Mexico and ICRISAT collections. *Plant Foods Hum. Nutr.* 73, 122–129. doi: 10.1007/s11130-018-0661-6
- Reddy, M. R. I., Latha, P. B., Vijaya, T., and Rao, D. S. (2012). The saponin-rich fraction of a *Gymnema sylvestre* R. Br. Aqueous leaf extract reduces cafeteria and high-fat diet-induced obesity. *Z. Naturforsch. C Biosci.* 67, 39–46. doi: 10.1515/znc-2012-1-206
- Rehman, H. M., Nawas, M. A., Shah, Z. H., Yang, S. H., and Chung, G. (2018). Characterization of naturally occurring wild soybean mutant (*sg-5*) lacking astringent saponins using whole genome sequencing approach. *Plant Sci.* 267, 148–156. doi: 10.1016/j.plantsci.2017.11.014
- Reinhard, H., Rupp, H., Sager, F., Streule, M., and Zoller, O. (2006). Quinolizidine alkaloids and phomopsins in lupin seeds and lupin containing food. *J. Chromatogr. A* 1112, 353–360. doi: 10.1016/j.chroma.2005.11.079

- Roberts, R. L., Green, J., and Lewis, B. (2009). Lutein and zeaxanthin in eye and skin health. *Clin. Dermatol.* 27, 195–201. doi: 10.1016/j.clindermatol.2008.01.011
- Romeo, F. V., Fabroni, S., Ballistreri, G., Muccilli, S., Spina, A., and Rapisarda, P. (2018). Characterization and antimicrobial activity of alkaloid extracts from seeds of different genotypes of *Lupinus* spp. *Sustainability* 10:788. doi: 10.3390/su10030788
- Rossi, A., Serraino, I., Dugo, P., Di Paola, R., Mondello, L., Genovese, T., et al. (2012). Protective effects of anthocyanins from blackberry in a rat model of acute lung inflammation. *Free Radic. Res.* 37, 891–900. doi: 10.1080/1071576031000112690
- Rowan, D. D., Cao, M., Lin-Wang, K., Cooney, J. M., Jensen, D. J., Austin, P. T., et al. (2009). Environmental regulation of leaf colour in red 35S:PAP1 *Arabidopsis thaliana*. *New Phytol.* 182, 102–115. doi: 10.1111/j.1469-8137.2008.02737.x
- Rubiales, D., and Fernández-Aparicio, M. (2012). Innovations in parasitic weeds management in legume crops. A review. *Agron. Sustain. Dev.* 32, 433–449. doi: 10.1007/s13593-011-0045-x
- Rychel, S., and Książkiewicz, M. (2019). Development of gene-based molecular markers tagging low alkaloid pauper locus in white lupin (*Lupinus albus* L.). *J. Appl. Genet.* 60, 269–281. doi: 10.1007/s13353-019-00508-9
- Saito, K., Suzuki, H., Takamatsu, S., and Murakoshi, I. (1992). Acyltransferases for lupin alkaloids in *Lupinus hirsutus*. *Phytochemistry* 32, 87–91. doi: 10.1016/0031-9422(92)80112-R
- Sauvaire, Y., Baissac, Y., Leconte, O., Petit, P., and Ribes, G. (1996). “Steroid saponins from fenugreek and some of their biological properties” in *Saponins used in food and agriculture. Advances in experimental medicine and biology*. eds. G. R. Waller and K. Yamasaki (Boston, MA: Springer).
- Sawai, S., and Saito, K. (2011). Triterpenoid biosynthesis and engineering in plants. *Front. Plant Sci.* 2:25. doi: 10.3389/fpls.2011.00025
- Sekhon, J., Grewal, S. K., Singh, I., and Kaur, J. (2017). Evaluation of nutritional quality and antioxidant potential of pigeonpea genotypes. *J. Food Sci. Technol.* 54, 3598–3611. doi: 10.1007/s13197-017-2818-y
- Sepiol, C. J., Yu, J., and Dhaubhadel, S. (2017). Genome-wide identification of *Chalcone Reductase* gene family in soybean: insight into root-specific *GmCHRs* and *Phytophthora sojae* resistance. *Front. Plant Sci.* 8:2073. doi: 10.3389/fpls.2017.02073
- Sharma, A., Shahzad, B., Rehman, A., Bhardwaj, R., Landi, M., and Zheng, B. (2019). Response of phenylpropanoid pathway and the role of polyphenols in plants under abiotic stress. *Molecules* 24:2452. doi: 10.3390/molecules24132452
- Shi, J., Arunasalam, K., Yeung, D., Kakuda, Y., Mittal, G., and Jiang, Y. (2004). Saponins from edible legumes: chemistry, processing, and health benefits. *J. Med. Food* 7, 67–78. doi: 10.1089/109662004322984734
- Shi, Y., Guo, R., Wang, X., Yuan, D., Zhang, S., Wang, J., et al. (2014). The regulation of alfalfa saponin extract on key genes involved in hepatic cholesterol metabolism in hyperlipidemic rats. *PLoS One* 9:e88282. doi: 10.1371/journal.pone.0088282
- Shim, S. I., Jun, W. J., and Kang, B. H. (2003). Evaluation of nutritional and antinutritional components in Korean wild legumes. *Plant Foods Hum. Nutr.* 58, 1–11. doi: 10.1023/B:QUAL.0000041166.10069.6f
- Shiraishi, T., Oku, H., Tsuji, Y., and Ouchi, S. (1978). Inhibitory effect of pisatin on infection process of *Mycosphaerella pinodes* on pea. *Ann. Phytopath. Soc. Japan* 44, 641–645.
- Shiraiwa, M., Kudo, S., Shimoyamada, M., Harada, K., and Okubo, K. (1991). Composition and structure of “group a saponin” in soybean seed. *Agric. Biol. Chem.* 55, 315–322. doi: 10.1080/00021369.1991.10870574
- Singh, M., Bisht, I. S., Kumar, S., Dutta, M., Bansal, K. C., Karale, M., et al. (2014). Global wild annual *Lens* collection: a potential resource for lentil genetic base broadening and yield enhancement. *PLoS One* 9:e107781. doi: 10.1371/journal.pone.0107781
- Singh, B., Jatinder, S. P., Kaur, A., and Singh, N. (2017). Phenolic composition and antioxidant potential of grain legume seeds: a review. *Food Res. Int.* 101, 1–16. doi: 10.1016/j.foodres.2017.09.026
- Sinha, R., Bala, M., Kumar, M., Sharma, T. R., and Singh, A. K. (2020). “Methods for screening legume crops for abiotic stress tolerance through physiological and biochemical approaches” in *Legume genomics: Methods and protocols, methods in molecular biology*. eds. M. Jain and R. Garg (New York: Springer US).
- Smallwood, C. J., Nyinyi, C. N., Kopsell, D. A., Sams, C. E., West, D. R., Chen, P., et al. (2014). Detection and confirmation of quantitative trait loci for soybean seed isoflavones. *Crop Sci.* 54, 595–606. doi: 10.2135/cropsci2013.05.0340
- Smýkal, P., Vernoud, V., Blair, M. W., Soukup, A., and Thompson, R. D. (2014). The role of the testa during development and in establishment of dormancy of the legume seed. *Front. Plant Sci.* 5:351. doi: 10.3389/fpls.2014.00351
- Soda, K. (2010). Polyamine intake, dietary pattern, and cardiovascular disease. *Med. Hypotheses* 75, 299–301. doi: 10.1016/j.mehy.2010.03.008
- Song, J., Liu, Z., Hong, H., Ma, Y., Tian, L., Li, X., et al. (2016). Identification and validation of loci governing seed coat color by combining association mapping and bulk segregation analysis in soybean. *PLoS One* 11:e0159064. doi: 10.1371/journal.pone.0159064
- Spaink, H. P. (1995). The molecular basis of infection and nodulation by rhizobia: the ins and outs of symbiogenesis. *Annu. Rev. Phytopathol.* 33, 345–368. doi: 10.1146/annurev.py.33.090195.002021
- Stagnari, F., Maggio, A., Galieni, A., and Pisante, M. (2017). Multiple benefits of legumes for agriculture sustainability: an overview. *Chem. Biol. Technol. Agric.* 4:2. doi: 10.1186/s40538-016-0085-1
- Sugiyama, A., Shitan, N., and Yazaki, K. (2007). Involvement of a soybean ATP-binding cassette-type transporter in the secretion of genistein, a signal flavonoid in legume-Rhizobium symbiosis. *Plant Physiol.* 144, 2000–2008. doi: 10.1104/pp.107.096727
- Sukumaran, A., McDowell, T., Chen, L., Renaud, J., and Dhaubhadel, S. (2018). Isoflavonoid-specific prenyltransferase gene family in soybean: GmPTO1, a pterocarpan 2-dimethylallyltransferase involved in glyceollin biosynthesis. *Plant J.* 96, 966–981. doi: 10.1111/tpj.14083
- Sundaramoorthy, J., Palaniswamy, S., Park, G. T., Son, H. R., Tsukamoto, C., Lee, J. -D., et al. (2019). Characterization of a new sg-5 variant with reduced biosynthesis of group a saponins in soybean (*Glycine max* (L.) Merr.). *Mol. Breed.* 39:144. doi: 10.1007/s11032-019-1066-4
- Sundaramoorthy, J., Park, G. T., Lee, J. -D., Kim, J. H., Seo, H. S., and Song, J. T. (2015). Genetic and molecular regulation of flower pigmentation in soybean. *J. Korean Soc. Appl. Biol. Chem.* 58, 555–562. doi: 10.1007/s13765-015-0077-z
- Sweetingham, M., and Kingwell, R. (2008). “Lupins—reflections and future possibilities.” in *Proceedings 12th International Lupin Conference*; September 14–18, 2008; Geraldton: International Lupin Association, 514–524.
- Swiecicki, W., and Swiecicki, W. K. (1995). Domestication and breeding improvement of narrow-leaved lupin (*L. angustifolius* L.). *J. Appl. Genet.* 36, 155–167.
- Takahashi, Y., Li, X. -H., Tsukamoto, C., and Wang, K. -J. (2016). Identification of a novel variant lacking group a soyasaponin in a Chinese wild soybean (*Glycine soja* Sieb. & Zucc.): implications for breeding significance. *Plant Breed.* 135, 607–613. doi: 10.1111/pbr.12403
- Tanaka, Y. (2006). Flower colour and cytochromes P450. *Phytochem. Rev.* 5, 283–291. doi: 10.1007/s11011-006-9003-7
- Tanaka, Y., Sasaki, N., and Ohmiya, A. (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 54, 733–749. doi: 10.1111/j.1365-3113X.2008.03447.x
- Taylor, J. L., De Angelis, G., and Nelson, M. N. (2020). “How have narrow-leaved lupin genomic resources enhanced our understanding of lupin domestication” in *The lupin genome*. eds. K. B. Singh, L. G. Kamphuis and M. N. Nelson (Cham, Switzerland: Springer International Publishing), 95–108.
- Tee, E. S., Goh, A. H., and Khor, S. C. (1995). Carotenoid composition and content of legumes, tubers and starchy roots by HPLC. *Malays. J. Nutr.* 1, 63–74.
- Timmerman-Vaughan, G. M., Frew, T. J., Butler, R., Murray, S., Gilpin, M., Fallon, K., et al. (2004). Validation of quantitative trait loci for *Ascochyta* blight resistance in pea (*Pisum sativum* L.), using populations from two crosses. *Theor. Appl. Genet.* 109, 1620–1631. doi: 10.1007/s00122-004-1779-5
- Todd, J. J., and Vodkin, L. O. (1993). Pigmented soybean (*Glycine max*) seed coats accumulate proanthocyanidins during development. *Plant Physiol.* 102, 663–670. doi: 10.1104/pp.102.2.663
- Todd, J. J., and Vodkin, L. O. (1996). Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. *Plant Cell* 8, 687–699. doi: 10.1105/tpc.8.4.687
- Tomooka, N., Kashiwaba, K., Vaughan, D. A., Ishimoto, M., and Egawa, Y. (2000). The effectiveness of evaluating wild species: searching for sources of resistance to bruchid beetles in the genus *Vigna* subgenus *Ceratotropis*. *Euphytica* 115, 27–41. doi: 10.1023/A:1003906715119

- Toon, A., Cook, L. G., and Crisp, M. D. (2014). Evolutionary consequences of shifts to bird-pollination in the Australian pea-flowered legumes (Mirbeliaceae and Bossiaceae). *BMC Evol. Biol.* 14:43. doi: 10.1186/1471-2148-14-43
- Troll, H. J. (1958). Erbgänge des Alkaloidgehaltes und Beobachtungen über Heterosiswirkung bei *Lupinus albus*. *Zeitschrift für Pflanzenzüchtung* 39, 35–46.
- Tsukamoto, C., Kikuchi, A., Harada, K., Kitamura, K., and Okubo, K. (1993). Genetic and chemical polymorphisms of saponins in soybean seed. *Phytochemistry* 34, 1351–1356. doi: 10.1016/0031-9422(91)80028-y
- Tsuzuki, S., Handa, Y., Takeda, N., and Kawaguchi, M. (2016). Strigolactone-induced putative secreted protein 1 is required for the establishment of symbiosis by the arbuscular mycorrhizal fungus *Rhizophagus irregularis*. *Mol. Plant-Microbe Interact.* 29, 277–286. doi: 10.1094/MPMI-10-15-0234-R
- Tuteja, J. H., Clough, S. J., Chan, W., and Vodkin, L. O. (2004). Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in *Glycine max*. *Plant Cell* 16, 819–835. doi: 10.1105/tpc.021352.1
- Tuteja, J. H., Zabala, G., Varala, K., Hudson, M., and Vodkin, L. O. (2009). Endogenous, tissue-specific short interfering RNAs silence the chalcone synthase gene family in *Glycine max* seed coats. *Plant Cell* 21, 3063–3077. doi: 10.1105/tpc.109.069856
- Uzoh, I. M., Igwe, C. A., Okebalama, C. B., and Babalola, O. O. (2019). Legume-maize rotation effect on maize productivity and soil fertility parameters under selected agronomic practices in a sandy loam soil. *Sci. Rep.* 9:8539. doi: 10.1038/s41598-019-43679-5
- Van den Heuvel, J., and Glazener, J. A. (1975). Comparative abilities of fungi pathogenic and nonpathogenic to bean (*Phaseolus vulgaris*) to metabolize phaseollin. *Neth. J. Plant Pathol.* 81, 125–137. doi: 10.1007/BF01976803
- Vimolmangkang, S., Han, Y., Wei, G., and Korban, S. S. (2013). An apple MYB transcription factor, *MdMYB3*, is involved in regulation of anthocyanin biosynthesis and flower development. *BMC Plant Biol.* 13:176. doi: 10.1186/1471-2229-13-176
- von Sengbusch, R. (1942). Sweet lupins and oil lupins. The history of the origin of some new crop plants. *Landwirtsch. Jahrb.* 91, 719–880.
- Voutilainen, S., Nurmi, T., Mursu, J., and Rissanen, T. H. (2006). Carotenoids and cardiovascular health. *Am. J. Clin. Nutr.* 83, 1265–1271. doi: 10.1093/ajcn/83.6.1265
- Walton, J. D. (1997). “Biochemical plant pathology” in *Plant biochemistry*. eds. P. M. Dey and J. B. Harborne (London, UK: Academic Press), 487–502.
- Wan, L., Li, B., Pandey, M. K., Wu, Y., Lei, Y., Yan, L., et al. (2016). Transcriptome analysis of a new peanut seed coat mutant for the physiological regulatory mechanism involved in seed coat cracking and pigmentation. *Front. Plant Sci.* 7:1491. doi: 10.3389/fpls.2016.01491
- Wang, M., Li, W., Fang, C., Xu, F., Liu, Y., Wang, Z., et al. (2018). Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nat. Genet.* 50, 1435–1441. doi: 10.1038/s41588-018-0229-2
- Wang, C., Wang, Y., Shi, Y., Yan, X., He, Y., and Fan, W. (2011). Effects of alfalfa saponins on the lipid metabolism, antioxidation and immunity of weaned piglets. *Acta Pratacul. Sin.* 20, 210–218.
- Wang, C., Zhao, T., and Gai, J. (2010). Genetic variability and evolutionary peculiarity of isoflavone content and its components in soybean germplasm from China. *Sci. Agric. Sin.* 43, 3919–3929.
- Wiesner, M., Hanschen, F. S., Maul, R., Neugart, S., Schreiner, M., and Baldermann, S. (2017). “Nutritional quality of plants for food and fodder” in *Encyclopedia of applied plant sciences*. eds. B. Thomas and D. J. Murphy (Cambridge, Massachusetts: Academic Press), 285–291.
- Wink, M. (2013). Evolution of secondary metabolites in legumes (*Fabaceae*). *S. Afr. J. Bot.* 89, 164–175. doi: 10.1016/j.sajb.2013.06.006
- Wroth, J. M. (1999). Evidence suggests that *Mycosphaerella pinodes* infection of *Pisum sativum* is inherited as a quantitative trait. *Euphytica* 107, 193–204. doi: 10.1023/A:1003688430893
- Xie, M., Chung, C. Y., Li, M., Wong, F., Wang, X., Liu, A., et al. (2019). A reference-grade wild soybean genome. *Nat. Commun.* 10:1216. doi: 10.1038/s41467-019-09142-9
- Yang, K., Jeong, N., Moon, J. K., Lee, Y. H., Lee, S. H., Kim, H. M., et al. (2010). Genetic analysis of genes controlling natural variation of seed coat and flower colors in soybean. *J. Hered.* 101, 757–768. doi: 10.1093/jhered/esq078
- Yazaki, K. (2005). Transporters of secondary metabolites. *Curr. Opin. Plant Biol.* 8, 301–307. doi: 10.1016/j.pbi.2005.03.011
- Yoneyama, K., Akashi, T., and Aoki, T. (2016). Molecular characterization of soybean pterocarpan 2-dimethylallyltransferase in glyceollin biosynthesis: local gene and whole-genome duplications of prenyltransferase genes led to the structural diversity of soybean prenylated isoflavonoids. *Plant Cell Physiol.* 57, 2497–2509. doi: 10.1093/pcp/pcw178
- Yoneyama, K., Xie, X., Sekimoto, H., Takeuchi, Y., Ogasawara, S., Akiyama, K., et al. (2008). Strigolactones, host recognition signals for root parasitic plants and arbuscular mycorrhizal fungi, from *Fabaceae* plants. *New Phytol.* 179, 484–494. doi: 10.1111/j.1469-8137.2008.02462.x
- Yoshikawa, M., Yamauchi, K., and Masago, H. (1978). Glyceollin: its role in restricting fungal growth in resistant soybean hypocotyls infected with *Phytophthora megasperma* var. *sojae*. *Physiol. Plant Pathol.* 12, 73–82. doi: 10.1016/0048-4059(78)90020-6
- Zhang, Q., Blaylock, L. A., and Harrison, M. J. (2010). Two *Medicago truncatula* half-ABC transporters are essential for arbuscule development in arbuscular mycorrhizal symbiosis. *Plant Cell* 22, 1483–1497. doi: 10.1105/tpc.110.074955
- Zhang, J., Ge, Y., Han, F., Li, B., Yan, S., Sun, J., et al. (2014). Isoflavone content of soybean cultivars from maturity group 0 to VI grown in northern and Southern China. *J. Am. Oil Chem. Soc.* 91, 1019–1028. doi: 10.1007/s11746-014-2440-3
- Zhang, Y., Pechan, T., and Chang, S. K. C. (2018). Antioxidant and angiotensin-I converting enzyme inhibitory activities of phenolic extracts and fractions derived from three phenolic-rich legume varieties. *J. Funct. Foods* 42, 289–297. doi: 10.1016/j.jff.2017.12.060
- Zhou, Y., Huang, J., Zhang, X., Zhu, L., Wang, X., Guo, N., et al. (2018). Overexpression of chalcone isomerase (CHI) increases resistance against *Phytophthora sojae* in soybean. *J. Plant Biol.* 61, 309–319. doi: 10.1007/s12374-018-0017-7
- Zhu, H., Wan, J., Wang, Y., Li, B., Xiang, C., He, J., et al. (2014). Medicinal compounds with antiepileptic/anticonvulsant activities. *Epilepsia* 55, 3–16. doi: 10.1111/epi.12463

Disclaimer: Any opinions, findings, conclusions or recommendations expressed in this publication do not reflect the views of the Government of the Hong Kong Special Administrative Region or the Innovation and Technology Commission.

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

Copyright © 2020 Ku, Contador, Ng, Yu, Chung and Lam. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Identification of QTLs for Domestication-Related Traits in Zombi Pea [*Vigna vexillata* (L.) A. Rich], a Lost Crop of Africa

Kitiya Amkul¹, Prakrit Somta^{1,2*}, Kularb Laosatit¹ and Lixia Wang^{3*}

¹ Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, Thailand, ² Center of Excellence on Agricultural Biotechnology: (AG-BIO/PERDO-CHE), Bangkok, Thailand, ³ Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

OPEN ACCESS

Edited by:

Gaofeng Zhou,
Department of Primary Industries
and Regional Development
of Western Australia (DPIRD),
Australia

Reviewed by:

Pei Xu,
China Jiliang University, China
Longbiao Guo,
Chinese Academy of Agricultural
Sciences, China

*Correspondence:

Prakit Somta
agrps@ku.ac.th;
pksomta@gmail.com
Lixia Wang
wanglixia03@caas.cn

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 29 April 2020

Accepted: 06 July 2020

Published: 18 September 2020

Citation:

Amkul K, Somta P, Laosatit K and
Wang L (2020) Identification of QTLs
for Domestication-Related Traits
in Zombi Pea [*Vigna vexillata* (L.) A.
Rich], a Lost Crop of Africa.
Front. Genet. 11:803.
doi: 10.3389/fgene.2020.00803

Zombi pea [*Vigna vexillata* (L.) A. Rich] is a legume crop found in Africa. Wild zombi pea is widely distributed throughout the tropical and subtropical regions, whereas domesticated zombi pea is rarely cultivated. Plant domestication is an evolutionary process in which the phenotypes of wild species, including seed dormancy, pod shattering, organ size, and architectural and phenological characteristics, undergo changes. The molecular mechanism underlying the domestication of zombi pea is relatively unknown. In this study, the genetic basis of the following 13 domestication-related traits was investigated in an F₂ population comprising 198 individuals derived from a cross between cultivated (var. *macrosperma*) and wild (var. *vexillata*) zombi pea accessions: seed dormancy, pod shattering, days-to-flowering, days-to-maturity, stem thickness, stem length, number of branches, leaf area, pod length, 100-seed weight, seed width, seed length, and seeds per pod. A genetic map containing 6,529 single nucleotide polymorphisms constructed for the F₂ population was used to identify quantitative trait loci (QTLs) for these traits. A total of 62 QTLs were identified for the 13 traits, with 1–11 QTLs per trait. The major QTLs for days-to-flowering, stem length, number of branches, pod length, 100-seed weight, seed length, and seeds per pod were clustered in linkage group 5. In contrast, the major QTLs for seed dormancy and pod shattering belonged to linkage groups 3 and 11, respectively. A comparative genomic analysis with the cowpea [*Vigna unguiculata* (L.) Walp.] genome used as the reference sequence (i.e., the genome of the legume species most closely related to zombi pea) enabled the identification of candidate genes for the major QTLs. Thus, we revealed the genomic regions associated with domestication-related traits and the candidate genes controlling these traits in zombi pea. The data presented herein may be useful for breeding new varieties of zombi pea and other *Vigna* species.

Keywords: zombi pea, *Vigna vexillata*, domestication syndrome, QTL, legume

INTRODUCTION

Plant domestication is a process in which wild forms of crop (cultivated) plants evolve and adapt to agricultural conditions and human needs via conscious and unconscious selections by humans and natural selection due to environmental factors. Domestication is one of the most important technological innovations in human history (Doebley et al., 2006). Although the processes and spatiotemporal requirements for domestication vary among crops, the processes generally lead to similar results regarding morphological and phenological changes to the wild forms. The changes from wild species to domesticated/cultivated species are collectively referred to as the “domestication syndrome” (Hammer, 1984). Compared with wild forms, domesticated forms exhibit a lack of or decreased seed dormancy and seed dispersal, produce larger organs (fruit, seed, leaf, and stem), flower and mature earlier, and have a more robust and firm plant architecture. During the last 20 years, plant domestication studies have become a major part of crop evolution and adaptation research. In fact, domestication is the earliest form of plant breeding (Doebley et al., 2006). The information generated in investigations of the evolutionary relationships between crops and their wild progenitors may be useful for exploiting wild plants for crop improvement (Vaughan et al., 2007). The genetic basis of the domestication syndrome of several plant species, but mainly cereals and legumes, including rice, maize, wheat, sorghum, soybean, common bean, and cowpea, has been investigated via molecular genetic analyses, especially quantitative trait locus (QTL) mapping. Several genes controlling the domestication-related traits of some of these crops have been identified and cloned (Doebley et al., 2006; Vaughan et al., 2007; Ishimaru et al., 2013).

Zombi pea [*Vigna vexillata* (L.) A. Rich] is an underutilized legume crop. Wild zombi pea is widely distributed in Africa, Asia, Australia, and the Americas (Dachapak et al., 2017). Because of its wide distribution, several wild forms of zombi pea have been identified and described (Maxted et al., 2004). Additionally, there are two types of cultivated zombi pea (i.e., seed and tuberous root) (Ferguson, 1954; Karuniawan et al., 2006). The cultivated seed type is grown in Africa (Ferguson, 1954), whereas the cultivated tuberous root type exists only in Bali, Indonesia, and East Timor (Karuniawan et al., 2006). The former is grown principally for its seeds, although the tuberous roots are sometimes consumed, whereas the latter is cultivated mainly for its tuberous roots. An analysis of molecular diversity based on microsatellite (i.e., simple sequence repeat) markers revealed that these two forms were domesticated independently (Dachapak et al., 2017). There exists a strong crossing barrier between the cultivated tuberous root type and the cultivated seed type/wild forms of zombi pea.

The genus *Vigna* includes as many as nine cultivated/domesticated species and several wild species that are used as human food, animal feed, and cover crops. Several comparative genomic analyses revealed that the genomes of *Vigna* species are highly conserved (Chaitieng et al., 2006; Isemura et al., 2007, 2010; Dachapak et al., 2018; Amkul et al., 2019; Yundaeng et al., 2019). We previously developed

linkage maps of zombi pea and determined that its genome is highly conserved with those of cowpea [*Vigna unguiculata* (L.) Walp.], mung bean [*Vigna radiata* (L.) Wilczek], azuki bean [*Vigna angularis* (Willd.) Ohwi and Ohashi], and rice bean [*Vigna umbellata* (Thunb.) Ohwi and Ohashi] (Dachapak et al., 2018; Dachapak et al., 2019; Amkul et al., 2019). A recent study proved that the zombi pea genome and the genomes of common bean (*Phaseolus vulgaris* L.) and soybean [*Glycine max* (L.) Merr.] (Amkul et al., 2019) are also highly conserved. Additionally, QTLs controlling domestication-related traits have been identified in *Vigna* species, including mung bean (Isemura et al., 2012), azuki bean (Isemura et al., 2007), rice bean (Isemura et al., 2010), cowpea (Kongjaimun et al., 2012; Lo et al., 2018), and zombi pea (Dachapak et al., 2018). Although these studies indicated that the QTLs for the same traits are generally conserved among *Vigna* species, the major QTLs for some important traits, such as seed size and weight, vary between species.

The genetic basis of zombi pea domestication is particularly interesting because the two cultivated types were likely domesticated from the same or different wild zombi pea species via different processes (Dachapak et al., 2017). The seed-type zombi pea plant produces relatively few branches and exhibits a determinate growth habit. In contrast, the tuberous root-type zombi pea is a highly branched, vine-like plant with an indeterminate growth habit. Dachapak et al. (2018) identified QTLs for the domestication of the tuberous root form of zombi pea. However, because of the sterility of the mapping population, the QTLs of key domestication-related traits, including pod dehiscence (pod shattering) and seed dormancy, were not detected (Dachapak et al., 2018). Therefore, identifying the QTLs controlling domestication-related traits in the seed-type zombi pea may provide important insights regarding the genetic basis of the domestication of this species and other *Vigna* species, which may facilitate the identification of new candidate genes for crop improvement. The objectives of this study were to (i) detect the QTLs associated with domestication-related traits and (ii) identify candidate genes for these traits in the seed-type zombi pea.

MATERIALS AND METHODS

Mapping Population

An F₂ population comprising 198 individuals developed from a cross between TVNu 240 and TVNu 1623 was used in this study. Specifically, TVNu 240 is a cultivated zombi pea (*V. vexillata* var. *macrosperma*) accession from the Central African Republic, whereas TVNu 1623 is a wild zombi pea (*V. vexillata* var. *vexillata*) accession from Nigeria. The F₂ population was previously used for developing high-density linkage maps and for the mapping of QTLs for bruchid (*Callosobruchus* spp.) resistance (Amkul et al., 2019). Briefly, the F₂ plants and 10 plants of each parent were grown under field conditions, with 0.75 m × 0.75 m spacing, at Kasetsart University, Kamphaeng Sean Campus, Nakhon Pathom, Thailand between December 2017 and March 2018.

Measurement of Domestication-Related Traits

Thirteen domestication-related traits were assessed in the F₂ and parental plants (Table 1). All of the traits, except for leaf area (LFA), were analyzed in previous studies on *Vigna* species, including mung bean (Isemura et al., 2012), azuki bean (Isemura et al., 2007), rice bean (Isemura et al., 2010), cowpea (Kongjaimun et al., 2012), and zombi pea (Dachapak et al., 2018). The LFA of nine fully expanded leaves was measured with the LI-3100C area meter (LI-COR Biosciences, United States) at 50 days after sowing. Stem length (STL) and stem thickness (STT) were recorded at harvest. The number of days from sowing to first flowering (FLD) and the number of days from sowing to harvesting of the first pod (PDDM) were also recorded. The mature pods and seeds of each plant were harvested separately to analyze the pod-related and seed-related traits. Specifically, pod length (PDL) was measured and the number of seeds per pod (SDNPPD) was determined based on 10 pods. Pod shattering was evaluated by counting the number of twists along the pod (PDT) of five pods maintained at room temperature. The seed length and width (SDL and SDW, respectively) were recorded as the average value of five seeds. The 100-seed weight (SD100WT) was measured twice using intact seeds. Seed dormancy was assessed by measuring seed water absorption (SDWA). Specifically, 40 intact seeds harvested from each plant were immersed in water in a plastic container and incubated at 25°C for 7 days. The number of seeds that imbibed water was recorded.

Heritability Estimation and Correlation Analysis

The broad-sense heritability (H^2) (Allard, 1999) of each trait was calculated using the following formula:

$$\frac{\sigma_{F_2}^2}{(\sigma_{TVNu\ 240}^2 + \sigma_{TVNu1623}^2)/2}$$

where $\sigma_{F_2}^2$, $\sigma_{TVNu240}^2$ and $\sigma_{TVNu1623}^2$ are the variances of the F₂ population, TVNu 240, and TVNu 1623, respectively. The correlation between traits was calculated with the R program (version 10.12.0).

Quantitative Trait Loci Mapping

The F₂ population was previously used for constructing a single nucleotide polymorphism (SNP)-based linkage map and for mapping QTLs for bruchid resistance (Amkul et al., 2019). The SNP markers were generated via a specific locus amplified fragment sequencing method (Sun et al., 2013). The map consisted of 11 linkage groups (LGs) and comprised 6,529 SNP markers (Amkul et al., 2019). The linkage map and genotypic (SNP marker) data were used for analyzing the QTLs of domestication-related traits in the present study.

The QTLs were analyzed according to the inclusive composite interval mapping (ICIM) method (Li et al., 2007) of the QTL IciMapping 4.2 software¹, with 0.1 cM steps. A probability of

stepwise regression of 0.001 was used for the ICIM. A logarithm of odds of 3.0 was used as the threshold for assessing the significance of the QTLs.

Identification of Candidate Genes for Domestication-Related Traits

Candidate genes at the major QTLs [i.e., phenotypic variance explained (PVE) $\geq 10\%$] for domestication-related traits were identified based on the nearest DNA markers flanking the QTLs on the cowpea reference genome sequence (Lonardi et al., 2019)². All of the annotated genes between the markers were identified, after which the genes with functions potentially associated with the specific trait were selected as candidate genes.

RESULTS

Variation and Heritability of Domestication-Related Traits

The data for the domestication-related traits of the F₂ population (mean and range) and the traits of the parents (mean) are summarized in Table 2. The values for the following traits were higher for the cultivated parent (TVNu 240) than for the wild parent (TVNu 1623): germination (SDWA), organ size (SD100WT, LFA, SDL, PDL, and STL), yield (SDNPPD), and earliness (FLD and PDDM). In contrast, the STT and branching (BRNPP) values were higher for the wild parent than for the cultivated parent. The parents had very similar PDT and SDL values. The SDWA, SD100WT, LFA, SDW, STT, BRNPP, and SDNPPD mean values of the F₂ population were between those of the parents (Table 2). However, the PDT and SDL mean values were similar among the F₂ population and both parents. The F₂ population mean values for PDL and STL were almost the same as those of the wild parent and cultivated parent, respectively, whereas the SDNPPD mean value was lower for the F₂ population than for either parent (Table 2). All traits exhibited a continuous distribution (Figure 1), indicative of a quantitative trait inheritance. Specifically, SDNPPD exhibited a bimodal distribution, suggesting that this trait may be controlled by a few loci. Transgressive segregation was clearly detected for several traits, including PDT, SDWA, LFA, PDL, STT, STL, BRNPP, FLD, PDDM, and SDNPPD (Figure 1 and Table 2), implying that both parents possessed alleles that positively and negatively affected these traits. Most of the traits exhibited moderate to relatively high heritability (Table 2), with the heritability value exceeding 0.70 for seven traits.

The correlations among the examined traits are listed in Supplementary Table S2. There were significant and positive correlations between related traits, including SD100WT and SDW or SDL, PDL and SDNPPD, and FLD and PDDM. Additionally, SD100WT was positively correlated with PDT, STT, STL, BRNPP, and LFA, but negatively correlated with SDNPPD. Interestingly, LFA was significantly and positively correlated with all of the other traits, with the exception of SDWA

¹<http://www.isbreeding.net>

²https://phytozome-next.jgi.doe.gov/info/Vunguiculata_v1_2

TABLE 1 | Domestication-related traits in the F₂ population derived from a cross between TVNu 240 (cultivated) and TVNu 1623 (wild).

General attribute	Organ	Trait	Trait abbreviation	QTL	Evaluation
Pod dehiscence	Pod	Number of twists (count)	PDT	<i>Pdt</i>	Number of twists along the length of shattered pod
Seed dormancy	Seed	Water absorption (%)	SDWA	<i>Sdwa</i>	Seeds that absorbed water at 7 days after sowing at room temperature
Gigantism	Seed	100-seed weight (g)	SD100WT	<i>Sd100wt</i>	Weight of 100 seeds
		Length (mm)	SDL	<i>Sdl</i>	Maximum distance from top to bottom of the seed
		Width (mm)	SDW	<i>Sdw</i>	Maximum distance from hilum to its opposite side
	Pod	Length (mm)	PDL	<i>Pdl</i>	Length of straight pod
	Stem	Thickness (mm)	STT	<i>Stt</i>	Stem diameter under the primary leaf (measured at flowering stage)
Plant type	Leaf	Leaf area	LFA	<i>Lfa</i>	Distance from pulvinus to leaf tip
	Stem	Length (up to 10th node) (cm)	STL	<i>Stl</i>	Length from node on primary leaf to node 10 of trifoliate leaf
	Branch	Number of branches	BRNPP	<i>BRNPP</i>	Number of branches on the main stem
Earliness	Flower	Days to first flower (day)	FLD	<i>Fld</i>	Number of days from planting to 1st flowering
	Pod	Days to maturity of 1st pod (day)	PDDM	<i>Pddm</i>	Number of days from 1st flowering to harvesting of 1st pod
Yield potential	Seed	Number of seeds per pod (seeds/pod)	SDNPPD	<i>Sdnppd</i>	Number of seeds per pod

TABLE 2 | Mean, minimum, and maximum values as well as the broad-sense heritability (H^2) of domestication-related traits of the zombi pea F₂ population derived from a cross between TVNu 240 (cultivated) and TVNu 1623 (wild).

Trait ^a	Unit	TVNu 240	TVNu 1623	F ₂ population		H^2
		Mean	Mean	Mean	Min–Max	
PDT	count	1.85	1.99	1.81	0–3.3	0.72
SDWA	%	59.80	23.20	27.25	2.00–89.47	0.68
SD100WT	g	5.89	2.61	3.92	1.35–5.80	0.78
SDL	mm	4.59	4.51	4.54	3.21–5.60	0.40
SDW	mm	4.35	2.83	3.37	2.60–4.19	0.54
PDL	cm	14.55	11.13	11.24	7.12–25.86	0.92
STT	mm	4.76	7.78	6.45	2.26–13.84	0.52
STL	cm	27.50	20.25	27.12	5.50–72.00	0.83
BRNPP	count	0.00	4.20	1.66	0–8	0.42
FLD	day	59.40	67.40	55.93	40–91	0.75
PDDM	day	77.30	87.60	74.73	52–110	0.82
SDNPPD	count	20.54	15.15	12.38	2.75–21.8	0.81
LFA	mm ²	59.65	35.35	41.80	9.44–132.98	0.64

^aTrait abbreviations are shown in **Table 1**.

and PDT. Moreover, SDWA was significantly and negatively correlated with PDT, FLD, PDDM, and BRNPP. The data also revealed a significant correlation between BRNPP and all other analyzed traits.

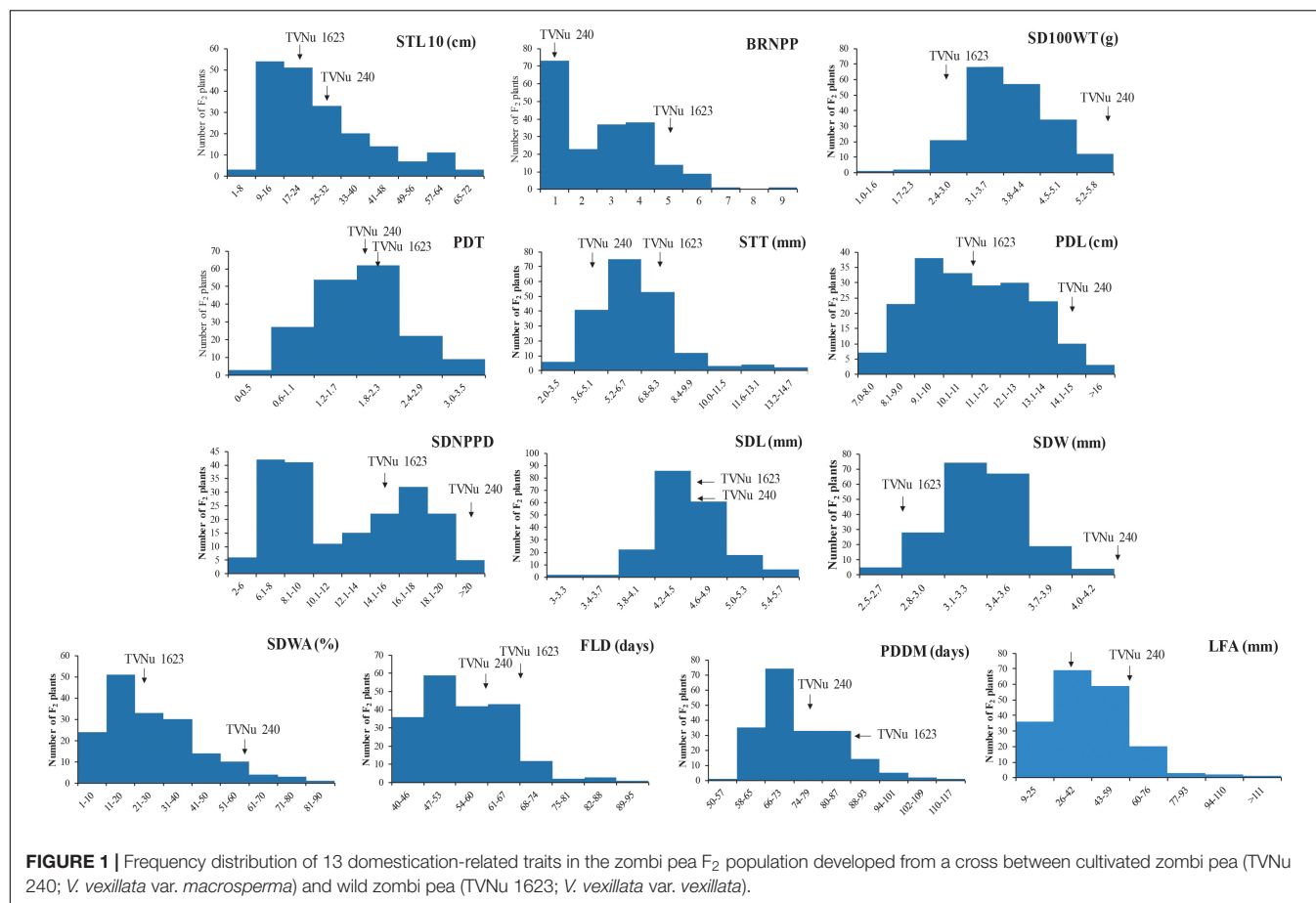
QTLs and Candidate Genes for Domestication-Related Traits

A total of 62 QTLs were identified for the 13 analyzed traits, with 1–11 loci per trait (**Table 3** and **Figure 2**). These 62 QTLs were

distributed in all LGs, except for LG10. The number, location, and effect of the QTLs identified for each trait are summarized in **Table 3**.

The wild and cultivated parents differed considerably regarding SDWA (23.20 and 59.80%, respectively). The three QTLs detected for this trait, *qSdwa1.1+*, *qSdwa3.1-*, and *qSdwa5.1+*, were in LGs 1, 3, and 5, respectively (**Table 3** and **Figure 2**). The QTL in LG3, *qSdwa3.1-*, had the largest effect (PVE of 22.2 %). Interestingly, at this QTL, the cultivated zombi pea alleles decreased the seed coat permeability (decreased germination). The *qSdwa1.1+* and *qSdwa5.1+* QTLs both accounted for about 9% of the total variation. At these two QTLs, the cultivated zombi pea alleles increased the seed coat permeability. The *qSdwa3.1-* region consisted of approximately 179.0 kb on cowpea chromosome 3 and included 16 genes (**Supplementary Table S3**). Of these genes, *Vigun03g339300*, encoding a calmodulin-binding protein-like protein, was identified as a candidate gene for this QTL. The *qSdwa5.1+* region spanned about 1.74 Mb on cowpea chromosome 5 and contained 38 genes (**Supplementary Table S3**), with *Vigun05g170600* and *Vigun05g170700*, both of which encode gibberellin 2-beta-dioxygenase 1, identified as candidate genes for this QTL.

The wild zombi pea exhibited extensive pod dehiscence, in contrast to the cultivated zombi pea (**Table 2**). Two QTLs were identified for this trait, with one each in LGs 5 and 11 (**Table 3** and **Figure 2**). The QTL in LG11, *qPdt11.1-*, had a PVE of 17.6%, whereas the QTL in LG5, *qPdt5.1-*, had a PVE of only 5.7%. At both QTLs, the cultivated zombi pea alleles decreased pod dehiscence. The *qPdt11.1-* and *qPdt5.1-* regions spanned approximately 406 and 72.7 kb on cowpea chromosomes 11 and 5, and included eight and



seven genes, respectively (**Supplementary Table S3**). These genes included *Vigun11g053700* [encoding the bifunctional UDP-glucose 4-epimerase and UDP-xylose 4-epimerase 1 (UGE1)] and *Vigun11g054200* [encoding glycosyltransferase family 14 (GT14)], which were considered as candidate genes for *qPdt11.1*-. The *Vigun05g241400* gene, encoding UDP-apiose/xylose synthase, was selected as a candidate gene for *qPdt5.1*-.

The cultivated zombi pea seeds were about 2.5-fold larger than the wild zombi pea seeds (**Table 2**). Regarding the seed size-related traits (SD100WT, SDL, and SDW), 4–10 QTLs were identified, with at least one QTL in all LGs, with the exception of LGs 10 and 11 (**Table 3** and **Figure 2**). At all QTLs, except for *qSd100tw4.1*- and *qSdw9.1*-, the cultivated zombi pea alleles were associated with increases in seed size. The QTLs with the largest effects on SD100WT and SDL were *qSd100wt5.1*+ (PVE of 17.9%) and *qSdl5.2*+ (PVE of 23.8%), respectively, both of which belonged to LG5 (**Table 3**). The QTL that contributed the most to SDW was *qSdw9.1*+ (PVE of 20.6%), which was in LG9. Additionally, LG5 and LG8 included 11 and 5 QTLs for seed size, respectively. Interestingly, *qSd100wt5.1*+ and *qSdl5.2*+ were mapped to the same position. Another large-effect QTL for SD100WT was *qSd100tw5.2*+ (PVE of about 10%). The *qSd100wt5.1*+ and *qSdl5.2*+ region covered 1.03 Mb on cowpea chromosome

5 and contained 38 genes (**Supplementary Table S3**). The identified candidate genes included *Vigun05g139100*, which encodes an auxin-responsive protein-related protein, and *Vigun05g140300*, which encodes a protein with an unknown function (DUF538). The *qSd100wt5.2*+ region comprised 217.9 kb on cowpea chromosome 5 and contained 30 genes (**Supplementary Table S3**). Among these genes, *Vigun05g277600*, encoding an IAA-amino acid hydrolase ILR1-like 1-related protein, was considered as a candidate gene for *qSd100wt5.2*+. Moreover, the *qSdw9.1*+ region covered a 130.0-kb region on cowpea chromosome 9 and comprised 12 genes (**Supplementary Table S3**). Both *Vigun09g215300* (WRKY transcription factor 40-related protein) and *Vigun09g215400* (WRKY DNA-binding protein) were selected as candidate genes for this QTL.

The leaves of the cultivated parent were larger than those of the wild parent. The only QTL detected for LFA, *qLfaA9.1*- (PVE of 18.2%), was mapped to LG9 (**Table 3** and **Figure 2**). Moreover, the cultivated parent alleles at this QTL decreased the LFA value. The *qLfa9.1*- region spanned 201.1kb on cowpea chromosome 9 and comprised 17 genes (**Supplementary Table S3**). Among these genes, *Vigun09g018300* and *Vigun09g017000*, encoding a WD40 repeat-containing protein and falz-related bromodomain-containing protein/transcription factor GTE1, respectively, were identified as candidate genes for this QTL.

TABLE 3 | Quantitative trait loci for 13 domestication-related traits in the zombi pea F_2 population derived from a cross between TVNu 240 (female cultivated) and TVNu 1623 (male wild).

Trait ¹	QTL name	LG	Position ²	LOD	PVE ³ (%)	Add ⁴	Dom ⁵
PDT	<i>qPdt5.1-</i>	5	137.4	3.1	5.68	-0.22	0.00
	<i>qPdt11.1-</i>	11	22.3	8.8	17.57	-0.07	-0.50
SDWA	<i>qSdwa1.1+</i>	1	142.3	3.2	6.34	1.98	7.97
	<i>qSdwa3.1-</i>	3	145.0	10.5	22.22	-10.66	-5.82
	<i>qSdwa5.1+</i>	5	88.3	4.1	8.71	6.84	-1.33
SD100WT	<i>qSd100wt1.1+</i>	1	135.1	6.9	5.46	0.25	0.06
	<i>qSd100wt3.1+</i>	3	29.3	3.8	2.89	0.19	0.01
	<i>qSd100wt3.1+</i>	3	164.0	3.4	2.57	0.16	0.08
	<i>qSd100wt4.1-</i>	4	146.2	3.6	2.76	-0.02	0.26
	<i>qSd100wt5.1+</i>	5	78.6	19.6	17.92	0.05	0.65
	<i>qSd100wt5.2+</i>	5	150.1	11.6	9.54	0.32	0.10
	<i>qSd100wt8.1+</i>	8	7.7	3.4	2.56	0.19	-0.02
	<i>qSd100wt8.2+</i>	8	112.6	9.7	7.84	0.30	0.09
LFA	<i>qLfa9.1-</i>	9	1.1	8.1	18.24	-8.69	6.75
SDL	<i>qSdl5.1-</i>	5	30.9	3.5	4.79	-0.11	-0.01
	<i>qSdl5.2+</i>	5	78.6	15.2	23.84	0.03	0.36
	<i>qSdl5.3+</i>	5	155.2	6.4	9.07	0.17	-0.04
	<i>qSdl8.1+</i>	8	154.2	5.6	7.87	0.14	0.04
SDW	<i>qSdw2.1+</i>	2	109.2	4.9	4.00	0.03	0.10
	<i>qSdw3.1+</i>	3	163.0	6.0	5.11	0.09	0.02
	<i>qSdw4.1+</i>	4	51.3	6.6	5.67	0.01	-0.15
	<i>qSdw4.2+</i>	4	65.8	11.1	10.10	0.02	0.18
	<i>qSdw5.1+</i>	5	56.3	5.4	4.51	0.03	0.12
	<i>qSdw5.2+</i>	5	156.5	8.2	7.18	0.11	0.01
	<i>qSdw8.1+</i>	8	58.1	6.8	5.81	0.10	-0.02
	<i>qSdw8.2+</i>	8	154.2	4.0	3.29	0.07	0.02
	<i>qSdw9.1+</i>	9	135.3	20.6	20.56	0.19	0.05
	<i>qSdw9.2-</i>	9	157.8	12.0	10.82	-0.14	0.03
PDL	<i>qPdl3.1+</i>	3	121.1	3.3	3.30	0.26	0.71
	<i>qPdl5.1-</i>	5	80.4	28.0	37.72	-0.03	-2.75
	<i>qPdl6.1-</i>	6	83.1	6.0	6.05	-0.68	-0.46
	<i>qPdl7.1+</i>	7	39.3	5.2	5.37	0.74	-0.16
	<i>qPdl9.1-</i>	9	48.3	4.5	4.57	-0.63	0.22
STT	<i>qStt5.1-</i>	5	64.5	3.4	5.52	-0.58	-0.10
	<i>qStt8.1-</i>	8	144.6	4.5	7.43	-0.28	-0.98
	<i>qStt9.1+</i>	9	2.9	16.2	30.63	1.48	0.24
	<i>qStt9.2-</i>	9	8.3	31.1	72.41	-2.25	-0.11
STL	<i>qStl1.1+</i>	1	4.5	3.2	5.78	1.92	6.21
	<i>qStl5.1-</i>	5	62.9	4.9	9.01	-5.82	0.17
	<i>qStl8.1-</i>	8	1.5	4.4	7.97	-2.24	-7.78
	<i>qStl9.1-</i>	9	24.2	5.6	10.13	-5.83	4.27
BRNPP	<i>qBrnpp1.1-</i>	1	23.5	4.1	5.99	-0.52	0.27
	<i>qBrnpp2.1-</i>	2	62.3	4.8	7.18	-0.58	-0.05
	<i>qBrnpp5.1+</i>	5	56.2	7.5	11.51	0.04	-1.17
	<i>qBrnpp11.1-</i>	11	20.7	3.9	5.75	-0.56	0.17
FLD	<i>qFld1.1-</i>	1	56.0	4.3	6.62	-3.59	-0.55
	<i>qFld2.1-</i>	2	16.4	10.1	16.47	-4.98	-2.48
	<i>qFld5.1-</i>	5	35.1	11.1	18.61	-5.34	2.02
	<i>qFld5.2+</i>	5	127.7	3.6	5.54	3.39	0.41
PDDM	<i>qPddm2.1-</i>	2	12.0	5.5	10.02	-4.10	-1.97
	<i>qPddm5.1-</i>	5	38.7	5.2	9.47	-3.84	0.36
	<i>qPddm9.1-</i>	9	3.0	3.2	5.24	-2.28	2.83
	<i>qPddm11.1-</i>	11	111.6	3.2	5.56	-2.96	-2.10

TABLE 3 | Continued

Trait ¹	QTL name	LG	Position ²	LOD	PVE ³ (%)	Add ⁴	Dom ⁵
SDNPPD	<i>qSdnppd1.1-</i>	1	2.8	5.5	1.90	-0.80	0.68
	<i>qSdnppd2.1-</i>	2	137.2	4.4	1.83	-0.83	0.18
	<i>qSdnppd3.1-</i>	3	71.4	31.6	15.00	-0.40	-3.57
	<i>qSdnppd5.1-</i>	5	80.4	47.3	27.76	-0.40	-4.88
	<i>qSdnppd7.1-</i>	7	43.3	38.8	21.05	-3.20	-0.18
	<i>qSdnppd7.2+</i>	7	51.5	50.9	33.02	3.92	-0.13
	<i>qSdnppd8.1-</i>	8	77.3	4.3	1.50	-0.71	0.54
	<i>qSdnppd9.1-</i>	9	48.2	4.0	1.39	-0.72	0.30
	<i>qSdnppd9.2+</i>	9	106.8	5.5	1.99	0.08	1.33

¹See **Table 1** for trait description. ²Position (centimorgan) on the linkage map.

³Phenotypic variance explained by the QTL. ⁴Additive effect. ⁵Dominant effect.

The cultivated zombi pea pods were approximately 3.5 cm longer than the wild zombi pea pods (**Table 2**). Five QTLs related to PDL were detected on different LGs (**Table 3** and **Figure 2**). At QTLs *qPdl3.1+* and *qPdl7.1+*, the cultivated zombi pea alleles increased the PDL value. Additionally, at QTLs *qPdl5.1-*, *qPdl6.1-*, and *qPdl9.1-*, the wild zombi pea alleles increased the PDL value. The *qPdl5.1-* QTL in LG5 had a PVE of 37.7%, which was considerably higher than the PVE for the other four QTLs (<5%). The *qPdl5.1-* region spanned 860.4 kb on cowpea chromosome 5 and contained 41 genes (**Supplementary Table S3**). The *Vigun05g125800* gene, encoding pectate lyase 1, was chosen as a candidate gene for *qPdl5.1-*.

The cultivated zombi pea plants had shorter stems (the first 10 internodes), but more branches, than the wild zombi pea plants. The four QTLs for STL belonged to different LGs (**Table 3** and **Figure 2**). The PVE of each of the QTLs for STL was $\leq 10\%$, but was highest for *qStl9.1-* in LG9. The cultivated parent alleles at three of the four QTLs decreased the internode length. Similarly, four QTLs on different LGs were detected for BRNPP, and the PVE of each QTL was $\leq 12\%$, but was highest for *qBrnpp5.1-* in LG5 (**Table 3** and **Figure 2**). The cultivated parent alleles at three of the four QTLs decreased the number of branches. Only one marker flanking *qStl9.1-*, Marker293030, matched a sequence in the cowpea reference genome. The *Vigun09g037900* gene, encoding a MOB-like protein phoein (MOB1), which was detected approximately 113 kb from Marker293030, was considered to be a candidate gene for *qStl9.1-*. The *qBrnpp5.1-* region was less than 1.0 kb on cowpea chromosome 5 and included *Vigun05g082200*, which encodes an auxin efflux carrier family protein (**Supplementary Table S3**).

Domestication generally leads to cultivated crops that flower and mature earlier than their wild progenitors. The cultivated zombi pea plants flowered and matured approximately 10 and 8 days earlier, respectively, than the wild zombi pea plants (**Table 2**). Four QTLs on three LGs were identified for FLD (**Table 3** and **Figure 2**). The effects of QTLs *qFld2.1-* in LG2 and *qFld5.1-* in LG5 on FLD were similar (PVE of 16.5 and 18.6%, respectively). With the exception of *qFld5.2+*, the cultivated parent alleles at these QTLs enhanced flowering. The *qFld2.1-* region spanned 391.5 kb on cowpea chromosome 2 and comprised 28 genes



FIGURE 2 | Quantitative trait locus plots for 13 domestication-related traits in the zombi pea F_2 population developed from a cross between cultivated zombi pea (TVNu 240; *V. vexillata* var. *macrosperma*) and wild zombi pea (TVNu 1623; *V. vexillata* var. *vexillata*). The x-axis indicates the linkage groups, whereas the y-axis indicates the logarithm of odds (LOD) scores. The gray line horizontal to the y-axis indicates the LOD significance threshold.

(Supplementary Table S3). Both *Vigun02g051100* (WRKY transcription factor 57) and *Vigun02g052800* (ring finger and CCH-type zinc finger domain-containing protein) were selected as candidate genes for *qFld2.1*-. An analysis of *qFld2.1*- revealed that this QTL spanned 90.5 kb on cowpea chromosome 5 and contained eight genes (Supplementary Table S3), of which *Vigun05g043100*, encoding a bHLH transcription factor PRE1-related protein, was identified as a candidate gene.

Similarly to the FLD analysis, four QTLs on different LGs were identified for PDDM (Table 3 and Figure 2). The

qPddm2.1- (LG2) and *qPddm5.1*- (LG5) QTLs had the largest effects on this trait (PVE of about 10%). At all four QTLs, the cultivated parent alleles accelerated maturity. Additionally, *qFld2.1*- and *qPddm2.1*- were located close to each other, whereas *qFld2.1*- was mapped close to *qPddm5.1*-. The *qPddm2.1*- region covered 83.3 kb on cowpea chromosome 2 and contained seven genes, including *Vigun02g050200*, encoding the 14-3-3-like protein GF14 IOTA, which was chosen as a candidate gene. The *qPddm5.1*- region spanned only 19.0 kb on cowpea chromosome 5 and contained four genes

(**Supplementary Table S3**). The *Vigun05g050300* gene, which encodes 3-deoxy-8-phosphooctulonate synthase/phospho-2-keto-3-deoxyoctonate aldolase, was designated as a candidate gene for *qPddm5.1*–.

The cultivated parent produced more seeds per pod than the wild parent. Nine QTLs were detected for SDNPPD (**Table 3** and **Figure 2**). The *qSdnppd3.1*–, *qSdnppd5.1*–, *qSdnppd7.1*–, and *qSdnppd7.2*+ QTLs had larger effects on SDNPPD (PVE of 15.0, 27.8, 21.1, and 33.0%, respectively) than the other QTLs (PVE of only about 2.0%). Regarding the *qSdnppd7.2*+ and *qSdnppd9.2*+ QTLs, the cultivated zombi pea alleles increased the seed yield. In contrast, at the other QTLs, the cultivated zombi pea alleles had the opposite effect on seed yield. The *qSdnppd3.1*– region covered about 221.6 kb on cowpea chromosome 3 and comprised five genes, including *Vigun03g187300*, which encodes abscisic acid (ABA)-insensitive 5-like protein 6. The *qSdnppd5.1*– and *qSdnppd7.2*+ regions spanned approximately 860.4 kb and 2.38 Mb on cowpea chromosomes 5 and 7 and contained 41 and 60 genes, respectively. Moreover, *Vigun05g126900*, encoding the MALE STERILE 5 protein, was identified as a candidate gene for *qSdnppd5.1*– (**Supplementary Table S3**). Both *Vigun07g094201* and *Vigun07g096300*, encoding the B3-like transcription factor and the B3 DNA-binding domain protein, respectively, were selected as candidate genes for *qSdnppd7.2*+ (**Supplementary Table S3**). An analysis of *qSdnppd7.1*–, revealed only one flanking marker (Marker225965) on cowpea chromosome 7. Furthermore, *Vigun07g074900*, encoding the transcription factor TCP21, and located approximately 81.6 kb from Marker225965, was identified as a candidate gene for *qSdnppd7.1*–.

DISCUSSION

Compared with wild zombi pea, cultivated zombi pea has undergone phenotypic changes due to domestication. In this study, we identified 62 QTLs controlling 13 domestication-related traits in zombi pea. The QTLs were distributed in all LGs, except for LG10. The number of QTLs identified for individual traits varied from one to ten, with an average of five QTLs (**Table 3**). Large-effect QTLs (PVE > 15%) were detected for all examined traits, except for STL, PDDM, and BRNPP (**Table 3**). These results imply that the zombi pea domestication-related traits are controlled by one or only a few major QTLs along with a small number of minor QTLs. This is consistent with the results of earlier investigations of other *Vigna* species, including mung bean (Isemura et al., 2012), azuki bean (Isemura et al., 2007), rice bean (Isemura et al., 2010), and cowpea (Kongjaimun et al., 2012; Lo et al., 2018). However, in a recent study, Dachapak et al. (2018) identified an average of 2.3 QTLs for zombi pea domestication-related traits, which is about twofold less than the number of QTLs detected in our study. This difference may have been due to the sterility of the mapping population used in the earlier study.

In *Vigna* species, the major QTLs for domestication-related traits, especially organ size, have always been mapped to the same LG (Isemura et al., 2007, 2010, 2012; Kongjaimun et al., 2012; Lo et al., 2018; Yundaeng et al., 2019). In our study,

the major QTLs for the domestication-related traits of zombi pea, including SD100WT, SDL, PDL, STL, BRNPP, FLD, PDDM, and SDNPPD, were clustered in LG5, corresponding to cowpea chromosome 5. In an earlier study of cowpea, which is the legume crop most closely related to zombi pea, the major QTLs for domestication-related traits, such as pod shattering, seed dormancy, seed size (weight, width, and length), pod size (length and width), stem size (thickness and length), and leaf size (width and length), were clustered in LG7 (Kongjaimun et al., 2012), which corresponds to chromosome 5. These findings suggest that the genetics underlying the effects of domestication on the organ size (e.g., seed size, pod length, and stem length) of these two African *Vigna* crops may be the same. However, the genetic basis of seed dormancy and pod shattering, which are key domestication-related traits, varies between these two species.

Domestication results in a decrease in or a loss of seed dormancy, enabling the uniform germination of seeds, which is relevant for agricultural production. The seed dormancy of cultivated zombi pea is suppressed. In this study, we identified one major and two minor QTLs for the seed dormancy-related trait (**Table 3**). Specifically, *qSdwa3.1*– was detected as the major QTL, accounting for 22.2% of the phenotypic variation. Because the wild alleles at *qSdwa3.1*– positively affected SDWA (i.e., decreased dormancy), their introgression into cultivated zombi pea may be useful for decreasing seed dormancy. The *Vigun03g339300* gene, encoding a calmodulin-binding protein-like protein, was identified at *qSdwa3.1*–. Calmodulin (CaM) and CaM-like (CML) proteins are the major Ca²⁺-binding proteins. The CaM/CML signals are involved in the ABA-induced inhibition of seed germination and seedling growth (Zhou et al., 2018). In *Arabidopsis thaliana* (Arabidopsis), IQM4, which is a CaM-binding protein, was recently reported to affect seed dormancy by modulating ABA biosynthesis and ABA signaling during seed maturation and germination (Zhou et al., 2018). In cowpea, *Sdp1.1*+ is the major QTL related to seed dormancy, explaining about 40% of the trait variation (Kongjaimun et al., 2012). The markers flanking *Sdp1.1*+ suggest that this QTL is located on chromosome 3 between positions 45,527,961 and 51,968,469. Thus, *qSdwa3.1*– and *Sdp1.1*+ are different loci. The minor QTL *qSdwa5.1*+ accounted for about 9% of the phenotypic variation, and included *Vigun05g170600* and *Vigun05g170700*, both of which encode gibberellin 2-beta-dioxygenase 1 (GA2OX1). Gibberellin is a key plant hormone that regulates seed dormancy and germination by antagonistically suppressing ABA-triggered seed dormancy (Shu et al., 2016). Loss-of-function mutations to *GA2ox2*, *GA2ox7*, and *GA2ox8*, which encode GA2OX, reportedly enhance seed germination in Arabidopsis (Schomburg et al., 2003; Yamauchi et al., 2007).

Preventing or limiting pod dehiscence (pod shattering) may minimize pre-harvest yield losses, resulting in a more efficient harvest. Thus, pod indehiscence may be an advantageous trait during the harvesting of seeds, making it an important consideration during crop domestication. A major QTL in LG11 and a minor QTL in LG5 were related to pod twisting (**Table 3**). The major QTL, *qPdt11.1*+, explained 17.6% of the dehiscence variation, and included two genes, *Vigun11g053700*, encoding UGE1, and *Vigun11g054200*, encoding the glycosyltransferase

GT14. Earlier studies confirmed that UGE is a nucleotide sugar interconversion enzyme that mediates the interconversion of UDP-D-glucose and UDP-D-galactose and contributes to cell wall biosynthesis (Barber et al., 2006; Rösti et al., 2007). Glycosyltransferases in plants are mainly involved in the biosynthesis of cell wall polysaccharides and glycoproteins (Hansen et al., 2012). Although the functions of plant GT14 remain uncharacterized, two *GT14* genes identified in a hybrid aspen species (*Populus tremula* × *P. tremuloides*) reportedly exhibit xylem-specific expression (Aspeborg et al., 2005). Accordingly, GT14 may participate in secondary cell wall biosynthesis. In cowpea, three major QTLs for pod shattering have been identified (Kongjaimun et al., 2012; Suanum et al., 2016; Lo et al., 2018). Candidate genes for these QTLs include *Vigun02g095200* (cellulose synthase), *Vigun03g306000* (NAC domain transcription factor), *Vigun03g302600* (C2H2 zinc finger protein), *Vigun05g262100* (MYB46 transcription factor), *Vigun05g273500* (MYB26 transcription factor), and *Vigun05g266600* (beta glucosidase) (Suanum et al., 2016; Lo et al., 2018; Takahashi et al., 2019; Watcharatpong et al., 2020). Thus, the major QTLs/genes which control pod shattering differ between zombi pea and cowpea.

In most cases, domestication increases organ sizes, especially the harvested or consumed or exploited organs. Seed size is a key determinant of seed yield in legume crops as well as evolutionary fitness. It is a polygenic trait affected by various cellular processes. Twenty-two QTLs were identified for seed size-related traits, of which the QTLs with the largest effects on seed weight (*qSd100wt5.1+*) and seed length (*qSdl5.1+*) were mapped to the same position of LG5 (Table 3). Two genes were identified at these two QTLs, namely *Vigun05g139100*, encoding an auxin-responsive protein-related protein, and *Vigun05g140300*, encoding DUF538. Auxin is important for seed development, with specific effects on the embryo, endosperm, and seed coat, all of which influence seed size and yield (Cao et al., 2020). Genes encoding auxin-related proteins control seed size (Schruoff et al., 2006; Ishimaru et al., 2013; Liu et al., 2015; Sun et al., 2017). In Arabidopsis mutant line F08314, the overexpression of *RAFL06-12-J05* (encoding DUF538) results in seeds that are larger than the wild-type seeds (Akiyama et al., 2014)³. In addition to *qSd100wt5.1+*, *qSd100wt5.2+* also substantially affects seed weight. We identified *Vigun05g277600*, which encodes an IAA-amino acid hydrolase ILR1-like 1-related protein, as a candidate gene for *qSd100wt5.2+*. Earlier studies revealed ILR1 is involved in auxin conjugation and affects auxin homeostasis (Rampey et al., 2004; and see Cao et al., 2020 for review). In rice, *TGW3*, encoding an IAA-glucose hydrolase, controls the rice grain weight and yield (Ishimaru et al., 2013). The QTL with the largest effect on seed length, *qSdl9.1+*, was included in LG9. Both *Vigun09g215300* (WRKY transcription factor 40-related protein) and *Vigun09g215400* (WRKY DNA-binding protein) were detected as candidate genes for *qSdl9.1+*. In plants, WRKY proteins have diverse biological functions related to seed development as well as developmental and hormone-controlled processes (Bakshi and Oelmüller, 2014). For example,

MINISEED3, which encodes WRKY10, regulates seed size in Arabidopsis (Luo et al., 2005) and *Loose Panicle1*, which encodes a WRKY transcription factor, regulates seed size by increasing the seed length and width in foxtail millet [*Setaria italica* (L.) P. Beauv.] (Xiang et al., 2017). Additionally, *SoyWRKY15a*, which encodes WRKY15, is associated with changes to seed size due to the domestication of soybean [*Glycine max* (L.) Merr.], with expression levels that vary significantly between domesticated and wild soybeans (Gu et al., 2017).

In the current study, *qLfa9.1-*, with a PVE of 18.2%, was the only QTL identified for leaf size (Table 3). The *Vigun09g018300* gene (WD40 repeat-containing protein) was detected as a candidate gene for this QTL. In cucumber, *LITTLELEAF*, which encodes a WD40 repeat domain-containing protein, regulates the size of various organs, including leaves (Yang et al., 2018). We also identified *Vigun09g017000*, encoding the bromodomain-containing protein GTE1, as a candidate gene. In Arabidopsis, *GTE4* and *GTE6* encode bromodomain-containing proteins that regulate leaf shape (and thus leaf size). Specifically, *GTE4* controls the mitotic cell cycle during plant development (Airoidi et al., 2010), whereas *GTE6* induces the expression of *ASYMMETRIC LEAVES1*, encoding a MYB-domain protein that regulates proximodistal leaf patterns (Chua et al., 2005).

Legume domestication results in increased pod length. Five QTLs were identified for pod length, but *qPdl5.1-* was the only major QTL, accounting for nearly 38% of the trait variation (Table 3). Pod elongation requires changes to fiber cell structures and components. The *Vigun05g125800* gene, encoding pectate lyase 1, was selected as a candidate gene for *qPdl5.1-*. Pectin is a major polymer in the plant primary cell wall. Pectate lyase is a pectin-degrading enzyme that can degrade demethylesterified homogalacturonan (HG), which is a pectic polymer (Palusa et al., 2007). The methylesterification status of HG can regulate cellular growth and cell shape, thereby influencing plant growth and development (Wolf et al., 2009; Peaucelle et al., 2011). A previous study regarding common bean (*Phaseolus vulgaris* L.), which is a legume closely related to zombi pea, revealed that the major changes in cell wall polysaccharides during pod development involve pectic polymers and that the deposition of galactose-rich pectic polymers in cell walls increases in elongating pods (Stolle-Smits et al., 1999). Moreover, during pod elongation, the HG level steadily increases. Accordingly, pectate lyase may help mediate the elongation of zombi pea pods. In rice, a mutation to *DEL1*, which encodes a pectate lyase precursor, decreases the total pectate lyase activity, leading to an increase in the methylesterified HG content and altered cell wall compositions and structures, with implications for plant height, root length, flag leaf length, grain size, and panicle and internode lengths. In cotton, the downregulated expression of pectate lyase genes *GhPEL48_Dt* and *GhPEL76* decreases fiber length by impeding pectin degradation (Wang et al., 2009; Sun et al., 2020).

Most domesticated legume crops exhibit greater determinate growth, with less branching, than their wild ancestors. The *qStl9.1-* QTL explained about 10% of the variability in the stem length (plant height) (Table 3). We identified *Vigun09g037900*, encoding MOB1, as a candidate gene for this QTL. Arabidopsis

³<http://range-v2.psc.riken.jp/cdna/RAFL06-12-J05>

MOB1 proteins are critical for plant development because they promote auxin signaling (Cui et al., 2016) and regulate jasmonate accumulation (Guo et al., 2020). Mutations to the *MOB1* gene in *Arabidopsis* negatively affect plant height (Xiong et al., 2016; Guo et al., 2020). We identified *Vigun05g082200*, encoding an auxin efflux carrier protein, as a candidate gene for *qBrnpp5.1*—, which controls the number of branches in zombi pea. The *Vigun05g082200* sequence is similar to that of the PIN-LIKES 6 transporter (AT5G01990). The PIN proteins form a family of auxin efflux carriers. Auxin is well known for its ability to regulate numerous plant growth and developmental processes such as meristem formation and lateral organ formation and patterning (Wang and Jiao, 2018). For example, PIN6 contributes to auxin signaling-mediated developmental processes, including lateral/adventitious root organogenesis, primary/lateral root development and growth, and shoot apical dominance (Cazzonelli et al., 2013; Nisar et al., 2014; Simon et al., 2016). In rice, *PIN1* and *PIN2* are involved in tillering (Xu et al., 2005; Chen et al., 2012).

Generally, wild ancestors of domesticated crops are sensitive to short-day conditions. In this study, we identified four QTLs related to days-to-flowering in zombi pea, of which two are large-effect QTLs, *qFld2.1*— and *qFld5.1*— (PVE of 16.5 and 18.6%, respectively). We selected *Vigun02g051100* (WRKY57 transcription factor) and *Vigun02g052800* (RING finger and CCCH-type zinc finger protein) as candidate genes for *qFLD2.1*— and *qFLD5.1*—, respectively. The WRKY transcription factors are involved in physiological changes and responses to biotic and abiotic stresses. In *Arabidopsis*, WRKY12, WRKY13, WRKY71, and WRKY75 regulate flowering time (Li W. et al., 2016; Yu et al., 2016; Zhang et al., 2018). Additionally, WRKY75 interacts with DELLA proteins, which are the major components involved in gibberellic acid perception and signaling (Zhang et al., 2018). A mutation to *WRKY75* reportedly delays flowering, whereas the overexpression of this gene has the opposite effect (Zhang et al., 2018). Plant zinc finger proteins comprise a large protein family and have diverse functions related to plant development and resistance to biotic and abiotic stresses. The *Arabidopsis* genes *AtC3H17*, *AtZFP1*, *AtKHZ1*, and *AtKHZ2* as well as the *Medicago sativa* gene *MsZFN*, encoding CCCH zinc finger proteins, influence flowering time (Chao et al., 2014; Seok et al., 2016; Yan et al., 2017; Wang et al., 2018).

Four QTLs were detected for days-to-flowering in zombi pea, of which *qPddm2.1*— and *qPddm5.1*— had a similar PVE of about 10% (Table 3). We identified *Vigun02g050200*, encoding the 14-3-3-like protein GF14 IOTA, and *Vigun05g050300*, encoding 3-deoxy-8-phosphooctulonate synthase/phospho-2-keto-3-deoxyoctonate aldolase, as the candidate genes for *qPddm2.1*— and *qPddm5.1*—, respectively. The 14-3-3 (GF14) proteins are critical components of various cellular signaling processes and are important regulators of several physiological processes in plants. The *GF14* genes are associated with fruit development and/or ripening in banana (*Musa acuminata* L.) (Li et al., 2012; Li M. et al., 2016) and pear (*Pyrus pyrifolia* Nakai.) (Shi and Zhang, 2014). Phospho-2-keto-3-deoxyoctonate aldolase is also known as

3-deoxy-d-manno-2-octulosonic acid-8-phosphate synthase (KDO8P), which has a key function related to the biosynthesis of KDO, which is a sugar present only in the rhamnogalacturonan II pectic fraction of the primary cell walls of higher plants (Delmas et al., 2003). In the common bean, which is a legume closely related to zombi pea, pod elongation and maturation are accompanied by changes to rhamnogalacturonan (Stolle-Smits et al., 1999).

The number of seeds per pod is one of the yield determinants of legumes. Some domesticated legumes have more seeds per pod than their wild ancestors. Up to nine QTLs were detected for this trait in zombi pea (Table 3). Among these QTLs, *qSdnppd3.1*—, *qSdnppd5.1*—, *qSdnppd7.1*—, and *qSdnppd7.2+* had relatively large effects (PVE between 15 and 33%). Because the wild alleles at *qSdnppd3.1*—, *qSdnppd5.1*—, and *qSdnppd7.1*— were observed to increase the number of seeds per pod, they may be useful for increasing the seed yield of cultivated zombi pea. We identified *Vigun03g187300* (ABA-insensitive 5-like protein 6) as a candidate gene for *qSdnppd3.1*—. The encoded protein is an ABA-responsive element (ABRE)-binding factor that regulates ABRE-dependent gene expression (Nakashima and Yamaguchi-Shinozaki, 2013). In *Arabidopsis*, an ABA-deficiency reportedly decreases the number of seeds per silique (Cheng et al., 2014). The *Vigun05g126900* gene, encoding MALE STERILE 5, was selected as a candidate gene for *qSdnppd5.1*—. In an earlier investigation on *Arabidopsis*, mutations to *MS5*, which encodes MALE STERILE 5, resulted in the formation of “polyads” (i.e., tetrads with more than four pools of chromosomes following male meiosis) (Glover et al., 1998). Plants that are homozygous for the *ms5* recessive allele reportedly exhibit stunted growth and produce empty siliques, whereas in plants that are heterozygous for *MS5*, silique elongation and seed set are less inhibited (Glover et al., 1998). We identified *Vigun07g094201* (B3-like transcription factor) and *Vigun07g096300* (B3 DNA-binding domain protein) as the candidate genes for *qSdnppd7.1*—. The B3 domain family proteins are associated with ovule development (Skinner and Gasser, 2009). A recent study on *Arabidopsis* investigated the roles of three *Reproductive Meristem (REM)* genes, *REM34*, *REM35*, and *REM36*, which encode B3 domain family transcription factors affecting female and male gametophyte development (Caselli et al., 2019). This previous study proved that the silencing of one or more of these genes via RNA interference increases the number of unfertilized ovules, thereby decreasing the number of seeds per silique. Additionally, the *Vigun07g094201* sequence is highly similar to that of *Arabidopsis* *REM39* (AT3G18990). In the current study, *Vigun07g074900*, which encodes the transcription factor TCP21, was selected as a candidate gene for *qSdnppd7.2+*. In *Arabidopsis*, several TCP transcription factors, including TCP21, are involved in ovule development via their interactions with SPOROCTELESS (Chen et al., 2014; Wei et al., 2015), which is required for sporogenesis in both male and female organs (Yang et al., 1999). An earlier study involving cowpea identified a gene encoding TCP5 as a candidate gene for the number of seeds per pod (Lo et al., 2018).

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

PS and LW conceptualized the study and secured research funding. PS designed and supervised the study and revised the manuscript. KA and KL developed population and conducted data collection and analysis. PS and KA wrote the manuscript. All authors read and approved the accepted manuscript.

FUNDING

This research was funded by the Thailand Research Fund (Grant No. RSA5880051), China Agriculture Research System (CARS-08), and the Agricultural Science and Technology Innovation Program (ASTIP) of Chinese Academy of Agricultural Sciences.

REFERENCES

- Airolidi, C. A., Rovere, F. D., Falasca, G., Marino, G., Kooiker, M., Altamura, M. M., et al. (2010). The *Arabidopsis* BET bromodomain factor GTE4 is involved in maintenance of the mitotic cell cycle during plant development. *Plant Physiol.* 152, 1320–1334. doi: 10.1104/pp.109.150631
- Akiyama, K., Kurotani, A., Iida, K., Kuromori, T., Shinozaki, K., and Sakurai, T. (2014). RARGE II: an integrated phenotype database of *Arabidopsis* mutant traits using a controlled vocabulary. *Plant Cell Physiol.* 55:e4. doi: 10.1093/pcp/pct165
- Allard, R. W. (1999). *Principles of Plant Breeding, 2nd Edition*. New York, NY: John Wiley & Sons.
- Amkul, K., Wang, L., Somta, P., Wang, S., and Cheng, X. (2019). Construction of a high density linkage map and genome dissection of bruchid resistance in zombi pea (*Vigna vexillata* (L.) A. Rich). *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-48239-5
- Aspeborg, H., Schrader, J., Coutinho, P. M., Stam, M., Kallas, A., Djerbi, S., et al. (2005). Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen. *Plant Physiol.* 137, 983–997. doi: 10.1104/pp.104.055087
- Bakshi, M., and Oelmüller, R. (2014). WRKY transcription factors: jack of many trades in plants. *Plant Signal Behav.* 9:e27700. doi: 10.4161/psb.27700
- Barber, C., Rösti, J., Rawat, A., Findlay, K., Roberts, K., and Seifert, G. J. (2006). Distinct properties of the five UDP-D-glucose/UDP-D-galactose 4-epimerase isoforms of *Arabidopsis thaliana*. *J. Biol. Chem.* 281, 17276–17285. doi: 10.1074/jbc.M512727200
- Cao, J., Li, G., Qu, D., Li, X., and Wang, Y. (2020). Into the seed: auxin controls seed development and grain yield. *Int. J. Mol. Sci.* 21:1662. doi: 10.3390/ijms21051662
- Caselli, F., Beretta, V. M., Mantegazza, O., Petrella, R., Leo, G., Guazzotti, A., et al. (2019). REM34 and REM35 control female and male gametophyte development in *Arabidopsis thaliana*. *Front. Plant Sci.* 10:1351. doi: 10.3389/fpls.2019.01351
- Cazzonelli, C. I., Vanstraelen, M., Simon, S., Yin, K., Carron-Arthur, A., Nisar, N., et al. (2013). Role of the *Arabidopsis* PIN6 auxin transporter in auxin homeostasis and auxin-mediated development. *PLoS One* 8:e70069. doi: 10.1371/journal.pone.0070069
- Chaitieng, B., Kaga, A., Tomooka, N., Isemura, T., Huroda, Y., and Vaughan, D. A. (2006). Development of black gram [*Vigna mungo* (L.) Hepper] linkage map and its comparison with an azuki bean [*Vigna angularis* (Wild.) Ohwi and

ACKNOWLEDGMENTS

We are thankful to the Joint Legume Research Center between Chinese Academy of Agriculture Sciences and Kasetsart University for molecular lab facilities. We thank Liwen Bianji, Edanz Editing China (www.liwenbianji.cn/ac) for editing the English text of a draft of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00803/full#supplementary-material>

TABLE S1 | Number of markers, marker length, and distance between markers in each linkage group of the zombi pea genetic map used for detecting the QTLs of domestication-related traits in the F₂ population derived from a cross between TVNu 240 and TVNu 1623. The markers are SNPs generated by specific locus amplified fragment sequencing. The map was constructed by Amkul et al. (2019).

TABLE S2 | Correlations among domestication-related traits in the zombi pea F₂ population derived from a cross between TVNu 240 and TVNu 1623.

TABLE S3 | Candidate genes for each QTL.

- Ohashi] linkage map. *Theor Appl Genet.* 13, 1261–1269. doi: 10.1007/s00122-006-0380-5
- Chao, Y., Zhang, T., Yang, Q., Kang, J., Sun, Y., Gruber, M. Y., et al. (2014). Expression of the alfalfa CCH-type zinc finger protein gene *MsZFN* delays flowering time in transgenic *Arabidopsis thaliana*. *Plant Sci.* 215, 92–99. doi: 10.1016/j.plantsci.2013.10.012
- Chen, G. H., Sun, J. Y., Liu, M., Liu, J., and Yang, W. C. (2014). SPOROCTELESS is a novel embryophyte-specific transcription repressor that interacts with TPL and TCP proteins in *Arabidopsis*. *J. Genet. Genomics* 41, 617–625. doi: 10.1016/j.jgg.2014.08.009
- Chen, Y., Fan, X., Song, W., Zhang, Y., and Xu, G. (2012). Over-expression of *OsPIN2* leads to increased tiller numbers, angle and shorter plant height through suppression of *OsLAZY1*. *Plant Biotechnol. J.* 10, 139–149. doi: 10.1111/j.1467-7652.2011.00637.x
- Cheng, Z. J., Zhao, X. Y., Shao, X. X., Wang, F., Zhou, C., Liu, Y. G., et al. (2014). Absciscic acid regulates early seed development in *Arabidopsis* by ABI5-mediated transcription of *SHORT HYPOCOTYL UNDER BLUE1*. *Plant Cell* 26, 1053–1068. doi: 10.1105/tpc.113.121566
- Chua, Y. L., Channelière, S., Mott, E., and Gray, J. C. (2005). The bromodomain protein GTE6 controls leaf development in *Arabidopsis* by histone acetylation at ASYMMETRIC LEAVES1. *Genes Dev.* 19, 2245–2254. doi: 10.1101/gad.352005
- Cui, X., Guo, Z., Song, L., Wang, Y., and Cheng, Y. (2016). NCP1/AtMOB1A plays key roles in auxin-mediated *Arabidopsis* development. *PLoS Genet.* 12:e1005923. doi: 10.1371/journal.pgen.1005923
- Dachapak, S., Somta, P., Naito, K., Tomooka, N., Kaga, A., and Srinives, P. (2019). Detection of quantitative trait loci for salt tolerance in zombi pea [*Vigna vexillata* (L.) A. Rich]. *Euphytica* 215:208. doi: 10.1007/s10681-019-2530-2
- Dachapak, S., Somta, P., Poonchaivilaisak, S., Yimram, T., and Srinives, P. (2017). Genetic diversity and structure of the zombi pea (*Vigna vexillata* (L.) A. Rich) gene pool based on SSR marker analysis. *Genetica* 145, 189–200. doi: 10.1007/s10709-017-9957-y
- Dachapak, S., Tomooka, N., Somta, P., Naito, K., Kaga, A., and Srinives, P. (2018). QTL analysis of domestication syndrome in zombi pea (*Vigna vexillata*), an underutilized legume crop. *PLoS One* 13:e0200116. doi: 10.1371/journal.pone.0200116
- Delmas, F., Petit, J., Joubès, J., Séveno, M., Paccalet, T., Hernould, M., et al. (2003). The gene expression and enzyme activity of plant 3-deoxy-D-manno-2-octulosonic acid-8-phosphate synthase are preferentially associated with cell

- division in a cell cycle-dependent manner. *Plant Physiol.* 133, 348–360. doi: 10.1104/pp.103.026872
- Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321. doi: 10.1016/j.cell.2006.12.006
- Ferguson, H. (1954). The food crops of the SUDAN and Their relation to environment. *Paper Presented at the 1953 Conference of the Philosophical Society of the Sudan on "Food and Society in the Sudan", Philosophical Society of the Sudan Khartoum*. London: McCorquodale and Co. (Sudan), Ltd.
- Glover, J., Grelon, M., Craig, S., Chaudhury, A., and Dennis, E. (1998). Cloning and characterization of *MS5* from *Arabidopsis*: a gene critical in male meiosis. *Plant J.* 15, 345–356. doi: 10.1046/j.1365-3113x.1998.00216.x
- Gu, Y., Li, W., Jiang, H., Wang, Y., Gao, H., Liu, M., et al. (2017). Differential expression of a *WRKY* gene between wild and cultivated soybeans correlates to seed size. *J. Exp. Bot.* 68, 2717–2729. doi: 10.1093/jxb/erx147
- Guo, Z., Yue, X., Cui, X., Song, L., and Cheng, Y. (2020). *AtMOB1* genes regulate jasmonate accumulation and plant development. *Plant Physiol.* 182, 1481–1493. doi: 10.1104/pp.19.01434
- Hammer, K. (1984). Das domestikations syndrom. *Genet Res Crop Evol.* 32, 11–34.
- Hansen, S. F., Harholt, J., Oikawa, A., and Scheller, H. V. (2012). Plant glycosyltransferases beyond CAZy: a perspective on DUF families. *Front. Plant Sci.* 3:59. doi: 10.3389/fpls.2012.00059
- Isemura, T., Kaga, A., Konishi, S., Ando, T., Tomooka, N., Han, O. K., et al. (2007). Genome dissection of traits related to domestication in azuki bean (*Vigna angularis*) and their comparison with other warm season legumes. *Ann. Bot.* 100, 1053–1071. doi: 10.1093/aob/mcm155
- Isemura, T., Kaga, A., Tabata, S., Somta, P., Srinives, P., Shimizu, T., et al. (2012). Construct of genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). *PLoS One* 7:e41304. doi: 10.1371/journal.pone.0041304
- Isemura, T., Kaga, A., Tomooka, N., Shimizu, T., and Vaughan, D. A. (2010). The genetics of domestication of rice bean, *Vigna umbellata*. *Ann. Bot.* 106, 927–944. doi: 10.1093/aob/mcq188
- Ishimaru, K., Hirotsu, N., Madoka, Y., Murakami, N., Hara, N., Onodera, H., et al. (2013). Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. *Nat. Genet.* 45, 707–711. doi: 10.1038/ng.2612
- Karuniawan, A., Iswandi, A., Kale, P. R., Heinzemann, J., and Grüneberg, W. J. (2006). *Vigna vexillata* (L.) A. Rich. cultivated as a root crop in Bali and Timor. *Genet Resour. Crop Evol.* 53, 213–217. doi: 10.1007/s10722-005-1654-5
- Kongjaimun, A., Kaga, A., Tomooka, N., Somta, P., Vaughan, D. A., and Srinives, P. (2012). The genetics of domestication of yardlong bean, *Vigna unguiculata* (L.) Walp. ssp. *unguiculata* cv.-gr. *sesquipedalis*. *Ann. Bot.* 109, 1185–2000. doi: 10.1093/aob/mcs048
- Li, H., Ye, G., and Wang, J. (2007). A modified algorithm for the improvement of composite interval mapping. *Genetics* 175, 361–374. doi: 10.1534/genetics.106.066811
- Li, M., Ren, L., Xu, B., Yang, X., Xia, Q., He, P., et al. (2016). Genome-wide identification, phylogeny, and expression analyses of the 14-3-3 family reveal their involvement in the development, ripening, and abiotic stress response in banana. *Front. Plant Sci.* 7:1442. doi: 10.3389/fpls.2016.01442
- Li, M. Y., Xu, B. Y., Liu, J. H., Yang, X. L., Zhang, J. B., Jia, C. H., et al. (2012). Identification and expression analysis of four 14-3-3 genes during fruit ripening in banana (*Musa acuminata* L. AAA group, cv. Brazilian). *Plant Cell Rep.* 31, 369–378. doi: 10.1007/s00299-011-1172-1
- Li, W., Wang, H., and Yu, D. (2016). *Arabidopsis* *WRKY* transcription factors *WRKY12* and *WRKY13* oppositely regulate flowering under short-day conditions. *Mol. Plant* 9, 1492–1503. doi: 10.1016/j.molp.2016.08.003
- Liu, L., Tong, H., Xiao, Y., Che, R., Xu, F., Hu, B., et al. (2015). Activation of *Big Grain1* significantly improves grain size by regulating auxin transport in rice. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11102–11107. doi: 10.1073/pnas.1512748112
- Lo, S., Muñoz-Amatriain, M., Boukar, O., Herniter, I., Cisse, N., Guo, Y.-N., et al. (2018). Identification of genetic factors controlling domestication-related traits in cowpea (*Vigna unguiculata* L. Walp.). *Sci. Rep.* 19:6261. doi: 10.1038/s41598-018-24349-4
- Lonardi, S., Muñoz-Amatriain, M., Liang, Q., Shu, S., Wanamaker, S. I., Lo, S., et al. (2019). The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant J.* 98, 767–782. doi: 10.1111/tpj.14349
- Luo, M., Dennis, E. S., Berger, F., Peacock, W. J., and Chaudhury, A. (2005). *MINISEED3 (MINI3)*, a *WRKY* family gene, and *HAIKU2 (IKU2)*, a leucine-rich repeat (*LRR*) *KINASE* gene, are regulators of seed size in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 17531–17536.
- Maxted, N., Mabuzza-Diamini, P., Moss, H., Padulosi, S., Jarvis, A., and Guarino, L. (2004). *An Ecogeographic Study: African Vigna, Systematic and Ecogeographic Studies of Crop Genepool*. Rome: International Plant Genetic Resources Institutes.
- Nakashima, K., and Yamaguchi-Shinozaki, K. (2013). ABA signaling in stress-response and seed development. *Plant Cell Rep.* 32, 959–970. doi: 10.1007/s00299-013-1418-1
- Nisar, N., Cuttriss, A. J., Pogson, B. J., and Cazzonelli, C. I. (2014). The promoter of the *Arabidopsis* *PIN6* auxin transporter enabled strong expression in the vasculature of roots, leaves, floral stems and reproductive organs. *Plant Signal Behav.* 9:e27898. doi: 10.4161/psb.27898
- Palusa, S. G., Golovkin, M., Shin, S. B., Richardson, D. N., and Reddy, A. S. (2007). Organ-specific, developmental, hormonal and stress regulation of expression of putative pectate lyase genes in *Arabidopsis*. *New Phytol.* 174, 537–550. doi: 10.1111/j.1469-8137.2007.02033.x
- Peaucelle, A., Braybrook, S. A., Le, G. L., Bron, E., Kuhlemeier, C., and Höfte, H. (2011). Pectin-induced changes in cell wall mechanics underlie organ initiation in *Arabidopsis*. *Curr. Biol.* 21, 1720–1726. doi: 10.1016/j.cub.2011.08.057
- Rampey, R. A., LeClere, S., Kowalczyk, M., Ljung, K., Sandberg, G., and Bartel, B. (2004). A family of auxin-conjugate hydrolases that contributes to free indole-3-acetic acid levels during *Arabidopsis* germination. *Plant Physiol.* 135, 978–988. doi: 10.1104/pp.104.039677
- Rösti, J., Barton, C. J., Albrecht, S., Dupree, P., Pauly, M., Findlay, K., et al. (2007). UDP-glucose 4-epimerase isoforms UGE2 and UGE4 cooperate in providing UDP-galactose for cell wall biosynthesis and growth of *Arabidopsis thaliana*. *Plant Cell* 19, 1565–1579. doi: 10.1105/tpc.106.049619
- Schomburg, F. M., Bizzell, C. M., Lee, D. L., Zeevaert, J. A. D., and Amasino, R. M. (2003). Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. *Plant Cell* 15, 151–163. doi: 10.1105/tpc.005975
- Schruff, M. C., Spielman, M., Tiwari, S., Adams, S., Fenby, N., and Scott, R. J. (2006). The *AUXIN RESPONSE FACTOR 2* gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. *Development* 133, 251–261. doi: 10.1242/dev.02194
- Seok, H.-Y., Woo, D.-H., Park, H.-Y., Lee, S.-Y., Tran, H. T., Lee, E.-H., et al. (2016). *AtC3H17*, a non-tandem CCH zinc finger protein, functions as a nuclear transcriptional activator and has pleiotropic effects on vegetative development, flowering and seed development in *Arabidopsis*. *Plant Cell Physiol.* 57, 603–615. doi: 10.1093/pcp/pcw013
- Shi, H., and Zhang, Y. (2014). Pear *14-3-3a* gene (*Pp14-3-3a*) is regulated during fruit ripening and senescence, and involved in response to salicylic acid and ethylene signaling. *J. Genet.* 93, 747–753. doi: 10.1007/s12041-014-0447-z
- Shu, K., Liu, X. D., Xie, Q., and He, Z. H. (2016). Two faces of one seed: hormonal regulation of dormancy and germination. *Mol. Plant* 9, 34–45. doi: 10.1016/j.molp.2015.08.010
- Simon, S., Sküpa, P., Viaene, T., Zwiewka, M., Tejos, R., Klíma, P., et al. (2016). *PIN6* auxin transporter at endoplasmic reticulum and plasma membrane mediates auxin homeostasis and organogenesis in *Arabidopsis*. *New Phytol.* 211, 65–74. doi: 10.1111/nph.14019
- Skinner, D. J., and Gasser, C. S. (2009). Expression-based discovery of candidate ovule development regulators through transcriptional profiling of ovule mutants. *BMC Plant Biol.* 9:29. doi: 10.1186/1471-2229-9-29
- Stolle-Smits, T., Beekhuizen, J. G., Kok, M. T. C., Pijnenburg, M., Recourt, K., Derksen, J., et al. (1999). Changes in cell wall polysaccharides of green bean pods during development. *Plant Physiol.* 121, 363–372. doi: 10.1104/pp.121.2.363
- 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.* 36:80.

- Sun, H., Hao, P., Gu, L., Cheng, S., Wang, H., Wu, A., et al. (2020). Pectate lyase-like gene *GhPEL76* regulates organ elongation in *Arabidopsis* and fiber elongation in cotton. *Plant Sci.* 293:110395. doi: 10.1016/j.plantsci.2019.110395
- Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., et al. (2013). SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS One* 8:e58700. doi: 10.1371/journal.pone.0058700
- Sun, Y., Wang, C., Wang, N., Jiang, X., Mao, H., Zhu, C., et al. (2017). Manipulation of auxin response factor 19 affects seed size in the woody perennial *Jatropha curcas*. *Sci. Rep.* 7:40844.
- Takahashi, Y., Kongjaimun, A., Muto, C., Kobayashi, Y., Kumagai, M., Sakai, H., et al. (2019). Genetic factor for twisting legume pods identified by fine-mapping of shattering-related traits in azuki bean and yard-long bean. *bioRxiv*. [Preprint]. doi: 10.1101/774844
- Vaughan, D. A., Balázs, E., and Heslop-Harrison, J. S. (2007). From crop domestication to super-domestication. *Ann. Bot.* 100, 893–901. doi: 10.1093/aob/mcm224
- Wang, H., Guo, Y., Lv, F., Zhu, H., Wu, S., Jiang, Y., et al. (2009). The essential role of *GhPEL* gene, encoding a pectate lyase, in cell wall loosening by depolymerization of the de-esterified pectin during fiber elongation in cotton. *Plant Mol. Biol.* 72, 397–406. doi: 10.1007/s11103-009-9578-7
- Wang, W., Zheng, H., Wang, Y., Han, G., and Sui, N. (2018). Overexpression of *CCCH* zinc finger protein gene delays flowering time and enhances salt tolerance in *Arabidopsis* by increasing fatty acid unsaturation. *Acta Physiol. Plant* 40:196. doi: 10.1007/s11738-018-2775-8
- Wang, Y., and Jiao, Y. (2018). Auxin and above-ground meristems. *J. Exp. Bot.* 69, 147–154. doi: 10.1093/jxb/erx299
- Watcharatpong, P., Kaga, A., Chen, X., and Somta, P. (2020). Narrowing down a major qtl region conferring pod fiber contents in yardlong bean (*Vigna unguiculata*), a vegetable cowpea. *Genes* 11:363. doi: 10.3390/genes11040363
- Wei, B., Zhang, J., Pang, C., Yu, H., Guo, D., Jiang, H., et al. (2015). The molecular mechanism of sporocytelless/nozzle in controlling *Arabidopsis* ovule development. *Cell Res.* 25, 121–134. doi: 10.1038/cr.2014.145
- Wolf, S., Mouille, G., and Pelloux, J. (2009). Homogalacturonan methylesterification and plant development. *Mol. Plant.* 2, 851–860. doi: 10.1093/mp/ssp066
- Xiang, J., Tang, S., Zhi, H., Jia, G., Wang, H., and Diao, X. (2017). Loose *Panicle1* encoding a novel WRKY transcription factor, regulates panicle development, stem elongation, and seed size in foxtail millet [*Setaria italica* (L.) P. Beauv.]. *PLoS One* 12:e0178730. doi: 10.1371/journal.pone.0178730
- Xiong, J., Cui, X., Yuan, X., Yu, X., Sun, J., and Gong, Q. (2016). The Hippo/STE20 homolog SIK1 interacts with MOB1 to regulate cell proliferation and cell expansion in *Arabidopsis*. *J. Exp. Bot.* 67, 1461–1475. doi: 10.1093/jxb/erv538
- Xu, M., Zhu, L., Shou, H., and Wu, P. (2005). A PIN1 family gene, *OsPIN1*, involved in auxin-dependent adventitious root emergence and tillering in rice. *Plant Cell Physiol.* 46, 1674–1681. doi: 10.1093/pcp/pci183
- Yamauchi, Y., Takeda-Kamiya, N., Hanada, A., Ogawa, M., Kuwahara, A., Seo, M., et al. (2007). Contribution of gibberellin deactivation by AtGA2ox2 to the suppression of germination of dark-imbibed *Arabidopsis thaliana* seeds. *Plant Cell Physiol.* 48, 555–561. doi: 10.1093/pcp/pcm023
- Yan, Z., Jia, J., Yan, X., Shi, H., and Han, Y. (2017). *Arabidopsis* KHZ1 and KHZ2, two novel non-tandem CCCH zinc-finger and K-homolog domain proteins, have redundant roles in the regulation of flowering and senescence. *Plant Mol. Biol.* 95, 549–565. doi: 10.1007/s11103-017-0667-8
- Yang, L., Liu, H., Zhao, J., Pan, Y., Cheng, S., Lietzow, C. D., et al. (2018). *LITTLELEAF* (*LL*) encodes a WD40 repeat domain-containing protein associated with organ size variation in cucumber. *Plant J.* 95, 834–847. doi: 10.1111/tpj.13991
- Yang, W. C., Ye, D., Xu, J., and Sundaresan, V. (1999). The *SPOROCTELESS* gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes Dev.* 13, 2108–2117. doi: 10.1101/gad.13.16.2108
- Yu, Y., Liu, Z., Wang, L., Kim, S. G., Seo, P. J., et al. (2016). WRKY71 accelerates flowering via the direct activation of FLOWERING LOCUS T and LEAFY in *Arabidopsis thaliana*. *Plant J.* 85, 96–106. doi: 10.1111/tpj.13092
- Yundaeng, C., Somta, P., Amkul, K., Kongjaimun, A., Kaga, A., and Tomooka, N. (2019). Construction of genetic linkage map and genome dissection of domestication-related traits of moth bean (*Vigna aconitifolia*), a legume crop of arid areas. *Mol. Genet. Genomics* 294, 621–635. doi: 10.1007/s00438-019-01536-0
- Zhang, L., Chen, L., and Yu, D. (2018). Transcription factor WRKY75 interacts with DELLA proteins to affect flowering. *Plant Physiol.* 176, 790–803. doi: 10.1104/pp.17.00657
- Zhou, Y. P., Wu, J. H., Xiao, W. H., Chen, W., Chen, Q. H., Fan, T., et al. (2018). *Arabidopsis* IQM4, a novel calmodulin-binding protein, is involved with seed dormancy and germination in *Arabidopsis*. *Front. Plant Sci.* 9:721. doi: 10.3389/fpls.2018.00721

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

Copyright © 2020 Amkul, Somta, Laosatit and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Impacts of Domestication and Agricultural Practices on Legume Nutrient Acquisition Through Symbiosis With Rhizobia and Arbuscular Mycorrhizal Fungi

Ailin Liu^{††}, Yee-Shan Ku[†], Carolina A. Contador[†] and Hon-Ming Lam^{*}

Centre for Soybean Research of the State Key Laboratory of Agrobiotechnology and School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China

OPEN ACCESS

Edited by:

Gaofeng Zhou,
Department of Primary Industries
and Regional Development
of Western Australia, Australia

Reviewed by:

Camilla Fagorzi,
University of Florence, Italy
Alexandra Stoll,
Universidad Católica del Norte, Chile

*Correspondence:

Ailin Liu
merestarry@gmail.com
Hon-Ming Lam
honming@cuhk.edu.hk

^{††} These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 16 July 2020

Accepted: 08 September 2020

Published: 30 September 2020

Citation:

Liu A, Ku Y-S, Contador CA and
Lam H-M (2020) The Impacts
of Domestication and Agricultural
Practices on Legume Nutrient
Acquisition Through Symbiosis With
Rhizobia and Arbuscular Mycorrhizal
Fungi. *Front. Genet.* 11:583954.
doi: 10.3389/fgene.2020.583954

Legumes are unique among plants as they can obtain nitrogen through symbiosis with nitrogen-fixing rhizobia that form root nodules in the host plants. Therefore they are valuable crops for sustainable agriculture. Increasing nitrogen fixation efficiency is not only important for achieving better plant growth and yield, but it is also crucial for reducing the use of nitrogen fertilizer. Arbuscular mycorrhizal fungi (AMF) are another group of important beneficial microorganisms that form symbiotic relationships with legumes. AMF can promote host plant growth by providing mineral nutrients and improving the soil ecosystem. The trilateral legume-rhizobia-AMF symbiotic relationships also enhance plant development and tolerance against biotic and abiotic stresses. It is known that domestication and agricultural activities have led to the reduced genetic diversity of cultivated germplasms and higher sensitivity to nutrient deficiencies in crop plants, but how domestication has impacted the capability of legumes to establish beneficial associations with rhizospheric microbes (including rhizobia and fungi) is not well-studied. In this review, we will discuss the impacts of domestication and agricultural practices on the interactions between legumes and soil microbes, focusing on the effects on AMF and rhizobial symbioses and hence nutrient acquisition by host legumes. In addition, we will summarize the genes involved in legume-microbe interactions and studies that have contributed to a better understanding of legume symbiotic associations using metabolic modeling.

Keywords: legume–microbe interaction, arbuscular mycorrhizal fungi, symbiotic nitrogen fixation, rhizobia, domestication, metabolic modeling, metabolic profiling

INTRODUCTION

Major macronutrients such as nitrogen (N), phosphorus (P), and potassium (K) play important roles in crop development. Lack of soil fertility is a vital constraint on crop production. To avoid intensifying the use of chemical fertilizers which may cause irreversible environmental damages including soil salinization and eutrophication of local lakes and rivers from fertilizer runoffs, plant growth-promoting rhizobacteria (PGPR) are regarded as efficient biofertilizers (Singh et al., 2016).

The application of biofertilizers to the soil increases the biodiversity in the soil, and improves soil fertility through N_2 fixation, as well as P and K mobilization in the form of organic acids (Egamberdiyeva and Höflich, 2004; Sulieman and Tran, 2014; Itelima et al., 2018).

Legumes are important crops in agriculture, not only because they are a protein-rich food source, but also because of their contribution to soil fertilization through symbiotic nitrogen fixation (SNF) as a result of their association with nitrogen-fixing PGPR, i.e., rhizobia (Zahran, 1999). Each year, legumes can provide more than 70 million tonnes of N to soil (Brockwell et al., 1995). The symbiosis between legume and rhizobium is responsible for a substantial part of global N flux in which atmospheric N_2 is fixed to form ammonia, nitrate, and organic nitrogen. Rhizobial species (including the genera of *Mesorhizobium*, *Bradyrhizobium*, *Azorhizobium*, *Allorhizobium*, and *Sinorhizobium*) infect legume roots and induce the formation of root nodules (Stacey, 2007; Gopalakrishnan et al., 2015). The infection is initiated by the flavonoids released by legume hosts that induce the expressions of nodulation (*nod*) genes in rhizobia, which in turn trigger root cell divisions by producing lipo-chitooligosaccharide (LCO) signals (Dakora, 1995, 2003). Nodules can generally be classified into determinate and indeterminate ones. Determinate nodules are round shaped, with a well-defined homogeneous central fixation zone containing infected rhizobia-filled cells surrounded by uninfected cells (Schultze and Kondorosi, 1998). Indeterminate nodules have a gradient of developmental stages from the nodule tip to the root as a result of a persistent meristem that generates new cells continuously (Crespi and Gálvez, 2000), which is absent in determinate nodules. Well-studied model species with indeterminate nodules include *Medicago truncatula* (barrelclover), *Medicago sativa* (alfalfa), and *Pisum sativum* (pea), while *Glycine max* (soybean), *Vicia faba* (fava bean), and *Lotus japonicus* (birdsfoot trefoil) are typical model species for the study of determinate nodules (Gage, 2004).

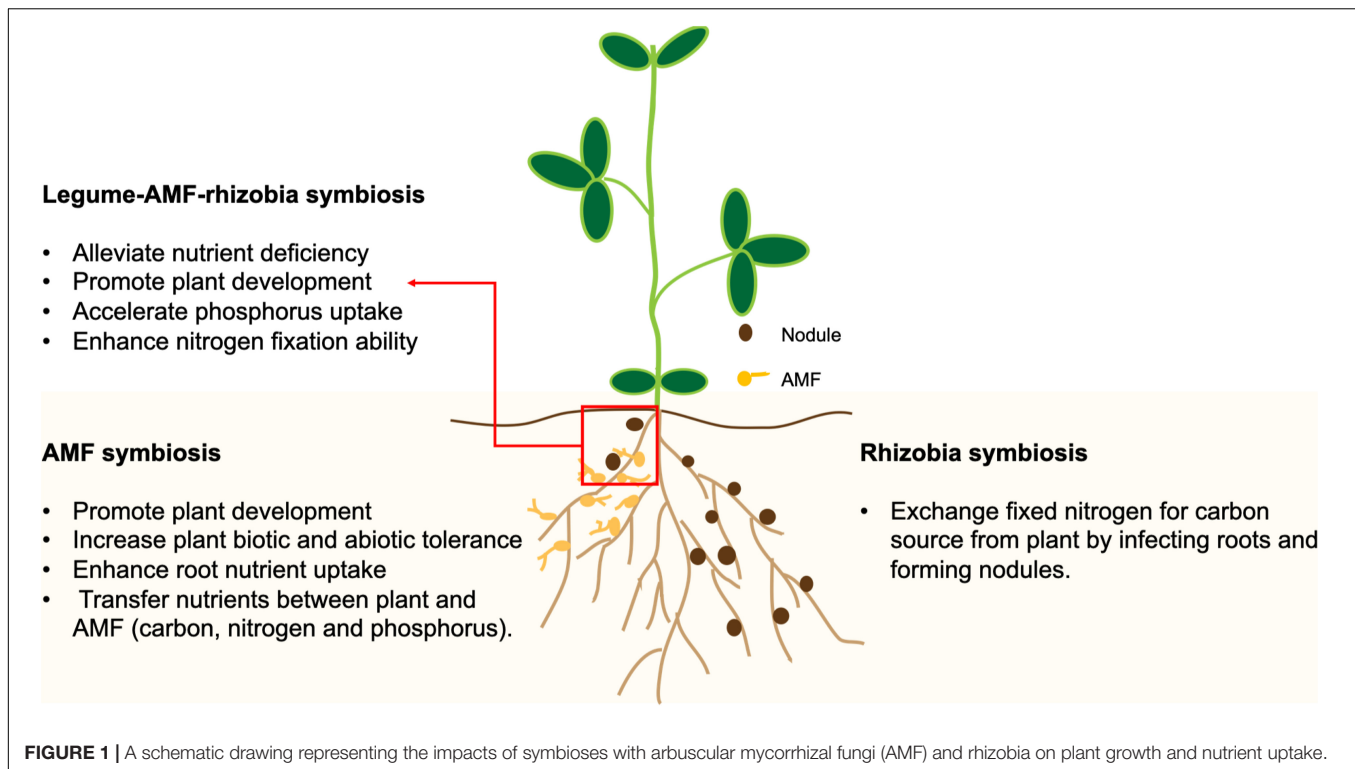
Besides rhizobia, legumes form symbiosis with soil fungi such as arbuscular mycorrhizal fungi (AMF) and *Trichoderma* spp. (TR). The symbioses with AMF and TR are considered to be beneficial for plant growth (Brundrett, 1991; Woo et al., 2014). AMF colonize plant roots and radiate their hyphae into the surrounding soil, complementing the host's root functions (Jansa et al., 2011). The enhanced root performance due to AMF has been shown in *G. max* (Wang et al., 2011), *V. faba* (Jia et al., 2004), *M. sativa* (Jansa et al., 2011), and *Phaseolus vulgaris* (common bean) (Tajini et al., 2012). Moreover, nodules normally have a high demand for inorganic phosphate (Pi) as nitrogenase functions in the bacteroid are highly ATP-consuming (Jakobsen, 1985; Liu et al., 2018). In such cases, AMF have the capability to accelerate the Pi uptake. The amount of Pi delivered to the plant varies with specific AMF (Ianson and Linderman, 1993; Valdenegro et al., 2001). On the other hand, TR-based biofertilizers have also been reported to enhance N, P, and K uptakes (Amaresan et al., 2020). TR communicates with the plant root by chemical signals such as auxins and small peptides. They could also colonize plant roots by penetrating the outer layers of the root tissue. One of the well-known beneficial

effects of TR to plants is the solubilization of Pi by acidification, chelation or redox activities to improve the Pi availability to the plant. The benefits of AMF and TR to crop growth have been comprehensively reviewed (Szczalba et al., 2019).

Domestication is the conversion of wild plant species to cultivated ones through human selection and breeding of desirable characteristics over many generations, but it often results in the loss of genetic diversity (Gepts, 2010). It is well-documented that the domestication of legumes has emphasized the selection of favorable aboveground traits including larger seed size, palatability, reduced seed dormancy and heritability of other desirable agronomic traits (Abbo et al., 2014; Pérez-Jaramillo et al., 2016). Such a process of artificial selection has reduced the self-sustaining capability and increased sensitivities to diseases, abiotic stresses and nutrient depletion in cultivated legumes. Furthermore, artificial selection has also influenced underground traits including root architecture and root exudate composition, often unintentionally (Pérez-Jaramillo et al., 2016). This may affect the establishment of rhizospheric microbial communities in the soil where the legumes are grown. In this review, we will discuss how domestication and agricultural practices have influenced the interactions between legumes and rhizospheric microbes (rhizobia and fungi), with a focus on the potential impacts of such altered interactions on macronutrient acquisition by the host plant. We will also summarize the identified genes that regulate plant-rhizobia interactions in both plants and rhizobia, and the new study approaches using metabolic modeling and metabolic profiling, with the aim of contributing to a better understanding of legume-microbe symbiotic associations.

DOMESTICATION AND AGRICULTURAL PRACTICES INFLUENCE THE DIVERSITY OF RHIZOBIA AND SOIL FUNGI

Legume-rhizobium mutualism contributes to the vast majority of non-anthropogenically fixed N in terrestrial ecosystems (Cleveland et al., 1999), and it is especially susceptible to anthropogenic environmental changes (Kiers et al., 2002). Through mutualistic symbiosis, legumes receive ammonia fixed from atmospheric N_2 by rhizobia in the root nodule, in exchange for providing rhizobia with a carbon source obtained through photosynthesis (Figure 1). The efficiency and intensity of SNF is dependent on both rhizobia and plant hosts under a specific set of conditions (Lau et al., 2012). A study comparing the SNF efficiencies between wild (local) varieties and cultivars in major legume crops showed that the wild populations of alfalfa, pea and feugreek (*Trigonella foenumgraecum* L.) had better performance than their cultivated counterparts (Provorov and Tikhonovich, 2003). On the other hand, a study using wild and cultivated soybeans inoculated with the slow-growing *Bradyrhizobium japonicum* strain USDA110 or the fast-growing *Sinorhizobium fredii* strain CCBAU45436, indicated the SNF efficiency was improved by the domestication process (Muñoz et al., 2016). Moreover, the activity and diversity of rhizobia can be affected by human activities including the use of N fertilizer, cropping system and soil management system (Ferreira et al., 2000;



Bizarro et al., 2011). The utilization of high levels of N fertilizer across soil ecosystems may result in the breakdown of nitrogen-fixing symbiosis between legumes and rhizobia (Foley et al., 2005). It was suggested that when the resources exchanged in mutualistic symbiosis are abundant in the soil and the host is less reliant on the microsymbiont, the plant-microbe interaction could shift from mutualism to parasitism (Bronstein, 2001). The genetic diversity of *rhizobium* in nodules of *P. vulgaris* was decreased partially due to the N fertilization (Caballero-Mellado and Martinez-Romero, 1999). A study showed that *Trifolium* species inoculated with long-term N fertilizer-treated rhizobial strains had less biomass and chlorophyll contents than plants inoculated with non-fertilizer treated rhizobia, implying an elevated N supply resulted in the evolution of less-mutualistic rhizobia that provide fewer growth benefits to plant hosts in the *Trifolium*-rhizobium symbiosis (Weese et al., 2015).

Arbuscular mycorrhizal fungi belong to the subphylum *Glomeromycotina*, which are asexual obligate biotrophs that infect roots and establish a mutualistic relationship with the host plant to complete their life cycle. The mutualism is characterized by the transfer of nutrients including P, N, and carbon (C) between the fungi and the plants (Figure 1). Moreover, AMF play an important role in improving N₂ fixation by providing a favorable environment to facilitate the infection of plant roots by rhizobia (Mohammadi et al., 2012). It is believed that the responsiveness of crops to AMF is not under intentional selection by breeders during domestication (Bennett et al., 2013). However, agricultural activities such as soil disturbance, grazing, fertilizer application and monocropping may have reduced the AMF diversity and

their community structure in the soil (Jansa et al., 2006). Plants having sufficient P supplies are also less responsive to AMF for symbiosis (Jansa et al., 2006). Moreover, soil types can also influence the structure of soil fungal communities (Chang et al., 2020). By simulating the monocropping of cultivated and wild soybeans in the greenhouse, it was shown that monocropping resulted in a stable fungal community in the rhizospheres of both types of soybeans, but with different community structures (Tian et al., 2020). Therefore, modern agricultural practices may have resulted in the reduced AMF diversity and the inadvertent selection of crop genotypes with reduced responsiveness to AMF for symbiosis (Bennett et al., 2013).

WILD LEGUMES RECRUIT MORE DIVERSE RHIZOBIA AND SOIL FUNGI

It is documented that domesticated legumes tend to have fewer compatible symbionts than wild legumes (Mutch and Young, 2004). Rhizobia varies on their ability to nodulate various legume species. Some strains show high nodulation specificities and only form nodules with a limited number of legume species, while others can have a wide range of hosts (Pueppke and Broughton, 1999). *G. max* is one of the most important cultivated legumes. *G. soja* is the wild progenitor species of *G. max*, with high genetic diversity and better resistance to environmental stresses (Muñoz et al., 2017). Both *G. soja* and *G. max* can be nodulated with rhizobia, including *Sinorhizobium fredii*, *Bradyrhizobium japonicum* and

Bradyrhizobium elkanii (Wang et al., 2009). A study on bacterial isolates from *G. soja* nodules collected from different ecoregions in China showed that the biodiversity of rhizobia is associated with the geographical distribution of the particular ecotype of *G. soja* (Wu et al., 2011). Moreover, by comparing the rhizospheric bacteria between wild and cultivated soybeans in different soil types, wild soybean has the ability to recruit a higher abundance of *Bradyrhizobium* strains than cultivated soybean (Chang et al., 2019). A study of the impacts of domestication on the chickpea-*Mesorhizobium* symbiosis using 10 accessions of chickpea showed that the wild chickpea progenitor (*Cicer reticulatum*) could associate with more diverse *Mesorhizobium* populations than the cultivated chickpea (*C. arietinum*) (Kim et al., 2014).

By comparing the rhizospheric fungal communities of wild soybeans, ZYQ95 and 01-289, to those of cultivated soybeans, Williams 82 and Zhonghuang, the wild and cultivated germplasms were found to recruit different fungi in the rhizosphere (Chang et al., 2020). The wild soybeans recruited soil fungi with diverse potential functions while the cultivated soybeans mainly recruited those soil fungi which can enhance nutrient uptake by the plant (Chang et al., 2020). In another study conducted in the greenhouse, it was found that the genus *Paraglomus* was more enriched in the rhizosphere of wild soybean ZYQ95 while *Funnelformis* and *Rhizophagus* were more enriched in the rhizosphere of cultivated soybean Williams 82 (Zhang et al., 2019). In the field, no AMF could be found in the rhizosphere of Williams 82 while only *Paraglomus* was found in the rhizosphere of ZYQ95 under drought condition (Zhang et al., 2019). The reduced diversity of symbiotic rhizobia and fungi in modern legume cultivars compared to the wild relatives is likely a result of domestication. Agricultural practices such as monocropping might have played a role. Moreover, domesticated crops are usually grown in a confined cultivation area. As a result, the selection force on legume-associated soil microbe diversity in complex natural habitats might be absent.

AMF IMPROVE THE EFFICIENCY OF RHIZOBIUM-MEDIATED NITROGEN ASSIMILATION IN BOTH WILD AND DOMESTICATED LEGUMES

In a survey of 27 crops including legumes such as chickpea, soybean, grass pea, lentil, white clover and common bean, the domesticated crops and their wild relatives were inoculated with *Rhizophagus irregularis* (Blaszk, Wubet, Renker & Buscot) C. Walker & A. Schüssler strain EEZ 58 (Gi), which is a common strain of AMF in wild and agricultural lands. The inoculation led to increases in leaf Pi concentrations across all domesticated crops and wild relatives tested (Martín-Robles et al., 2018). In the same study, for the 14 non-leguminous crops such as barley, corn and tomato, when Pi availability in the soil was high, the growth benefits to the domesticated crops by the arbuscular mycorrhizae were reduced

but not for their wild relatives (Martín-Robles et al., 2018). In another study, wild soybean (*G. soja*) or cultivated soybean (*G. max*) was inoculated with both *Scutellospora heterogama* (*S. heterogama*) and their own rhizobial cells (Eom et al., 1994). It was found that *S. heterogama* stimulated the triple symbiosis among the wild soybean plant, *S. heterogama* and the rhizobial cells (Eom et al., 1994; **Figure 1**). Furthermore, it was demonstrated that unimproved soybean accessions showed a greater mycorrhizal dependency (MD) than improved accessions (Khalil et al., 1994). When inoculated with the vascular-arbuscular mycorrhizal fungi, *Gigaspora margarita* or *Glomus intraradices*, under low soil Pi condition, the positive growth effects due to the inoculation were greater in the unimproved accessions than in the improved ones (Khalil et al., 1994). A follow-up study showed that the unimproved soybean accession having a high MD value, *G. soja* Sieb. & Zucc. PI 468916, developed symbiosis with *G. margarita* more quickly than *G. max* (L.) Merr. cv. Mandarin (unimproved) and *G. max* cv. Swift (improved), which have intermediate and low MD values, respectively (Khalil et al., 1999). Among the three soybean accessions, *G. soja* Sieb. & Zucc. PI 468916 also had higher phosphatase activities and higher percentage increases in phosphatase activities when inoculated with *G. margarita* (Khalil et al., 1999). In contrast, in a study on cowpea using three cultivated accessions, Katumani 80, KenKunde 1 and Kunde Mboga, and one wild accession, SP 219, it was found that the ability to respond to mycorrhizal inoculation was retained in the cultivated accessions, which are the common modern cultivars grown by smallholding farmers in Kenya (Oruru et al., 2017). In these experiments, the cowpea accessions were grown in sterilized soil inoculated with a filtrate from non-sterilized soil and a commercial mycorrhizal inoculum comprising *Funnelformis mosseae*, *Glomus aggregatum*, *Glomus etunicatum*, and *Rhizophagus irregularis* (Oruru et al., 2017). AMF root colonization was observed in all the accessions in both non-sterilized and sterilized soil. In all the accessions, the inoculation of the mycorrhizal inoculum improved the nodule number, dry weights of nodules, roots and shoots, and both the nitrogen and Pi levels in the shoots of all the wild and cultivated accessions. However, the increase in biomass of nodules, roots and shoots were more significant in the cultivated accessions than in the wild counterpart (Oruru et al., 2017). Katumani 80 and Kunde Mboga also had higher percentages of mycorrhizal colonization compared to the wild accession while Katumani 80 and KenKunde 1 had more nodules than the wild SP 219 (Oruru et al., 2017).

Although several studies showed that wild legumes benefited more from the inoculation with AMF than cultivated counterparts (Eom et al., 1994; Khalil et al., 1994, 1999; Martín-Robles et al., 2018), the opposite effect of AMF was also reported (Oruru et al., 2017). It is important to note that different legumes and AMF species were tested in these studies. It has been pointed out that the responses to soil fungi could be genotype-specific, instead of wild versus cultivated accessions of the legume in question (Prashar and Vandenberg, 2017). As discussed above, the response to soil fungi is probably not intentionally selected

for during domestication. Therefore, it may not be appropriate to generalize the observations to all wild and domesticated legumes.

THE CAPABILITY TO RESPOND TO AMF AND *Trichoderma* spp. (TR) IS RETAINED IN MODERN LEGUME CULTIVARS

During plant breeding, varieties which are more resistant to diseases are usually selected for. It has been hypothesized that such selections may also have influenced the susceptibility of these varieties to beneficial microbes such as AMF. When different varieties of chickpea were grown in paddock soil containing naturally occurring AMF, there was a correlation between the level of resistance to *Phytophthora* root rot (PRR) and the degree of AMF colonization on the root (Plett et al., 2016), where the PRR-susceptible variety, Sonali, had a significantly higher root AMF colonization level than the PRR-resistant varieties, Yorker and PBA HatTrick, although such a correlation was absent when the varieties were grown in sterilized soil inoculated with *Funneliformis mosseae* (Plett et al., 2016). A possible explanation for such a phenomenon is that there is a certain degree of overlap between genes which confer resistance against beneficial AMF and those against pathogenic microbes (Plett et al., 2016). Although there are few studies on the response of wild legumes to TR, current findings also suggest that the capability to respond to TR is retained in domesticated legumes. In a study on 23 wild and cultivated lentil accessions inoculated with commercial TR inocula, RootShield® (RS) and RootShield® Plus (RSP), based on *Trichoderma harzianum* T22 and *Trichoderma virens* G41, respectively, the wild accession, *Lens tomentosus* PI 572390, had enhancements in most of the tested agronomic traits such as root and shoot biomass (Prashar and Vandenberg, 2017). Positive effects on the cultivated accessions by such inoculations were also observed (Prashar and Vandenberg, 2017). However, in some accessions, be they cultivated or wild, the inoculation had no or negative effects on the same agronomic traits (Prashar and Vandenberg, 2017). This showed that the TR responses were genotype-specific rather than distinguished along the line of wild versus domesticated accessions (Prashar and Vandenberg, 2017).

It was hypothesized that new high-yield crops may be less responsive to AMF compared to their ancestors (Lehmann et al., 2012). However, a meta-analysis of the data on 320 different crops, including those in the families *Poaceae*, *Fabaceae*, *Pedaliaceae*, *Asteraceae*, and *Cucurbitaceae*, from 39 published studies suggested that the new crops have not lost the ability to respond to AMF compared to their ancestors (Lehmann et al., 2012). Wild and domesticated legume accessions usually have distinct root system architectures. For example, in a study on the quantitative trait locus (QTL) of root system architecture in soybean, a recombinant inbred (RI) population was produced by crossing a *G. max* parent, V71-370, with a *G. soja* parent, PI407162 (Prince et al., 2015). The domesticated parent *G. max* V71-370 had a larger root system than the wild parent. The resulting RI population displayed a transgressive segregation of the root traits, including taproot length and total root length,

from the domesticated and wild parents (Prince et al., 2015). In another study on the QTL governing the root architecture of soybean, a cultivated parent, *G. max* Dunbar, and a wild parent, *G. soja* PI 326582A, were used to construct the RI population (Manavalan et al., 2015). The cultivated parent had longer tap root and more lateral roots than the wild parent (Manavalan et al., 2015). In common bean (*P. vulgaris*), it was found that the wild accessions, A1 and A2, have longer specific root lengths (root length: root dry weight) than the cultivated accessions, M1, M2, M3, M4, and M5, and a lower root density (root dry weight: root volume) than the cultivated M2 (Pérez-jaramillo et al., 2017). A link between domestication, specific root morphology and rhizobacterial community assembly was observed (Pérez-jaramillo et al., 2017). It is therefore convenient to suggest that the different root architecture of wild and domesticated legumes may result in different responses to soil fungi. However, based on the published data on root architecture and mycorrhizal growth responses, a meta-analysis suggested that there is no strong correlation between the mycorrhizal growth response and root diameter, root hair length or root hair density (Maherali, 2014). It appears that the differential responses to AMF colonization by different genotypes/accessions are influenced by factors other than the physical attributes of the root system. Similar to the improved efficiency of rhizobium-mediated nitrogen assimilation by AMF which happens in both wild and cultivated legumes, the capability to respond to AMF and TR is also retained in modern legume cultivars. As the symbiosis with AMF and TR are beneficial for plant growth (Brundrett, 1991; Woo et al., 2014), the capability of legumes to interact with these beneficial soil fungi should be have been selected by breeders during domestication.

FUNCTIONAL GENES INVOLVED IN LEGUME-RHIZOBIUM-AMF SYMBIOSIS

Symbiosis with rhizobia and AMF play a key role in plant nutrient acquisition. The mechanisms of these interactions and the genes driving these processes have been investigated for decades. In these endeavors, mutant libraries have been highly valuable resources for the research on symbiotic nitrogen fixation in legumes. They have been used for the identification and functional analyses of essential SNF-related genes. Over the years, legume mutants have been produced using physical/chemical mutagenesis, insertional mutagenesis and targeted genome editing tools (Le Signor et al., 2009; Urnov et al., 2010; Cui et al., 2013), including zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and CRISPR-Cas systems (Wang L. et al., 2017).

Genes and/or functions required for rhizobium and AMF symbiosis are summarized in **Table 1**. In *L. japonicus*, the biological functions of leghemoglobins were validated using RNAi and the CRISPR/Cas9 system (Ott et al., 2005; Wang et al., 2016, 2019). Leghemoglobins are plant hemoglobins required to maintain a low-oxygen environment inside root nodules for the efficient functioning of nitrogenases in bacteroids (Vázquez-Limón et al., 2012). The *Ljlb1*, *Ljlb2*, and *Ljlb3* mutants had

TABLE 1 | Functional genes involved in legume symbiosis.

Gene	Function	Organism	References
<i>LjLB1</i>	Leghemoglobin	<i>Lotus japonicus</i>	Ott et al., 2005; Wang et al., 2016, 2019
<i>LjLB2</i>	Leghemoglobin	<i>Lotus japonicus</i>	Ott et al., 2005; Wang et al., 2016, 2019
<i>LjLB3</i>	Leghemoglobin	<i>Lotus japonicus</i>	Ott et al., 2005; Wang et al., 2016, 2019
<i>nifV</i>	Homocitrate synthase	<i>Azorhizobium caulinodans</i> ORS571 and <i>Bradyrhizobium</i>	Nouwen et al., 2017
<i>Fen1</i>	Homocitrate synthase	<i>Lotus japonicus</i>	Imaizumi-Anraku et al., 1997
<i>Rj4</i>	Symbiotic partnership specificity	<i>Glycine max</i>	Hayashi et al., 2014; Tang et al., 2016
<i>MtNFS1</i>	Symbiotic partnership specificity	<i>Medicago truncatula</i> Jemalong A17	Wang Q. et al., 2017; Yang S. et al., 2017
<i>MtNFS2</i>	Symbiotic partnership specificity	<i>Medicago truncatula</i> Jemalong A17	Wang Q. et al., 2017; Yang S. et al., 2017
<i>SYMRK</i>	Infection thread initiation and nodule development by rhizobia	<i>Lotus japonicus</i>	Kistner et al., 2005
<i>CASTOR</i>	Infection thread initiation and nodule development by rhizobia	<i>Lotus japonicus</i>	Kistner et al., 2005
<i>POLLUX</i>	Infection thread initiation and nodule development by rhizobia	<i>Lotus japonicus</i>	Kistner et al., 2005
<i>SYM3</i>	Infection thread initiation and nodule development by rhizobia	<i>Lotus japonicus</i>	Kistner et al., 2005
<i>SYM6</i>	Infection thread initiation	<i>Lotus japonicus</i>	Kistner et al., 2005
<i>SYM15</i>	Infection thread initiation and nodule development by rhizobia	<i>Lotus japonicus</i>	Kistner et al., 2005
<i>SYM24</i>	Infection thread initiation and nodule development by rhizobia	<i>Lotus japonicus</i>	Kistner et al., 2005
	Leghemoglobin	<i>Medicago truncatula</i>	Liese et al., 2017
	Nicotianamine synthase-like protein	<i>Medicago truncatula</i>	Liese et al., 2017
	Nodule-specific cysteine-rich peptides	<i>Medicago truncatula</i>	Liese et al., 2017
<i>GmEXLB1</i>	Regulation of root development and responses to abiotic stress	<i>Glycine max</i>	Kong et al., 2019
<i>GmACP1</i>	Acid phosphatase	<i>Glycine max</i>	Zhang et al., 2014

elevated free-oxygen concentrations in infected zones and N₂ fixation was abolished (Ott et al., 2005; Wang et al., 2019). Transcript analyses of bacteroid genes from the nodules of these mutants revealed lower expression levels of *nif* and *fix* genes compared to those from the wild type (Ott et al., 2005, 2009). In particular, the NifH protein was not detected despite the presence of *nifH* transcripts in the mutant nodules (Ott et al., 2005).

Efficient nitrogenase activities also depend on other conditions besides the microaerobic environment (Kuzma et al., 1993). Homocitrate is a component of the iron-molybdenum cofactor of nitrogenases (Rubio and Ludden, 2008). For rhizobia lacking the homocitrate synthase-encoding gene (*nifV*), homocitrate supplied by the host plant is a key element of successful SNF, as demonstrated using an SNF-defective *L. japonicus* mutant, *fen1* (Imaizumi-Anraku et al., 1997). A homocitrate synthase gene, *FEN1*, in the host plant was identified as responsible for compensating for the lack of *nifV* in rhizobia (Hakoyama et al., 2009). Additionally, an external supply of homocitrate was shown to restore the nitrogen-fixing capacity in the ineffective nodules.

On the other hand, the editing of genes responsible for the recognition mechanism underlying legume-rhizobium specificity enables nodulation with unspecific strains. In soybean, the disruption of the dominant allele *Rj4* allowed nodulation with some strains of *Bradyrhizobium japonicum* and *Bradyrhizobium*

elkanii (Hayashi et al., 2014; Tang et al., 2016). *Rj4* encodes a thaumatin-like protein that usually restricts nodulation with these ineffective rhizobial strains. In *M. truncatula* Jemalong A17, the editing of two nodule cysteine-rich (NCR) peptide-encoding genes, *MtNFS1* and *MtNFS2*, allowed nodulation by *S. meliloti* strain Rm41 (Wang Q. et al., 2017; Yang S. et al., 2017). NCR peptides control and regulate the symbiotic partnership specificity in *M. truncatula*. *MtNFS1* and *MtNFS2* were identified using a recombinant inbred line (RIL) population of Jemalong A17 and DZA315. Parental genotypes were segregated according to their ability to form functional nodules with *S. meliloti* strain Rm41 or not (Wang Q. et al., 2017; Yang S. et al., 2017).

Seven *L. japonicus* genes (*SYMRK*, *CASTOR*, *POLLUX*, *SYM3*, *SYM6*, *SYM15*, and *SYM24*) are required for infection thread initiation for both fungal and bacterial symbiosis (Kistner et al., 2005). Mutant phenotypes showed that these genes are involved in the intracellular infection by fungal and bacterial symbionts. Additionally, these genes are required for nodule development by rhizobia, specifically for the formation of nodule primordia, except *SYM6*, since small nodule-like structures could still be observed on *sym6* mutants during the interaction with *Mesorhizobium loti* but not on the other six mutants. After inoculation with AMF, *sym* mutant roots had a limited ability to form arbuscules, indicating the key role of the symbiosis-related genes for intracellular infection by fungal partners (Kistner et al., 2005).

Symbiotic nitrogen fixation in nodules is also affected when the growth is slow due to nutrient deficiency. An RNA-Seq transcriptome profiling of the nodules of *M. truncatula* inoculated with *S. meliloti* under P-deficient conditions revealed lower nitrogenase activities (Liese et al., 2017). The reduction in SNF efficiency was due to a downregulation of genes encoding leghemoglobins, nicotianamine synthase-like proteins and NCR peptides. Also, low-Pi conditions reduced growth, nodule numbers and P concentrations in stem, root and nodule tissues compared to the control (Liese et al., 2017). However, the P concentration in nodules was relatively higher than in stem and root tissues. In soybean, *GmEXLB1*, encoding an expansin, was identified as a key gene involved in the response to low-P stress through transcription profiling (Kong et al., 2019). Expansins are involved in the regulation of root development and responses to abiotic stresses in plants (Cosgrove, 2015). It was observed that *GmEXLB1* expression was induced in the lateral roots of soybean under P starvation conditions. The ectopic expression of *GmEXLB1* in transgenic *Arabidopsis* promoted changes to the root architecture by increasing the number and length of lateral roots, thus improving P acquisition under low-P conditions (Kong et al., 2019). The acid phosphatase-encoding gene, *GmACP1*, has also been associated with P efficiency in soybean (Zhang et al., 2014). *GmACP1* is located in the QTL, *qPE8*, related to soybean P efficiency, on chromosome 8. Transgenic soybean hairy roots overexpressing *GmACP1* showed a 2.3-fold increase in acid phosphatase activity and increased the efficiency of P usage by 11.2–20.0% relative to the control (Zhang et al., 2014).

METABOLIC MODELING AND METABOLIC PROFILING STUDIES ON LEGUME-RHIZOBIUM SYMBIOSIS

Legume-rhizobium symbiosis involves biological nitrogen fixation as well as the exchange of nutrients. Transport systems and carbon-nitrogen metabolism are coordinated to provide nutrients for effective SNF. Each host plant-rhizobium interaction is association-specific and determines the accumulation of carbon and energy storage metabolites inside the microsymbionts (Terpolilli et al., 2012). Metabolic modeling provides new insights into the metabolism in plant systems. An emerging approach to studying the process of SNF in legumes is constraint-based modeling, which relies on network topology and the integration of genomic and high-throughput data for analyzing the metabolic capabilities of organisms (Feist et al., 2009; Heirendt et al., 2019). Among the most frequently used frameworks is flux balance analysis using genome-scale models (Orth et al., 2010). This approach has been successfully used to study the process of SNF carried out in mature nodules by rhizobia. Metabolic network reconstructions have been reported for *S. meliloti* (Zhao et al., 2012; diCenzo et al., 2016, 2018), *Bradyrhizobium diazoefficiens* (Yang Y. et al., 2017), the symbiotic forms of *Rhizobium etli* (Resendis-Antonio et al., 2007, 2012), and *Sinorhizobium fredii* (Contador et al., 2020). These studies included the nutrient requirements for each N₂-fixing

relationship and performed preliminary analyses of SNF such as flux analyses of relevant biochemical pathways (e.g., the TCA cycle, the pentose phosphate pathway, and the production of cofactors for nitrogenase) and determination of the genes required for an efficient SNF through computational simulations. However, these analyses simplified the plant metabolism by only including inputs from the host plant to the microsymbiont metabolism during symbiosis.

Recently, the impact of wild and cultivated soybeans on nitrogen fixation was assessed in the *S. fredii* model (Contador et al., 2020). Transcriptome profiles of *S. fredii* in symbiotic conditions were used to capture the differential nitrogen-fixing capacity of *S. fredii* strain CCBAU45436 in symbiosis with *G. max* C08 and *G. soja* W05, respectively. The bacteroid model quantified and predicted a higher nitrogen fixation activity and C/N ratio of *S. fredii* with the cultivated soybean than with the wild accession. This is consistent with previous experimental observations that this soybean cultivar (C08) outperformed the wild accession (W05) in a series of nitrogen fixation-related traits including nodule number, total nitrogen and total ureide accumulation, and new QTLs for ureide content and nodule fresh weight have been identified using RIL populations (Muñoz et al., 2016). An examination of these QTLs revealed a very low diversity in some regions among cultivated soybeans. Domestication process was suggested as the responsible for the selection of these traits (Muñoz et al., 2016).

The representation of whole-rhizobium metabolic networks has been achieved for *S. meliloti* and *B. diazoefficiens* USDA110 (diCenzo et al., 2016; Yang Y. et al., 2017). These reconstructions were used to build context-specific models to characterize the metabolic capabilities of rhizobia in bulk soil, rhizosphere and nodule bacteroids. However, metabolic models have been published for only two legume species, *G. max* and *M. truncatula*, despite the significance of legumes to sustainable agriculture. A compartmentalized reconstruction for *G. max* was used to construct a multi-organ model to represent the reserve mobilization in the cotyledon and hypocotyl/root tissues (Moreira et al., 2019). This reconstruction can also be used to represent other tissues and conditions since the whole set of metabolic reactions in soybean were used to construct the metabolic model. Annotations of the *G. max* genome were used to define the gene-protein-reaction associations in the model. On the other hand, the *M. truncatula* model was used to capture both biomass production during day and night conditions and host-microsymbiont interactions (Pfau et al., 2018). This was the first attempt to elucidate the effects of symbiosis on metabolic fluxes and plant growth. Computational analyses revealed the costs and benefits of the symbiotic system for biomass growth when external ammonium is not available. Recently, the *M. truncatula* model was built to represent multi-organism metabolic interactions in different developmental zones within the nodule during nitrogen fixation (diCenzo et al., 2020). The multi-organism model consisted of *M. truncatula* nodulated by *S. meliloti* Rm1021. The plant system was represented by a multi-compartmental reconstruction that includes root and shoot tissues together with five nodule zones (apical meristem, distal, proximal, interzone, and nitrogen fixation zones). The

accuracy of the metabolic reconstruction for *S. meliloti* was improved over the previous versions of the model. The integrated model was used to evaluate the use of nutrients by N₂-fixing bacteroids, N₂ fixation efficiency and costs related to the nitrogen fixation process.

The metabolic states of the legume-rhizobium association can also be assessed by performing metabolic profiling (Roessner et al., 2001; Rambla et al., 2015), which can provide a comprehensive picture of a particular sample or a genetically manipulated system. Metabolic profiling has been used to study metabolic shifts in SNF associations. Recently, metabolic profiling were performed on effective and ineffective nodules of soybean to identify the active processes according to the partners involved in the symbiosis (Agtuca et al., 2020). Metabolite levels were determined to characterize the different plant phenotypes and interactions with rhizobia. Similar comparative studies have been performed in *Medicago* and *L. japonicus* to investigate metabolites and related pathways that influence nodule metabolism during nitrogen fixation (Desbrosses et al., 2005; Ye et al., 2013; Gemperline et al., 2015). Other metabolomic studies have characterized the differentiation process from free-living rhizobia into bacteroids to detect differences in the metabolite composition of these two physiological conditions (Vauclare et al., 2013). Metabolite profiling of root hairs was also performed to characterize the early stage of rhizobial infection, by identifying those metabolites the accumulation of which was regulated in response to rhizobial inoculation (Brechenmacher et al., 2010).

CONCLUSION AND PERSPECTIVE

Legumes engage in mutualistic relationships with rhizobia and mycorrhizal fungi, enabling them to obtain essential nutrients, promote growth, and enhance biotic and abiotic stress resistance. There is extensive evidence to support that AMF affect the establishment and functions of rhizobial nodulation in both wild and cultivated legumes and enhance nitrogen acquisition in the host plant through SNF. However, domestication has reduced the genetic diversity in host plants and impacted their compatibility with rhizobial strains and AMF. And the intensive applications of chemical fertilizers and large-scale monoculture

farming practices during domestication actually have a negative impact on AMF community structures and their colonization of legume roots. It is important for researchers to engineer or screen for rhizobium strains, which are both competent for nodulation and capable of high nitrogen fixation efficiency, and AMF, which can effectively improve the nutrient acquisition by the host plants. Meanwhile, it is also vital to improve host plant's ability to interact with the best beneficial mutualistic symbionts. Understanding the genetic control of the symbiosis specificity between wild and cultivated legumes and identifying the key genetic factors controlling the symbiotic interaction will provide useful information for improving legume-soil microbe symbiosis. Moreover, utilizing advanced approaches such as metabolic modeling and metabolic profiling to investigate the molecular mechanisms of metabolite exchanges among legumes, AMF and rhizobia under nutrient-deficient conditions will help researchers better understand legume-microbe symbiosis, and ultimately contribute to the implementation of sustainable agriculture.

AUTHOR CONTRIBUTIONS

H-ML and AL planned and coordinated the writing and finished the final draft. AL and Y-SK put together the first complete draft. CC contributed to the literature search and writing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Hong Kong Research Grants Council Area of Excellence Scheme (AoE/M-403/16) and the Lo Kwee-Seong Biomedical Research Fund. Any opinions, findings, conclusions, or recommendations expressed in this publication do not reflect the views of the Government of the Hong Kong Special Administrative Region or the Innovation and Technology Commission.

ACKNOWLEDGMENTS

Jee-Yan Chu edited the manuscript.

REFERENCES

- Abbo, S., Pinhasi van-Oss, R., Gopher, A., Saranga, Y., Ofner, I., and Peleg, Z. (2014). Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends Plant Sci.* 19, 351–360. doi: 10.1016/j.tplants.2013.12.002
- Agtuca, B. J., Stopka, S. A., Evans, S., Samarah, L., Liu, Y., Xu, D., et al. (2020). Metabolomic profiling of wild-type and mutant soybean root nodules using laser-ablation electrospray ionization mass spectrometry reveals altered metabolism. *Plant J.* 103, 1937–1958. doi: 10.1111/tpj.14815
- Amarasena, N., Senthilkumar, M., and Krishna Kumar, A. S. (2020). *Microbial Mitigation of Stress Response of Food Legumes*. Boca Raton, FL: CRC PRESS (Taylor & Francis Group).
- Bennett, A. E., Daniell, T. J., and White, P. J. (2013). “Benefits of breeding crops for yield response to soil organisms,” in *Molecular Microbial Ecology of the Rhizosphere*, ed. F. J. de Bruijn (Hoboken, NJ: John Wiley & Sons, Inc), 17–27. doi: 10.1002/9781118297674.ch3
- Bizarro, M. J., Giongo, A., Vargas, L. F. W., Gano, K. A., de Sá, E. L. S., et al. (2011). Genetic variability of soybean *Bradyrhizobia* populations under different soil managements. *Biol. Fertil. Soils* 47, 357–362. doi: 10.1007/s00374-010-0512-6
- Brechenmacher, L., Lei, Z., Libault, M., Findley, S., Sugawara, M., Sadowsky, M. J., et al. (2010). Soybean metabolites regulated in root hairs in response to the symbiotic bacterium *Bradyrhizobium japonicum*. *Plant Physiol.* 153, 1808–1822. doi: 10.1104/pp.110.157800
- Brockwell, J., Bottomley, P. J., and Thies, J. E. (1995). “Manipulation of rhizobia microflora for improving legume productivity and soil fertility: a critical assessment,” in *Management of Biological Nitrogen Fixation for the Development of More Productive and Sustainable Agricultural Systems*, eds J. K. Ladha and M. B. Peoples (Dordrecht: Springer), doi: 10.1007/978-94-011-0055-7_7

- Bronstein, J. L. (2001). The exploitation of mutualisms. *Ecol. Lett.* 4, 277–287. doi: 10.1046/j.1461-0248.2001.00218.x
- Brundrett, M. (1991). Mycorrhizas in natural ecosystems. *Adv. Ecol. Res.* 21, 171–313. doi: 10.1016/S0065-2504(08)60099-9
- Caballero-Mellado, J., and Martínez-Romero, E. (1999). Soil fertilization limits the genetic diversity of *Rhizobium* in bean nodules. *Symbiosis* 26, 111–121.
- Chang, C., Chen, W., Luo, S., Ma, L., Li, X., and Tian, C. (2019). Rhizosphere microbiota assemblage associated with wild and cultivated soybeans grown in three types of soil suspensions. *Arch. Agron. Soil Sci.* 65, 74–87. doi: 10.1080/03650340.2018.1485147
- Chang, C., Zhang, J., Liu, T., Song, K., Xie, J., Luo, S., et al. (2020). Rhizosphere fungal communities of wild and cultivated soybeans grown in three different soil suspensions. *Appl. Soil Ecol.* 153:103586. doi: 10.1016/j.apsoil.2020.103586
- Cleveland, C. C., Townsend, A. R., Schimel, D. S., Fisher, H., Howarth, R. W., Hedin, L. O., et al. (1999). Global patterns of terrestrial biological nitrogen (N₂) fixation in natural ecosystems. *Global Biogeochem. Cycles* 13, 623–645. doi: 10.1002/(ISSN)1944-9224
- Contador, C. A., Lo, S.-K., Chan, S. H. J., and Lam, H.-M. (2020). Metabolic analyses of nitrogen fixation in the soybean microsymbiont *Sinorhizobium fredii* using constraint-based modeling. *mSystems* 5, e516–e519.
- Cosgrove, D. J. (2015). Plant expansins: diversity and interactions with plant cell walls. *Curr. Opin. Plant Biol.* 25, 162–172. doi: 10.1016/j.pbi.2015.05.014
- Crespi, M., and Gálvez, S. (2000). Molecular mechanisms in root nodule development. *J. Plant Growth Regul.* 19, 155–166. doi: 10.1007/s003440000023
- Cui, Y., Barampuram, S., Stacey, M. G., Hancock, N., Findley, S., Mathieu, M., et al. (2013). Tnt1 retrotransposon mutagenesis: a tool for soybean functional genomics. *Plant Physiol.* 161, 36–37. doi: 10.1104/pp.112.205369
- Dakora, F. D. (1995). Plant flavonoids: biological molecules for useful exploitation. *Aust. J. Plant Physiol.* 22, 87–99. doi: 10.1071/PP950087
- Dakora, F. D. (2003). Defining new roles for plant and rhizobial molecules in sole and mixed plant cultures involving symbiotic legumes. *New Phytol.* 158, 39–49. doi: 10.1046/j.1469-8137.2003.00725.x
- Desbrosses, G. G., Kopka, J., and Udvardi, M. K. (2005). *Lotus japonicus* metabolic profiling. Development of gas chromatography-mass spectrometry resources for the study of plant-microbe interactions. *Plant Physiol.* 137, 1302–1318. doi: 10.1104/pp.104.054957
- diCenzo, G. C., Benedict, A. B., Fondi, M., Walker, G. C., Finan, T. M., Mengoni, A., et al. (2018). Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium *Sinorhizobium meliloti*. *PLoS Genet.* 14:e1007357. doi: 10.1371/journal.pgen.1007357
- diCenzo, G. C., Checucci, A., Bazzicalupo, M., Mengoni, A., Viti, C., Dziewit, L., et al. (2016). Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat. Commun.* 7, 1–10. doi: 10.1038/ncomms12219
- diCenzo, G. C., Tesi, M., Pfau, T., Mengoni, A., and Fondi, M. (2020). Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. *Nat. Commun.* 11:2574.
- Egamberdiyeva, D., and Höflich, G. (2004). Effect of plant growth-promoting bacteria on growth and nutrient uptake of cotton and pea in a semi-arid region of Uzbekistan. *J. Arid Environ.* 56, 293–301. doi: 10.1016/S0140-1963(03)00050-8
- Eom, A. H., Lee, S. S., Ahn, T. K., and Lee, M. W. (1994). Ecological roles of arbuscular mycorrhizal fungi in two wild legume plants. *Mycoscience* 35, 69–75. doi: 10.1007/bf02268531
- Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L., and Palsson, B. Ø. (2009). Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* 7, 129–143. doi: 10.1038/nrmicro1949
- Ferreira, M. C., De, S., Andrade, D., Ligia, L. M., Takemura, S. M., and Hungria, M. (2000). Tillage method and crop rotation effects on the population sizes and diversity of *bradyrhizobia* nodulating soybean. *Soil Biol. Biochem.* 32, 627–637. doi: 10.1016/S0038-0717(99)00189-3
- 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
- Gage, D. J. (2004). Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol. Mol. Biol. Rev.* 68, 280–300. doi: 10.1128/mmbr.68.2.280-300.2004
- Gemperline, E., Jayaraman, D., Maeda, J., Ané, J.-M., and Li, L. (2015). Multifaceted investigation of metabolites during nitrogen fixation in *Medicago* high resolution MALDI-MS Imaging and ESI-MS. *J. Am. Soc. Mass Spectrom.* 26, 149–158. doi: 10.1007/s13361-014-1010-0
- Gepts, P. (2010). “Crop domestication as a long-term selection experiment,” in *Plant Breeding Reviews*, eds R. Singh and P. Jauhar (Oxford: John Wiley & Sons, Inc), 1–44. doi: 10.1002/9780470650288.ch1
- Gopalakrishnan, S., Sathya, A., Vijayabharathi, R., Varshney, R. K., Gowda, C. L. L., and Krishnamurthy, L. (2015). Plant growth promoting rhizobia: challenges and opportunities. *3 Biotech* 5, 355–377. doi: 10.1007/s13205-014-0241-x
- Hakoyama, T., Niimi, K., Watanabe, H., Tabata, R., Matsubara, J., Sato, S., et al. (2009). Host plant genome overcomes the lack of a bacterial gene for symbiotic nitrogen fixation. *Nature* 462, 514–517. doi: 10.1038/nature08594
- Hayashi, M., Shiro, S., Kanamori, H., Mori-Hosokawa, S., Sasaki-Yamagata, H., Sayama, T., et al. (2014). A thaumatin-like protein, Rj4, controls nodule symbiotic specificity in soybean. *Plant Cell Physiol.* 55, 1679–1689. doi: 10.1093/pcp/pcu099
- Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., et al. (2019). Creation and analysis of biochemical constraint-based models using the COBRA toolbox v.3.0. *Nat. Protoc.* 14, 639–702. doi: 10.1038/s41596-018-0098-2
- Ianson, D., and Linderman, R. (1993). Variation in the response of nodulating pigeonpea (*Cajanus cajan*) to different isolates of mycorrhizal fungi. *Symbiosis* 15, 105–119.
- Imaizumi-Anraku, H., Kawaguchi, M., Koiwa, H., Akao, S., and Syôno, K. (1997). Two ineffective-nodulating mutants of *Lotus japonicus*-different phenotypes caused by the blockage of endocytotic bacterial release and nodule maturation. *Plant Cell Physiol.* 38, 871–881. doi: 10.1093/oxfordjournals.pcp.a029246
- Itelima, J., Bang, W., Onyimba, I., Sila, M., and Egbere, O. (2018). Bio-fertilizers as key player in enhancing soil fertility and crop productivity: a review. *J. Microbiol. Biotechnol. Rep.* 1.
- Jakobsen, I. (1985). The role of phosphorus in nitrogen fixation by young pea plants (*Pisum sativum*). *Physiol. Plant.* 64, 190–196. doi: 10.1111/j.1399-3054.1985.tb02334.x
- Jansa, J., Finlay, R., Wallander, H., Smith, F. A., and Smith, S. E. (2011). Role of mycorrhizal symbioses in phosphorus cycling. *Soil Biol.* 26, 137–168. doi: 10.1007/978-3-642-15271-9_6
- Jansa, J., Wiemken, A., and Frossard, E. (2006). “The effects of agricultural practices on arbuscular mycorrhizal fungi,” in *Function of Soils for Human Societies and the Environment*, eds E. Frossard, W. E. H. Blum, and B. P. Warkentin (London: Geological Society), 89–115. doi: 10.1144/gsl.sp.2006.266.01.08
- Jia, Y., Gray, V. M., and Straker, C. J. (2004). The influence of rhizobium and arbuscular mycorrhizal fungi on nitrogen and phosphorus accumulation by *Vicia faba*. *Ann. Bot.* 94, 251–258. doi: 10.1093/aob/mch135
- Khalil, S., Loynachan, T. E., and Tabatabai, M. A. (1994). Mycorrhizal dependency and nutrient uptake by improved and unimproved corn and soybean cultivars. *Agron. J.* 86, 949–958. doi: 10.2134/agronj1994.00021962008600060005x
- Khalil, S., Loynachan, T. E., and Tabatabai, M. A. (1999). Plant determinants of mycorrhizal dependency in soybean. *Agron. J.* 91, 135–141. doi: 10.2134/agronj1999.00021962009100010021x
- Kiers, E. T., West, S. A., and Denison, R. F. (2002). Mediating mutualisms: farm management practices and evolutionary changes in symbiont co-operation. *J. Appl. Ecol.* 39, 745–754. doi: 10.1046/j.1365-2664.2002.00755.x
- Kim, D. H., Kaashyap, M., Rathore, A., Das, R. R., Parupalli, S., Upadhyaya, H. D., et al. (2014). Phylogenetic diversity of *Mesorhizobium* in chickpea. *J. Biosci.* 39, 513–517. doi: 10.1007/s12038-014-9429-9
- Kistner, C., Winzer, T., Pitzschke, A., Mulder, L., Sato, S., Kaneko, T., et al. (2005). Seven *Lotus japonicus* genes required for transcriptional reprogramming of the root during fungal and bacterial symbiosis. *Plant Cell* 17, 2217–2229. doi: 10.1105/tpc.105.032714
- Kong, Y., Wang, B., Du, H., Li, W., Li, X., and Zhang, C. (2019). *GmEXLB1*, a soybean expansin-like B gene, alters root architecture to improve phosphorus acquisition in Arabidopsis. *Front. Plant Sci.* 10:808. doi: 10.3389/fpls.2019.00808
- Kuzma, M. M., Hunt, S., and Layzell, D. B. (1993). Role of oxygen in the limitation and inhibition of nitrogenase activity and respiration rate in individual soybean nodules. *Plant Physiol.* 101, 161–169. doi: 10.1104/pp.101.1.161

- Lau, J. A., Bowling, E. J., Gentry, L. E., Glasser, P. A., Monarch, E. A., Olesen, W. M., et al. (2012). Direct and interactive effects of light and nutrients on the legume-rhizobia mutualism. *Acta Oecol.* 39, 80–86. doi: 10.1016/j.actao.2012.01.004
- Le Signor, C., Savoie, V., Aubert, G., Verdier, J., Nicolas, M., Pagny, G., et al. (2009). Optimizing TILLING populations for reverse genetics in *Medicago truncatula*. *Plant Biotechnol. J.* 7, 430–441. doi: 10.1111/j.1467-7652.2009.00410.x
- Lehmann, A., Barto, E. K., Powell, J. R., and Rillig, M. C. (2012). Mycorrhizal responsiveness trends in annual crop plants and their wild relatives - a meta-analysis on studies from 1981 to 2010. *Plant Soil* 355, 231–250. doi: 10.1007/s11104-011-1095-1
- Liese, R., Schulze, J., and Cabeza, R. A. (2017). Nitrate application or P deficiency induce a decline in *Medicago truncatula* N₂-fixation by similar changes in the nodule transcriptome. *Sci. Rep.* 7:46264.
- Liu, A., Contador, C. A., Fan, K., and Lam, H. M. (2018). Interaction and regulation of carbon, nitrogen, and phosphorus metabolisms in root nodules of legumes. *Front. Plant Sci.* 9:1860. doi: 10.3389/fpls.2018.01860
- Maherali, H. (2014). Is there an association between root architecture and mycorrhizal growth response?? *New Phytol.* 204, 192–200. doi: 10.1111/nph.12927
- Manavalan, L. P., Prince, S. J., Musket, T. A., Chaky, J., Deshmukh, R., Vuong, T. D., et al. (2015). Identification of novel QTL governing root architectural traits in an interspecific soybean population. *PLoS One* 10:e0120490. doi: 10.1371/journal.pone.0120490
- Martín-Robles, N., Lehmann, A., Seco, E., Aroca, R., Rillig, M. C., and Milla, R. (2018). Impacts of domestication on the arbuscular mycorrhizal symbiosis of 27 crop species. *New Phytol.* 218, 322–334. doi: 10.1111/nph.14962
- Mohammadi, K., Sohrabi, Y., Heidari, G., Khalesro, S., and Majidi, M. (2012). Effective factors on biological nitrogen fixation. *Afr. J. Agric. Res.* 7, 1782–1788. doi: 10.5897/ajarx11.034
- Moreira, T. B., Shaw, R., Luo, X., Ganguly, O., Kim, H.-S., Ferreira Coelho, L. G., et al. (2019). A genome-scale metabolic model of soybean (*Glycine max*) highlights metabolic fluxes in seedlings. *Plant Physiol.* 180, 1912–1929. doi: 10.1104/pp.19.00122
- Muñoz, N., Liu, A., Kan, L., Li, M. W., and Lam, H. M. (2017). Potential uses of wild germplasm of grain legumes for crop improvement. *Int. J. Mol. Sci.* 18:328. doi: 10.3390/ijms18020328
- Muñoz, N., Qi, X., Li, M.-W., Xie, M., Gao, Y., Cheung, M.-Y., et al. (2016). Improvement in nitrogen fixation capacity could be part of the domestication process in soybean. *Heredity* 117, 84–93. doi: 10.1038/hdy.2016.27
- Mutch, L. A., and Young, J. P. W. (2004). Diversity and specificity of *Rhizobium leguminosarum* biovar *viciae* on wild and cultivated legumes. *Mol. Ecol.* 13, 2435–2444. doi: 10.1111/j.1365-294X.2004.02259.x
- Nouwen, N., Arrighi, J.-F., Cartieaux, F., Chaintreuil, C., Gully, D., Klopp, C., et al. (2017). The role of rhizobial (NiFv) and plant (FEN1) homocitrate synthases in *Aeschynomene*/photosynthetic *Bradyrhizobium* symbiosis. *Sci. Rep.* 7:448. doi: 10.1038/s41598-017-00559-0
- Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010). What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248. doi: 10.1038/nbt.1614
- Oruru, M. B., Njeru, E. M., Pasquet, R., and Runo, S. (2017). Response of a wild-type and modern cowpea cultivars to arbuscular mycorrhizal inoculation in sterilized and non-sterilized soil. *J. Plant Nutr.* 41, 90–101. doi: 10.1080/01904167.2017.1381728
- Ott, T., Sullivan, J., James, E. K., Fletmetakis, E., Günther, C., Gibon, Y., et al. (2009). Absence of symbiotic leghemoglobins alters bacteroid and plant cell differentiation during development of *Lotus japonicus* root nodules. *Mol. Plant Microbe Interact.* 22, 800–808. doi: 10.1094/mpmi-22-7-0800
- Ott, T., van Dongen, J. T., Günther, C., Krusell, L., Desbrosses, G., Vigeolas, H., et al. (2005). Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Curr. Biol.* 15, 531–535. doi: 10.1016/j.cub.2005.01.042
- Pérez-jaramillo, J. E., Carrión, V. J., Bosse, M., Ferrão, L. F. V., de Hollander, M., García, A. A. F., et al. (2017). Linking rhizosphere microbiome composition of wild and domesticated *Phaseolus vulgaris* to genotypic and root phenotypic traits. *ISME J.* 11, 2244–2257. doi: 10.1038/ismej.2017.85
- Pérez-jaramillo, J. E., Mendes, R., and Raaijmakers, J. M. (2016). Impact of plant domestication on rhizosphere microbiome assembly and functions. *Plant Mol. Biol.* 90, 635–644. doi: 10.1007/s11103-015-0337-7
- Pfau, T., Christian, N., Masakapalli, S. K., Sweet, L. J., Poolman, M. G., and Ebenhö, O. (2018). The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. *Sci. Rep.* 8:12504. doi: 10.1038/s41598-018-30884-x
- Plett, J. M., Plett, K. L., Bithell, S. L., Mitchell, C., Moore, K., Powell, J. R., et al. (2016). Improved *Phytophthora* resistance in commercial chickpea (*Cicer arietinum*) varieties negatively impacts symbiotic gene signalling and symbiotic potential in some varieties. *Plant Cell Environ.* 39, 1858–1869. doi: 10.1111/pce.12757
- Prashar, P., and Vandenberg, A. (2017). Genotype-specific responses to the effects of commercial Trichoderma formulations in lentil (*Lens culinaris* ssp. *culinaris*) in the presence and absence of the oomycete pathogen *Aphanomyces euteiches*. *Biocontrol Sci. Technol.* 27, 1123–1144. doi: 10.1080/09583157.2017.1376035
- Prince, S. J., Song, L., Qiu, D., Maldonado dos Santos, J. V., Chai, C., Joshi, T., et al. (2015). Genetic variants in root architecture-related genes in a *Glycine soja* accession, a potential resource to improve cultivated soybean. *BMC Genomics* 16:132. doi: 10.1186/s12864-015-1334-6
- Provorov, N. A., and Tikhonovich, I. A. (2003). Genetic resources for improving nitrogen fixation in legume-rhizobia symbiosis. *Genet. Resour. Crop Evol.* 50, 89–99. doi: 10.1023/A:1022957429160
- Pueppke, S. G., and Broughton, W. J. (1999). *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. *Mol. Plant Microbe Interact.* 12, 293–318. doi: 10.1094/MPMI.1999.12.4.293
- Rambla, J. L., López-Gresa, M. P., Bellés, J. M., and Granell, A. (2015). “Metabolomic profiling of plant tissues,” in *Plant Functional Genomics*, eds J. Alonso and A. Stepanova (New York, NY: Humana Press).
- Resendis-Antonio, O., Hernández, M., Mora, Y., and Encarnación, S. (2012). Functional modules, structural topology, and optimal activity in metabolic networks. *PLoS Comput. Biol.* 8:e1002720. doi: 10.1371/journal.pcbi.1002720
- Resendis-Antonio, O., Reed, J. L., Encarnación, S., Collado-Vides, J., and Palsson, B. (2007). Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. *PLoS Comput. Biol.* 3:e192. doi: 10.1371/journal.pcbi.0030192
- Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., et al. (2001). Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* 13, 11–29. doi: 10.2307/3871150
- Rubio, L. M., and Ludden, P. W. (2008). Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annu. Rev. Microbiol.* 62, 93–111. doi: 10.1146/annurev.micro.62.081307.162737
- Schultze, M., and Kondorosi, A. (1998). Regulation of symbiotic root nodule development. *Annu. Rev. Genet.* 32, 33–57. doi: 10.1146/annurev.genet.32.1.33
- Singh, D. P., Singh, H. B., and Prabha, R. (2016). *Microbial Inoculants in Sustainable Agricultural Productivity: Vol. 1: Research Perspectives*. Delhi: Springer, doi: 10.1007/978-81-322-2647-5
- Stacey, G. (2007). “The rhizobium-legume nitrogen-fixing symbiosis,” in *Biology of the Nitrogen Cycle*, eds H. Bothe, S. J. Ferguson, and W. E. Newton (Amsterdam: Elsevier), doi: 10.1016/B978-0-44452857-5.50011-4
- Suliman, S., and Tran, L.-S. P. (2014). Symbiotic nitrogen fixation in legume nodules: metabolism and regulatory mechanisms. *Int. J. Mol. Sci.* 15, 19389–19393. doi: 10.3390/ijms151119389
- Szczalba, M., Kopta, T., Gaštol, M., and Sękara, A. (2019). Comprehensive insight into arbuscular mycorrhizal fungi, *Trichoderma* spp. and plant multilevel interactions with emphasis on biostimulation of horticultural crops. *J. Appl. Microbiol.* 127, 630–647. doi: 10.1111/jam.14247
- Tajini, F., Trabelsi, M., and Drevon, J. J. (2012). Combined inoculation with *Glomus intraradices* and *Rhizobium tropici* CIAT899 increases phosphorus use efficiency for symbiotic nitrogen fixation in common bean (*Phaseolus vulgaris* L.). *Saudi J. Biol. Sci.* 19. doi: 10.1016/j.sjbs.2011.11.003
- Tang, F., Yang, S., Liu, J., and Zhu, H. (2016). *Rj4*, a gene controlling nodulation specificity in soybeans, encodes a Thaumatin-Like protein but not the one previously reported. *Plant Physiol.* 170, 26–32. doi: 10.1104/pp.15.01661
- Terpolilli, J. J., Hood, G. A., and Poole, P. S. (2012). What determines the efficiency of N₂-fixing rhizobium-legume symbioses? *Adv. Microb. Physiol.* 60, 325–389. doi: 10.1016/B978-0-12-398264-3.00005-X
- Tian, L., Shi, S., Ma, L., Tran, L. P., and Tian, C. (2020). Community structures of the rhizospheric microbiomes of cultivated and wild soybeans in their continuous cropping. *Microbiol. Res.* 232:126390. doi: 10.1016/j.micres.2019.126390

- Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* 11, 636–646. doi: 10.1038/nrg2842
- Valdenegro, M., Barea, J. M., and Azcón, R. (2001). Influence of arbuscular-mycorrhizal fungi, *Rhizobium meliloti* strains and PGPR inoculation on the growth of *Medicago arborea* used as model legume for re-vegetation and biological reactivation in a semi-arid mediterranean area. *Plant Growth Regul.* 34, 233–240. doi: 10.1023/A:1013323529603
- Vaulaire, P., Bligny, R., Gout, E., and Widmer, F. (2013). An overview of the metabolic differences between *Bradyrhizobium japonicum* 110 bacteria and differentiated bacteroids from soybean (*Glycine max*) root nodules: an in vitro ¹³C- and ³¹P-nuclear magnetic resonance spectroscopy study. *FEMS Microbiol. Lett.* 343, 49–56. doi: 10.1111/1574-6968.12124
- Vázquez-Limón, C., Hoogewijs, D., Vinogradov, S. N., and Arredondo-Peter, R. (2012). The evolution of land plant hemoglobins. *Plant Sci.* 191–192, 71–81. doi: 10.1016/j.plantsci.2012.04.013
- Wang, H., Man, C. X., Wang, E. T., and Chen, W. X. (2009). Diversity of rhizobia and interactions among the host legumes and rhizobial genotypes in an agricultural-forestry ecosystem. *Plant Soil* 314, 169–182. doi: 10.1007/s11104-008-9716-z
- Wang, L., Rubio, M. C., Xin, X., Zhang, B., Fan, Q., Wang, Q., et al. (2019). CRISPR/Cas9 knockout of leghemoglobin genes in *Lotus japonicus* uncovers their synergistic roles in symbiotic nitrogen fixation. *New Phytol.* 224, 818–832. doi: 10.1111/nph.16077
- Wang, L., Wang, L., Tan, Q., Fan, Q., Zhu, H., Hong, Z., et al. (2016). Efficient inactivation of symbiotic nitrogen fixation related genes in *Lotus japonicus* using CRISPR-Cas9. *Front. Plant Sci.* 7:1333. doi: 10.3389/fpls.2016.01333
- Wang, L., Wang, L., Zhou, Y., and Duanmu, D. (2017). Use of CRISPR/Cas9 for symbiotic nitrogen fixation research in legumes. *Prog. Mol. Biol. Transl. Sci.* 149, 187–213. doi: 10.1016/bs.pmbts.2017.03.010
- Wang, Q., Yang, S., Liu, J., Terecskei, K., Abrahám, E., Gombár, A., et al. (2017). Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* 114, 6854–6859.
- Wang, X., Pan, Q., Chen, F., Yan, X., and Liao, H. (2011). Effects of co-inoculation with arbuscular mycorrhizal fungi and rhizobia on soybean growth as related to root architecture and availability of N and P. *Mycorrhiza*. 21, 173–181. doi: 10.1007/s00572-010-0319-1
- Weese, D. J., Heath, K. D., Dentinger, B. T. M., and Lau, J. A. (2015). Long-term nitrogen addition causes the evolution of less-cooperative mutualists. *Evolution* 69, 631–642. doi: 10.1111/evo.12594
- Woo, S. L., Ruocco, M., Vinale, F., Nigro, M., Marra, R., Lombardi, N., et al. (2014). *Trichoderma*-based products and their widespread use in agriculture. *Open Mycol. J.* 8. doi: 10.2174/1874437001408010071
- Wu, L. J., Wang, H. Q., Wang, E. T., Chen, W. X., and Tian, C. F. (2011). Genetic diversity of nodulating and non-nodulating rhizobia associated with wild soybean (*Glycine soja* Sieb. & Zucc.) in different ecoregions of China. *FEMS Microbiol. Ecol.* 76, 439–450. doi: 10.1111/j.1574-6941.2011.01064.x
- Yang, S., Wang, Q., Fedorova, E., Liu, J., Qin, Q., Zheng, Q., et al. (2017). Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* 114, 6848–6853.
- Yang, Y., Hu, X. P., and Ma, B. G. (2017). Construction and simulation of the *Bradyrhizobium diazoefficiens* USDA110 metabolic network: a comparison between free-living and symbiotic states. *Mol. Biosyst.* 13, 607–620. doi: 10.1039/C6MB00553E
- Ye, H., Gempferline, E., Venkateshwaran, M., Chen, R., Delaux, P.-M., Howes-Podoll, M., et al. (2013). MALDI mass spectrometry-assisted molecular imaging of metabolites during nitrogen fixation in the *Medicago truncatula* -*Sinorhizobium meliloti* symbiosis. *Plant J.* 75, 130–145. doi: 10.1111/tpj.12191
- Zahran, H. H. (1999). Rhizobium-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. *Microbiol. Mol. Biol. Rev.* 63, 968–989. doi: 10.1128/mmbr.63.4.968-989.1999
- Zhang, D., Song, H., Cheng, H., Hao, D., Wang, H., Kan, G., et al. (2014). The acid phosphatase-encoding gene *GmACP1* contributes to soybean tolerance to low-phosphorus stress. *PLoS Genet.* 10:e1004061. doi: 10.1371/journal.pgen.1004061
- Zhang, J., Zhang, J., Zhang, P., He, P., Song, H., Zhang, J., et al. (2019). Rhizosphere fungal diversity of wild and cultivated soybeans in field and greenhouse experiments. *Int. J. Agric. Biol.* 21, 479–485. doi: 10.17957/IJAB/15.0918
- Zhao, H., Li, M., Fang, K., Chen, W., and Wang, J. (2012). In silico insights into the symbiotic nitrogen fixation in *Sinorhizobium meliloti* via metabolic reconstruction. *PLoS One* 7:e31287. doi: 10.1371/journal.pone.0031287

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

Copyright © 2020 Liu, Ku, Contador and Lam. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Modification of Circadian Clock Components in Soybean During Domestication and Improvement

Man-Wah Li¹ and Hon-Ming Lam^{1,2*}

¹ Center for Soybean Research of the State Key Laboratory of Agrobiotechnology and School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China, ² Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China

OPEN ACCESS

Edited by:

Peerasak Srinives,
Kasetsart University, Thailand

Reviewed by:

Jian Sun,
Shenyang Agricultural University,
China

Suk-Ha Lee,
Seoul National University,
South Korea

Jun Abe,
Hokkaido University, Japan

*Correspondence:

Hon-Ming Lam
honming@cuhk.edu.hk

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 10 June 2020

Accepted: 19 August 2020

Published: 30 September 2020

Citation:

Li M-W and Lam H-M (2020) The
Modification of Circadian Clock
Components in Soybean During
Domestication and Improvement.
Front. Genet. 11:571188.
doi: 10.3389/fgene.2020.571188

Agricultural production is greatly dependent on daylength, which is determined by latitude. Living organisms align their physiology to daylength through the circadian clock, which is made up of input sensors, core and peripheral clock components, and output. The light/dark cycle is the major input signal, moderated by temperature fluctuations and metabolic changes. The core clock in plants functions mainly through a number of transcription feedback loops. It is known that the circadian clock is not essential for survival. However, alterations in the clock components can lead to substantial changes in physiology. Thus, these clock components have become the *de facto* targets of artificial selection for crop improvement during domestication. Soybean was domesticated around 5,000 years ago. Although the circadian clock itself is not of particular interest to soybean breeders, specific alleles of the circadian clock components that affect agronomic traits, such as plant architecture, sensitivity to light/dark cycle, flowering time, maturation time, and yield, are. Consequently, compared to their wild relatives, cultivated soybeans have been bred to be more adaptive and productive at different latitudes and habitats for acreage expansion, even though the selection processes were made without any prior knowledge of the circadian clock. Now with the advances in comparative genomics, known modifications in the circadian clock component genes in cultivated soybean have been found, supporting the hypothesis that modifications of the clock are important for crop improvement. In this review, we will summarize the known modifications in soybean circadian clock components as a result of domestication and improvement. In addition to the well-studied effects on developmental timing, we will also discuss the potential of circadian clock modifications for improving other aspects of soybean productivity.

Keywords: circadian clock, domestication, early flowering 3, *Gigantea*, pseudo response regulator 3, *J* locus, soybean

INTRODUCTION

Studies have shown that plants with circadian clocks synchronized to their environmental conditions gain growth advantage over those not synchronized (Dodd et al., 2005). The circadian clock not only plays roles in growth and development, it is also involved in metabolism and stress responses (Fukushima et al., 2009; Dodd et al., 2015; Nakamichi et al., 2016; Nitschke et al., 2016;

Kim et al., 2017; Gil and Park, 2019). Current knowledge of the plant circadian clock mostly came from studies on *Arabidopsis*, which is a relatively simple model plant. The *Arabidopsis* circadian clock consists of mainly three transcription feedback loops, namely the central loop, the morning loop, and the evening loop (Staiger et al., 2013; Foo et al., 2016; Gil and Park, 2019). The central loop is made up of a feedback loop made up of circadian clock associated 1/late elongated hypocotyl (CCA1/LHY) and timing of CAB (TOC1; also known as pseudo response regulator 1, PRR1), which are mutual repressors of each other (Alabadi et al., 2001). On the other hand, in the morning loop, CCA1/LHY activates PRR5, PRR7, and PRR9, which in turn inhibit CCA1/LHY in a negative feedback loop (Mizuno and Nakamichi, 2005). Early flowering 3 (ELF3), early flowering 4 (ELF4), and LUX ARRHYTHMO (LUX) make up the evening complex (EC) that represses PRRs, while they themselves are repressed by CCA1/LHY (Huang and Nusinow, 2016). This delicate configuration, together with the input sensors and other peripheral components, allows the clock to oscillate in a cycle of approximately 24 h. Detailed discussions on the *Arabidopsis* circadian clock could be found in some recent reviews (Staiger et al., 2013; Foo et al., 2016; Shim et al., 2017; Gil and Park, 2019).

Studies on some crop species suggested that domestication has imposed significant changes on the circadian clock, in terms of the phase, amplitude and period (Shor and Green, 2016). Shifting the circadian phase of important biological processes could have big impacts. For example, a 3-bp deletion in the Phytochrome A-associated F-box protein-encoding gene, *EID1* (*Empfindlicher im dunkelroten Licht protein 1*), in the cultivated tomato has led to the lengthening of the clock period (Muller et al., 2016). This transition allowed the day-neutral tomatoes that originated from the equatorial Andean region of South America to adapt to longer daylengths in Mesoamerica and Europe (Muller et al., 2016). The deceleration of the clock reoriented the phase of some biological processes such as photosynthesis with the longer photoperiods in the higher latitudes to maximize the productivity of these processes. Other examples include the *EAM8* from Barley (Faure et al., 2012), a homolog of *Arabidopsis* *ELF3*, and *Eps3* from wheat (Gawronski et al., 2014), a homolog of *Arabidopsis* *LUX*. Alleles of these two genes have played important roles in affecting flowering time of the respective crops and altering the phase, period and amplitude of their circadian clocks (Faure et al., 2012; Gawronski et al., 2014). However, the physiological consequences of the altered clock in these two crops as a result of changes in these two genes are still largely unknown.

Up to now, specific studies on soybean circadian rhythm have been limited, but they can be dated back to the 1950s (Bunning, 1954; Brest, 1970). In those days, it has already been discovered that the circadian movement of trifoliate leaves of soybean, as measured with a kymograph, could last for two weeks after shifting to continuous light in a roughly 24-h fashion relying on the internal clock (Brest, 1970). A much more recent study examined the circadian fluctuations of chlorophyll contents using non-destructive multispectral imaging under drought conditions (Pan et al., 2015). Interestingly, circadian rhythm was not only observed in the aerial parts of the soybean plant. Expressing

the luciferase reporter gene driven by the soybean *LHY-CCA1-LIKE b2* (*GmLCLb2*) promoter or soybean *PRR9b2* promoter in transgenic soybean hairy roots revealed that the circadian rhythm of the soybean root is out of sync with the clock in the leaf (Wang Y. et al., 2020), suggesting that the circadian clock may play different roles in different organs of the soybean plant. A recent circadian study on soybean observed that the free-running circadian period lengths of elite soybean cultivars are positively correlated to the latitudes associated with the maturity groups that these cultivars belong to (Greenham et al., 2017). Although the study did not involve any wild soybean, since maturity groups are largely the results of domestication and its subsequent diversification (Jiang et al., 2014; Zhou et al., 2015; Wang Y. et al., 2016; Miladinovic et al., 2018), the correlation between period length and latitude/maturity group suggests that alterations in the soybean circadian clock could be a by-product of these processes.

So far, there are a few transcriptomic studies regarding the soybean circadian clock. One study observed that only about 1.8% of the transcriptome from *Glycine max* cv. Williams 82 developing seeds cycled under constant light condition (Hudson, 2010), which was much lower than 6–40% in *Arabidopsis* seedlings (Covington et al., 2008). Another study identified 3,695 time-indicating genes in the unifoliate leaves of Williams 82, which amount to less than 10% of the total protein-coding genes in the soybean genome (Li M. et al., 2019). The discrepancy between the two species was probably due to organ-specific effects of the clock. However, we cannot rule out the possibility that the oscillation of the clock is normally less robust in cultivated varieties (Shor and Green, 2016), which makes rhythmic genes less likely to be detected.

In addition to the well-known input signals of light and temperature, stresses are also input signals that could greatly alter the clock. It has been found that mild drought, heat shock, iron deficiency, and alkaline stress could change the period and phase of the expression of core clock genes in soybean (Li M. et al., 2019). Under alkaline stress, the leaf movement of wild soybean (*Glycine soja*) was roughly aligned with the clock gene expressions which showed an advancement of phase (Li M. et al., 2019). However, the leaf movements under drought and heat stress, and in iron-deficient plants, went out of sync with the core clock gene expressions, suggesting that different stresses may have different physiological effects (Li M. et al., 2019). Treatment of soybean with arsenate (As[V]) and arsenite (As[III]) could lead to changes in the amplitude of diurnal expressions of *GmLCL1*, *GmPRR9*, *GmELF4*, and *GmGI* (*Gigantea*), depending on the tissue and the form of arsenic compounds (Vezza et al., 2020). Stomatal movements and expressions of antioxidative enzymes were also altered by the As(V) and As(III) treatments. Yet, the link between the changes in core clock gene expressions and physiological changes remain elusive (Vezza et al., 2020). All in all, the direct physiological consequences of the alteration of the soybean circadian clock as a result of stresses are mostly uncharacterized.

During the domestication and improvement processes, breeders have selected for new soybean varieties with the aim to improve yield by mainly modifying the plant architecture and

ensuring the soybean plant will reach maturity within a desirable timeframe, especially when adapting to a new cultivation region or habitat. It is known that artificial selection has enriched some minor alleles compared to the gene pool of the progenitor populations (Lam et al., 2010). Using the latest technologies in genomic studies, it is discovered that breeders, without prior knowledge of the circadian clock, have inadvertently introduced a few important and dominant modifications to the core clock components. The soybean genome has recently undergone two rounds of whole-genome duplication. Some of the clock component genes have diversified in functions after duplication, which further complicates the studies on their functions. Genetic manipulation of the circadian clock components of soybean using the well-studied *Arabidopsis* circadian clock model as a guide could help further improve soybean productivity. With the recent population genomic data, three circadian clock components, *GmPRR3*, *GmELF3*, and *GmGla* from soybean were well-supported to be under artificial selection. In the following sections, we have summarized the recent discoveries in the modifications of these three circadian clock components in soybean during domestication.

MODIFICATIONS IN CIRCADIAN CLOCK COMPONENT GENES IN THE CULTIVATED SOYBEAN

Mutations in a Pair of *Pseudo Response Regulator 3 (PRR3)* Genes Resulted in Early Maturation

Earlier studies on the molecular controls of soybean maturation mainly focused on the *E* locus (Cober and Voldeng, 2001; Liu et al., 2008; Watanabe et al., 2009; Cober et al., 2010; Watanabe et al., 2011; Xia et al., 2012; Dissanayaka et al., 2016; Zhao et al., 2016; Lu et al., 2017; Samanfar et al., 2017; Yue et al., 2017). It was only recently that two major-effect quantitative trait loci (QTLs) on chromosomes 11 and 12 controlling maturation time and flowering time were found in wild soybeans or landraces (Qi et al., 2014; Lu et al., 2016, 2020; Fang et al., 2017; Pan et al., 2018; Li M. W. et al., 2019; Li C. et al., 2020; Ogiso-Tanaka et al., 2019). Dissection of these two QTLs identified two *Pseudo Response Regulator 3* genes, *GmPRR3a/Tof11* and *GmPRR3b/Tof12* on chromosomes 11 and 12, respectively, as the causal genes for controlling these two functions (Li M. W. et al., 2019; Li Y. et al., 2019; Li C. et al., 2020; Li Y. H. et al., 2020; Lu et al., 2020).

There are five *PRR* genes in *Arabidopsis*, including *PRR9*, *PRR7*, *PRR5*, *PRR3*, and *TOC1/PRR1*, which are all expressed in a circadian fashion and their expressions peak sequentially (Staiger et al., 2013; Gil and Park, 2019). These *PRR* proteins are typified with a pseudo receiver (PR) domain, an ERF-associated amphiphilic repression (EAR) motif, and a CONSTANTS, CO-like and TOC1 (CCT) domain (Gendron et al., 2012). Compared with other *Arabidopsis* *PRR* proteins, two non-synonymous substitutions naturally occurred on the EAR motif and CCT domain, respectively, of *AtPRR3*. These two substitutions are

basically fixed in all *Arabidopsis thaliana* ecotype (Alonso-Blanco et al., 2016). These substitutions made *AtPRR3* somehow behaves differently from other *PRRs* (Para et al., 2007). Unlike other *PRRs* which play direct roles in transcript repression, *AtPRR3* probably participates in the circadian clock through stabilizing *TOC1/PRR1* but without directly interacting with DNA through the CCT domain (Para et al., 2007). However, the soybean genome preserved functional alleles of *PRR3* with intact EAR motif and CCT domain (Li C. et al., 2020; Lu et al., 2020), making soybean a good model for the study of this clock component.

A recent investigation of the growth period QTLs on chromosomes 11 (*Gp11*) and 12 (*Gp12*) using a cultivated-wild soybean recombinant inbred lines (RIL) population identified major-effect mutations in *GmPRR3a* and *GmPRR3b* (Li M. W. et al., 2019). A frameshift mutation was found in *GmPRR3a* (*GmPRR3a*^{C08}) and a nonsense mutation in *GmPRR3b* in the cultivated parent C08 (*GmPRR3b*^{C08}) by comparing the reference genome of the soybean cultivar Williams 82 to that of the wild soybean parent W05 (*GmPRR3a*^{W05} and *GmPRR3b*^{W05}) (Li M. W. et al., 2019). Both mutations bring about a pre-mature stop codon to the coding sequence and lead to the loss of the C-terminal CCT domain in the encoded proteins (Li M. W. et al., 2019). The end result was the shortening of the growth period of the soybean plant. Analysis of resequencing data suggested that the mutations are almost fixed in the improved cultivars, inferring that the mutations were strongly selected for and played important roles in domestication or during improvement (Li M. W. et al., 2019).

Genome-wide association mapping using 279 landraces discovered 16 flowering time and maturation time QTLs (Li C. et al., 2020). Linkage disequilibrium block analysis narrowed down the QTL on chromosome 12 to a single gene, *GmPRR3b*, confirming it to be the causal gene of the maturation and flowering time phenotypes (Li C. et al., 2020). In the same study, eight haplotypes of *GmPRR3b* were discovered (H1–H8). H1 has a stop-gain near the start codon (C43T), hence appearing to be a complete knockout of the *GmPRR3b* gene. H2, H3, and H5 have non-synonymous single-nucleotide polymorphisms (SNPs) outside of the conserved domains, which are assumed to be fully functional. H4 is the full-length *GmPRR3b* that is the same as *GmPRR3b*^{W05} and is the dominant allele in wild soybean. H6 is the sole haplotype in the elite cultivar which has an ochre mutation and the loss of the CCT domain (equivalent to *GmPRR3b*^{C08}). Both H7 and H8 have a non-synonymous mutation in the PR domain (Li C. et al., 2020). Studies on the flowering time of landraces carrying the H4–H8 haplotypes showed that alleles with mutations in the PR domain (H7 and H8) or with a truncated CCT domain (H6) have an early flowering phenotype (Li C. et al., 2020).

A more recent genome-wide association mapping using a panel of 424 soybean accessions also identified *GmPRR3a* and *GmPRR3b* in the QTLs *Tof11* and *Tof12* as the major causal genes controlling flowering and maturity (Lu et al., 2020). Molecular dating suggested that both alleles of *GmPRR3* in *tof11* and *tof12* lost their CCT domain ~8,000 and ~10,500 years ago, respectively (Lu et al., 2020). While only a tiny fraction of the

tof11 allele occurred alone, it is proposed that the *tof11* allele arose from the *tof12* background (Lu et al., 2020).

Based on the above studies, it is agreed that the domestication has selected for the alleles of *GmPRR3* without the CCT domain, leading to the early flowering phenotype and possibly altering the plant architecture related to yield (Figure 1) (Li M. W. et al., 2019; Li C. et al., 2020; Lu et al., 2020). *GmPRR3*s without the CCT domain probably promote flowering through derepressing the flower-promoting *Florigen T* gene, *GmFT2a*, under long-day conditions. It is interesting that a null allele with an 8.6% allele frequency in the wild soybean population was not selected (Li C. et al., 2020). This null allele was completely wiped out in the

landrace (Li C. et al., 2020). It is highly possible that a complete loss of *GmPRR3* functions may be detrimental. Knocking out the full-length *GmPRR3b* in the ZGDD background (equivalent to *GmPRR3b^{H4}*) by CRISPR/cas9 led to a significant early flowering phenotype under long-day conditions (Wang L. et al., 2020). However, knocking out the *GmPRR3b^{C08}/GmPRR3b^{H6}/tof12* allele in the Jack and Tianlong backgrounds did not result in any significant change in flowering time (Wang L. et al., 2020) and a late-flowering phenotype with diminished yield components (Li C. et al., 2020), respectively. Furthermore, the overexpression of full-length *GmPRR3b* in Williams 82 generated a late-flowering phenotype (Lu et al., 2020), and the overexpression of

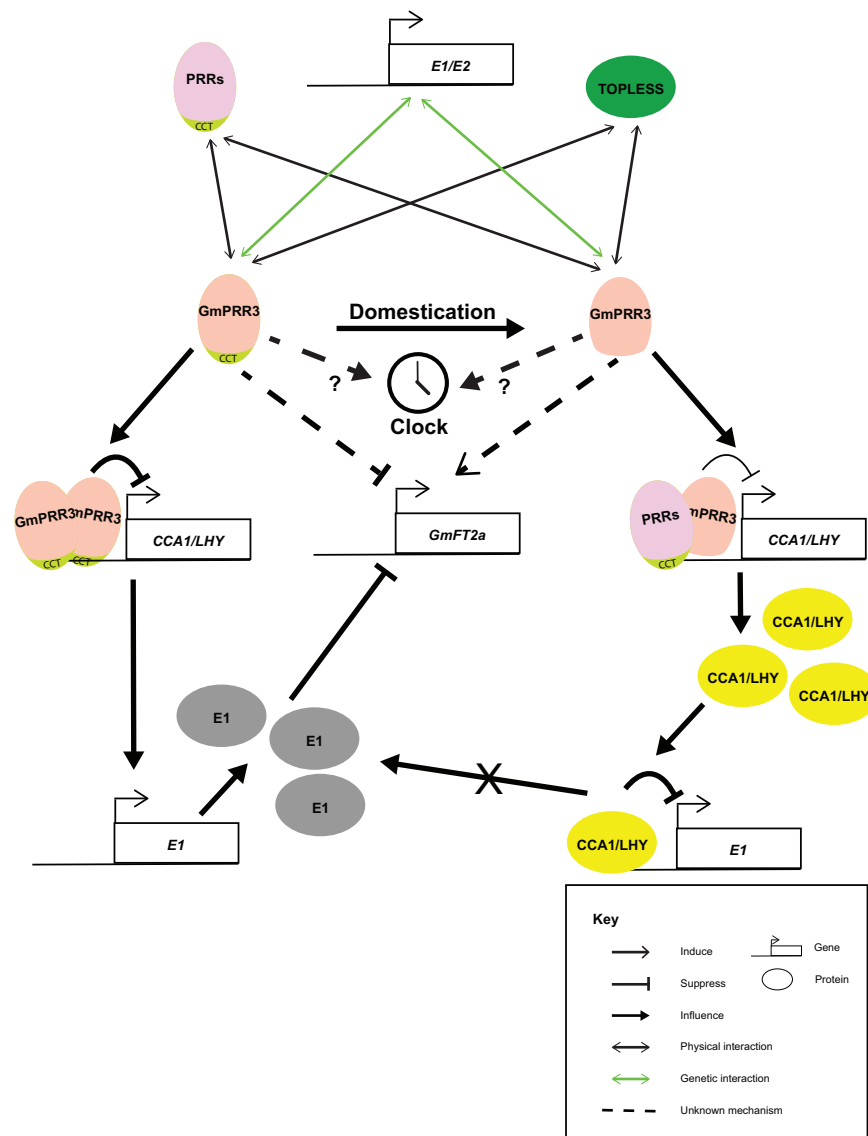


FIGURE 1 | Cartoon summarizing the known interactions of *GmPRR3a* and *GmPRR3b*. *GmPRR3*s are genetically interacting with *E1* or *E2* loci in controlling flowering time and maturation of soybean probably through the regulation of *GmFT2a* expression. During domestication and soybean improvement, alleles of *GmPRR3*s without CCT domain were selected. *GmPRR3*s without CCT domain have weakened suppression of the expression of *CCA1/LHY* genes. The interaction of *GmPRR3*s without CCT domain may rely on other PRR proteins. Thus, the derepression of *CCA1/LHY* will in turn suppress the expression of *E1*. The suppression of *E1* maybe one of the routes that *GmPRR3*s without CCT domain derepress *GmFT2a*.

GmPRR3b^{C08}/*GmPRR3b*^{H6}/*tof12* in Tianlong (Li C. et al., 2020) also produced a late-flowering phenotype and improved yield components. These observations suggest that the loss of the CCT domain in *GmPRR3* in domesticated soybean might have gained new functions for the protein that are related to yield and plant architect which is not present in the null allele.

The actual link between *GmPRR3* and flowering time is largely unknown. There are also some controversies with the existing models. Expressions of *GmPRR3a* and *GmPRR3b* cycle robustly in both long-day and short-day conditions (Li M. W. et al., 2019; Li C. et al., 2020) similar to the other *PRR* genes. The expressions peak at the middle of the day no matter whether it was a long-day (the 8th hour in a 16/8 hour light/dark cycle) or a short-day regime (the 4th hour in a 8/16 hour light/dark cycle) (Li M. W. et al., 2019; Li C. et al., 2020). The amplitude of oscillation seems to be slightly affected by the *E3* and *E4* loci (Lu et al., 2020). It has been shown that *GmPRR3* may be genetically interacting exclusively with either *E1* (Lu et al., 2020) or *E2* (Li M. W. et al., 2019), while transcriptomic network analyses suggested that *GmPRR3* may interact with both (Pan et al., 2018). Overexpressed *GmPRR3b*^{H6}, the allele without a CCT domain, was shown to bind to the *GmCCA1a* promoter and dampen the expression of *GmCCA1a* (Li C. et al., 2020), suggesting that *GmPRR3b* is playing a direct role in the soybean circadian clock. However, the CCT domain was known to be essential for *PRR* proteins to directly interact with DNA (Gendron et al., 2012). Thus, the interaction of *GmPRR3b*^{H6} with DNA may depend on its interactions with other *PRR* proteins or TOPLESS-like transcription factors (Li M. W. et al., 2019). On the other hand, the overexpression of *GmPRR3b*^{H4}, the allele with a CCT domain, resulted in stronger suppression of *GmCCA1a* expression (Li C. et al., 2020), which could be the result of direct interactions with the *GmCCA1a* promoter. *Tof11* was also shown to be potentially able to bind to the promoters of *LHY/CCA1s* and suppress their expressions (Lu et al., 2020). Thus, Lu et al. (2020) proposed that *GmPRR3* with a functional CCT domain suppresses *GmCCA1/LHY* expression, which leads to the derepression of *E1*. The high *E1* level in turn suppresses *GmFT* expressions, thus delaying flowering. Therefore, the mutated version of *GmPRR3* without a CCT domain will lead to a higher expression of *GmCCA1/LHY* and the subsequent early flowering phenotype (Lu et al., 2020). However, this was soon questioned by the observation of Li et al. as the complete knockout of *GmPRR3b*^{H6} had led to higher expressions of *GmCCA1a* and *GmFT2a* and a lower expression of *E1* (Li C. et al., 2020). Although this expression pattern is consistent with the assumption that *GmPRR3b* represses the expression of *LHY/CCA1s*, which act as repressors for *E1* (Lu et al., 2020), but the lower *E1* expression is not tally with the observed late-flowering phenotype (Li C. et al., 2020). Two possibilities on the delayed flowering were proposed. One is the involvement of pathway that is not mediated by *CCA1/LHY-E1-FT2a* module and the other is the consequence of alternated growth of *PRR3b* mutant (Li C. et al., 2020). The first possibility could be likely as *E1* is a legume specific protein, while rice *OsPRR37*, a far homolog of *GmPRR3*, can influence the expression of *Hd3a(FT)* in the absent of *E1* like protein (Matsubara et al., 2014). More

delicate genetic studies will be needed to resolve these obstacles. Furthermore, studies suggested that *GmPRR3* mainly regulated the expression of *GmFT2a* but not *GmFT5a* (Li M. W. et al., 2019; Li C. et al., 2020), which resembled the regulation of *E2* (Lu et al., 2020). Nevertheless, *GmFT5a* may also be a target under certain circumstances. Therefore, further investigations into the interactions between *GmPRR3* and *GmFT5a* is required to solve this puzzle. Actually, the rice *OsPRR37* can either promote or suppress flowering depending on the genetic background (Zhang et al., 2019). Thus, controversies of *GmPRR3* functions in previous studies were due to the inadequate study of these genes.

Live Long in the Tropics With Alterations in the *J* Locus

Soybean is a strict short-day species. Soybean plant normally requires staying vegetative for around 2 months to accumulate enough biomass to generate reasonable yield. Yet, a short-day photoperiod will trigger premature flowering of soybean plants and eventually result in low yield (Board and Hall, 1984). Thus, seasonal and latitudinal shifting of day-length have limited the growing season and acreage of soybean cultivation. However, according to model prediction, it is likely that more soybean will be grown in the tropics in 2100 (Fodor et al., 2017) where daylength is inductive for flowering in soybean.

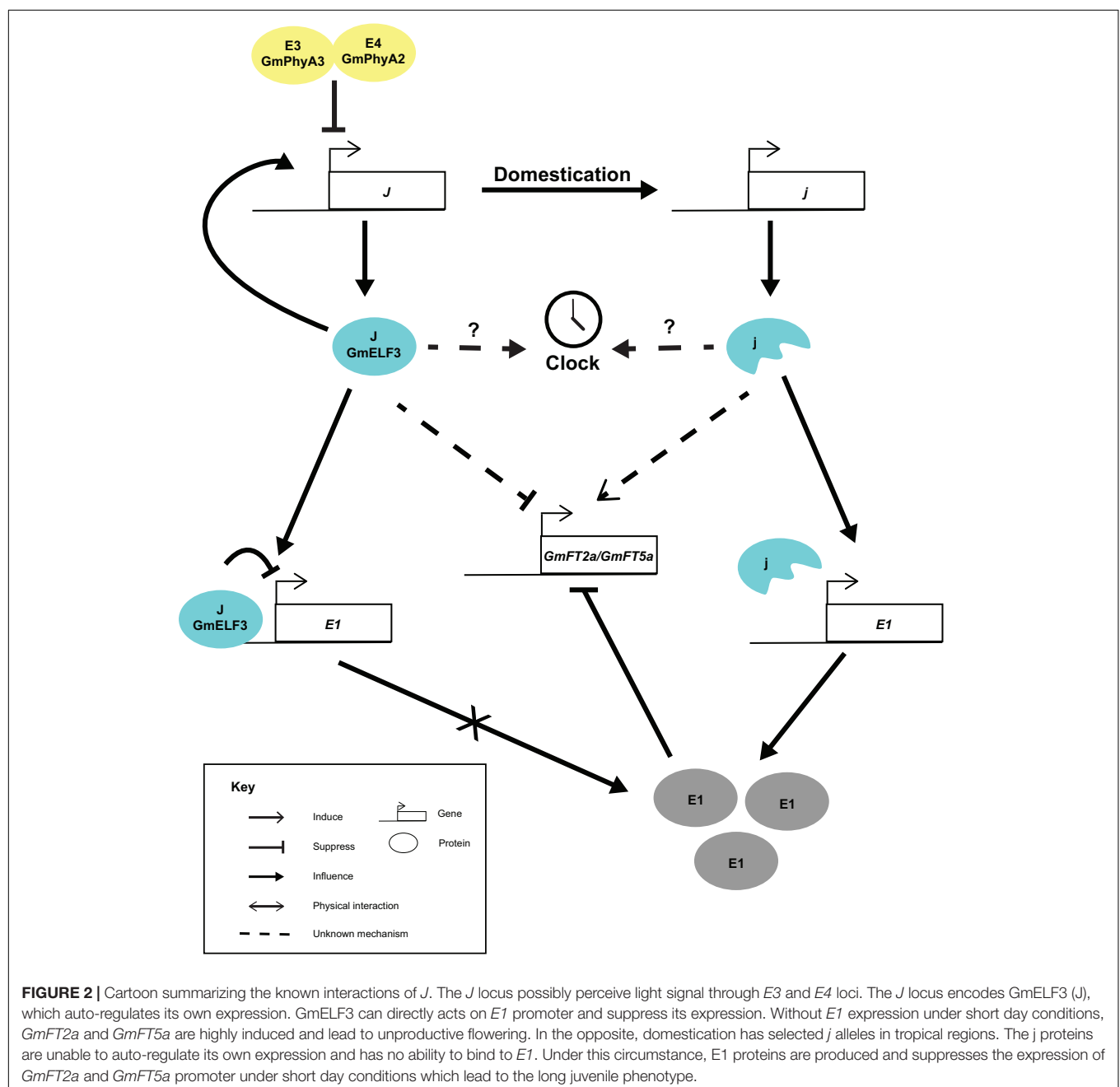
Some late-flowering soybean cultivars in tropical regions were described as early as 1970s (Hartwig, 1970; Hartwig and Kiihl, 1979), especially in Brazil. These varieties could maintain a long juvenile (LJ) phenotype under short-day conditions, which enabled them to generate adequate seeds in the scale suitable for agricultural production (Destro et al., 2001; Carpentieri-Pipolo et al., 2002). The *E6* (Bonato and Vello, 1999) and *J* (Ray et al., 1995) loci were independently identified to control the LJ phenotype and were later found to be possibly closely linked genes (Cober, 2011; Li et al., 2017). Although there are other potential loci controlling the LJ phenotype (Destro et al., 2001; Carpentieri-Pipolo et al., 2002), the *J* locus is the best characterized.

Study of the *J* locus using a RIL population from a conventional juvenile variety, Zhonghuang 24, and an LJ variety, Huaxia 3, narrowed the causal gene to an indel (AT > A) in Glyma.04G050200 (Yue et al., 2017). Glyma.04G050200 (*GmELF3*) encodes a homolog of the Arabidopsis Early Flowering 3 (AtELF3). The deletion of the nucleotide (AT > A) results in truncation of the protein corresponding to the recessive *j* in the LJ variety (Yue et al., 2017). Constitutive expression of the full-length *GmELF3* promoted flowering in transgenic plants in the LJ parent, Huaxia 3, background (Yue et al., 2017). Another study identified a 10-bp deletion and a deletion of a cytosine in the *ELF3* of BR121 and PI 159925, respectively, to be the causal mutations of the LJ phenotype in these varieties (Lu et al., 2017).

GmELF3 shows a circadian expression pattern and peaks at dusk (Lu et al., 2017), resembling the expression pattern of *AtELF3*. Interestingly, the loss-of-function allele of *GmELF3* did not oscillate diurnally, suggesting that an intact *GmELF3* *per se* is essential to sustain its own expression pattern (Lu et al., 2017). Nevertheless, the effect of this *GmELF3* mutation on soybean

circadian rhythm is not yet characterized. Like other *E* loci which are inter-dependent, the expression of *GmELF3* is slightly suppressed by *E3E4* (*GmPHYA3* and *GmPHYA2*) compared to the *e3e4* lines in the Harosoy background (Figure 2) (Lu et al., 2017). Yet, it is not known whether the higher expression of *GmELF3* in the *e3e4* background was due to the alteration in light input to the clock or due to a direct effect of the loss of *GmPHYA* functions. On the other hand, the functions of the *J* locus has been shown genetically to rely on *E1* (Lu et al., 2017). Only an intact *GmELF3* and not the truncated protein can physically interact with the promoter of *E1*, probably as part of the evening complex (Lu et al., 2017). The recessive *j* allele was thought to

suppress *FT* expression through *E1*, as the expression of *E1* was derepressed in the near-isogenic line having a recessive *j* (*NIL-j*) with reference to *NIL-J* under inductive short-day conditions (Lu et al., 2017). Thus, although *j* allele was able to delay flowering in the *e1^{as}* (a weak allele of *E1*) background, the effect was less prominent than in the *E1* background (Lu et al., 2017). But still, it is possible that *GmELF3* can alter flowering through alternative pathways. Taking rice *Hd17*(*OsELF3*) as an example. *Hd17* mediates the expression of *OsPRRs*, *OsLHY*, *OsGI*, *Ghd7*, and *Hd1*(*CO*), which are acting in multiple pathways that affect *Hd3a*(*FT*) and flowering (Zhao et al., 2012; Yang et al., 2013; Matsubara et al., 2014).



Unlike other genes artificially selected, *J* is not widespread among soybean cultivars. Rather, *J* is tailored for soybean cultivation in tropical regions far away from the temperate origin of wild soybean (Lu et al., 2017). For instance, the Huaxia 3 allele (AT > A) of *J* can only be found in southern China (Yue et al., 2017) and eight other loss-of-function alleles (*j-1* to *j-8*) were all found in low-latitude regions (Lu et al., 2017). One of these alleles was the result of induced mutation (Lu et al., 2017), which unintentionally mutated a core clock gene.

There is trade-off for having the recessive *j* alleles. A recent study has found that near isogenic lines with a recessive *j* were more sensitive to salt stress compared to those having the dominant *J* (Cheng et al., 2020). Transient overexpression of *J* in soybean hairy roots produced composite plants with higher salt tolerance, as expected (Cheng et al., 2020). A transcriptomic study comparing NIL-*j* and NIL-*J* showed that NIL-*j* had significantly lower expressions of ~95 stress-related transcription factor genes (Cheng et al., 2020). Higher expressions of *GmWRKY12*, *GmWRKY27*, *GmWRKY54*, *GmNAC11*, and *GmSIN1* were confirmed in *J*-overexpressing transgenic hairy roots (Cheng et al., 2020). It is not known whether *J* targeted the promoters of these genes directly, or if *J* altered the circadian clock, which led to different stress responses (Park et al., 2016; Coyne et al., 2019).

Gigantea Is a Big Target for Modification During Domestication and Improvement

Gigantea (GI) is not directly involved in the circadian clock transcription feedback loop in the model plant, Arabidopsis. However, it has been found playing crucial roles in clock functions, seasonal flowering, and many other important biological processes (Mishra and Panigrahi, 2015). In brief, GI assisted the maturation and accumulation of ZEITLUPE (ZTL), an F-box E3 ligase that is responsible for the degradation of clock components including TOC1/PRR1, PRR5 and CCA1 HIKING EXPEDITION (CHE) (Cha et al., 2017; Lee et al., 2019), so as to derepress *LHY/CCA1* and bring about the progression of the circadian rhythm. On the other hand, under inductive long-day conditions, the diurnal expression of Arabidopsis *GI* coincides with the circadian-controlled *FLAVIN-BINDING, KELCH REPEAT, F-BOX 1 (FKF1)* (Mishra and Panigrahi, 2015). The GI-FKF1 thus is able to form a complex to degrade the flowering suppressors, cycling DOF factors (CDFs), allowing the transcription of *CONSTANTS (CO)*. CO will then promote the expression of *FT* and result in flowering (Mishra and Panigrahi, 2015).

There are three copies of *GI* in the soybean genome, *GmGla*, *GmGlb*, and *GmGlc*. *GmGla* (Glyma.10g221500) was known as *E2* (Watanabe et al., 2011), which played major roles in the flowering and maturation of soybean (Wang Y. et al., 2016). The nucleotide diversity of *Glb* and *Glc* were not different between wild and cultivated soybeans (Wang Y. et al., 2016), suggesting that they were not targets for human selection. Analyses of 104 wild soybeans and 203 Chinese landraces identified 47 haplotypes (H1–H47) of *GmGla* in total. Due to artificial selection, there were only three haplotypes (H1–H3) of *GmGla* in the 203

landraces. Furthermore, the *GmGla* in cultivated soybeans retained only 4.7% nucleotide diversity and 2.9% nucleotide polymorphism compared to 66% nucleotide diversity and 49% nucleotide polymorphism of the entire genome, respectively (Hyten et al., 2006; Wang Y. et al., 2016), implying a strong bottleneck in this gene.

H2 and H3 encode the full-length, 1,177 (or 1,170; Watanabe et al., 2011) amino acids of the *GmGla* protein and are different in a single amino acid substitution from V₂₂₀ to I₂₂₀ (Wang Y. et al., 2016). The substitution appeared to have no significant effect on the GI function on flowering. H1 (*e2*) has a 66.95% frequency in the Chinese landraces compared to 4.81% in wild soybeans. Due to a premature stop codon in the 10th exon, H1 encodes a protein with only 527 (or 521; Watanabe et al., 2011) amino acids and contributes to the early flowering and maturation phenotype (Wang Y. et al., 2016). So far, no Korean or Japanese wild soybean bearing the H1 haplotype has been found (Wang Y. et al., 2016). Therefore, it is proposed that H1 (*e2*) originated from China and was later introduced into the other East Asian regions (Kim et al., 2018). On the other hand, another haplotype with a premature stop codon in the 2nd exon has been specifically described in Korean early flowering varieties. This suggests the mutation arose independently and only spread locally (Kim et al., 2018).

Although *GmGla* is well known as a gene controlling flowering and maturity, with the discrepancies between the short-day soybean and the long-day Arabidopsis, the actual molecular mechanism of how *GmGla* regulates flowering and maturation is largely unknown. It is believed that the function of *GmGla* has diverged from the GIs in other plant species. For instance, the full-length *GmGla* serves as a flowering suppressor while its Arabidopsis counterpart acts as a flowering promoter (Watanabe et al., 2011). Interestingly, full-length *OsGI* from rice, a short-day monocot, serves as flowering suppressor (Hayama et al., 2003). Seemingly, GI functions in long-day and short-day plants are different. The expression of *GmGla* was found to be correlated with the expression of *GmFT2a* but not that of *GmFT5a* under natural light conditions (Figure 3) (Watanabe et al., 2011). Yet, the mechanistic link between *GmGla* and *GmFT2a* expressions is still missing. The expression of *GmGla* is suppressed in the *e3e4* background or by the overexpression of *GmCOL1a* (Cao et al., 2015). A combination of *E2* and *E3* alleles can synergistically improve seed yield in July sowing in Japan (Kawasaki et al., 2018). Under long-day conditions, the expression of *GmGla* oscillates following a circadian rhythm but is slightly out of sync with the expressions of the two *GmFKF1* genes (Li et al., 2013), while under short-day conditions, the expression of *GmGla* is dampened and no longer oscillates (Li et al., 2013). This is something different from the expression of *AtGI* (Sawa et al., 2008) and *OsGI* (Hayama et al., 2003). Thus, the GI-FKF1 coincident model in Arabidopsis cannot be directly applied to soybean. Ectopic expressions of H2 and H3 in Arabidopsis delayed flowering while only the constitutive expression of H1 (a truncated *GmGla*), but not those of H2 and H3, can rescue the early flowering phenotype of the *gi-2* Arabidopsis mutant (Wang Y. et al., 2016). This may be explained by the fact that *GmFKF1* could only interact with the N-terminus but not the full-length *GmGla* in yeast two-hybrid experiments

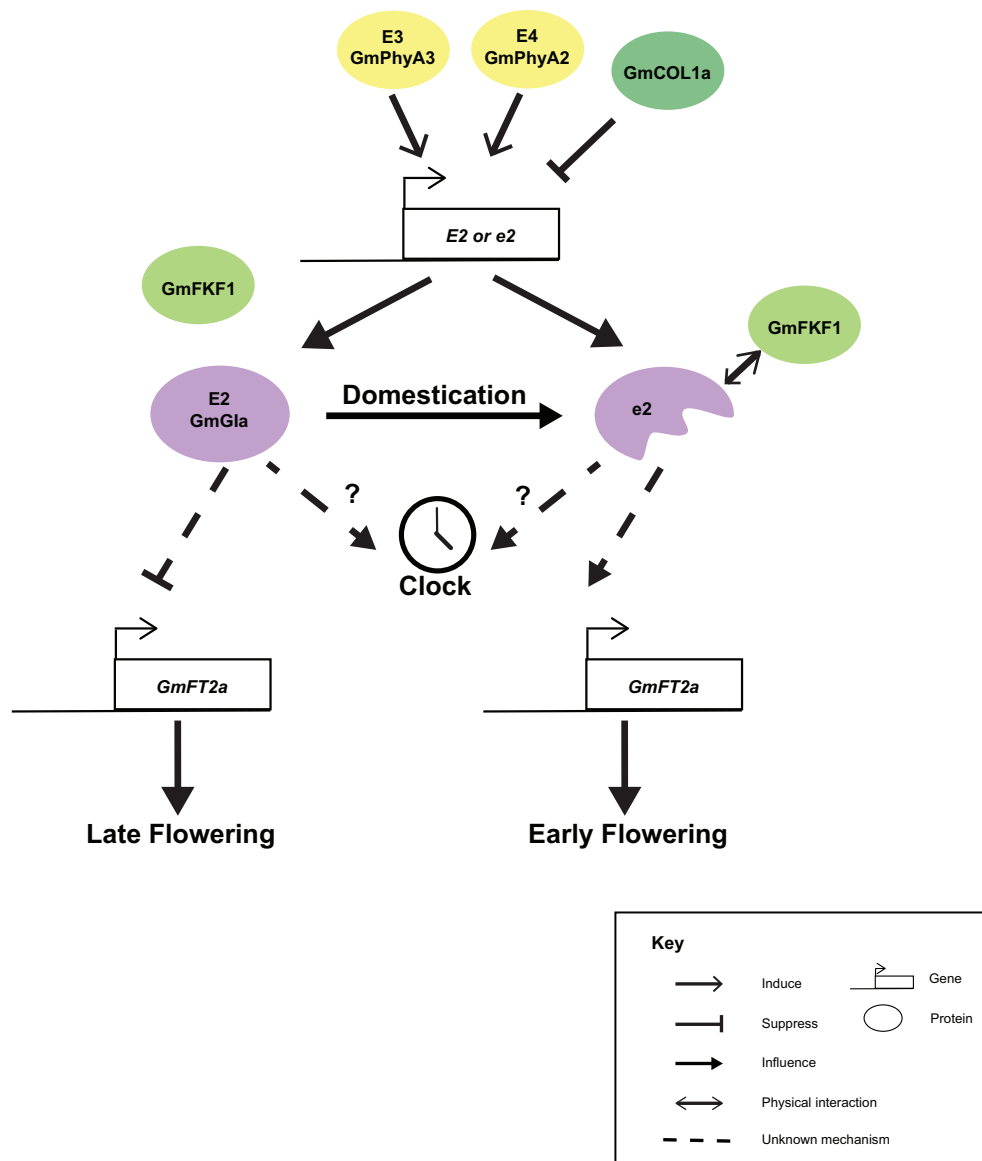


FIGURE 3 | Cartoon summarizing the known interactions of *GmGla*. Expression of *GmGla* (*E2*) is controlled by *E3*, *E4* loci and *GmCOL1a*. During domestication, a truncated *GmGla* gene (*e2*) was selected. The truncated protein but not the full-length protein can interact with *GmFKF1* in yeast 2 hybrid assay. The full-length *GmGla* (*E2*) protein can suppress the expression of *GmFT2a* and leading late flowering phenotype. In the opposite, *GmFT2a* is derepressed in *e2* background allowing early flowering phenotype.

(Li et al., 2013). It is possible that the Arabidopsis FKF1 can only be stabilized by the truncated but not the full-length *GmGla*. Thus, ectopic expressions of H2 and H3 could not rescue the late-flowering phenotype of *gi-2* (Wang Y. et al., 2016). Furthermore, *GmGla* has also been found to mediate the level of *miR172a*, which in turn regulates the stability of an *AP2/TOE1* mRNA that affects flowering time (Wang T. et al., 2016).

Mutations in *GI* in other species have been shown to affect salt stress tolerance, oxidative stress tolerance, and water use efficiency (Kurepa et al., 1998; Kim et al., 2018; Simon et al., 2020), but it is still largely unknown whether the alteration of *GmGla* during domestication has also contributed

to stress-related phenotypes in addition to the well-known effects on flowering and maturation.

DISCUSSION

The modifications of *GmPRR3*, *GmELF3*, and *GmGla* as a result of artificial selection seem to complement one another. The truncated alleles of *GmPRR3a* and *GmPRR3b* are widespread in landraces and are almost fixed in improved cultivars. This may be due to the possibility that the alteration of *GmPRR3* confers an overall advantage in growth period as well as yield. On the

other hand, the combination of different alleles of *E2* with other *E* loci partially define the maturity groups of soybean cultivars. Maturity groups are useful for selecting the appropriate cultivars at different latitudes to maximize yield. However, the effective application of *E2* is confined to the temperate and subtropical zones. In the tropics, *GmELF3* serves a similar function to *E2* for selecting the best adapted cultivars for that zone.

At the molecular level, the modifications of *GmPPR3* and *GmGla* are interesting processes. It involved more than simply altering the gene expressions or completely knocking out the gene functions because knockout alleles also exist in wild accessions. Instead, domestication has introduced new functions to these two proteins. Mutation by deletion of the CCT domain of both copies of *GmPPR3* would not have happened purely by chance, since they showed a certain degree of redundancy, only modification of both could maximize the effect (Li M. W. et al., 2019). Similar mutations or deletions of the CCT domain in *PRR37* alleles have also occurred during rice and sorghum domestication (Murphy et al., 2011; Koo et al., 2013). The plants having these CCT domain deletions in *GmPPR3* behave differently from those complete *GmPPR3*-knockout lines (Li C. et al., 2020; Wang L. et al., 2020). The loss of the CCT domain abolished the DNA-binding ability of *GmPPR3* but retained its protein–protein interaction capability. In this case, the truncated *GmPPR3* may serve as an inhibitor in protein complexes and alter their functions. In the case of *GmGla*, the truncation of the protein allows its interaction with *GmFKF1*. Interestingly, the morphologies of different *Arabidopsis gi* mutants differ drastically, implying that different parts of the *GI* protein are involved in diverse biological functions, and mutations in different parts of this protein could thus lead to quite different phenotypes (Mishra and Panigrahi, 2015). Based on this information, strategic modifications of *GmGla* by genome editing may result in other possible changes to soybean as what researchers have done on the rice *waxy* gene (Chen and Mas, 2019; Huang et al., 2020). Although we have gained some knowledge of the soybean circadian clock components, in that they have crucial impacts on growth and development, the mechanisms of their actual involvement in the clock is still largely unknown. While the clock outputs are involved in diverse biological processes, ignorance of the mechanisms of how the clock components function while conducting genetic manipulation of the crop plant may result in undesired outcomes such as the alteration of metabolic profiles and stress responses. Besides, so far, there are many more circadian clock components identified in the soybean genome than the three discussed in this review (Liu et al., 2009; Liew et al., 2017; Marcolino-Gomes et al., 2017; Li M. et al., 2019). For example, in addition to *GmPPR3a* and *GmPPR3b*, there are 12 *PRR* homolog encoding genes in the soybean genome (Li M. W. et al., 2019). Nonetheless, their roles in soybean domestication and improvement are either less prominent or the genes have experienced less significant modification.

Modification of the circadian clock during domestication and improvement was entirely serendipitous. Selection was usually based on phenotypes at macroscopic levels. Thus, targets for

modification are limited to those causing obvious phenotypic changes. However, with advanced knowledge in the circadian clock, soybean genomics, and genome editing technology, a second wave of modifications of the circadian clock components at the molecular level has become possible.

For example, isoflavones, a sub-group of flavonoids, are secondary metabolites unique to legumes and are beneficial to human either as food or nutraceuticals. Contrary to the model plant *Arabidopsis* and many other crop species, soybean and other legumes are able to fix nitrogen through a symbiotic relationship with rhizobia. The initiation of the host-symbiont interaction requires flavonoid signaling. Precursors of flavonoids are produced through the phenylpropanoid pathway, which is heavily regulated by the circadian clock at the transcriptional level (Harmer et al., 2000). Thus, well-designed genetic alterations of the circadian clock components could improve the production of these compounds, which raises the commercial values and also the nitrogen-fixing ability of soybean.

Although photosynthesis is dependent on daylength, light intensity, and temperature, the utilization of photosynthates at night is largely controlled by the clock (Graf and Smith, 2011). Since nitrogen fixation relies heavily on the availability of photosynthates supplied by the plant host, the ability of the microsymbiont to fix nitrogen could be impacted by the alteration in the circadian clock of the plant. At present, this is a completely unexplored area.

In addition to spreading soybean cultivation to lower latitudes, agriculture is also moving indoors. Advanced agricultural systems such as vertical farms or hydroponics are gaining popularity in more urban areas. Synchronization of the artificial light-dark cycle with the internal clock of the crop would be beneficial for maximizing crop production using these systems (Belbin et al., 2019). To achieve this goal, more in-depth knowledge of the circadian clock in soybean and other crop plants would be desirable.

In the past decade, since the official release of the first soybean genome assembly in 2010, identification of variants of soybean circadian clock components heavily relied on genome resequencing based on a single reference genome. Such strategy has limited power in the discovery of new genes and structural variations. With the advance of soybean genomics, more and more high-quality soybean genome assemblies are available (Shen et al., 2018; Valliyodan et al., 2019; Xie et al., 2019). A high-quality pan-genome of wild and cultivated soybeans are also reported recently (Liu et al., 2020). These genomes allow direct comparison of sequence of clock components from different soybean accessions which resulted in more accurate discovery of variants. Furthermore, better reference genomes also facilitate epigenetic study. It has been demonstrated that histone modifications have played crucial roles in regulating the core clock components in plant (reviewed in Chen and Mas, 2019; Du et al., 2019). Nevertheless, information of the roles of epigenetic on soybean domestication and soybean circadian clock function is scarce. Thus, exploration of the modification of circadian clock components and their regulation at epigenetic level could be the next chapter of soybean research.

AUTHOR CONTRIBUTIONS

M-WL drafted the manuscript. M-WL and H-ML edited and proofread the final manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Hong Kong Research Grants Council Area of Excellence Scheme (AoE/M-403/16), the Lo Kwee-Seong Biomedical Research Fund, and Shenzhen Virtual

University Park Support Scheme to CUHK Shenzhen Research Institute (YFJGJS1.0). Any opinions, findings, conclusions, or recommendations expressed in this publication do not reflect the views of the Government of the Hong Kong Special Administrative Region or the Innovation and Technology Commission.

ACKNOWLEDGMENTS

This manuscript was copy-edited by Jee-Yan Chu.

REFERENCES

- Alabadi, D., Oyama, T., Yanovsky, M. J., Harmon, F. G., Mas, P., and Kay, S. A. (2001). Reciprocal regulation between *TOC1* and *LHY/CCA1* within the *Arabidopsis* circadian clock. *Science* 293, 880–883. doi: 10.1126/science.1061320
- Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K. M., et al. (2016). 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166, 481–491.
- Belbin, F. E., Fraser, D. P., Comben, N. E., and Dodd, A. N. (2019). “Plant circadian rhythms and vertical agriculture,” in *Plant Factory Using Artificial Light: Adapting to Environmental Disruption and Clues to Agricultural Innovation*, eds M. Anpo, H. Fukuda, and T. Wada (Amsterdam: Elsevier), 79–88. doi: 10.1016/b978-0-12-813973-8.00007-5
- Board, J. E., and Hall, W. (1984). Premature flowering in soybean yield reductions at nonoptimal planting dates as influenced by temperature and photoperiod. *Agron. J.* 76, 700–704. doi: 10.2134/agronj1984.00021962007600040043x
- Bonato, E. R., and Vello, N. A. (1999). *E6*, a dominant gene conditioning early flowering and maturity in soybeans. *Genet. Mol. Biol.* 22, 229–232. doi: 10.1590/s1415-47571999000200016
- Brest, D. E. (1970). Circadian leaf movements in Biloxi soybeans. *Plant Physiol.* 45, 647–648. doi: 10.1104/pp.45.5.647
- Bunning, E. (1954). Der verlauf der endogenen tagersrhythmik bei photoperiodischen storlicht-versuchen mit soja. *Physiol. Plant* 7, 538–547. doi: 10.1111/j.1399-3054.1954.tb07599.x
- Cao, D., Li, Y., Lu, S. J., Wang, J. L., Nan, H. Y., Li, X. M., et al. (2015). *GmCOL1a* and *GmCOL1b* function as flowering repressors in soybean under long-day conditions. *Plant Cell Physiol.* 56, 2409–2422. doi: 10.1093/pcp/pcv152
- Carpentieri-Pipolo, V., De Almeida, L. A., and Kihl, R. A. D. (2002). Inheritance of a long juvenile period under short-day conditions in soybean. *Genet. Mol. Biol.* 25, 463–469. doi: 10.1590/s1415-47572002000400016
- Cha, J. Y., Kim, J., Kim, T. S., Zeng, Q., Wang, L., Lee, S. Y., et al. (2017). GIGANTEA is a co-chaperone which facilitates maturation of ZEITLUPE in the *Arabidopsis* circadian clock. *Nat. Commun.* 8:3.
- Chen, Z. J., and Mas, P. (2019). Interactive roles of chromatin regulation and circadian clock function in plants. *Genome Biol.* 20:62.
- Cheng, Q., Gan, Z. R., Wang, Y. P., Lu, S. J., Hou, Z. H., Li, H. Y., et al. (2020). The soybean gene *J* contributes to salt stress tolerance by up-regulating salt-responsive genes. *Front. Plant Sci.* 11:272. doi: 10.3389/fpls.2020.00272
- Cober, E. R. (2011). Long juvenile soybean flowering responses under very short photoperiods. *Crop Sci.* 51, 140–145. doi: 10.2135/cropsci2010.05.0262
- Cober, E. R., Molnar, S. J., Charette, M., and Voldeng, H. D. (2010). A new locus for early maturity in soybean. *Crop Sci.* 50, 524–527. doi: 10.2135/cropsci2009.04.0174
- Cober, E. R., and Voldeng, H. D. (2001). A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*. *Crop Sci.* 41, 698–701. doi: 10.2135/cropsci2001.413698x
- Covington, M. F., Maloof, J. N., Straume, M., Kay, S. A., and Harmer, S. L. (2008). Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biol.* 9:R130. doi: 10.1186/gb-2008-1189-1188-r1130
- Coyne, K., Davis, M. M., Mizoguchi, T., and Hayama, R. (2019). Temporal restriction of salt inducibility in expression of salinity-stress related gene by the circadian clock in *Solanum lycopersicum*. *Plant Biotechnol.* 36, 195–200. doi: 10.5511/plantbiotechnology.19.0703a
- Destro, D., Carpentieri-Pipolo, V., Afonso, R., Kihl, S., and Almeida, L. (2001). Photoperiodism and genetic control of the long juvenile period in soybean: a review. *Crop Breed. Appl. Biotechnol.* 1, 72–92. doi: 10.13082/11984-17033.v13001n13001a13010
- Dissanayaka, A., Rodriguez, T. O., Di, S., Yan, F., Githiri, S. M., Rodas, F. R., et al. (2016). Quantitative trait locus mapping of soybean maturity gene *E5*. *Breed. Sci.* 66, 407–415. doi: 10.1270/jsbbs.15160
- Dodd, A. N., Bebin, F. E., Frank, A., and Webb, A. A. R. (2015). Interactions between circadian clocks and photosynthesis for the temporal and spatial coordination of metabolism. *Front. Plant Sci.* 6:245. doi: 10.3389/fpls.2015.00245
- Dodd, A. N., Salathia, N., Hall, A., Kevei, E., Toth, R., Nagy, F., et al. (2005). Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* 309, 630–633. doi: 10.1126/science.1115581
- Du, S. X., Chen, L., Ge, L. F., and Huang, W. (2019). A novel loop: mutual regulation between epigenetic modification and the circadian clock. *Front. Plant Sci.* 10:22. doi: 10.3389/fpls.2019.00022
- Fang, C., Ma, Y. M., Wu, S. W., Liu, Z., Wang, Z., Yang, R., et al. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* 18:161.
- Faure, S., Turner, A. S., Gruszka, D., Christodoulou, V., Davis, S. J., Von Korff, M., et al. (2012). Mutation at the circadian clock gene *EARLY MATURITY 8* adapts domesticated barley (*Hordeum vulgare*) to short growing seasons. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8328–8333. doi: 10.1073/pnas.1120496109
- Fodor, N., Challinor, A., Droukas, I., Ramirez-Villegas, J., Zabel, F., Koehler, A. K., et al. (2017). Integrating plant science and crop modeling: assessment of the impact of climate change on soybean and maize production. *Plant Cell Physiol.* 58, 1833–1847. doi: 10.1093/pcp/pcx141
- Foo, M., Somers, D. E., and Kim, P. J. (2016). Kernel architecture of the genetic circuitry of the *Arabidopsis* circadian system. *PLoS Comput. Biol.* 12:e1004748. doi: 10.1371/journal.pcbi.1004748
- Fukushima, A., Kusano, M., Nakamichi, N., Kobayashi, M., Hayashi, N., Sakakibara, H., et al. (2009). Impact of clock-associated *Arabidopsis* pseudo-response regulators in metabolic coordination. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7251–7256. doi: 10.1073/pnas.0900952106
- Gawronski, P., Ariyadasa, R., Himmelbach, A., Poursarebani, N., Kilian, B., Stein, N., et al. (2014). A distorted circadian clock causes early flowering and temperature-dependent variation in spike development in the *Eps-3Am* mutant of einkorn wheat. *Genetics* 196, 1253–1261. doi: 10.1534/genetics.113.158444
- Gendron, J. M., Prunedo-Paz, J. L., Doherty, C. J., Gross, A. M., Kang, S. E., and Kay, S. A. (2012). *Arabidopsis* circadian clock protein, *TOC1*, is a DNA-binding transcription factor. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3167–3172.
- Gil, K. E., and Park, C. M. (2019). Thermal adaptation and plasticity of the plant circadian clock. *New Phytol.* 221, 1215–1229. doi: 10.1111/nph.15518
- Graf, A., and Smith, A. M. (2011). Starch and the clock: the dark side of plant productivity. *Trends Plant Sci.* 16, 169–175. doi: 10.1016/j.tplants.2010.12.003
- Greenham, K., Lou, P., Puzey, J. R., Kumar, G., Arnevik, C., Farid, H., et al. (2017). Geographic variation of plant circadian clock function in

- natural and agricultural settings. *J. Biol. Rhythms* 32, 26–34. doi: 10.1177/0748730416679307
- Harmer, S. L., Hogenesch, L. B., Straume, M., Chang, H. S., Han, B., Zhu, T., et al. (2000). Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* 290, 2110–2113. doi: 10.1126/science.290.5499.2110
- Hartwig, E. E. (1970). Growth and reproductive characteristics of soybeans [*Glycine max* (L) Merr.] grown under short-day conditions. *Trop. Sci.* 12, 47–53.
- Hartwig, E. E., and Kihl, R. A. S. (1979). Identification and utilization of a delayed flowering character in soybeans for short-day conditions. *Field Crops Res.* 2, 145–151. doi: 10.1016/0378-4290(79)90017-0
- Hayama, R., Yokoi, S., Tamaki, S., Yano, M., and Shimamoto, K. (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* 422, 719–722. doi: 10.1038/nature01549
- Huang, H., and Nusinow, D. A. (2016). Into the evening: complex interactions in the *Arabidopsis* circadian clock. *Trends Genet.* 32, 674–686. doi: 10.1016/j.tig.2016.08.002
- Huang, L. C., Li, Q. F., Zhang, C. Q., Chu, R., Gu, Z. W., Tan, H. Y., et al. (2020). Creating novel *Wx* alleles with fine-tuned amylose levels and improved grain quality in rice by promoter editing using CRISPR/Cas9 system. *Plant Biotechnol. J.* doi: 10.1111/pbi.13391 [Epub ahead of print].
- Hudson, K. A. (2010). The circadian clock-controlled transcriptome of developing soybean seeds. *Plant Genome* 3, 3–13. doi: 10.3835/plantgenome2009.08.0025
- Hyten, D. L., Song, Q. J., Zhu, Y. L., Choi, I. Y., Nelson, R. L., Costa, J. M., et al. (2006). Impacts of genetic bottlenecks on soybean genome diversity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16666–16671. doi: 10.1073/pnas.0604379103
- Jiang, B. J., Nan, H. Y., Gao, Y. F., Tang, L. L., Yue, Y. L., Lu, S. J., et al. (2014). Allelic combinations of soybean maturity loci *E1*, *E2*, *E3* and *E4* result in diversity of maturity and adaptation to different latitudes. *PLoS One* 9:e106042. doi: 10.1371/journal.pone.0106042
- Kawasaki, Y., Yamazaki, R., Katayama, K., Yamada, T., and Funatsuki, H. (2018). Effects of maturity genes *E2* and *E3* on yield formation in soybean cultivar Enrei in warm region, Fukuyama in Japan. *Plant Prod. Sci.* 21, 387–397. doi: 10.1080/1343943x.2018.1530063
- Kim, J. A., Kim, H. S., Choi, S. H., Jang, J. Y., Jeong, M. J., and Lee, S. I. (2017). The importance of the circadian clock in regulating plant metabolism. *Int. J. Mol. Sci.* 18:2680. doi: 10.3390/ijms18122680
- Kim, S. K., Kim, E. S., Kim, K. H., Jeong, N., Lee, J. S., and Kang, S. (2018). Genetic variance for flowering time conferring *E2* gene in photoperiod-insensitive early-maturing soybean accessions and topological distribution in Korea peninsula. *Mol. Breed.* 38:148. doi: 10.1007/s11032-11018-10887-x
- Koo, B. H., Yoo, S. C., Park, J. W., Kwon, C. T., Lee, B. D., An, G., et al. (2013). Natural variation in *OsPRR37* regulates heading date and contributes to rice cultivation at a wide range of latitudes. *Mol. Plant* 6, 1877–1888. doi: 10.1093/mp/sst088
- Kurepa, J., Smalle, J., Van Montagu, M., and Inze, D. (1998). Oxidative stress tolerance and longevity in *Arabidopsis*: the late-flowering mutant *gigantea* is tolerant to paraquat. *Plant J.* 14, 759–764. doi: 10.1046/j.1365-313x.1998.01068.x
- Lam, H. M., Xu, X., Liu, X., Chen, W. B., Yang, G. H., Wong, F. L., et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059. doi: 10.1038/ng.715
- Lee, C. M., Li, M. W., Feke, A., Liu, W., Saffer, A. M., and Gendron, J. M. (2019). GIGANTEA recruits the UBP12 and UBP13 deubiquitylases to regulate accumulation of the ZTL photoreceptor complex. *Nat. Commun.* 10:3750.
- Li, C., Li, Y. H., Li, Y., Lu, H., Hong, H., Tian, Y., et al. (2020). A domestication-associated gene *GmPRR3b* regulates the circadian clock and flowering time in soybean. *Mol. Plant* 13, 745–759. doi: 10.1016/j.molp.2020.01.014
- Li, Y. H., Li, D. L., Jiao, Y. Q., Schnable, J. C., Li, Y. F., Li, H. H., et al. (2020). Identification of loci controlling adaptation in Chinese soya bean landraces via a combination of conventional and bioclimatic GWAS. *Plant Biotechnol. J.* 18, 389–401. doi: 10.1111/pbi.13206
- Li, F., Zhang, X. M., Hu, R. B., Wu, F. Q., Ma, J. H., Meng, Y., et al. (2013). Identification and molecular characterization of *FKF1* and *GI* homologous genes in soybean. *PLoS One* 8:e79036. doi: 10.1371/journal.pone.0079036
- Li, M., Cao, L. J., Mwimba, M., Zhou, Y., Li, L., Zhou, M., et al. (2019). Comprehensive mapping of abiotic stress inputs into the soybean circadian clock. *Proc. Natl. Acad. Sci. U.S.A.* 116, 23840–23849. doi: 10.1073/pnas.1708508116
- Li, M. W., Liu, W., Lam, H. M., and Gendron, J. M. (2019). Characterization of two growth period QTLs reveals modification of *PRR3* genes during soybean domestication. *Plant Cell Physiol.* 60, 407–420. doi: 10.1093/pcp/pcy215
- Li, X. M., Fang, C., Xu, M. L., Zhang, F. G., Lu, S. J., Nan, H. Y., et al. (2017). Quantitative trait locus mapping of soybean maturity gene *E6*. *Crop Sci.* 57, 2547–2554. doi: 10.2135/cropsci2017.02.0106
- Li, Y., Dong, Y. S., Wu, H. Y., Hu, B., Zhai, H., Yang, J. Y., et al. (2019). Positional cloning of the flowering time QTL *qFT12-1* reveals the link between the clock related *PRR* homolog with photoperiodic response in soybeans. *Front. Plant Sci.* 10:1303. doi: 10.3389/fpls.2019.01303
- Liew, L. C., Singh, M. B., and Bhalla, P. L. (2017). A novel role of the soybean clock gene *LUX ARRHYTHMO* in male reproductive development. *Sci. Rep.* 7:10605. doi: 10.1038/s41598-017-10823-y
- Liu, B., Kanazawa, A., Matsumura, H., Takahashi, R., Harada, K., and Abe, J. (2008). Genetic redundancy in soybean photoreponses associated with duplication of the phytochrome a gene. *Genetics* 180, 995–1007. doi: 10.1534/genetics.108.092742
- Liu, H., Wang, H. G., Gao, P. F., Xu, J. H., Xu, T. D., Wang, J. S., et al. (2009). Analysis of clock gene homologs using unifoliolates as target organs in soybean (*Glycine max*). *J. Plant Physiol.* 166, 278–289. doi: 10.1016/j.jplph.2008.06.003
- Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., et al. (2020). Pan-genome of wild and cultivated soybeans. *Cell* 182, 162–176.e13.
- Lu, S., Dong, L., Fang, C., Liu, S., Kong, L., Cheng, Q., et al. (2020). Stepwise selection on homeologous *PRR* genes controlling flowering and maturity during soybean domestication. *Nat. Genet.* 52, 428–436. doi: 10.1038/s41588-020-0604-7
- Lu, S. J., Li, Y., Wang, J. L., Nan, H. Y., Cao, D., Li, X. M., et al. (2016). Identification of additional QTLs for flowering time by removing the effect of the maturity gene *E1* in soybean. *J. Integr. Agric.* 15, 42–49. doi: 10.1016/s2095-3119(15)61046-2
- Lu, S. J., Zhao, X. H., Hu, Y. L., Liu, S. L., Nan, H. Y., Li, X. M., et al. (2017). Natural variation at the soybean *J* locus improves adaptation to the tropics and enhances yield. *Nat. Genet.* 49, 773–779. doi: 10.1038/ng.3819
- Marcolino-Gomes, J., Nakayama, T. J., Molinari, H. B. C., Basso, M. F., Henning, L. M. M., Fuganti-Pagliarini, R., et al. (2017). Functional characterization of a putative *Glycine max* *ELF4* in transgenic *Arabidopsis* and its role during flowering control. *Front. Plant Sci.* 8:618. doi: 10.3389/fpls.2017.00618
- Matsubara, K., Hori, K., Ogiso-Tanaka, E., and Yano, M. (2014). Cloning of quantitative trait genes from rice reveals conservation and divergence of photoperiod flowering pathways in *Arabidopsis* and rice. *Front. Plant Sci.* 5:193. doi: 10.3389/fpls.2014.00193
- Miladinovic, J., Ceran, M., Dordevic, V., Balesevic-Tubic, S., Petrovic, K., Dukic, V., et al. (2018). Allelic variation and distribution of the major maturity genes in different soybean collections. *Front. Plant Sci.* 9:1286. doi: 10.3389/fpls.2018.01286
- Mishra, P., and Panigrahi, K. C. (2015). GIGANTEA - an emerging story. *Front. Plant Sci.* 6:8. doi: 10.3389/fpls.2015.00008
- Mizuno, T., and Nakamichi, N. (2005). Pseudo-response regulators (PRRs) or true oscillator components (TOCs). *Plant Cell Physiol.* 46, 677–685. doi: 10.1093/pcp/pci087
- Muller, N. A., Wijnen, C. L., Srinivasan, A., Ryngajlo, M., Ofner, I., Lin, T., et al. (2016). Domestication selected for deceleration of the circadian clock in cultivated tomato. *Nat. Genet.* 48, 89–93. doi: 10.1038/ng.3447
- Murphy, R. L., Klein, R. R., Morishige, D. T., Brady, J. A., Rooney, W. L., Miller, F. R., et al. (2011). Coincident light and clock regulation of pseudoreponse regulator protein 37 (*PRR37*) controls photoperiodic flowering in sorghum. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16469–16474. doi: 10.1073/pnas.1106212108
- Nakamichi, N., Takao, S., Kudo, T., Kiba, T., Wang, Y., Kinoshita, T., et al. (2016). Improvement of *Arabidopsis* biomass and cold, drought and salinity stress tolerance by modified circadian clock-associated PSEUDO-RESPONSE REGULATORs. *Plant Cell Physiol.* 57, 1085–1097. doi: 10.1093/pcp/pcw057
- Nitschke, S., Cortleven, A., Iven, T., Feussner, I., Havaux, M., Riefler, M., et al. (2016). Circadian stress regimes affect the circadian clock and cause jasmonic acid-dependent cell death in cytokinin-deficient *Arabidopsis* plants. *Plant Cell* 28, 1616–1639.
- Ogiso-Tanaka, E., Shimizu, T., Hajika, M., Kaga, A., and Ishimoto, M. (2019). Highly multiplexed AmpliSeq technology identifies novel variation of flowering time-related genes in soybean (*Glycine max*). *DNA Res.* 26, 243–260.

- Pan, L. Y., He, J. B., Zhao, T. J., Xing, G. N., Wang, Y. F., Yu, D. Y., et al. (2018). Efficient QTL detection of flowering date in a soybean RIL population using the novel restricted two-stage multi-locus GWAS procedure. *Theor. Appl. Genet.* 131, 2581–2599. doi: 10.1007/s00122-018-3174-7
- Pan, W. J., Wang, X., Deng, Y. R., Li, J. H., Chen, W., Chiang, J. Y., et al. (2015). Nondestructive and intuitive determination of circadian chlorophyll rhythms in soybean leaves using multispectral imaging. *Sci. Rep.* 5:11108.
- Para, A., Farre, E. M., Imaizumi, T., Pruneda-Paz, J. L., Harmon, F. G., and Kay, S. A. (2007). PRR3 is a vascular regulator of TOC1 stability in the *Arabidopsis* circadian clock. *Plant Cell* 19, 3462–3473. doi: 10.1105/tpc.107.054775
- Park, H. J., Qiang, Z., Kim, W. Y., and Yun, D. J. (2016). Diurnal and circadian regulation of salt tolerance in *Arabidopsis*. *J. Plant Biol.* 59, 569–578. doi: 10.1007/s12374-016-0317-8
- Qi, X. P., Li, M. W., Xie, M., Liu, X., Ni, M., Shao, G. H., et al. (2014). Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. *Nat. Commun.* 5:4340.
- Ray, J. D., Hinson, K., Mankono, J. E. B., and Malo, M. F. (1995). Genetic-control of a long-juvenile trait in soybean. *Crop Sci.* 35, 1001–1006. doi: 10.2135/cropsci1995.0011183x003500040012x
- Samanfar, B., Molnar, S. J., Charette, M., Schoenrock, A., Dehne, F., Golshani, A., et al. (2017). Mapping and identification of a potential candidate gene for a novel maturity locus, *E10*, in soybean. *Theor. Appl. Genet.* 130, 377–390. doi: 10.1007/s00122-016-2819-7
- Sawa, M., Kay, S. A., and Imaizumi, T. (2008). Photoperiodic flowering occurs under internal and external coincidence. *Plant Signal. Behav.* 3, 269–271. doi: 10.4161/psb.3.4.5219
- Shen, Y., Liu, J., Geng, H., Zhang, J., Liu, Y., Zhang, H., et al. (2018). De novo assembly of a Chinese soybean genome. *Sci. China Life Sci.* 61, 871–884.
- Shim, J. S., Kubota, A., and Imaizumi, T. (2017). Circadian clock and photoperiodic flowering in *Arabidopsis*: CONSTANS is a hub for signal integration. *Plant Physiol.* 173, 5–15. doi: 10.1104/pp.16.01327
- Shor, E., and Green, R. M. (2016). The impact of domestication on the circadian clock. *Trends Plant Sci.* 21, 281–283. doi: 10.1016/j.tplants.2016.01.022
- Simon, N. M. L., Graham, C. A., Comben, N. E., Hetherington, A. M., and Dodd, A. N. (2020). The circadian clock influences the long-term water use efficiency of *Arabidopsis*. *Plant Physiol.* 183, 317–330. doi: 10.1104/pp.20.00030
- Staiger, D., Shin, J., Johansson, M., and Davis, S. J. (2013). The circadian clock goes genomic. *Genome Biol.* 14:208.
- Valliyodan, B., Cannon, S. B., Bayer, P. E., Shu, S. Q., Brown, A. V., Ren, L. H., et al. (2019). Construction and comparison of three reference-quality genome assemblies for soybean. *Plant J.* 100, 1066–1082. doi: 10.1111/tpj.14500
- Veza, M. E., Alderete, L. G. S., Agostini, E., and Talano, M. A. (2020). Expression of circadian clock genes and diurnal oscillations of key physiological events in response to AsV and AsIII in soybean plants. *Environ. Exp. Bot.* 174:104054. doi: 10.1016/j.envexpbot.2020.104054
- Wang, L., Sun, S., Wu, T., Liu, L., Sun, X., Cai, Y., et al. (2020). Natural variation and CRISPR/Cas9-mediated mutation in *GmPRR37* affect photoperiodic flowering and contribute to regional adaptation of soybean. *Plant Biotechnol. J.* 18, 1869–1881. doi: 10.1111/pbi.13346
- Wang, T., Sun, M. Y., Wang, X. S., Li, W. B., and Li, Y. G. (2016). Over-expression of GmGla-regulated soybean *miR172a* confers early flowering in transgenic *Arabidopsis thaliana*. *Int. J. Mol. Sci.* 17:645. doi: 10.3390/ijms17050645
- Wang, Y., Gu, Y. Z., Gao, H. H., Qiu, L. J., Chang, R. Z., Chen, S. Y., et al. (2016). Molecular and geographic evolutionary support for the essential role of GIGANTEA in soybean domestication of flowering time. *BMC Evol. Biol.* 16:79. doi: 10.1186/s12862-016-0653-9
- Wang, Y., Yuan, L., Su, T., Wang, Q., Gao, Y., Zhang, S. Y., et al. (2020). Light- and temperature-entrainable circadian clock in soybean development. *Plant Cell Environ.* 43, 637–648. doi: 10.1111/pce.13678
- Watanabe, S., Hideshima, R., Xia, Z. J., Tsubokura, Y., Sato, S., Nakamoto, Y., et al. (2009). Map-based cloning of the gene associated with the soybean maturity locus *E3*. *Genetics* 182, 1251–1262. doi: 10.1534/genetics.108.098772
- Watanabe, S., Xia, Z. J., Hideshima, R., Tsubokura, Y., Sato, S., Yamanaka, N., et al. (2011). A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. *Genetics* 188, 395–407. doi: 10.1534/genetics.110.125062
- Xia, Z. J., Watanabe, S., Yamada, T., Tsubokura, Y., Nakashima, H., Zhai, H., et al. (2012). Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2155–E2164.
- Xie, M., Chung, C. Y. L., Li, M. W., Wong, F. L., Wang, X., Liu, A. L., et al. (2019). A reference-grade wild soybean genome. *Nat. Commun.* 10:1216.
- Yang, Y., Peng, Q., Chen, G. X., Li, X. H., and Wu, C. Y. (2013). OsELF3 is involved in circadian clock regulation for promoting flowering under long-day conditions in rice. *Mol. Plant* 6, 202–215. doi: 10.1093/mp/sss062
- Yue, Y. L., Liu, N. X., Jiang, B. J., Li, M., Wang, H. J., Jiang, Z., et al. (2017). A single nucleotide deletion in *J* encoding *GmELF3* confers long juvenility and is associated with adaption of tropic soybean. *Mol. Plant* 10, 656–658. doi: 10.1016/j.molp.2016.12.004
- Zhang, B., Liu, H. Y., Qi, F. X., Zhang, Z. Y., Li, Q. P., Han, Z. M., et al. (2019). Genetic interactions among *Ghd7*, *Ghd8*, *OsPRR37* and *Hd1* contribute to large variation in heading date in rice. *Rice* 12:48. doi: 10.1186/s12284-12019-10314-x
- Zhao, C., Takeshima, R., Zhu, J. H., Xu, M. L., Sato, M., Watanabe, S., et al. (2016). A recessive allele for delayed flowering at the soybean maturity locus *E9* is a leaky allele of *FT2a*, a *FLOWERING LOCUS T* ortholog. *BMC Plant Biol.* 16:20. doi: 10.1186/s12870-016-0704-9
- Zhao, J. M., Huang, X., Ouyang, X. H., Chen, W. L., Du, A. P., Zhu, L., et al. (2012). *OsELF3-1*, an ortholog of *Arabidopsis* *EARLY FLOWERING 3*, regulates rice circadian rhythm and photoperiodic flowering. *PLoS One* 7:e43705. doi: 10.1371/journal.pone.0043705
- Zhou, Z. K., Jiang, Y., Wang, Z., Gou, Z. H., Lyu, J., Li, W. Y., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33, 408–414. doi: 10.1038/nbt.3096

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

Copyright © 2020 Li and Lam. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Construction of a High-Density Genetic Map and Its Application for QTL Mapping of Leaflet Shapes in Mung Bean (*Vigna radiata* L.)

Jie Wang¹, Jianling Li^{1,2}, Zhenxing Liu³, Xingxing Yuan⁴, Suhua Wang¹, Honglin Chen¹, Xin Chen⁴, Xuzhen Cheng^{1*} and Lixia Wang^{1*}

¹ Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, ² College of Life Science, Yangtze University, Jingzhou, China, ³ Tangshan Academy of Agricultural Sciences, Tangshan, China, ⁴ Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China

OPEN ACCESS

Edited by:

Jianke Li,
Institute of Apiculture Research (CAS),
China

Reviewed by:

Beide Fu,
Institute of Hydrobiology (CAS), China
Changwei Shao,
Yellow Sea Fisheries Research
Institute (CAFS), China

*Correspondence:

Xuzhen Cheng
chengxuzhen@caas.cn
Lixia Wang
wanglixia03@caas.cn

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 19 May 2020

Accepted: 11 August 2020

Published: 30 September 2020

Citation:

Wang J, Li J, Liu Z, Yuan X,
Wang S, Chen H, Chen X, Cheng X
and Wang L (2020) Construction of a
High-Density Genetic Map and Its
Application for QTL Mapping
of Leaflet Shapes in Mung Bean
(*Vigna radiata* L.).
Front. Genet. 11:1032.
doi: 10.3389/fgene.2020.01032

Mung bean (*Vigna radiata* L.) is an important but understudied food legume in Asia and now worldwide. Genetic studies may help to accelerate the exploitation of new genes for breeding in this crop. Here, we used a recombination inbred line population to construct an SNP genetic linkage map by genome sequencing technology. We obtained 21,508 high-quality SNP markers integrated into 1,946 bin markers that were mapped onto 11 linkage groups (LGs) with 99–258 bin markers per LG. The total genetic length of the map was 1060.2 cM (38.76–168.03 cM per LG), with an average distance between markers of 0.54 cM. However, there were 18 gaps >5 cM, distribution on LG1, 3, 5, 7, and 9. Gene mapping for lobed and indented leaflets was conducted using the map. A major quantitative trait locus (QTL) associated with indented leaflets was detected on chromosome 10, with phenotypic variation explained (PVE) values of 39.7% and 45.4% under two different environments. Several QTLs for lobed leaflets were detected and most of them were tightly linked together on Chromosome 3. However, only one major QTL, which explained the largest phenotypic variation (27.7–69.5%), was stably detected under two different environments using both R and Q methods. In the two main stable QTLs regions on chromosomes 3 and 10, candidate genes for regulating the molecular mechanism of different leaflet shapes were detected by functional annotation. The overlap of major QTLs under different environments indicated that the present map would be good enough for precisely mapping genes, and both the QTL analysis and gene prediction were useful for investigating the mechanism of leaf development in mung bean or legumes.

Keywords: *Vigna radiata*, high-density map, leaflet shape, QTL, SNP

INTRODUCTION

Mung bean (*Vigna radiata* L.), a cultivated species in genus *Vigna* ($2n = 2 \times = 22$), has a genome size of about 500 Mb (Yang et al., 2014). It is an important plant protein resource and a potentially healthy food (Nair et al., 2013; Shi et al., 2016) that has traditionally grown across Asia (Tomooka et al., 2002). Mung bean production has spread to most parts of the world in

recent decades. The worldwide demand of mung bean has greatly increased because of its healthy properties, especially the identification of physiological components in the seeds (Yao et al., 2013), that have attracted much interest. Therefore, breeding new varieties with high yield and high quality is urgently required. Genetic and genomic study can help to improve the efficiency of mining for new genes and breeding of important varieties, but mung bean has been less studied than other major crops.

Linkage map construction and gene mapping are important for genetic and genomic studies. A number of genetic linkage maps for mung bean, based mainly on restriction fragment length polymorphism (RFLPs) (Young et al., 1992; Chaitieng et al., 2002; Humphry et al., 2002) or simple sequence repeats (SSRs) (Kajonphol et al., 2012; Sompong et al., 2012; Wang et al., 2016; Singh et al., 2018), have been reported previously. However, most of these maps were low density because of the limited number of polymorphic markers, although primarily mappings of genes associated with bruchid resistance (Chen et al., 2012; Wang et al., 2016), domesticated traits (Isemura et al., 2012), and agronomic traits (Kajonphol et al., 2012) were reported. With the release of the whole genome sequence of mung bean (Yang et al., 2014), rapid progress on gene tagging and candidate genes associated with resistance to pests such as seed beetles (*Coleoptera: Bruchidae*) and disease also has been made (Chotechung et al., 2016; Kaewwongwal et al., 2017; Mathivathana et al., 2019). Single nucleotide polymorphism (SNP) maps, essential for precise gene mapping, have been constructed for species related to mung bean, such as black gram and wild cowpea for bruchid resistance (Amkul et al., 2019; Somta et al., 2019), cowpea for yield traits (Pan et al., 2017), and adzuki bean for flowering time (Liu et al., 2016). However, similar studies are rare in mung bean. To enhance the genetic study of mung bean, we constructed an SNP map using a recombinant inbred line (RIL) population. The original parents of the RILs were both cultivated genotypes with different major agronomic traits. To validate the usefulness of the SNP map, we conducted a quantitative trait locus (QTL) analysis of two rare leaflet shapes of mung bean and predicted the candidate genes associated with these traits. The results will help to accelerate the genetic research of mung bean and contribute to the breeding of improved varieties.

MATERIALS AND METHODS

Plant Material

Dahuaye, green seeded, is a landrace with lobed leaflet, and *Jilyu 9-1*, black-seeded, is a natural mutant with indented leaflet that was obtained from field tests, and the indented leaflet was stably inherit after several season's propagation. A cross between these two genotypes in 2012 in Beijing produced a hybrid that was sown in the following season. After harvest of the F₁ plant, each line of the F₂ generation was planted continually for several years by the single seed descend method, until an RIL₈ population was developed in 2017 for high-density map construction.

DNA Preparation, Library Construction, and Sequencing

The RIL population was continually planted in 2018 and 2019 with 10 seeds for each line. Young leaves of 10 plants from each of 190 randomly selected lines were collected at the 2019 season. DNA was extracted using the CTAB method (Doyle and Doyle, 1987). The quality of the extracted DNA was checked using a NanoPhotometer® spectrophotometer (IMPLEN, Westlake Village, CA, United States) and the concentration was determined using a Qubit® DNA Assay Kit and a Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA, United States).

A total of 1.5 µg genomic DNA per sample was fragmented by sonication to 350 bp, then end polished, A-tailed, and ligated with the full-length adapter. After PCR amplification, product purification (AMPure XP system, Beckman Coulter, Beverly, MA, United States), and size distribution assessment using an Agilent 2100 Bioanalyzer, the target fragments were quantified by real-time PCR. The prepared libraries were sequenced on an Illumina NovaSeq 6000 PE150 platform (Illumina, San Diego, CA, United States) and 150-bp paired-end reads were generated with insert sizes of about 350 bp.

Raw reads in *fasta* format were processed using in-house C scripts (Biomarker Technologies, Beijing, China) for quality control (QC) to ensure the reliability and that the reads were free of base-calling duplicates, adapter contamination, and other artificial biases. The QC process removed reads with ≥10% unidentified nucleotides, reads with >50% low-quality bases, reads with >10 nucleotides aligned to the adapter allowing ≤10% mismatches, and putative PCR duplicates generated by PCR amplification in the library construction process.

Mapping to the Reference Genome and SNPs Detection

The genome sequence of mung bean (GenBank: GCA_000741045.2) was used as the reference genome (Yang et al., 2014). The clean reads in each library were aligned against the reference genome (settings: mem -t 4 -k 32 -M-R) using the Burrows-Wheeler Aligner (Li and Durbin, 2009). To ensure the accuracy of the SNP detection, duplicates were removed using the Picard Mark Duplicates tool¹. The local realignment and SNP detection were performed using the Genome Analysis Toolkit (GATK) protocol² (McKenna et al., 2010) according to the position of the clean reads on the reference genome.

Genotyping and Effective Marker Detection

The SNPs that were polymorphic between the two parents were selected and classified into eight segregation patterns (ab × cd, ef × eg, hk × hk, lm × ll, nn × np, aa × bb, ab × cc, and cc × ab). Only SNPs with the aa × bb pattern were selected for further analysis. The SNPs with sequencing depth less than 4 were removed. To find recombination breakpoints, we used a

¹<http://sourceforge.net/projects/picard/>

²<https://www.broadinstitute.org/gatk/guide/best-practices?bpm=DNaseq#variant-discovery-ovv>

sliding window that contained 15 SNPs to scan the genome, one SNP per slide. The window was typed as aa when the number of aa-type or bb-type in the window was >11 ; in all other cases, the window was typed as ab. After genotyping, the SNPs without recombination were integrated into one bin marker. Markers that had an unusual appearance, low coverage, and/or obvious segregation distortions were filtered prior to linkage analysis. Only the bin markers longer than 10 Kb were reserved for analysis, and bin markers with abnormal base and segregation distortion ($P < 0.01$) were removed.

Genetic Linkage Map Construction

The selected bin markers were used to construct the genetic linkage map. The bin markers were divided in linkage groups (LGs) based on their locations on the mung bean reference genome; that is, 11 LGs were preliminary set before mapping. For each LG, the mapping data were formatted according to the instruction of Ici-Mapping. The marker order was inferred nnTwoOpt, in which nearest neighbor and two-opt were used for tour construction and improvement, respectively, and recombinant frequencies were calculated with a minimum logarithm of odd (LOD) threshold of 3. Genetic distances were estimated with the Kosambi's mapping function. Rippling was performed using SARF (sum of adjacent recombination frequencies) with a window size of 5 (Meng et al., 2015).

Observation and Measurement of Leaf Blade Characteristics

To obtain precise locations of the shapes in the mung bean genome, both the parents and the RILs were investigated for their leaflet shapes under two continually cropping seasons. In the summer of 2018, the RILs were planted in the test site of the Institute of Crop Sciences at the Chinese Academy of Agricultural Sciences (CAAS) (39.96°N and 116.33°E). After harvesting, they were planted in Yazhou (18.38°N and 109.19°E) in winter of 2018 and then in Shunyi (40.23°N and 116.56°E) in summer of 2019. During the flowering stage, a top leaflet of classical ternate compound leaf was collected from the middle part of each plant photographed. For each line, the top leaflets from five plants were evaluated. Because both the lobed and indented leaflets are not classical qualitative traits, to avoid the statistical error in evaluation by visual observation, for the QTL analysis, we counted the number of apexes around each leaf margin for the indented shape, and applied two indexes to evaluate the degree of the lobed leaflet. One index was the ratio of leaf areas, which was calculated using a self-written protocol in Matlab as $R = (V/A) \times 100$, where V is the actual area of the leaflet and A is the sum of the actual and lobed areas. The other index was the ratio of the length and width, which was calculated as $Q = (L_1/L_2) \times 100$, where L_1 is the distance between the intersection of the concave point on both sides of the leaf and the main vein intersection to the base of the leaf, and L_2 is the distance from the tip to the base of the leaf. Lines with leaflets that showed both lobed and indented characters were removed from the QTL analysis because, when photographed, the curled margins made them difficult to evaluate.

Analysis of QTL and Candidate Genes

Quantitative trait locus analysis was conducted using the inclusive composite interval mapping (ICIM) method implemented in IciMapping (Meng et al., 2015). R and Q values were used to locate QTL for lobed leaflet and the number of apexes (NA) was used to locate QTL for indented leaflet. The ICIM was performed every 0.1 cM with the probability in stepwise regression set as 0.001. For the LOD threshold $LOD = \log_{10} (P_1/P_2)$, P_1 is the probability that the site has a QTL and P_2 is the probability that the site has no QTL. Each QTL was determined by a 1000 permutation test at $P = 0.05$.

The bin markers were located in the mung bean reference genome using the re-sequencing data. According to the physical location of a bin marker on each side of a QTL, the target region in the reference genome was searched on NCBI website³. The genes in the target regions were obtained from the GFF3 file and the KOBAS 3.0 database⁴ was used for the gene ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis by inputting the gene list and selecting the "Gene-list enrichment" module. The KEGG pathway was considered to be significantly enriched when the corrected P -value was <0.05 .

RESULTS

Sequencing and Genotyping

Totally, 10.99 Gb and 13.35 Gb of clean bases for parents *Dahuaye* and *Jilyu9-1*, respectively, and 494.03 Gb for offspring were obtained. The average Q30 for parents and offspring were 92.86% and 92.19%, and the average GC contents were 34.30% and 35.18%, respectively. The ratios of mapped clean reads for parents were 95.90% (*Dahuaye*) and 95.71% (*Jilyu9-1*), and 94.05% for offspring (**Supplementary Table S1**).

A total of 564,449 SNP was identified between the parents, and 485,977 (*Dahuaye*) and 485,084 (*Jilyu9-1*) of them were homozygous. However, only 23,170 of these SNPs had the aa $\times$ bb segregation pattern. These SNPs were further combined into 2202 bin markers for map construction.

Construction of the Genetic Linkage Map

Among the 190 RIL lines, 15 lines had a fairly high ratio of abnormal SNPs ($>2\%$ of the total number of SNPs). Therefore, only 175 lines were used for map construction. After linkage analysis, a total of 1946 bin markers consisting of 21,508 SNPs were finally located onto the 11 LGs. According to the marker positions, there was a good agreement between the LGs and the chromosomes of the mung bean reference genome (**Supplementary Table S2**). The total length of the new map was 1060.17 cM with an average distance between adjacent markers at 0.54 cM. The number of mapped markers in each LG ranged from 99 (LG4) to 258 (LG5) with an average of 177. The average length of the 11 LGs was 96.38 cM, LG4 was the shortest (38.76 cM), and LG9 was the longest (168.03 cM). The density of markers on each LG varied greatly with LG2 (0.35 cM) $>$ LG4 (0.40 cM) $>$ LG5

³<https://www.ncbi.nlm.nih.gov/genome/?term=mungbean>

⁴http://kobas.cbi.pku.edu.cn/anno_iden.php

(0.44 cM) > LG3 (0.45 cM) > LG7 (0.48 cM) > LG8 (0.50 cM) > LG1 (0.51 cM) > LG6 (0.56 cM) > LG10 (0.73 cM) > LG11 (0.74 cM) > LG9 (1.02 cM). A total of 18 gaps (>5 cM) were located in LG1, 5, 7, and 9. Seven of the gaps were > 10 cM; four were on LG9 including the longest one (19.9 cM) (Table 1 and Figure 1).

Inheritance of Different Leaflet Shapes

The leaflets of all F₁ plants were semi-lobed, and the indented shape did not appear at the first generation (Figure 2), indicating that the lobed leaflet was semi-dominant and the indented shape recessive. Among their offspring, there were individuals that expressed indented and lobed shapes simultaneously (Figure 2), suggesting these two leaflets were not allelic and were inherited independently.

Based on visual observation of the phenotypes, the separation ratio of indented leaflet to ovate shape did not fit 1:1, whereas the separation ratio of lobed leaflet to ovate shape fit 1:1 by both R ($\chi^2 = 0.0548$, $P = 0.8149$) and Q ($\chi^2 = 1.39$, $P = 0.2386$) methods. Using the precise methods, both the frequency of distribution of both the indented leaflet by counting the number of apexes and the lobed shape using R and Q methods did not accord with classical normal distribution, under the 2018 and 2019 environments (Figure 3).

Identification of QTL for the Lobed and Indented Leaflet Traits

A series of tightly linked peaks on LG3 and LG10 were detected on the LOD contours for the lobed leaflet under the different environments (Figure 4), corresponding with six QTLs on LG3 and two on LG10 (Table 2 and Figure 5). Among them, one major QTL, Block18747–Block 19028 (*qLobL.1*) spanning 1.0 cM on LG3, was stably detected in the 2018 libraries with Q and R values, with phenotypic variation explained (PVE) values from 50.0 to 69.5%. This stable QTL was physically mapped onto chromosome 3 from 5471015 to 5544796 bp. The other five QTLs on LG3 were identified in the 2019 libraries using Q and R values, with PVE

values from 15.8 to 39.8%. These five QTLs were almost connected together from Block19028 to Block19223, with two overlapped confidence intervals from Block18747 to Block18748 (*qLobL.2*) and Block19219 to Block19223 (*qLobL.4*). Therefore, one confidence interval from Block18747 to Block18748 (*qLobL.2*) was the only repeated QTL that could be detected under the two different environments. Two other QTLs for lobed leaflet from Block7476 to Block7482 (*qLobL.7*) and Block7483 to Block7489 (*qLobL.8*) on LG10 were detected only in the 2019 libraries, with PVE values of 12.2% and 15.5%, respectively.

For indented leaflet, one common QTL (*qInL.1*) on LG10 was detected in the 2018 and 2019 libraries. It was located from Block7460 and Block7463 and spanned about 2.4 cM, with PVE values of 39.7% (2018) and 45.5% (2019), respectively. QTL *qInL.1* was tightly linked with *qLobL.7* on LG10.

Candidate Gene Annotation

We detected 112 genes in all QTL regions related with lobed leaflet, however, only 10 of them were located in the common Block18747–Block18748 region in LG3. Seven of the 10 genes had only non-synonymous SNPs between the parents (Table 3). The products of the homologous genes included polyadenylate-binding protein, late embryogenesis abundant protein, and different protein kinases. The KEGG analysis indicated that *LOC106757470* and *LOC106757471* were involved in four pathways, namely RNA degradation, ubiquitin-mediated proteolysis, RNA transport, and mRNA surveillance, and that *LOC106757470* was involved in three of these pathways; the exception was ubiquitin-mediated proteolysis.

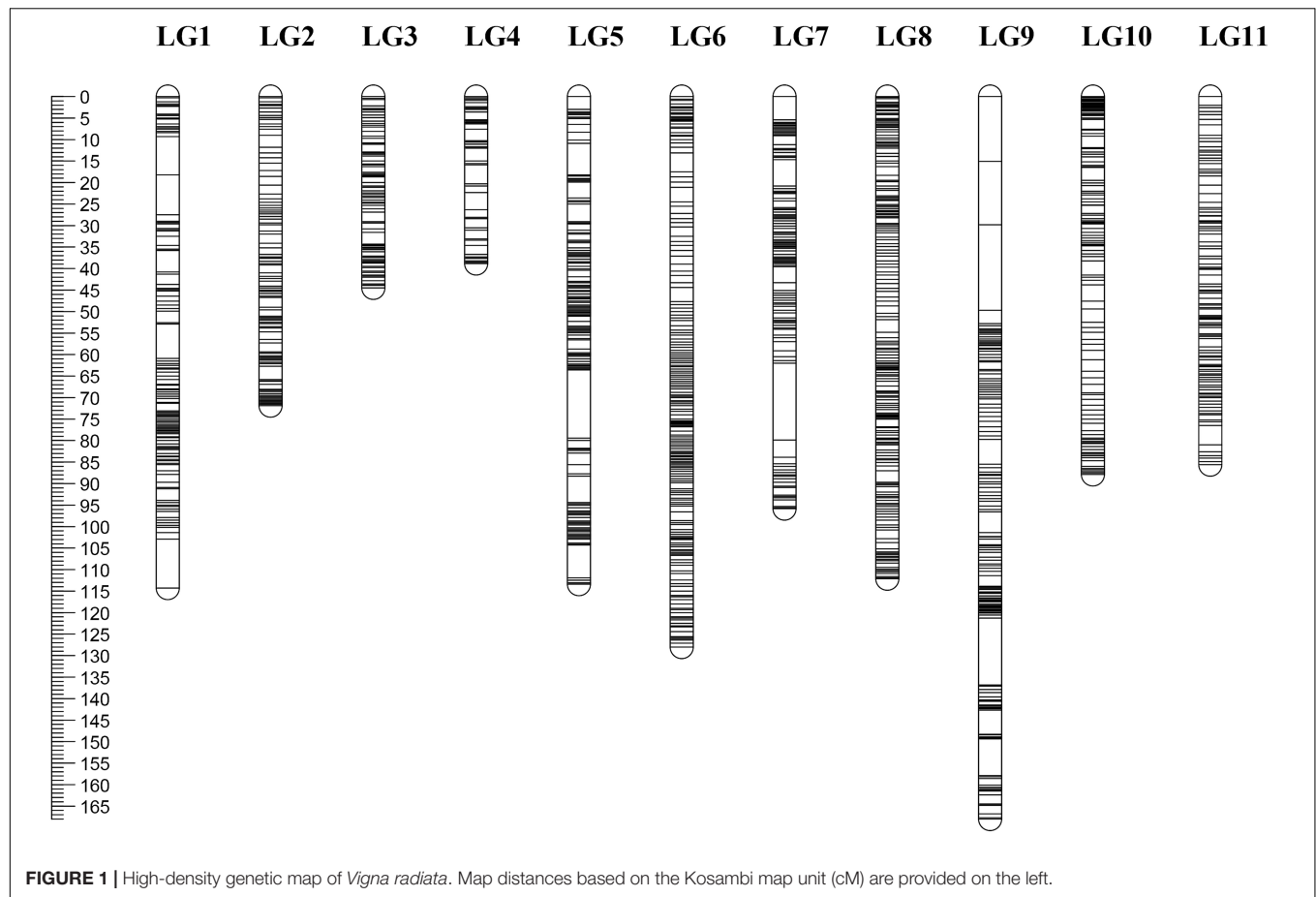
We detected 45 genes located in the QTL region related with the indented leaflet, and 35 of them encoded homologous products. However, only five had SNPs between the two parents. Four SNPs were in the non-coding regions of *LOC106774574*, *LOC106776045*, *LOC106774382*, and *LOC106776095*, and seven SNPs were synonymous mutations in *LOC106775793*. However, the only one SNP could cause an amino acid change, namely tryptophan (*Dahuaye*) to a stop codon (*Jilyu 9-1*).

DISCUSSION

Several maps have been constructed for mung bean with different marker types: RFLP (Young et al., 1992; Lambrides et al., 2000; Humphry et al., 2002), AFLP (Chaitieng et al., 2002), and SSR (Kajonphol et al., 2012; Sompong et al., 2012; Liu et al., 2016; Wang et al., 2016; Singh et al., 2018). The frequent development of SSR had made them popular both in map construction and gene tagging for more than a decade (Han et al., 2005; Wang et al., 2016). However, the densities of these maps were low because of the low level of variation of SSR markers (Wang et al., 2009). SNPs have gradually become more widely used for high density map construction, including in species related to mung bean (Liu et al., 2016; Pan et al., 2017; Amkul et al., 2019; Somta et al., 2019). With the release of the whole genome of

TABLE 1 | Information of the novel constructed genetic linkage map for mung bean.

LG	Num. of bin marker	Total distance (cM)	Average distance (cM)	Max gap (cM)	Num. of gaps >5 cM
LG1	223	114.27	0.51	11.4	4
LG2	209	71.90	0.35	3.2	0
LG3	101	44.55	0.45	2.6	0
LG4	99	38.76	0.4	4.4	0
LG5	258	113.37	0.44	15.8	4
LG6	229	127.99	0.56	4.4	0
LG7	202	95.76	0.48	17.9	3
LG8	223	112.09	0.5	2.9	0
LG9	165	168.03	1.02	19.9	7
LG10	121	87.85	0.73	3.8	0
LG11	116	85.60	0.74	4.5	0
Total	1946	1060.17	0.54	—	18

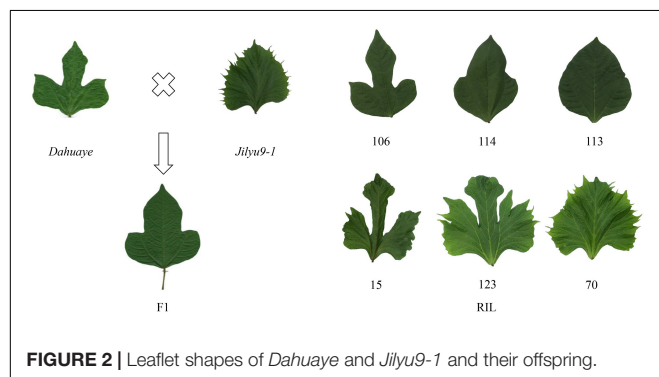


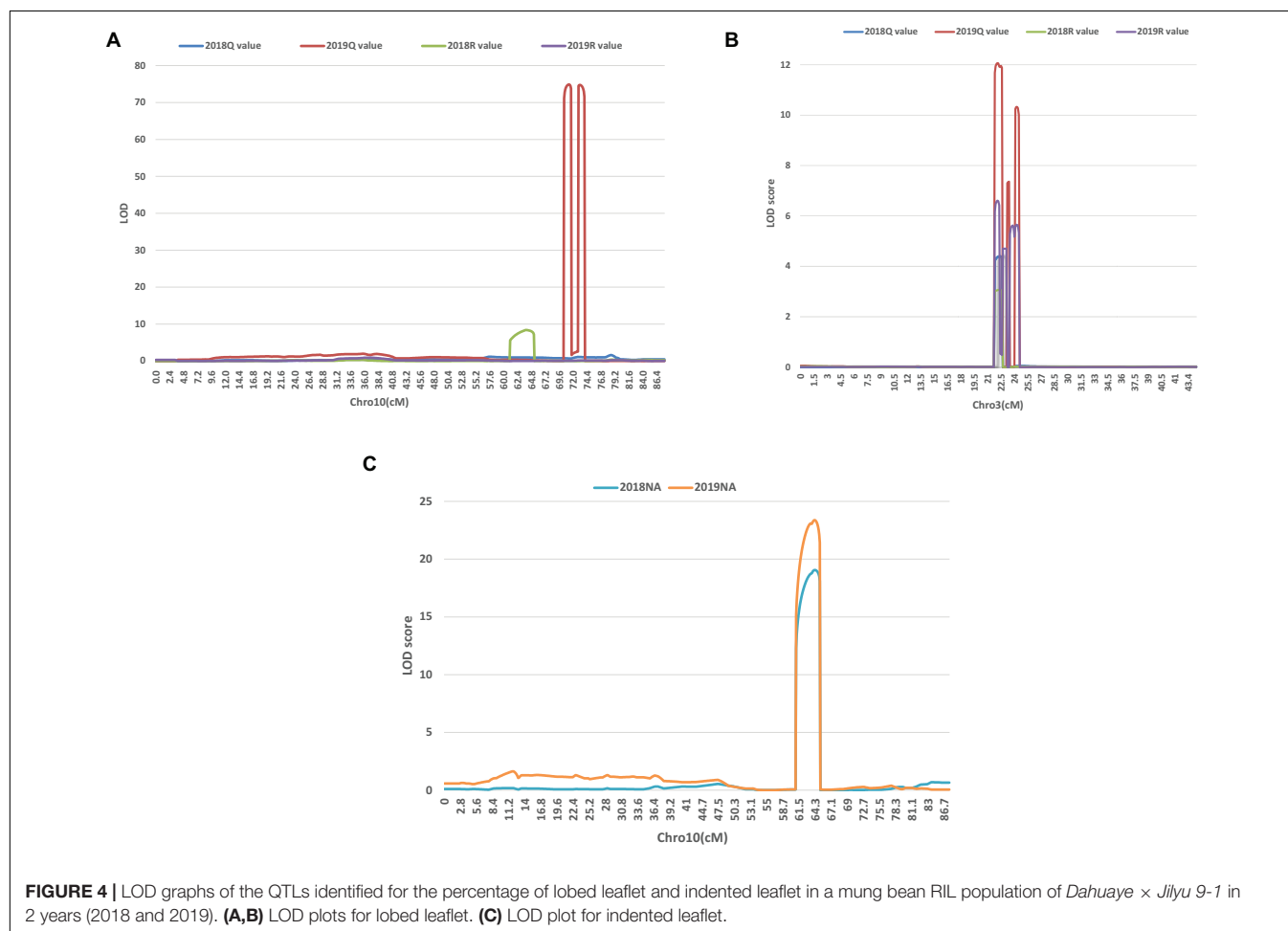
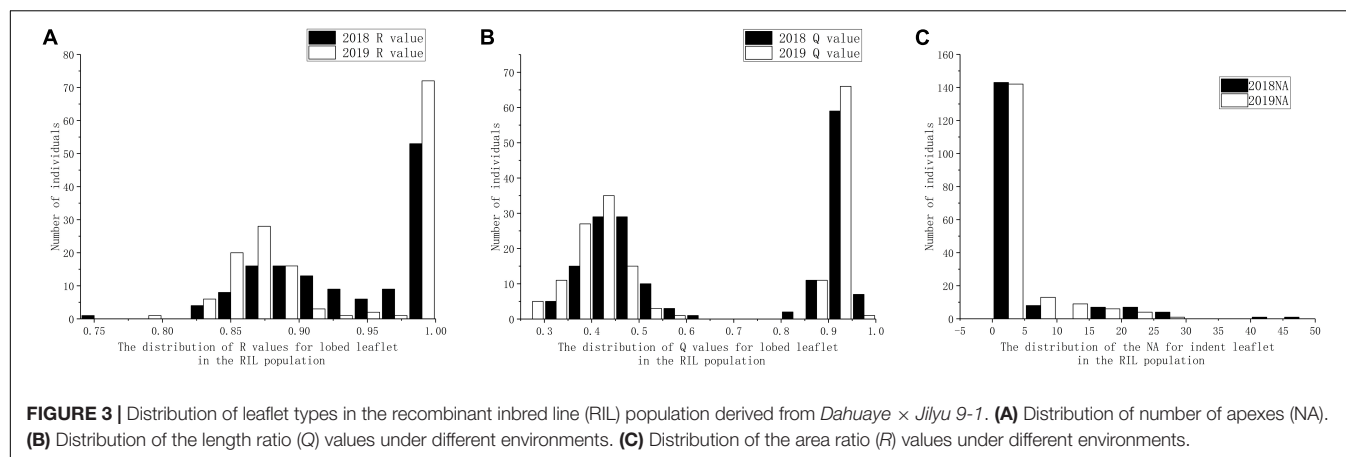
mung bean (Kang et al., 2014), two SNP maps were reported (Schafleitner et al., 2016; Mathivathana et al., 2019). However, because SNP map information is not easily shared among different labs, we constructed our own SNP map.

The order of the markers on LGs in our map agrees well with their locations on the reference genome of mung bean. We consider that the newly constructed map is an improvement over the map constructed by Schafleitner et al. (2016), because in their map, LG3 and LG4 were merged together. The other previous map also agreed well with the reference genome

(Mathivathana et al., 2019); however, our map has higher density (0.54 cM) than theirs (3.23 cM), mainly because of the different mapping populations and sequencing method.

The total length of our map was much shorter (1060.2 cM) than that of the previous map (1291.7 cM), and the length of the LGs for the two maps also was quite different. For example, in our map, LG9 was the longest (168.03 cM) and LG2 was the shortest (71.9 cM), while in the previous map, LG7 was the longest (149.1 cM) and LG3 the shortest (90.2 cM), whereas in the previous map, LG7 was the longest (149.1 cM) and LG3 the shortest (90.2 cM). In our map, LG9 was the lowest (1.02 cM), followed by LG11 (0.74 cM) and LG10 (0.73 cM), whereas in the previous map, LG8 had the lowest density (3.03 cM), followed by LG10 (3.02 cM) and LG9 (2.91 cM). However, the total trends of the two maps were similar, suggesting that the two maps were in a good agreement. Although the average interval distance was good enough for precise mapping, there were 18 large gaps in our map, especially on LG9, leading to the low density of this LG. Such gaps are seldom found in maps using traditional markers for several main reasons, including that we used a map population derived from two cultivated genotypes, and the genetic backgrounds were more similar than that between a wild and a cultivated genotype. Inaccuracies in the reference genome may be another reason for the gaps in our map.





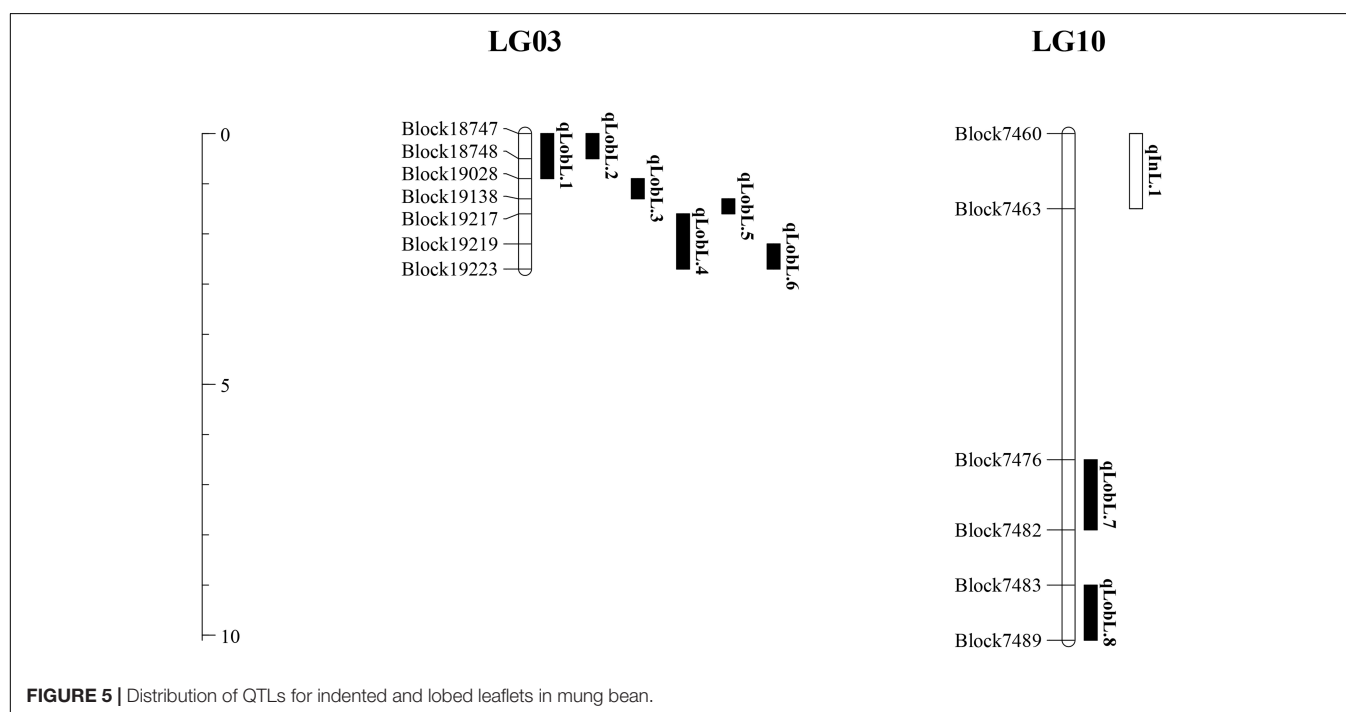
To validate the usefulness of the present map, we investigated QTL related with two types of leaflet shapes, lobed and indented. Lobed leaves are common in plants (Zhang et al., 2018), and they also have been reported in mung bean (Jiao et al., 2016). The phenotypes of F1 and F2 plants indicated that the lobed leaflet was semi-dominant and was controlled by multiple genes, and this finding was validated by the QTL analysis. Several

QTLs that were tightly linked together were detected on LG3 and LG10. However, only one was consistently found under different environments both with Q and R methods. This QTL spanned from *Block18747* and *Block 18748* on LG3, with PVE from 27.7 to 69.5%. The location of this QTL was not overlapped with the other QTLs on our map but was tightly linked with a QTL predicted by Jiao et al. (2016). Whether these two

TABLE 2 | Description of QTLs associated with the indented and lobed leaflet in mung bean.

Trait	Method	Environment	QTL name	Chro.	Left marker	Right marker	Physical position	Lod	PVE (%)
Indented leaflet	NA	2018	<i>qInL.1</i>	10	Block7460	Block7463	16431849–16788088	19.1	39.7
		2019	<i>qInL.1</i>	10	Block7460	Block7463	16431849–16788088	23.4	45.5
Lobed leaflet	<i>R</i> value	2018	<i>qLobL.1</i>	3	Block18747	Block19028	5471015–5941848	6.1	50.0
		2019	<i>qLobL.2</i>	3	Block18747	Block18748	5471015–5544796	66.1	27.7
			<i>qLobL.3</i>	3	Block19028	Block19138	5941848–6074518	47.0	16.7
			<i>qLobL.4</i>	3	Block19217	Block19223	6181985–6235477	56.2	19.5
	<i>Q</i> value	2018	<i>qLobL.1</i>	3	Block18747	Block19028	5471015–5941848	8.8	69.5
		2019	<i>qLobL.2</i>	3	Block18747	Block18748	5471015–5544796	120.7	29.1
			<i>qLobL.5</i>	3	Block19138	Block19217	6074518–6181985	73.7	15.8
			<i>qLobL.6</i>	3	Block19219	Block19223	6214542–6235477	103.3	17.4
			<i>qLobL.7</i>	10	Block7476	Block7482	17381168–17531051	74.9	12.2
			<i>qLobL.8</i>	10	Block7483	Block7489	17619040–17858382	74.8	14.5

NA, the number of apexes around each leaf margin for the indented shape; PVE, phenotypic variation explained value for a QTL.

**FIGURE 5** | Distribution of QTLs for indented and lobed leaflets in mung bean.

QTL have different allelic genes needs to be further validated. The other QTLs were only found in 2019 library, including two on LG10, indicating these QTLs were easily affected by the environments.

Ten genes were located in the target region of the common QTL related with lobed leaflet and all 10 genes had SNPs between the parents; however, non-synonymous SNPs were found only in seven of these genes. The metabolic pathways involving these genes were mostly unknown, so it was difficult to identify candidate gene associated the development of lobed leaflet. However, overexpression of the gene encoding a glycosyltransferase has been reported to affect the leaf veins in tobacco (Ma et al., 2012), and this gene has been predicted to be homologous with *LOC106757416*. Therefore, we consider this gene is the most probable candidate gene associated with the lobed leaflet trait.

The indented leaflet is a rare mutant that we found recently in the test field; it has not been reported previously. The indented shape may be caused by the visually observed abnormal variation at the shoot apical meristem, and it also showed differences in

TABLE 3 | SNPs between parents in genes related with lobed leaflet in mung bean.

Gene name	Total SNPs	SNPs in exon	Non-synonymous SNPs
LOC106757470	32	14	11
LOC106757280	11	7	5
LOC106757335	10	4	3
LOC106757671	51	8	7
LOC106757416	12	4	2
LOC106757422	3	3	1
LOC106757412	14	4	2

the number of apexes around the leaflet margins. This trait was considered recessive because it was absent in F1 plants and, based on our observation of F2 plants, it may be controlled by multiple genes. However, we detected only one major QTL spanning Block7460–Block7463 on LG10, and explaining 39.7% and 45.5% of PVE under the 2018 and 2019 environments. Although 35 genes with homologous annotations were found within the target region of the major QTL related with indented leaflet, only five genes had 11 SNPs between the parents. Among these SNPs, some were located in non-coding regions and some led to synonymous mutations, only one SNP in the coding region of *LOC106775793*, resulted in a synonymous mutation. In addition, *LOC106775793* was homologous with the gene encoding transcription factor TCP2, which is expressed mainly in rapidly growing tissues or organs and may be related with the development of plant cells (Palatnik et al., 2003; Koyama et al., 2017; Bresso et al., 2018). Overexpression of *TCP* reduced the degree of indent leaflet (Koyama et al., 2017), and knocked out of *TCP* led to shrinkage of the leaf margin (Koyama et al., 2010). Therefore, *LOC106775793* may be the most probable candidate gene controlling the indented leaflet trait. The QTL analysis not only validated the novel map as useful for fine mapping of other traits but also provides a base for investigating the molecular mechanism controlling leaflet shapes in mung bean and other legumes.

CONCLUSION

We constructed a high-density genetic linkage map for mung bean using re-sequencing technology, and the map agreed with the reference genome of mung bean. QTLs related with different leaflet shapes were analyzed using this map. One major QTL for indented leaflet was detected on LG10 and several tightly linked QTLs on LG3 and LG10 related with lobed leaflet were detected, but only one of them was consistently detected under different environments with different methods. Genes involved in the molecular mechanism of leaf shape also were predicted. The present map will be useful for fine mapping of genes, and the

QTLs and candidate genes will shed light on the development of leaflet shapes in mung bean and other legumes.

DATA AVAILABILITY STATEMENT

We uploaded the raw data of the re-sequencing to NCBI: PRJNA656124.

AUTHOR CONTRIBUTIONS

LW conceptualized, designed, supervised the study, and revised the manuscript. JW and JL developed the population and data analysis. ZL, XY, HC, SW, and XiC conducted the data collection. LW and XuC secured the research funding. JW wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Key R&D Program of China (Nos. 2019YFD1000700 and 2019YFD1000701), the China Agriculture Research System (CARS-08), and the Agricultural Science and Technology Innovation Program (ASTIP) of the Chinese Academy of Agricultural Sciences.

ACKNOWLEDGMENTS

We thank Liwen Bianji, Edanz Editing China (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript. We also thank Dr. Jun Liu from Beijing Normal University for writing the procedure to measure the area of the lobed leaflet.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.01032/full#supplementary-material>

REFERENCES

- Amkul, K., Wang, L. X., Somta, P., Wang, S. H., and Cheng, X. Z. (2019). Construction of a high density linkage map and genome dissection of bruchid resistance in zombi pea (*Vigna vexillata* (L.) A. Rich). *Sci. Rep.* 9:11719.
- Bresso, E. G., Chorostecki, U., Rodriguez, R. E., Palatnik, J. F., and Schommer, C. (2018). Spatial control of gene expression by miR319-regulated TCP transcription factors in leaf development. *Plant Physiol.* 176, 1694–1708. doi: 10.1104/pp.17.00823
- Chaiteng, B., Kaga, A., Han, O. K., Wang, X. W., Wongkaew, S., Laosuwan, P., et al. (2002). Mapping a new source of resistance to powdery mildew in mung bean. *Plant Breed.* 121, 521–525.
- Chen, H. M., Ku, H. M., Schafleitner, R., Bains, T. S., George Kuo, C., Liu, C. A., et al. (2012). The major quantitative trait locus for mung bean yellow mosaic Indian virus resistance is tightly linked in repulsion phase to the major bruchid resistance locus in a cross between mung bean (*Vigna radiata* (L.) Wilczek) and its wild relative *Vigna radiata*. *Euphytica* 192, 205–216. doi: 10.1007/s10681-012-0831-9
- Chotechung, S., Somta, P., Chen, J., Yimram, T., Chen, X., and Srinives, P. (2016). A gene encoding a polygalacturonase-inhibiting protein (PGIP) is a candidate gene for bruchid (Coleoptera: bruchidae) resistance in mung bean (*Vigna radiata*). *Theor. Appl. Genet.* 129, 1673–1683. doi: 10.1007/s00122-016-2731-1
- Doyle, J., and Doyle, J. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Han, O. K., Kaga, A., Isemura, T., Wang, X. W., Tomooka, N., and Vaughan, D. A. (2005). A genetic linkage map for azuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi]. *Theor. Appl. Genet.* 111, 1278–1287. doi: 10.1007/s00122-005-0046-8
- Humphry, E., Konduri, V., Lambrides, J., Magner, T., McIntyre, L., Aitken, B., et al. (2002). Development of a mung bean (*Vigna radiata*) RFLP linkage map and its comparison with lablab (*Lablab purpureus*) reveals a high level of colinearity between the two genomes. *Theor. Appl. Genet.* 105, 160–166. doi: 10.1007/s00122-002-0909-1
- Isemura, T., Kaga, A., Tabata, S., Somta, P., Srivines, P., Shimiz, T., et al. (2012). Construction of a genetic linkage map and genetic analysis of domestication

- related traits in mung bean (*Vigna radiata*). *PLoS One* 7:e41304. doi: 10.1371/journal.pone.0041304
- Jiao, K., Li, X., Guo, W., Yuan, X., Cui, X., and Chen, X. (2016). Genome re-sequencing of two accessions and fine mapping the locus of lobed leaflet margins in mungbean. *Mol. Breed.* 36:128.
- Kaewwongwal, A., Chen, J., Somta, P., Kongjaimun, A., Yimram, T., Chen, X., et al. (2017). Novel alleles of two tightly linked genes encoding polygalacturonase-inhibiting proteins (VrPGIP1 and VrPGIP2) associated with the Br locus that confer bruchid (*Callosobruchus* spp.) resistance to mung bean (*Vigna radiata*) accession V2709. *Front. Plant Sci.* 8:1692. doi: 10.3389/fpls.2017.01692
- Kajonphol, T., Sangsiri, C., Somta, P., Toojinda, T., and Srinives, P. (2012). SSR map construction and quantitative trait loci (QTL) identification of major agronomic traits in mung bean (*Vigna radiata* (L.) Wilczek). *SABRAO J. Breed. Genet.* 44, 71–86. doi: 10.1556/019.70.2019.09
- Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B. K., et al. (2014). Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat. Commun.* 5:5443. doi: 10.1038/ncomms6443
- Koyama, T., Mitsuda, N., Seki, M., Shinozaki, K., and Ohme, T. M. (2010). TCP transcription factors regulate the activities of ASYMMETRIC LEAVES1 and miR164, as well as the auxin response, during differentiation of leaves in *Arabidopsis*. *Plant Cell* 22, 3574–3588. doi: 10.1105/tpc.110.075598
- Koyama, T., Sato, F., and Ohme-Takagi, M. (2017). Roles of miR319 and TCP transcription factors in leaf development. *Plant Physiol.* 175, 874–885. doi: 10.1104/pp.17.00732
- Lambides, C. J., Lawn, R. J., Godwin, I. D., Manners, J., and Imrie, B. C. (2000). Two genetic linkage maps of mung bean using RFLP and RAPD markers. *Aust. J. Agric. Res.* 51, 415–425. doi: 10.1071/ar99052
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Liu, C. Y., Fan, B. J., Cao, Z. M., Su, Q. Z., Wang, Y., Zhang, Z. X., et al. (2016). Development of a high-density genetic linkage map and identification of flowering time QTLs in adzuki bean (*Vigna angularis*). *Sci. Rep.* 6:39523. doi: 10.1038/srep39523
- Ma, C., Liu, X., Wang, C., Xu, X., Cheng, G., and Tan, Y. (2012). Construction and genetic transformation of glycosyltransferase sm-ngt1 deletion expression vector in tobacco. *Guangdong Agric. Sci.* 39, 146–148.
- Mathivathana, M. K., Murukarthick, J., Karthikeyan, A., Jang, W., Dhasarathan, M., Jagadeeshselvam, N., et al. (2019). Detection of QTL associated with mung bean yellow mosaic virus (MYMV) resistance using the interspecific cross of *Vigna radiata* × *Vigna umbellata*. *J. Appl. Genet.* 60, 255–268. doi: 10.1007/s13353-019-00506-x
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001
- Nair, R. M., Yang, R. Y., Easdown, W. J., Thavarajah, D., Thavarajah, P., Hughes, J., et al. (2013). Biofortification of mung bean (*Vigna radiata*) as a whole food to enhance human health. *J. Sci. Food Agric.* 93, 1805–1813. doi: 10.1002/jsfa.6110
- Palatnik, J. F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J. C., et al. (2003). Control of leaf morphogenesis by microRNAs. *Nature* 425, 257–263. doi: 10.1038/nature01958
- Pan, L., Wang, N., Wu, Z., Guo, R., Yu, X., Zheng, Y., et al. (2017). A high density genetic map derived from RAD sequencing and its application in QTL analysis of yield-related traits in *Vigna unguiculata*. *Front. Plant Sci.* 8:1544. doi: 10.3389/fpls.2017.01544
- Schafleitner, R., Huang, S. M., Chu, S. H., Yen, J. Y., Lin, C. Y., Yan, M. R., et al. (2016). Identification of single nucleotide polymorphism markers associated with resistance to bruchids (*Callosobruchus* spp.) in wild mung bean (*Vigna radiata* var. *sublobata*) and cultivated *V. radiata* through genotyping by sequencing and quantitative trait locus analysis. *BMC Plant Biol.* 16:159. doi: 10.1186/s12870-016-0847-8
- Shi, Z., Yao, Y., Zhu, Y., and Ren, G. (2016). Nutritional composition and antioxidant activity of twenty mung bean cultivars in China. *Crop J.* 4, 398–406. doi: 10.1016/j.cj.2016.06.011
- Singh, N., Mallick, J., Sagolsem, D., Mandal, N., and Bhattacharyys, S. (2018). Mapping of molecular markers linked with MYMIV and yield attributing traits in mung bean. *Ind. J. Genet. Plant Breed.* 78, 118–126. doi: 10.5958/0975-6906.2018.00014.7
- Sompong, U., Somta, P., Raboy, V., and Srinives, P. (2012). Mapping of quantitative trait loci for phytic acid and phosphorus contents in seed and seedling of mung bean (*Vigna radiata* (L.) Wilczek). *Breed. Sci.* 62, 87–92. doi: 10.1270/jsbbs.62.87
- Somta, P., Chen, J., Yundaeng, C., Yuan, X., Yimram, T., Tomooka, N., et al. (2019). Development of an SNP-based high-density linkage map and QTL analysis for bruchid (*Callosobruchus maculatus* F.) resistance in black gram (*Vigna mungo* (L.) Hepper). *Sci. Rep.* 9:3930. doi: 10.1038/s41598-019-40669-5
- Tomooka, N., Vaughan, D., Moss, H., and Maxted, N. (2002). *The Asian Vigna: Genus Vigna Subgenus Ceratotropis Genetic Resources*. London: Kluwer Academic Publishers.
- Wang, L. X., Cheng, X. Z., Wang, S. H., Liu, C. Y., and Liang, H. (2009). Transferability of SSR markers from adzuki bean into mungbean. *Acta Agron. Sin.* 35, 816–820. doi: 10.1016/S1875-2780(08)60083-8
- Wang, L. X., Wu, C. S., Zhong, M., Zhao, D., Mei, L., Chen, H. L., et al. (2016). Construction of an integrated map and location of a bruchid resistance gene in mung bean. *Crop J.* 4, 360–366. doi: 10.1016/j.cj.2016.06.010
- Yang, J. K., Kim, S. K., Kim, M. Y., Lestari, P., and Lee, S. H. (2014). Genome sequence of mung bean and insights into evolution within *Vigna* species. *Nat. Commun.* 5:5443. doi: 10.1038/ncomms6443
- Yao, Y., Yang, X., Tian, J., Liu, C., Cheng, X., and Ren, G. (2013). Antioxidant and antidiabetic activities of black mung bean (*Vigna radiata* L.). *J. Agric. Food Chem.* 61, 8104–8109. doi: 10.1021/jf401812z
- Young, N. D., Kumar, L., Menancio-Hautea, D., Danesh, D., Talekar, N. S., Shanmugasundaram, S., et al. (1992). RFLP mapping of a major bruchid resistance gene in mung bean (*Vigna radiata* L. Wilczek). *Theor. Appl. Genet.* 84, 839–844. doi: 10.1007/bf00227394
- Zhang, Y., Xu, A., Lang, L., Wang, Y., Liu, X., Liang, F., et al. (2018). Genetic mapping of a lobed-leaf gene associated with salt tolerance in *Brassica napus* L. *Plant Sci.* 269, 75–84. doi: 10.1016/j.plantsci.2018.01.005

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

Copyright © 2020 Wang, Li, Liu, Yuan, Wang, Chen, Chen, Cheng and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Enhanced Chlorophyll Degradation Triggers the Pod Degreening of “Golden Hook,” a Special Ecotype in Common Bean (*Phaseolus vulgaris* L.)

Bo Hu^{1,2}, Jinlong Zhu¹, Hongyan Wu¹, Kun Xu¹, Hong Zhai¹, Ning Guo¹, Yi Gao^{1,2}, Jiayin Yang³, Danhua Zhu⁴ and Zhengjun Xia^{1*}

OPEN ACCESS

Edited by:

Jiayin Pang,
University of Western Australia,
Australia

Reviewed by:

Longbiao Guo,
Chinese Academy of Agricultural
Sciences, China
Ailin Liu,
The Chinese University of Hong Kong,
China

*Correspondence:

Zhengjun Xia
xiazhj@iga.ac.cn

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 09 June 2020

Accepted: 18 August 2020

Published: 06 October 2020

Citation:

Hu B, Zhu J, Wu H, Xu K, Zhai H,
Guo N, Gao Y, Yang J, Zhu D and
Xia Z (2020) Enhanced Chlorophyll
Degradation Triggers the Pod
Degreening of “Golden Hook,” a
Special Ecotype in Common Bean
(*Phaseolus vulgaris* L.).
Front. Genet. 11:570816.
doi: 10.3389/fgene.2020.570816

¹ Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin, China, ² University of Chinese Academy of Sciences, Beijing, China, ³ Huaiyin Institute of Agricultural Science of Xuhuai Region, Huai'an, China, ⁴ Zhejiang Academy of Agricultural Sciences, Hangzhou, China

To reveal genetic factors or pathways involved in the pod degreening, we performed transcriptome and metabolome analyses using a yellow pod cultivar of the common bean “golden hook” ecotype and its green pod mutants yielded via gamma radiation. Transcriptional profiling showed that expression levels of red chlorophyll catabolite reductase (*RCCR*, *Phvul.008G280300*) involved in chlorophyll degradation was strongly enhanced at an early stage (2 cm long) in wild type but not in green pod mutants. The expression levels of genes involved in cellulose synthesis was inhibited by the pod degreening. Metabolomic profiling showed that the content of most flavonoid, flavones, and isoflavonoid was decreased during pod development, but the content of afzelechin, taxifolin, dihydrokaempferol, and cyanidin 3-O-rutinoside was remarkably increased in both wild type and green pod mutant. This study revealed that the pod degreening of the golden hook resulting from chlorophyll degradation could trigger changes in cellulose and flavonoids biosynthesis pathway, offering this cultivar a special color appearance and good flavor.

Keywords: common bean (*Phaseolus vulgaris* L.), degreening, pod, transcriptome, cellulose

Abbreviations: 1-MCP, 1-methylcyclopropene; ANR, anthocyanidin reductase; CESA, cellulose synthase A; *CHLI*, Mg chelatase I subunit; *CHS*, chalcone synthase; *CLH*, chlorophyllase; DEGs, differential expression genes; *DFR*, dihydroflavonol reductase; GO, Gene Ontology; GP, the green pod mutant M628; HCA, hierarchical cluster analysis; *HEMA*, glutamyl tRNA reductase; *HEMC*, hydroxymethylbilane synthase; HPLC, high-performance liquid chromatography; *IF7MAT6*, isoflavone 7-O-glucosyltransferase; KEGG, Kyoto Encyclopedia of Genes and Genomes; *Kor*, cellulase; *LAR*, leucoanthocyanidin reductase; M693, M756, two green pod mutant lines; M729G, one mutant line with green pods; M729Y, one mutant line with yellow pods; *NOL*, NYC1-like; *NYC1*-like, non-yellow coloring 1-like; *PAO*, pheophorbide a oxygenase; *PCA*, principal component analysis; *POR*, protochlorophyllide oxidoreductase; QTL, quantitative trait loci; *RCCR*, red chlorophyll catabolite reductase; RT-qPCR, real-time quantitative PCR; *SGR1*, stay green 1; *SuSy*, sucrose synthase; UDP-glucose, uridine diphosphate glucose; *VR4*, vestitone reductase; YP, wild type Dalong 1.

INTRODUCTION

The common bean (*Phaseolus vulgaris* L.), as one of the most important legume crops worldwide, provides a major source of dietary protein, complex carbohydrates, dietary fiber, numerous vitamins, and trace minerals (Tharanathan and Mahadevamma, 2003). The common bean is cultivated from 52°N to 32°S latitude and from near sea level to elevations of more than 3,000 m. Among a large number of cultivated varieties and landraces, there is a high level of diversification in nutrient content, taste, color, and texture of pod and seed in the common bean.

Plant pigments, including carotenoid, anthocyanin, and chlorophyll, are in charge of fruit peel or flesh color (Tanaka et al., 2008). For a specific fruit or vegetable, the content and proportion of different pigments are the decisive factors for the color appearance. Chlorophyll is an essential photosynthetic pigment in chloroplast of higher plants and performs complex processes of harvesting light energy and driving electron transfer. Many fruit flesh or peel and leaves with green color are caused by the high chlorophyll content (Xu et al., 2020). However, yellowing or senescence is often accompanied by chlorophyll degradation. Chlorophyll and carotenoids are related to the color variation from green to yellow. When carotenoids are masked by excessive chlorophyll, the fruit flesh or peel and leaves appear green. The yellow color of carotenoids is unmasked upon chlorophyll degradation during ripening or senescence. Chlorophyll breakdown is a complicated multistep enzymatic process. *CLH* is a crucial factor in the regulation of the chlorophyll reduction in the pericarp (Harpaz-Saad et al., 2007). In two pear cultivars, one turned yellow during ripening due to loss of chlorophylls a/b and carotenoids, while the other stayed green until fully ripen, which could be accounted for the lower the expression levels of the genes for chlorophyll degradation, including *CLH*, *PAO*, and *NYC1*-like (Charoenchongsuk et al., 2015). Treatment with 1-MCP could maintain intact chloroplasts with well-organized grana thylakoids and small plastoglobuli and delay chlorophyll degradation by suppressing the expression of *PAO*, *NYC*, *NOL*, and *SGR1* (Cheng et al., 2012). The *SGR* gene and its homolog, *SGR*-like, had been detected in various plant species; overexpression of *SGRL* reduced the chlorophyll content and promoted chlorophyll breakdown (Sakuraba et al., 2014). Synthesis and degradation of chlorophyll are under control of the coordinate regulatory cascade, in which the malfunction of key steps or genes would affect total chlorophyll content and the color appearance.

The key factor that restricts the quality of fresh pods is the cellulose content. The tender pods of common beans with low cellulose content have a pleasant taste. In order to take full advantage of the common bean in human diet, it is necessary to improve the quality and reduce cellulose content. Cellulose biosynthesis is a complex biochemical process, which includes various enzymes, such as *CESA*, *Kor*, and *SuSy*. Uridine diphosphate-glucose (UDP-glucose) is regarded as the immediate substrate for cellulose polymerization in higher plants. Photosynthetically fixed CO₂ is the ultimate source of C for the synthesis of nucleotide sugars, such as UDP-glucose, which

are the building blocks for synthesis of cell wall polysaccharides (Nakai et al., 1999). UDP-glucose can be derived from the cleavage of sucrose catalyzed by *SuSy* yielding UDP-glucose and fructose, demonstrating that *SuSy* had tight association with cellulose synthesis and the availability of sucrose in the cell would affect the rate of cellulose synthesis (Coleman et al., 2009).

Flavonoids, including flavone, flavonol, flavanone, isoflavone, and anthocyanin, constitute an important group of plant secondary metabolite, which can enable plants to adjust to environmental pressures (Kovinich et al., 2014). Recent researches showed that these compounds have physiological functions such as antioxidant, bacteriostatic, and anti-inflammatory, which are beneficial to human health. Especially isoflavonoid is predominantly synthesized in legumes plants. Anthocyanin, a class of flavonoids, localized in vacuoles, provided a wide range of colors ranging from orange/red to violet/blue. The content and variety of anthocyanins are the primary determinants of color in many fruit peel and flesh or flowers; the family of MYB and WD40 transcription factors and *DFR* and *CHS* had significant regulatory function on anthocyanin synthesis (Wang et al., 2019; Zhuang et al., 2019).

Among abundant germplasms, including landraces, ecotypes, and cultivars, there is a wide range diversification in pod and seed coat colors and patterns in the common bean. By fine mapping of QTL, the genes controlling the color of the seed coat, pod, stem, and flower were mapped on chromosomes 1 and 3; UDP flavonoid glycosyl transferase (*UGT*) was identified as the candidate gene for black seed coat in Adzuki Bean (Li et al., 2017). Through an Andean intragene pool recombinant inbred line (RIL) population, 23 QTLs for 6 pod traits were detected, and 4 QTLs for pod color were identified (Yuste-Lisbona et al., 2014). The R2R3 MYB transcription factor *TcMYB113* regulated green/red pod color in cacao (Motamayor et al., 2013). The characterization of the genetic variability of three stay-green common bean cultivars indicated high initial chlorophyll *a* content and reduced chlorophyll degradation throughout senescence. In the common bean, the pod color, especially purple, is well documented; Lablab with purple pods contained the pigments of anthocyanins and flavonol. The expression patterns of *LpPAL*, *LpF3H*, *LpF30H*, *LpDFR*, *LpANS*, and *LpPAP1* were significantly induced in purple pods compared to the green ones (Cui et al., 2016). In purple kidney bean pod, *PvMYB1*, *PvMYB2* (R2R3 MYB transcription factors), and *PvTT8-1* [basic helix-loop-helix (*bHLH*) transcription factors] might play a crucial role in transcriptional activation of most anthocyanin biosynthetic genes (Hu et al., 2015). A locus, *Prp* (purple pod), having five alleles affecting anthocyanin pigmentation of corolla and pod, was detailed in *Phaseolus vulgaris* L. The allele *Prp* produced dark purple corolla (Okonkwo and Clayberg, 1984). Chlorophyll degradation and chloroplast breakdown may be involved in degreening and be the foundation for the appearance of other colors; however, the regulatory mechanism at the molecular level has not been well studied in the common bean.

The construction of mutant libraries provides researchers and breeders with good resources for fundamental research or

breeding materials. Dalong 1, having unique yellow and tender pods, is an important common bean ecotype with superior nutrition and is a favorite vegetable in northeast China. This cultivar with a determined growth habit does not require trellises to climb. We constructed a gamma radiation mutant library of Dalong 1. Among various phenotypic changes such as leaf morphology, pod dimension, and sterility, a higher proportion of green pod mutations were noted. In order to reveal the physiology and molecules mechanism underlying the yellow to green mutation, we first investigated the chlorophyll and cellulose content and performed transcriptome and metabolome analyses at different pod developmental stages. Key genes or metabolites triggered by the degreening of pod were revealed, which would provide new insights into the detailed regulatory mechanism of degreening process in the common bean.

MATERIALS AND METHODS

Common Bean Ecotype of Golden Hook and Its Green Pod Mutation Lines

Common bean mutant population was constructed by exposing the seeds of cultivar Dalong 1 of ecotype of the golden hook to ^{60}Co γ radiations at a dosage of 200 Gy. Phenotype observation was performed at M_3 generation. A total of 76 mutant lines with green pods were noted in this mutant library. The wild-type Dalong 1 (yellow pod, YP) and the green pod mutant M628 (green pod, GP) were cultivated in Harbin City, Heilongjiang Province of China in 2018. After fertilization, the pods were sampled at different growth stages based on pod length: 2, 5, and 10 cm. In total, six treatments, e.g., YP-2, YP-5, YP-10, GP-2, GP-5, and GP-10, were sampled in triplicate. For further confirmation, we selected another two green pod mutant lines (M693 and M756) and one mutant line (M729) whose pod color was segregated (yellow pod named M729Y and green pod named M729G). Three biological replicates were collected per treatment. The samples were immediately frozen in liquid nitrogen and then stored at -80°C for the following experiments.

Chlorophyll Content Measurement and Ultrastructure of Chloroplast

Chlorophyll Determination

Two hundred milligrams of plant sample was extracted in 25 ml of 80% (v/v) ethyl alcohol in a dark at room temperature for 48 h. The absorbance (A) of the supernatant was measured at 665, 649, and 470 nm, recorded as A665, A649, and A470, respectively. The pigment concentration was calculated using the following formulas:

$$\text{Ca}(\text{mg/L}) = 13.95\text{A}_{665} - 6.88\text{A}_{649}$$

$$\text{Cb}(\text{mg/L}) = 24.96\text{A}_{649} - 7.32\text{A}_{665}$$

$$\text{Ct}(\text{mg/L}) = \text{Ca} + \text{Cb}$$

$$\text{Cxc}(\text{mg/L}) = (1000\text{A}_{470} - 2.05\text{Ca} - 114.8\text{Cb})/245$$

The pigment content was calculated as pigment concentration multiplied by extraction liquid product and divided by fresh weight.

Chloroplast Ultrastructure

The pod tissues, 2 mm^3 in volume, were fixed in 2.5% glutaraldehyde fixation solution at 4°C for more than 2 h. The tissues were washed with phosphate-buffered saline (PBS) at 4°C for 15 min, followed by washing with acetone of different concentrations (50, 70, 90, and 100%). After washing, the resin was soaked overnight and polymerized at a high temperature, sliced by a microtome (Leica, Germany), and was stained by uranium acetate and lead citrate. The copper mesh was covered on the sample drop for adsorbing for 5 min and stained with phosphotungstic acid for 30 s. The samples were examined with a transmission electron microscope (H-7650, Hitachi, Japan).

Determination of Cellulose Content

Cell Wall Extraction

Sample (0.3 g) was added with 1 mL 80% ethanol, rapidly homogenized at room temperature, and then incubated at 90°C in a water bath for 20 min. After cooling, samples were centrifuged at $6,000 \times g$ for 10 min, and the supernatant was discarded. The precipitate was washed using 1.5 ml 80% ethanol and acetone, respectively. One milliliter reagent I was added for soaking for 15 h to remove starch. After being centrifuged at $6,000 \times g$ for 10 min, the supernatant was discarded; the precipitate was dried, weighed, and regarded as the cell wall mass (CWM).

Extraction of Cellulose

Five milligrams of dried CWM was homogenized in 0.5 ml distilled water. Then, 0.75 ml of concentrated sulfuric acid was slowly added to tubes in an ice water bath, mixed well and then incubated in ice for 30 min. The tubes were centrifuged at $8,000 \times g$ at 4°C for 10 min; the supernatants were diluted with distilled water 20 times.

Cellulose Determination

One hundred fifty microliters of diluted supernatant was mixed with 35 μl working solution and 315 μl sulfuric acid. A blank tube was added with 150 μl distilled water, 35 μl working solution, and 315 μl sulfuric acid. After mixing, it was placed in water bath at 95°C for 10 min and cooled to room temperature. The absorbance values of blank and sample tubes were read at 620 nm: $\Delta A = A(\text{sample tube}) - A(\text{blank tube})$. W stands for the dry weight of the cell wall mass.

$$\text{Cellulose content (mg/g dry weight)} =$$

$$3.17 \times (\Delta A + 0.0043)/W.$$

RNA Extraction, cDNA Library Construction, and RNA Sequencing

Eighteen samples were used for the transcriptome analysis. Total RNA was extracted from three biological replicates of each sample using TRIzol reagent, according to the manufacturer's

instructions. RNA purity was checked using a NanoPhotometer spectrophotometer (IMPLEN, CA, United States). RNA concentration was measured using a Qubit RNA Assay Kit in Qubit 2.0 Fluorometer (Life Technologies, CA, United States). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, United States).

A total of 1.5 µg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext Ultra RNA Library Prep Kit (New England Biolabs, United States), and index codes were added to attribute sequences to each sample. The library preparations were sequenced on Illumina HiSeq platform, and paired-end reads were generated. Gene function annotation and DEGs were analyzed.

Gene function was annotated using the following databases: NCBI non-redundant protein sequences (Nr), Clusters of Orthologous Groups of protein (KOG), GO, and KEGG. All the unigenes were searched against KO, KOG, and KEGG databases using the BLASTX algorithm.

The clean reads were mapped back onto the assembled reference transcriptome (EnsemblPlants¹), and gene expression levels were estimated by RSEM (Li and Dewey, 2011) for each sample. Differential expression analysis of two conditions/groups was performed using the DESeq R package. DESeq provides statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted $P < 0.05$ found by DESeq were assigned as differentially expressed. For pathway enrichment analysis, all DEGs of each comparison group were mapped to pathways in the KEGG database by KOBAS software to identify significantly enriched KEGG pathways (Kanehisa et al., 2008).

Real-Time Quantitative PCR Validation

For RT-qPCR, oligonucleotide primers were designed according to each gene's transcript sequence with Primer 3 and Beacon Designer 7 software. *Actin11* (*Phvul.008G011000*) was used as the reference gene (Borges et al., 2012). RT-qPCR was carried out using SYBR Green-based PCR assay in LightCycler 96 (Roche, Switzerland). Each reaction mix contained 1.0 µl of complementary DNAs (cDNAs), 7.5 µl SYBR Premix ExTaq, 0.3 µl PCR forward primer (10 µmol L⁻¹), 0.3 µl PCR reverse primer (10 µmol L⁻¹), and 5.9 µl ddH₂O, to a final volume of 15 µl. The PCR conditions were 95°C for 30 s, followed by 40 cycles of 95°C for 10 s, and 60°C for 30 s. The melting curve conditions were 95°C for 15 s, 60°C for 60 s, and 95°C for 1 s. Each RT-qPCR analysis was performed in triplicate, and the mean was used for RT-qPCR analysis. The relative expression of the genes was calculated according to the method of $2^{-\Delta\Delta C_t}$, and SPSS was used to analyze the data.

¹<http://plants.ensembl.org/index.html>

Preparation of Metabolome Samples and Measurement by HPLC

The freeze-dried pod was homogenized using a mixer mill (MM 400, Retsch) with a zirconia bead for 1.5 min at 30 Hz. One hundred milligrams powder of pod was extracted overnight at 4°C with 1.0 ml 70% aqueous methanol. Following centrifugation at 10,000 × g for 10 min, the extracts were absorbed (CNWBOND Carbon-GCB SPE Cartridge, 250 mg, 3 mL; ANPEL, Shanghai, China²) and filtrated (SCAA-104, 0.22 µm pore size; ANPEL, Shanghai, China³) before liquid chromatography–mass spectrometry (LC-MS) analysis.

HPLC Conditions

The sample extracts were analyzed using an LC-electrospray ionization (ESI)-MS/MS system (HPLC, Shim-pack UFLC SHIMADZU CBM30A system⁴; MS, Applied Biosystems 6500 Q TRAP⁵). The analytical conditions were as follows: HPLC, column, Waters ACQUITY UPLC HSS T3 C18 (1.8 µm, 2.1 mm × 100 mm); solvent system, water (0.04% acetic acid):acetonitrile (0.04% acetic acid); gradient program, 100:0 V/V at 0 min, 5:95 V/V at 11.0 min, 5:95 V/V at 12.0 min, 95:5 V/V at 12.0 min, 95:5 V/V at 15.0 min; flowrate, 0.40 mL/min; temperature, 40°C; and injection volume, 2 µl. The effluent was alternatively connected to an ESI-triple quadrupole-linear ion trap (QTRAP)-MS.

ESI-Q TRAP-MS/MS

Linear ion trap (LIT) and triple quadrupole (QQQ) scans were acquired on a QTRAP mass spectrometer, API 6500 QTRAP LC/MS/MS System, equipped with an ESI turbo ion-spray interface, operating in a positive ion mode and controlled by Analyst 1.6 software (AB Sciex). The ESI source operation parameters were as follows: ion source, turbo spray; source temperature, 500°C; ion spray voltage (IS), 5,500 V; ion source, gas I (GSI), gas II (GSII), and curtain gas (CUR) were set at 55, 60, and 25.0 psi, respectively; the collision gas (CAD) was high. Instrument tuning and mass calibration were performed with 10 and 100 µmol/L polypropylene glycol solutions in QQQ and LIT modes, respectively. QQQ scans were acquired as multiple reaction monitoring (MRM) experiments with collision gas (nitrogen) set to 5 psi. DP and CE for individual MRM transitions were done with further DP and CE optimization. A specific set of MRM transitions was monitored for each period according to the metabolites eluted within this period.

Statistical Analysis

Each experiment was repeated three times. All data were recorded as the means ± standard deviation (SD) and analyzed using a one-way analysis of variance (ANOVA) with SPSS software 19.0. Statistical difference between samples (* $P \leq 0.05$ was considered to be significant; ** $P \leq 0.01$ was considered to be extremely significant) was determined by Duncan's test.

²www.anpel.com.cn/cnw

³<http://www.anpel.com.cn/>

⁴www.shimadzu.com.cn/

⁵www.appliedbiosystems.com.cn/

RESULTS

A High Rate of Yellow-to-Green Mutation Was Observed in a Gamma Radiation-Induced Mutant Library

A mutant library was constructed by exposing seeds of Dalong 1 to ^{60}Co γ radiations at a dosage of 200 Gy. Compared with the wild type, this mutant library displayed a wide range of phenotypic variations in traits, e.g., 100—seed weight, flower time, plant height, sterility, pod width, pod, and seed coat color among 3,500 plants in M_3 lines. In contrast to yellow pod of wild-type Dalong 1, there were 76 mutant lines showing green pods or green and yellow segregation in a total of 607 M_3 lines, much higher mutation rate than other traits such as sterility (33 mutant lines). Among the green pod mutant lines, the pod color was unanimously green; however, the green and yellow segregation displayed green and yellow pods segregation, approximately in a green to yellow ratio of 3:1 or 15:1.

The pod color of wild-type Dalong 1 was green at the beginning of pod development (2 cm long), which rapidly turned to yellow as pod extended to approximately 10 cm long. The pod color of green pod mutant lines stayed green until eventual maturation. The pods of Dalong 1 appeared a little crooked, but green pod mutant was relatively straight (**Figure 1A**). From the intersection, the tissues underneath the epidermis of the pod also appeared green in consistent with pod appearance (**Figures 1B,C**).

Initially, a rapid colorimetric assay was employed to measure the chlorophyll and carotenoid contents during pod development. As early as the developmental stage of 2-cm-long pods, the chlorophyll and carotenoid contents in the green pod mutants were higher than that in the wild type. During pod development, the chlorophyll content in the wild type decreased sharply. In contrast, the chlorophyll content in green pod mutant lines only slightly decreased (**Figures 1E–G**).

The chloroplast ultrastructure of pods at 10 cm long in green pod mutant and wild type was compared using TEM. In contrast to the wild type, the chloroplast of the mutant remained a well-organized grana thylakoids and had more starch grain and less plastoglobuli (**Figures 1H,I**), indicating that the chloroplasts were capable of making chlorophyll.

Transcriptome Profiles

In order to reveal the transcriptional abundance of genes involved in degreening process, transcriptome was performed using equal amounts of RNA from pods. After removing low-quality reads, adaptor sequences, and ribosomal RNA (rRNA) reads, we obtained 50407645 (YP-2), 48995896 (YP-5), and 54170305 (YP-10) and 59445277 (GP-2), 51527246 (GP-5), and 49269434 (GP-10) clean reads, respectively (**Supplementary Table S1**). Clean reads were then mapped to the common bean reference genome using HISTAT software. Comparing YP-2 to YP-5, there were 4,916 genes downregulated and 3,974 genes upregulated, while only 203 genes were downregulated and 136 genes upregulated when YP-5 was compared with YP-10. In the green pod mutant, the down- and upregulated genes were 780

and 406 for GP-2 vs. GP-5 and 3,228 and 1,700 for GP-5 vs. GP-10 (**Figure 2A**). Three compared combinations (YP-2 vs. YP-5, YP-5 vs. YP-10, and YP-2 vs. YP-10 and GP-2 vs. GP-5, GP-5 vs. GP-10, and GP-2 vs. GP-10) shared 114 and 150 DEGs, respectively (**Figure 2C**).

To further determine the function of the DEGs during pod development, KEGG analysis was performed. KEGG enrichment analysis revealed that limonene and pinene degradation, histidine metabolism, fatty acid elongation, and protein processing in the endoplasmic reticulum were the prominently enriched for YP-2 vs. GP-2 (**Figure 2D**). Comparing YP-5 to GP-5, many pathways showed enrichment, such as fructose and mannose metabolism, fatty acid metabolism, amino sugar, and nucleotide sugar metabolism. Starch and sucrose metabolism and isoflavonoid biosynthesis were also noticeably enriched (**Figure 2E**). Neuroactive ligand–receptor interaction and D-glutamine and D-glutamate metabolism were the two most obvious enrichment pathways for YP-10 vs. GP-10; other pathways, e.g., circadian rhythm plant, phenylpropanoid biosynthesis, and carotenoid biosynthesis, were also enriched to a certain extent (**Figure 2F**).

As the pods extended, many marked changes occurred in numerous crucial pathways, such as glycolysis, starch and sucrose metabolism, porphyrin and chlorophyll metabolism, carotenoid biosynthesis, phenylpropanoid biosynthesis, carbon metabolism, and plant hormone signal transduction. In the wild type, radical changes in expression levels occurred from YP-2 to YP-5, while in the green pod mutation line, contrasting changes in expression levels were observed from GP-5 to GP-10. Numerous pathways were activated at the early stages in the wild type; there were a large number of DEGs for YP-2 vs. YP-5 but fewer DEGs for YP-5 vs. YP-10. In contrast, more DEGs for GP-5 vs. GP-10 than GP-2 vs. GP-5 were detected (**Figure 2G**).

Key Genes Involved in Chlorophyll Synthesis and Degradation Were Revealed by Transcriptome

As the chlorophyll content decreased with the pods degreening (**Figure 1E**), we focused on porphyrin and chlorophyll metabolism pathway. It was evident from **Supplementary Figure S3** that the expression levels of the genes for chlorophyll synthesis generally showed a downward trend during pod development both in wild type and green pod mutant (**Supplementary Figure S3**). The genes for chlorophyll synthesis, e.g., as *HEMC* (*Phvul.002G034500*), glutamyl transfer RNA (tRNA) reductase (*HEMA*, *Phvul.002G216100*), *CHLI* (*Phvul.003G057600*, *Phvul.006G178400*), and *POR* (*Phvul.005G083700*, *Phvul.011G148900*), were decreasing throughout in the wild type (**Supplementary Figure S3A**), whereas the genes for chlorophyll synthesis in the green pod mutant increased slightly from GP-2 to GP-5 and thereafter decreased (**Supplementary Figure S3A**). In particular, the transcription of *POR* (*Phvul.005G083700*) showed the opposite pattern when the pods extended from 2 to 5 cm, which was increased in the green pod mutant but decreased steeply in wild type (**Supplementary Figure S3A**).

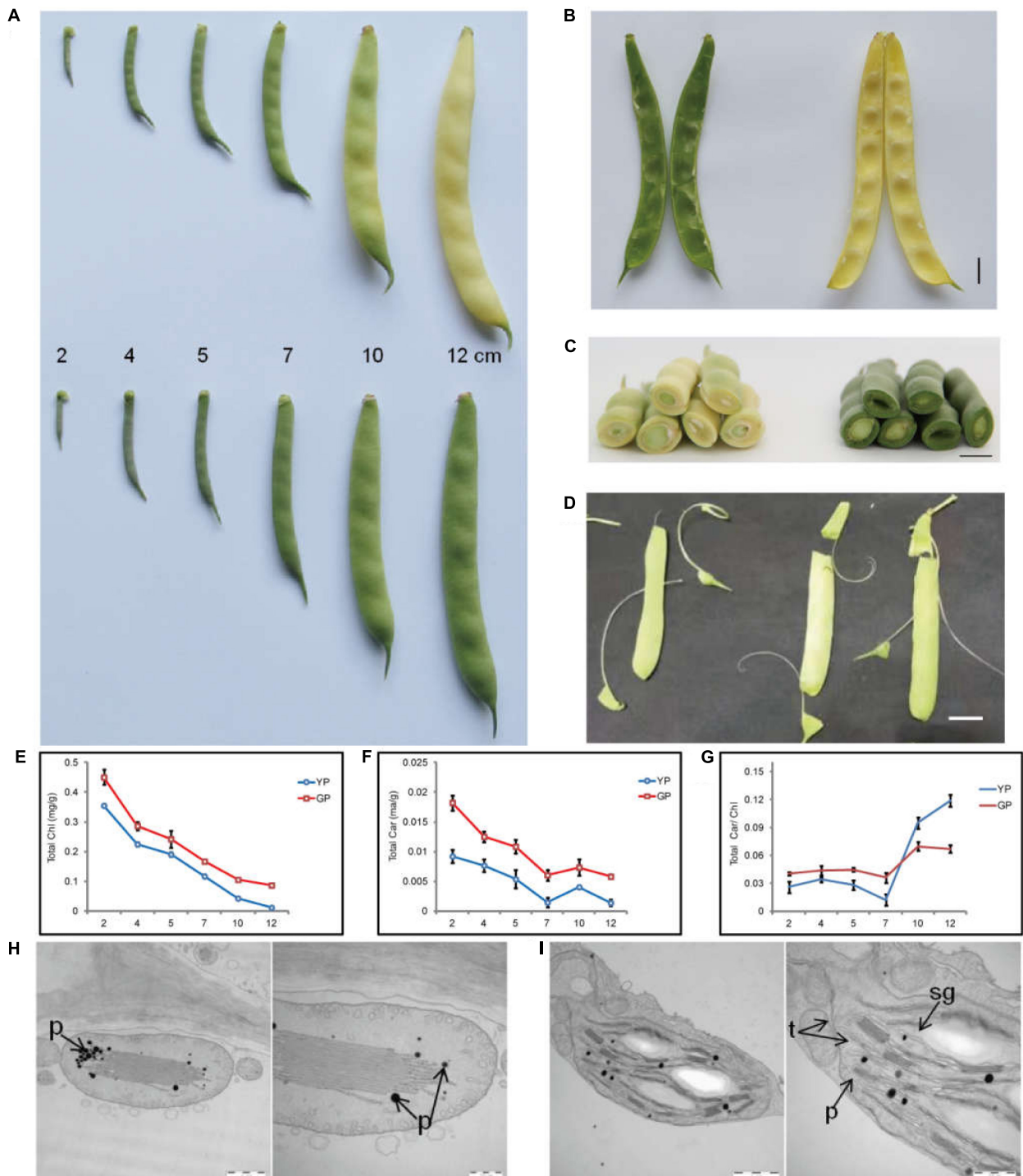


FIGURE 1 | Pod color conversion, pigment content changes, and chloroplast microstructure during pod development. **(A)** Phenotypes of common bean pods of wild type (YP) and green pod mutant (GP). **(B,C)** phenotypes of pod interior and cross-section of the wild type and green pod mutant. **(D)** Phenotypes of green pod mutant with high cellulose content in the pods. **(E–G)** Total chlorophyll and carotenoid contents and the Car/Chl ratio at all developmental stages. **(H,I)** Structure of chloroplasts from wild type **(H)** and mutant **(I)**; t, thylakoids, sg, starch grain, p, plastoglobuli. Scale bar: 1 μ m and 500 nm.

The expression levels of chlorophyll degradation-related genes started to accumulate at early stage of wild type; notably, *RCCR* (*Phvul.008G280300*) had high expression levels at YP-2

(**Figures 3A,B**) and then gradually decreased. Similar trends were observed for the genes of *PAO* (*Phvul.011G086700*) and *SGR2* (*Phvul.002G153100*) in the wild type (**Figure 3A**). In the green

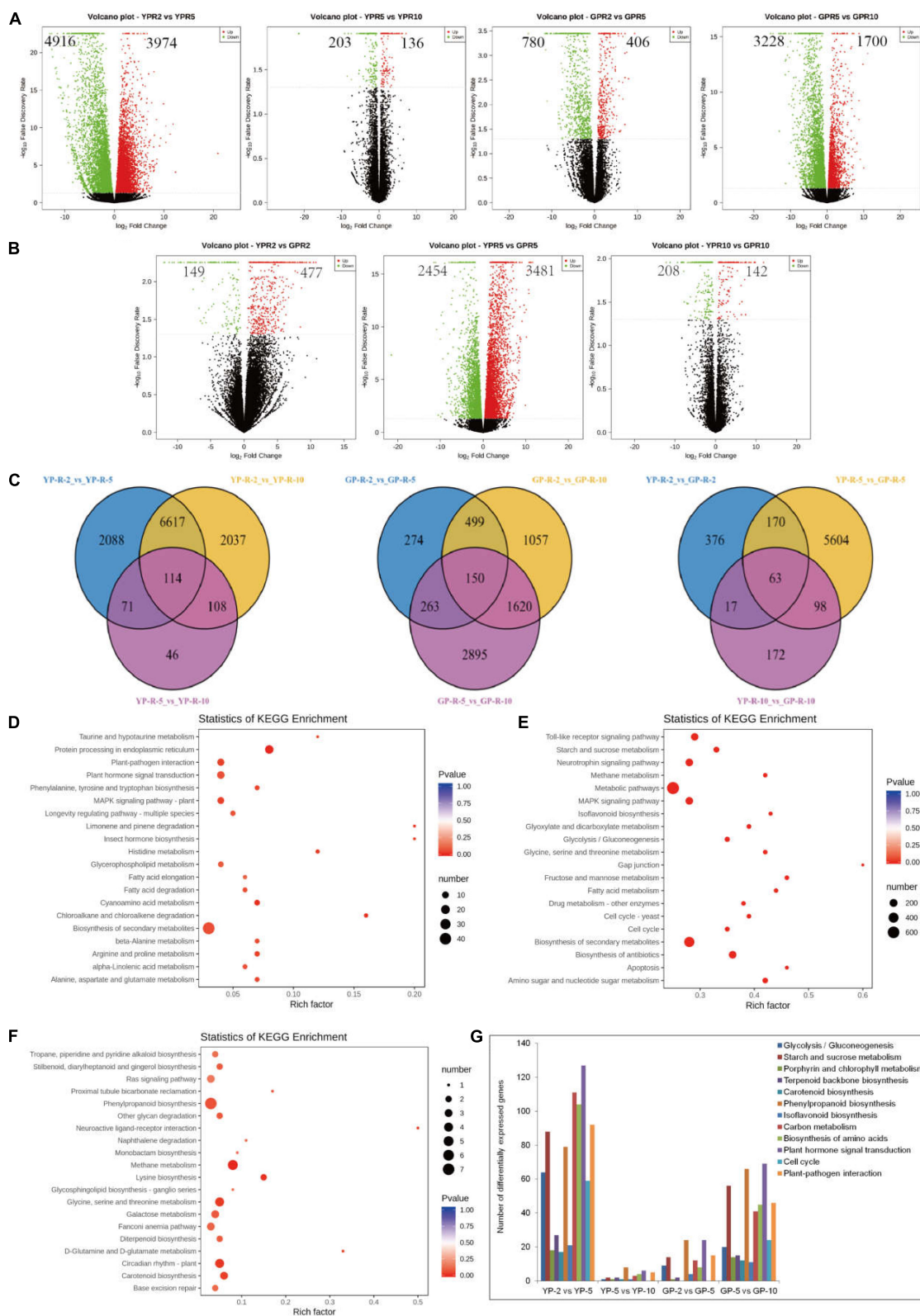


FIGURE 2 | The differential expression genes analysis between wild type and green pod mutant by transcriptome. **(A,B)** The volcano plot showing the number of differential expression genes between two comparative groups. **(C)** Venn diagram showing common and specific DEGs numbers from different combinations displayed in the overlapping and non-overlapping regions, respectively. **(D–F)** Top 20 significantly enriched KEGG pathways in **(D)** YP-2 vs. GP-2, **(E)** YP-5 vs. GP-5, and **(F)** YP-10 vs. GP-10. **(G)** The statistics of DEGs in some metabolic pathway at two developmental stages of wild type and green pod mutant.

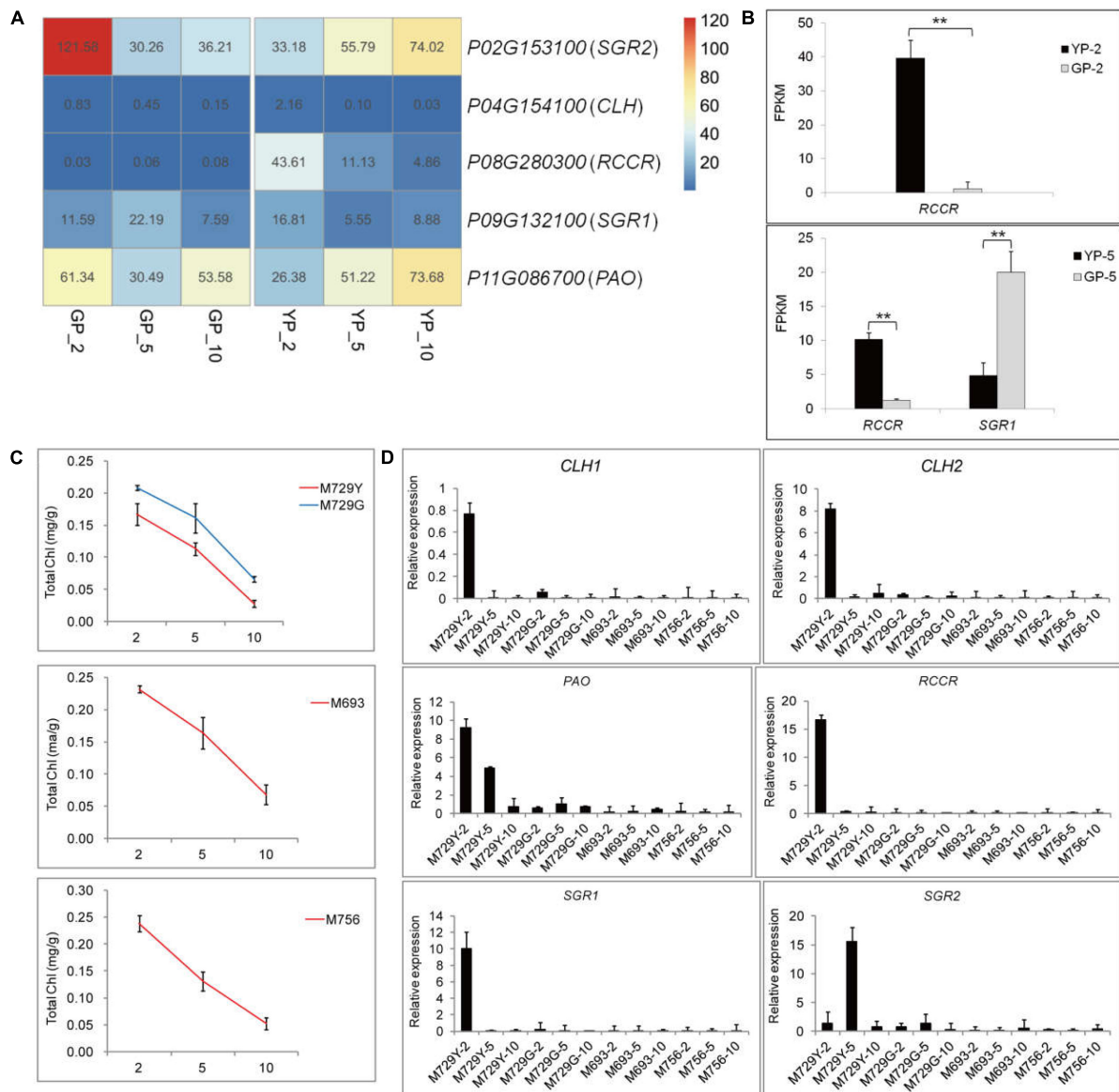


FIGURE 3 | The differential expression genes analysis in porphyrin and chlorophyll metabolism pathway of YP, GP, and the other green pod mutant lines. **(A)** The changes in gene expression level of degradation in YP and GP, respectively. **(B)** Expression levels of DEGs for chlorophyll degradation between YP-2 and GP-2 and YP-5 and GP-5, respectively. **(C)** The chlorophyll content in yellow pods (M729Y) and three green pod mutants (M729G, M693, and M756), respectively. **(D)** Expression levels of DEGs for chlorophyll degradation successively in the M729Y, M729G, M693, and M756 mutant lines. Error bars represent the standard deviation of three replicates. Data were analyzed using the *t*-test. **p* < 0.05, ***p* < 0.01.

pod mutant, the expression levels of *CLH* (*Phvul.007G278100*) increased gradually, and *SGR* (*Phvul.009G132100*) increased from GP-2 to GP-5 and then decreased sharply (Figure 3A). The expression levels of *RCCR* (*Phvul.008G280300*) in GP were extremely low throughout the pod development (Figures 3A,B). The aggravated degradation of chlorophyll together with the decreased synthesis resulted in the chlorophyll content in the pods decreased sharply, leading to the degreening phenotype.

The other four mutant lines (M729Y, M729G, M693, and M756) were also selected to analyze the variations of chlorophyll

content and expression levels of genes involved in chlorophyll metabolism pathway. Similarly, the chlorophyll content declined throughout in these mutant lines (Figure 3C). The similar expression pattern of the genes for chlorophyll synthesis was found in the other green pod mutant lines (Supplementary Figure S3B). The genes involved in chlorophyll degradation were highly expressed in the M729Y with yellow pods; especially the expression levels of *RCCR* were high at the initial stage. However, the expression levels of genes for chlorophyll degradation in other green pod mutant lines were low (M729G, M693, and M756) (Figure 3D).

Cellulose Contents and Transcriptional Abundance of Genes Involved in Cellulose Synthesis and Degradation

The tender pods of common bean are always consumed as vegetable, especially in Chinese recipes. In some cultivars, fibrosis in the pod leads to a hardened pod and loss of flavor. Therefore, eliminating fibrosis is a main target for breeding. After measuring, the cellulose content in the green pod mutant was higher than that in the wild type (Figure 4A). This result was also verified in other green pod mutant lines; the green pod mutant lines had higher cellulose content (Figures 4B,C). Therefore, we focused on cellulose synthesis and degradation and cellulose synthase genes. The genes of *Phvul.004G093300*, *Phvul.005G022100*, and *Phvul.009G242700* participating in cellulose synthesis had distinctly higher expression in GP than in YP samples (Figures 4D–F). The cellulose in the green pod mutant appeared to rapidly synthesized, leading to a higher cellulose content. There were also more genes involved in cellulose degradation in GP than in YP, which might be due to the degradation activity of cellulose being triggered by a higher cellulose content (Figures 4G–I).

Real-Time Quantitative PCR Validation

To validate the key DEGs, the DEGs that showed significant differences in expression levels were selected: 12 chlorophyll synthesis pathway genes, 6 chlorophyll degradation genes, and 5 and 4 genes for cellulose synthesis and degradation, respectively. To investigate the changes in expression levels of the genes related to chlorophyll pathway, six different development stages of pods were selected. The results of RNA-seq were confirmed by RT-qPCR. Particularly, some genes, such as *HEMA2*, *HEME1*, *HEMF*, *CHLI2*, *CAO*, and *POR*, had a high expression when the pods were 7 cm (Figure 5A). For chlorophyll degradation, *PAO* and *SGR* played a significant role. The expression levels of *RCCR* in the green pod mutant were low; however, they were very high in the early stage in the wild type (Figure 5B). The expression levels of cellulose synthesis and degradation genes showed a similar pattern to the transcriptome data (Figures 5C,D).

Meanwhile, the expression levels of the genes related to chlorophyll pathway were studied in different tissues (root, stem, leaf, and flower) using RT-qPCR (Figure 6). *CHLI2* and *POR* had a high expression in leaf, flower, and stem (Figures 6B,D,F); the expression of *HEMF* was higher in root (Figure 6H). *PAO* and *SGR2* had a high expression in leaf and flower (Figures 6C,E); the expression of *CLH2* was high in leaf and stem (Figures 6C,G). Compared to other genes, the expression levels of *CLH1*, *RCCR*, and *SGR1* were lower. Based on the above results, the mechanism model of pod degreening was speculated (Figure 7).

Transcriptome and Metabolome Variations in the Flavonoid Pathway

To determine the contents and composition variations of flavonoids during pod development, we performed metabolisms analysis by widely targeted metabolomics. A total of 154

flavonoid metabolites were identified from 18 samples (YP-2, YP-5, YP-10, GP-2, GP-5, and GP-10), each with three biological replicates. Setting $VIP \geq 1$ together with fold change ≥ 2 or ≤ 0.5 as thresholds for significant differences, there were 33, 42, and 33 differential metabolites in the three comparison groups: YP-2 vs. GP-2, YP-5 vs. GP-5, and YP-10 vs. GP-10, respectively. Comparing the three comparison groups, 11, 12, and 7 metabolites were upregulated, and 22, 30, and 36 metabolites were downregulated (Figure 8).

Hierarchical cluster analysis of the flavonoids showed three main clusters based on the relative differences in accumulation patterns (Figure 8A), which indicated the complex differences in the metabolite levels. In the PCA plot, PCA on all metabolites indicated that PC1, explaining 44.75% of the total variance, separated samples of three different stages. PC2, accounting for 16.43% of the total variance, separated samples of YP from those of GP (Figure 8B). In the Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) models, YP and GP samples were clearly separated (Figures 8C–E), indicating a major distinction in the metabolic profiles between the two different color groups.

In the comparison group, YP-2 vs. GP-2 (Table 1), naringin, vitexin 2''-O-beta-L-rhamnoside, and isotrifoliin were significantly higher in the YP-2, especially vitexin 2''-O-beta-L-rhamnoside (fold change = 0.0009, $VIP = 3.26$). Compared to YP-5 (Table 1), the content of three isoflavones (prunetin, biochanin A, and sissotrin) was higher in GP-5 and was 4.15-, 2.88-, and 8.81-fold, respectively. In YP-10, the content of 3,7-di-O-methylquercetin was significantly higher; other two flavonols (astragalin and kaempferin) were also higher in YP-10. Two flavonoids, luteolin and apigenin, were higher in GP-10. Compared to the wild type, isoflavones were largely accumulated in the green pod mutant, especially sissotrin. The content of five detected anthocyanin (oenin, pelargonin, cyanin, cyanidin, and keracyanin) was higher in the wild type at three different stages (Table 1).

Analysis of transcriptome data revealed that most DEGs involved in the flavonoids pathway showed to be downregulated (Supplementary Figures S4A,B); there were more DEGs in flavones and isoflavones pathways in the wild type. Many genes of YP and GP samples showed similar expression trends, such as *DFR2*, *ANR*, *LAR*, *DFR3*, *IF7MAT6*, *VR4* (Supplementary Figures S4A–D). There were more genes expressed in the pods at 5 cm long (YP-5, GP-5) (Supplementary Figure S5). By KEGG enrichment analysis, the DEGs (YP-5/GP-5) of isoflavonoid biosynthesis pathway were enriched. The genes involving in anthocyanin synthesis were rarely detected.

DISCUSSION

The Molecular Mechanism of Degreening Process

The loss of green color in senescent leaves and ripening fruits is a noticeable natural phenomenon. It is well known that the color conversion is generally accompanied by chlorophyll degradation; the chlorophyll content constantly decreased during

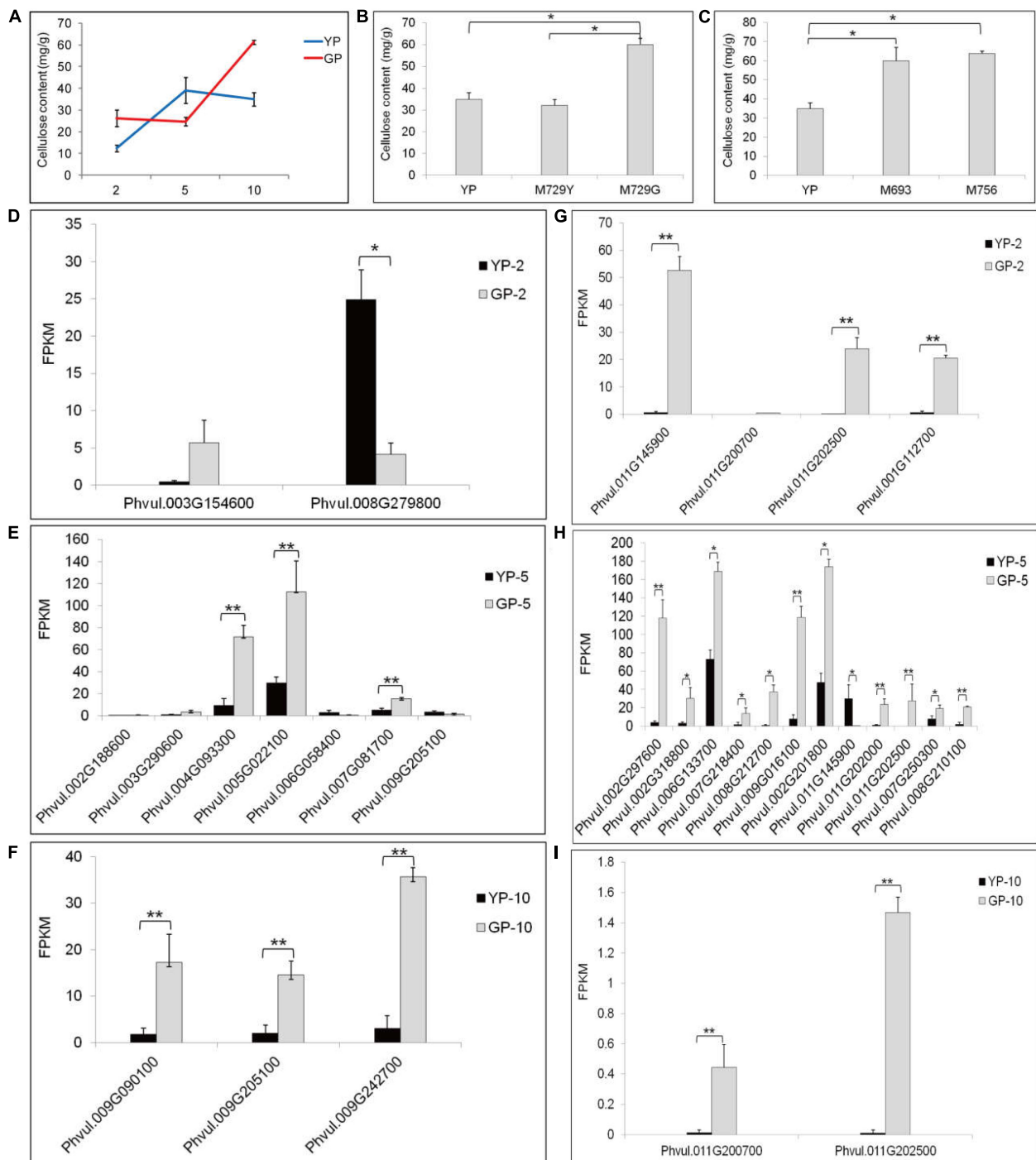


FIGURE 4 | The cellulose content and the differential expression genes analysis in starch and sucrose metabolism pathway of YP and GP. **(A)** The cellulose content of wild type and green pod mutant. **(B,C)** The cellulose content in the other green pod mutant lines, **(B)** M729Y and M729G and **(C)** M693 and M756. **(D–F)** Differential expression genes for cellulose synthesis between YP-2 and GP-2, YP-5 and GP-5, and YP-10 and GP-10, respectively. *Phvul.004G093300* (cellulose synthase A catalytic subunit 2, *CESA2*), *Phvul.005G022100* (*CESA3*), *Phvul.002G188600*, *Phvul.009G090100* and *Phvul.009G242700* (*CESA4*), *Phvul.007G081700* (*CESA6*), *Phvul.003G154600*, *Phvul.009G205100* (*CESA7*), *Phvul.003G290600* (cellulose synthase-like protein G2, *CSLG2*), *Phvul.008G279800*, and *Phvul.006G058400* (cellulose synthase-like protein E6, *CSLE6*). **(G–I)** Differential expression genes for cellulose degradation between YP-2 and GP-2, YP-5 and GP-5, and YP-10 and GP-10, respectively. *Phvul.011G145900*, *Phvul.011G200700*, *Phvul.011G202500*, *Phvul.011G112700*, *Phvul.002G201800*, *Phvul.011G145900*, *Phvul.011G202000*, *Phvul.011G202500*, *Phvul.007G250300*, and *Phvul.008G210100* (beta-glucosidase). *Phvul.002G297600*, *Phvul.002G318800*, *Phvul.006G133700*, *Phvul.007G218400*, *Phvul.008G212700*, and *Phvul.009G016100* (endoglucanase). Error bars represent the standard deviation of three replicates. Data were analyzed using the t-test. * $p < 0.05$, ** $p < 0.01$.

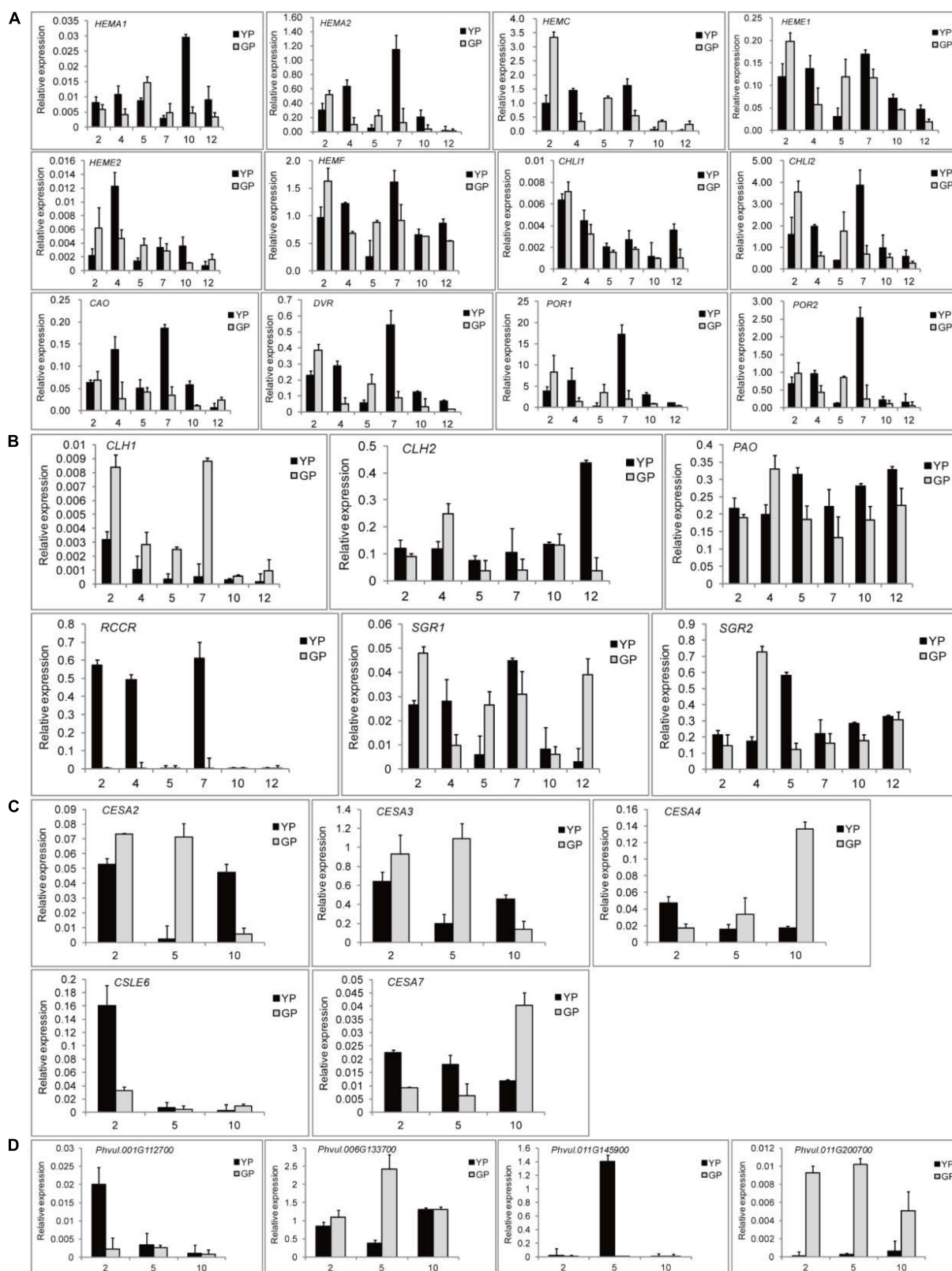


FIGURE 5 | Quantitative real-time PCR (RT-qPCR) validation of DEGs found throughout transcriptome between the wild type and mutant. **(A)** DEGs involved in chlorophyll synthesis. **(B)** DEGs involved in chlorophyll degradation. **(C)** DEGs involved in cellulose synthesis. **(D)** DEGs involved in cellulose degradation.

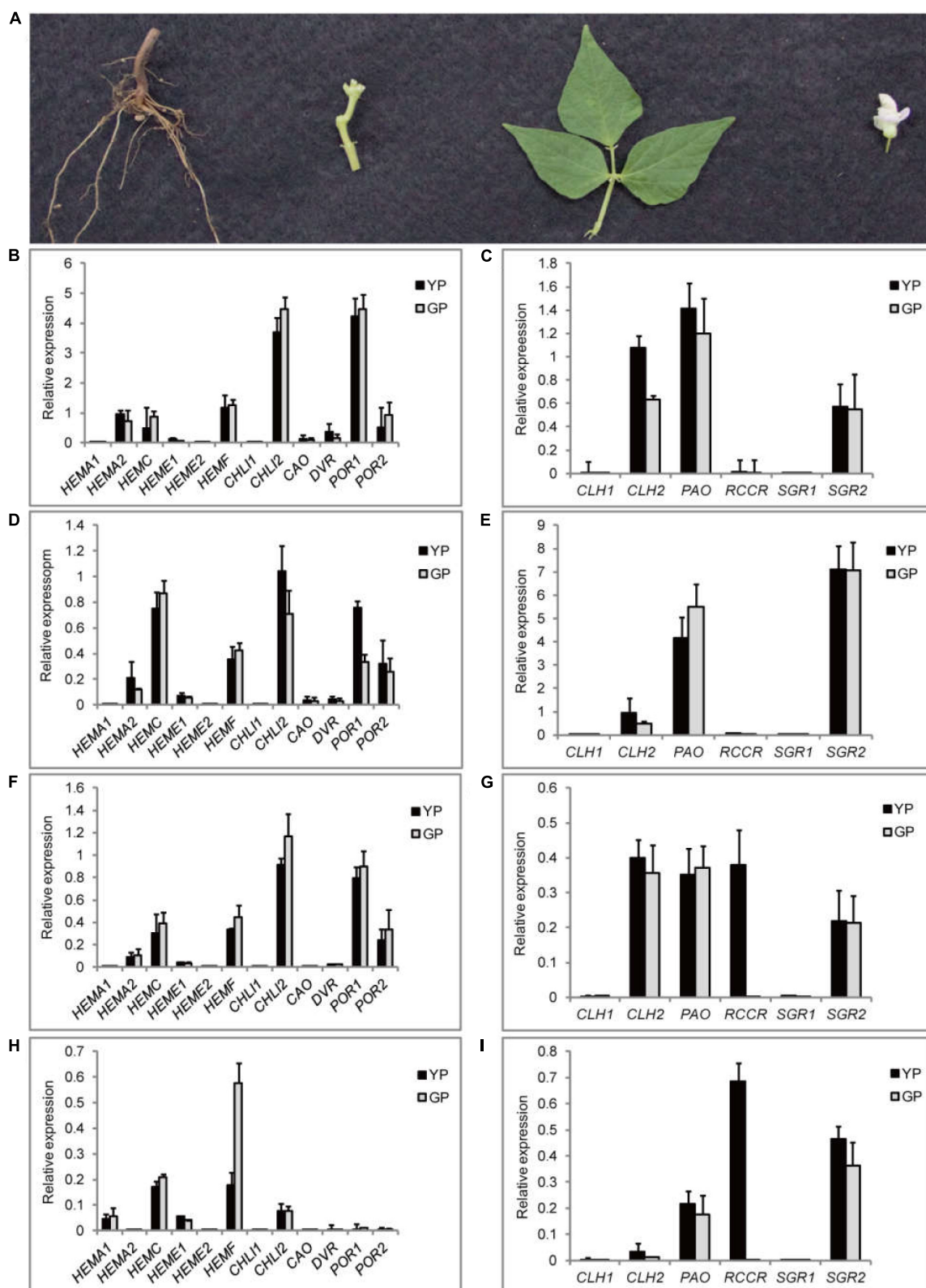
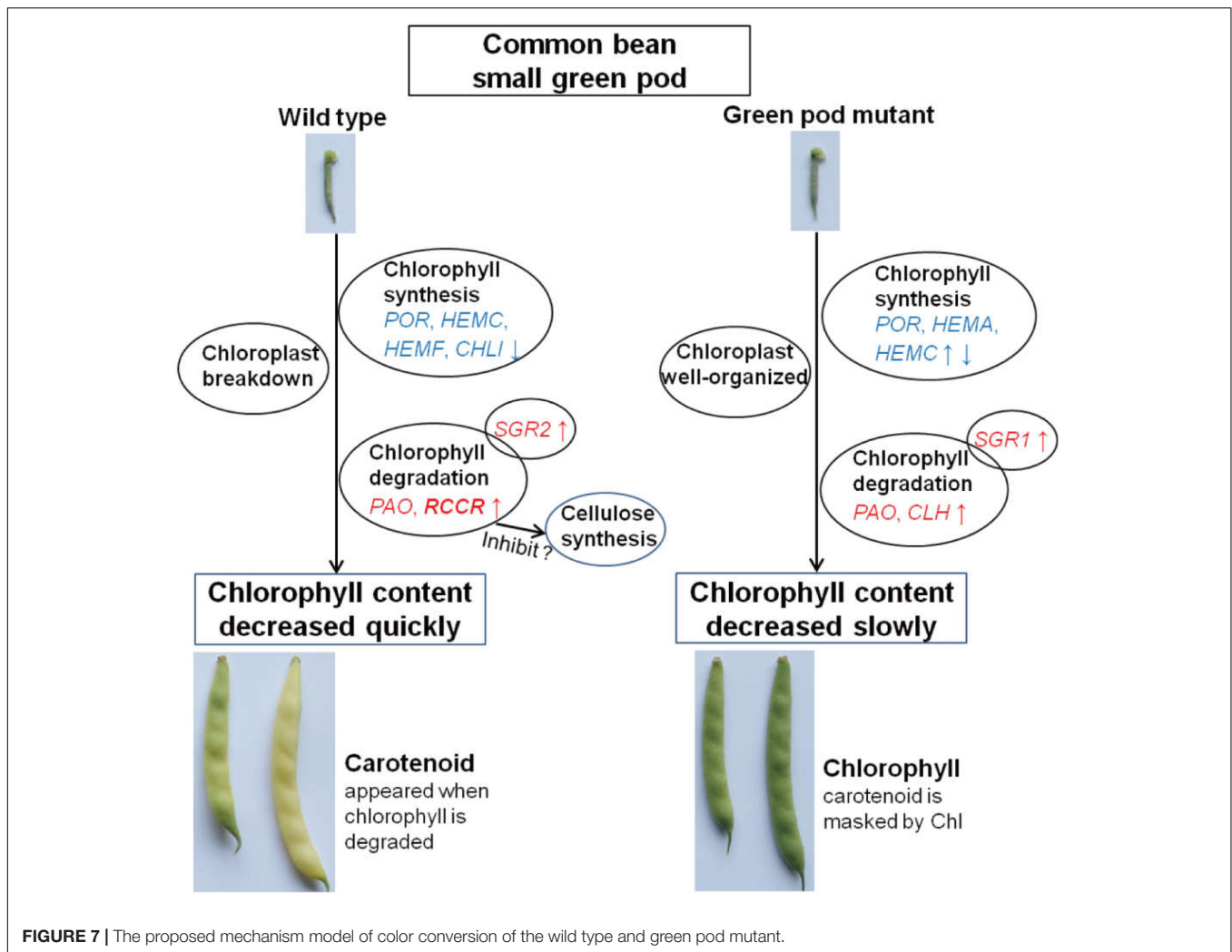


FIGURE 6 | Transcription levels of 18 DEGs involving in chlorophyll synthesis and degradation pathway in different tissues. **(A)** The pictures of different organs of common bean. **(B–I)** The relative transcript levels of chlorophyll synthesis and degradation, respectively, in four different organs: **(B,C)** leaf, **(D,E)** flower, **(F,G)** stem, and **(H,I)** root.



the green-peel fruit ripening (Gambi et al., 2018). During the pod development of the golden hook, the green pod color rapidly turned yellow, showing an inherent characteristic as an ecotype of the common bean. The yellow pods of the golden hook are tender with less cellulose and easy to cook.

In this study, the color transition from green to yellow resulted from the rapid degradation of chlorophyll and the appearance of carotenoids. The PAO degradation pathway of chlorophyll is a multistep enzymatic process, in which *CLH*, *PAO*, *RCCR*, and stay-green (*SGR*) genes have been documented (Tsuchiya et al., 1999; Pruzinska et al., 2003; Harpaz-Saad et al., 2007). *CLH*, the first step enzyme in the chlorophyll catabolic pathway, acts as a rate-limiting enzyme controlled via posttranslational regulation (Harpaz-Saad et al., 2007). In this study, *CLH1* showed low expression in both YP and GP, whereas *CLH2* was upregulated in the mutant during pod development. Previous researches had shown that *PAO* and *RCCR* participated in the chlorophyll catabolic pathway. *PAO* is a key enzyme and is upregulated during leaf senescence (Ginsburg et al., 1994; Hörtensteiner et al., 1995; Pruzinska et al., 2003). The *PAO* activity seemed to be responsible for the stay-green genotype

in pepper (Roca and Minguéz-Mosquera, 2006). The expression of *RCCR* is constitutive (Rodoni et al., 1997; Tsuchiya et al., 1999). In this study, the expression levels of *PAO* were high in both wild type and green pod mutant but particularly were increased in the wild type during pod development. The expression of *RCCR* was very high in small tender pod of the wild type and then declined; this means that *RCCR* acted as key enzyme for chlorophyll degradation in early stages in the wild type. There was an extremely low expression of *RCCR* in the green pod mutant lines. Some studies suggested that *SGR* may take part in the chloroplast breakdown process (Kusaba et al., 2007; Park et al., 2007). *SGR*, as a chloroplast protein, has an important role in chlorophyll degradation by catabolic enzymes and proteases through inducing LHCP II disassembly (Aubry et al., 2008). Our results support this notion; the expression levels of *SGR2* showed significant changes, but the expression levels of *SGR1* were low. The phytohormone ethylene and jasmonic acid could affect the chlorophyll degradation pathway; they could increase the expression of *Chl* catabolic genes (Fang et al., 2020; Lv et al., 2020). Compared to the green pod mutant, the expression levels of chlorophyll degradation genes in the

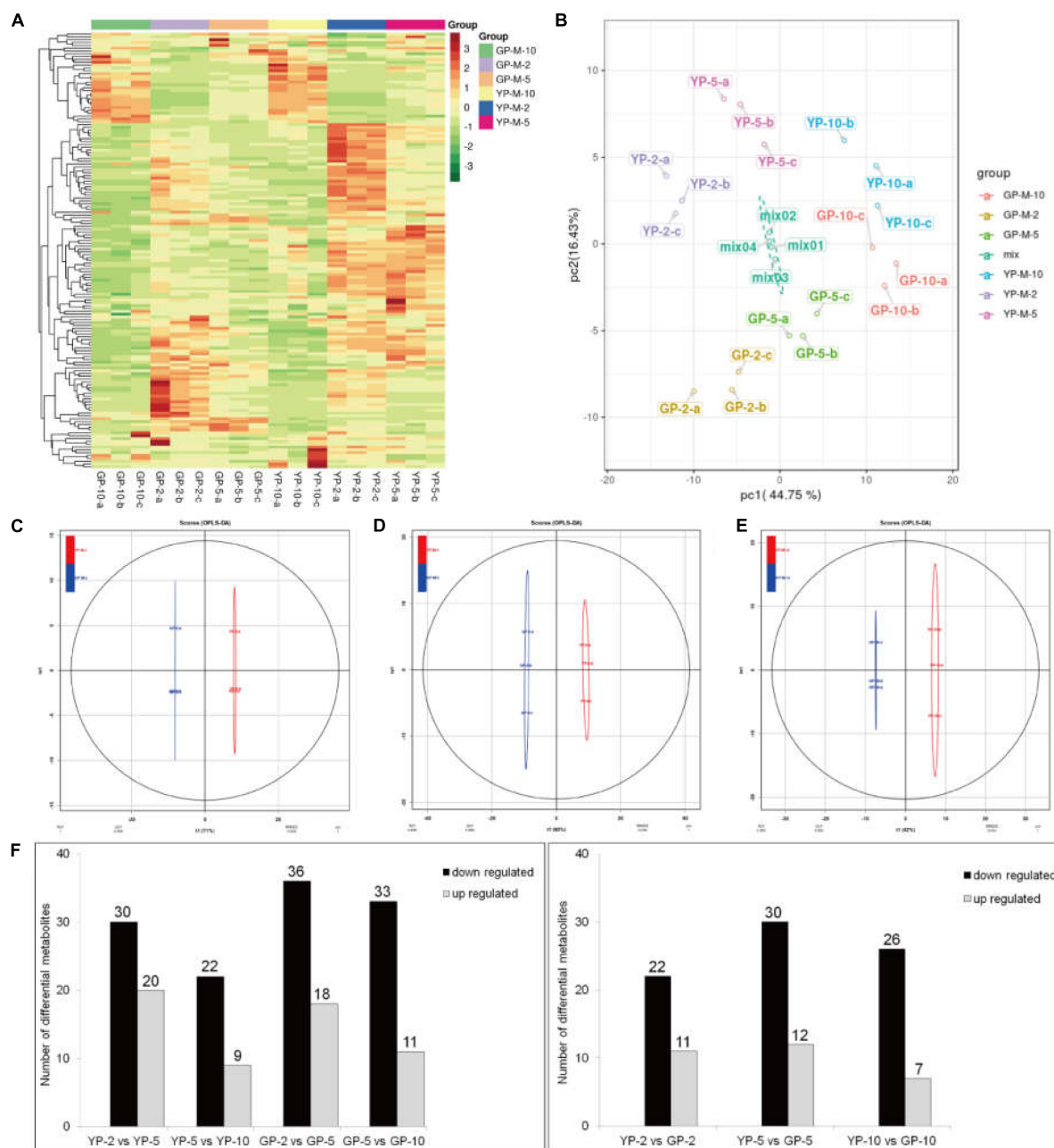


FIGURE 8 | Heat map visualization, PCA, and OPLS-DA of the relative difference in flavonoids metabolites of YP and GP. **(A)** Heat map of all detected metabolites. The content of each metabolite was normalized to complete linkage hierarchical clustering. Each example is visualized in a single column, and each metabolite is represented by a single row. Red indicates high abundance; green showed low relative content. **(B)** Score plots for principle components 1 and 2 within and between groups. **(C–E)** OPLS-DA model plots for YP versus GP at different comparison groups in the same period. **(F)** The differentially metabolism analysis between wild type and green pod mutant at different comparable groups.

wild type was earlier and higher. *PAO*, *RCCR*, and *SGR* might be the key genes in the golden hook involved in chlorophyll degradation; especially *RCCR* played a much more important role. Analysis of the resequencing data of the wild type and green pod mutant revealed that there were some no-sense mutations occurring in the genes involving in chlorophyll degradation (Supplementary Table S3).

In this study, comparing wild type to green pod mutant, only the pods turned yellow during pod development; there were no noticeable color differences in the other tissues. The expression levels of the genes for chlorophyll degradation were low in the leaf and stem. According to the results of this study, we developed a mechanic model describing degreening process (Figure 7), in which the chlorophyll content decreased, and the structure of

TABLE 1 | Differentially accumulated flavonoids compounds in the pod of wild type and green pod mutant.

YP-2 vs. GP-2	Index	Metabolite name	Class	Content		VIP	Fold change	Type
				YP-2	GP-2			
Flavonoid	pme0331	Naringin	Flavanone	79,200.00	38,166.67	1.03E + 00	4.82E − 01	Down
	pme3393	Fustin	Flavonol	18,166.67	43,833.33	1.16E + 00	2.41E + 00	Up
	pme0088	Luteolin	Flavone	219,333.33	593,666.67	1.16E + 00	2.71E + 00	Up
Flavone and flavonol	pme0363	Chrysoeriol	Flavone	265,333.33	633,333.33	1.09E + 00	2.39E + 00	Up
	pme3227	Vitexin 2''-O-beta-L-rhamnoside	Flavone	9,623.33	9.00	3.26E + 00	9.35E − 04	Down
	pme3212	Quercetin 3-O-glucoside	Flavonol	4,450,000.00	1,413,333.33	1.29E + 00	3.18E − 01	Down
Anthocyanin	pme0444	Malvidin 3-O-glucoside	Anthocyanins	196,333.33	90,933.33	1.04E + 00	4.63E − 01	Down
	pme1793	Pelargonin	Anthocyanins	330,666.67	89,533.33	1.38E + 00	2.71E − 01	Down
	pme1777	Cyanidin 3,5-O-diglucoside	Anthocyanins	49,966.67	23,966.67	1.02E + 00	4.80E − 01	Down
YP-5 vs. GP-5				YP-5	GP-5			
Flavonoid	pme0199	Quercetin	Flavonol	488,000.00	179,666.67	1.03E + 00	3.68E − 01	Down
	pme0088	Luteolin	Flavone	31,800.00	105,400.00	1.02E + 00	3.31E + 00	Up
	pme0196	Kaempferol	Flavonol	19,866.67	3,439.33	2.21E + 00	1.73E − 01	Down
Flavone and flavonol	pme3296	Kaempferin	Flavonol	12,933.33	3,746.67	1.20E + 00	2.90E − 01	Down
	pme1541	Acacetin	Flavone	9,696.67	38,300.00	1.22E + 00	3.95E + 00	Up
Isoflavonoid	pme3292	Prunetin	Isoflavone	11,070.00	45,933.33	1.27E + 00	4.15E + 00	Up
	pme3250	Biochanin A	Isoflavone	849.33	6,273.33	2.07E + 00	7.39E + 00	Up
	pme3399	Sissotrin	Isoflavone	10,406.67	91,633.33	1.59E + 00	8.81E + 00	Up
Anthocyanin	pme3609	Cyanidin	Anthocyanins	160,000.00	64,733.33	1.02E + 00	4.05E − 01	Down
YP-10 vs. GP-10				YP-10	GP-10			
Flavonoid	pme0088	Luteolin	Flavone	3,170.00	7,193.33	1.05E + 00	2.27E + 00	Up
	pme0379	Apigenin	Flavone	4,439.33	10,163.33	2.39E + 00	2.29E + 00	Up
Flavone and flavonol	pme3290	3,7-Di-O-methylquercetin	Flavonol	639.33	9.00	1.16E + 00	1.41E − 02	Down
	pmb0604	Astragalin	Flavonol	4,780,000.00	1,863,333.33	1.29E + 00	3.90E − 01	Down
	pme3296	Kaempferin	Flavonol	4,206.67	1,476.67	1.18E + 00	3.51E − 01	Down
Isoflavonoid	pme3502	Ononin	Isoflavone	27,300.00	10,116.67	1.30E + 00	3.71E − 01	Down
	pme3399	Sissotrin	Isoflavone	9.00	956.00	1.24E + 00	1.06E + 02	Up
Anthocyanin	pme1773	Cyanidin 3-O-rutinoside	Anthocyanins	327,333.33	134,266.67	1.11E + 00	4.10E − 01	Down

chloroplast breakdown in the wild type during pod development. The expression levels of chlorophyll degradation genes were higher and earlier in the pods of the wild type leading to rapid degreening, especially *RCCR* showed a significantly differential expression between the wild type and green pod mutant.

The Internal Mechanism of Phenotypic Changes in Cellulose Content

A common feature for different wall types of plants is the prevalence of cellulose, which consists of β -1,4 glucan chains that are synthesized by *CESA* protein complexes (CSCs). The CSC typically consists of different heterotrimeric *CESA* configurations. Different *CESA* has respective roles; during primary wall synthesis, the CSC contains *CESA1*, 3 and one *CESA6*-like subunit (Desprez et al., 2007; Persson et al., 2007). By contrast, the secondary wall-synthesizing CSCs contain *CESA4*, 7, and 8 (Taylor et al., 2003). In this study, some *CESA* showed differential expression. The genes (*Phvul.004G093300*, *Phvul.005G022100*, and *Phvul.009G242700*) participating in cellulose synthesis had distinct higher expression in GP than in YP samples. There were also more genes involving in cellulose degradation in GP; the cellulose in the green pod mutant showed

rapid synthesis. The high cellulose content in pods might be one of the reasons that the green pod mutant was much harder.

Cellulose biosynthesis is a complex biochemical process, which cannot be reproduced *in vitro*. Fiber content is a key index of common bean quality; the common bean with low fiber content has a good taste. Golden hook is a very popular ecotype of the common bean and has a very low cellulose content in fresh pods. There were also some genes for cellulose synthesis having relative low expression in the wild type. Conversely, there was relatively high cellulose content in green pod mutants at the same stage; MeJA treatment promoted peach fruit chlorophyll degradation and affected fruit softening through increasing the expression of cellulase (Wei et al., 2017). The genes for cellulose synthesis pathway might be downregulated by the genes involved in degreening of pods.

Sucrose is the main form of assimilated carbon that is produced during plant photosynthesis; UDP-glucose can be transformed from sucrose catalyzed by *SuSy*. UDP-glucose was just the immediate substrate for cellulose synthesis (Nakai et al., 1999; Verbancic et al., 2018). It was speculated that photosynthesis of green pod mutants might provide synthetic substrate for cellulose, but the intrinsic interaction need to be further studied.

Transcriptome and Metabolome Analyses Involved in Flavonoid Pathway

Flavonoids, important secondary metabolites found in plants, contribute to plant environmental adaptation, fruit development (Petroni and Tonelli, 2011; Li et al., 2019), and human health (Alipour et al., 2016). These compounds are accumulated in various tissues and are also the direct factors that cause color variations in many fruits or flowers. Combined metabolome and transcriptome analyses were carried out with *Ficus carica* L.; cyaniding *O*-malonylhexoside demonstrated a 3,992-fold increase; cyaniding 3-*O*-gglucoside, cyanidin-3, and 5-*O*-diglucoside were upregulated 100-fold, revealing the anthocyanins underlying the purple mutation (Wang et al., 2017). The pod color and pattern of the common bean is colorful; some varieties having red or purple patterns could be rich in anthocyanin. On the other hand, the common bean is a special legume variety that is abundant in flavonoids, especially isoflavones. In a purple kidney bean cultivar, malvidin 3,5-diglucoside was identified as the major anthocyanin in the pod skin by HPLC-ESI-MS (Hu et al., 2015). Naringin, vitexin 2''-*O*-beta-L-rhamnoside, and isotrifoliin demonstrated significantly higher contents in the YP-2. Compared to YP-5, the content of three isoflavonoid (prunetin, biochanin A, and sissotrin) was higher in GP-5. Only five anthocyanin were detected (oenin, pelargonin, cyanin, cyanidin, and keracyanin), and all these five anthocyanins were significantly higher (VIP > 1, fold change < 0.5) in the wild type than that in the green pod mutant. When the pod length reached to 10 cm, cyanidin 3-*O*-rutinoside (keracyanin) was the sole anthocyanin with datable difference between GP-10 and YP-10. Thus, it could be speculated that anthocyanins content could be influenced by the degreening to certain extent but not significantly.

The Whole Nutrition Value and Breeding Strategy in Common Bean

The common bean is a nutrient-rich food that contains nutrients essential for humans, such as proteins, minerals, vitamins, carbohydrates, and fiber. The protein quality in the common bean is high, and many cultivars have sufficiently high levels of essential and non-essential amino acids to meet daily nutritional needs (Ribeiro et al., 2007). Although the protein content of the common bean is approximately 50% of that of soybean, the digestibility of protein is higher (78.70%). In view of the high mineral content, common beans benefit health and can be used to cure a number of mineral deficiencies (Mesquita et al., 2007).

The yield potential of the common bean is determined by growth habit, pod number, maturity features, and seed characteristics. In addition to the yield, the external phenotype, taste, cook characteristic, and disease and abiotic stress tolerance are prime goals in breeding. As a high-quality variety of oil beans, Dalong 1 has high nutritional value, less cellulose, and good taste. In this study, we studied the color conversion of pods and flavonoids metabolism between the wild type and green pod mutants; it was helpful to understand the metabolic process of nutrients and provide a theoretical basis for breeding.

DATA AVAILABILITY STATEMENT

All the information regarding our sequencing data we deposited at National Genomics Data Center BIG Sub repository: Project No: PRJCA002296, GSAs: CRA002791, Experiment acc: CRX119315–CRX119332, Samples acc: SAMC138882–SAMC1388.

AUTHOR CONTRIBUTIONS

BH designed the experiments, carried out the research, and wrote and revised the manuscript. JY, DZ, HW, and NG prepared plant materials. JZ and YG helped to collected samples in the field. KX, HZ, and ZX gave important suggestions and assisted in data analysis. All authors participated in this study and approved the final version of the manuscript.

FUNDING

This work was supported by the National Key R&D Project (2016YFD0101900) from the Ministry of Science and Technology of China; Key Deployment Projects (ZDRW-ZS-2019-2) and Regional Key Project of Science and Technology Service (KFJ-STS-QYZD-037) of the Chinese Academy of Sciences; and by Programs of the National Natural Science Foundation of China (31771869, 31771818, and 31971970).

ACKNOWLEDGMENTS

We are grateful to Jiayin, Yang and Danhua, Zhu for their help in plant materials radiation and thank Dr. Zhengjun, Xia for giving significant comments and English language editing of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.570816/full#supplementary-material>

FIGURE S1 | The phenotype of pods of the wild type and green pod mutant.

FIGURE S2 | KEGG and GO enrichment analysis in wild type and green pod mutant.

FIGURE S3 | Heat map diagram of expression levels of the genes for chlorophyll synthesis in YP, GP and other green pod mutant lines.

FIGURE S4 | The heat map of DEGs in flavones and isoflavones pathway.

FIGURE S5 | The differentially expression genes analysis in flavonoids metabolism pathway between YP and GP.

TABLE S1 | Statistics for the number of clean data.

TABLE S2 | List of genes and primers for RT-qPCR.

TABLE S3 | Statistics of SNP variants of the genes for chlorophyll degradation.

REFERENCES

- Alipour, B., Rashidkhani, B., and Edalati, S. (2016). Dietary flavonoid intake, total antioxidant capacity and lipid oxidative damage: a cross-sectional study of Iranian women. *Nutrition* 32, 566–572. doi: 10.1016/j.nut.2015.11.011
- Aubry, S., Mani, J., and Hortensteiner, S. (2008). Stay-green protein, defective in Mendel's green cotyledon mutant, acts independent and upstream of pheophorbide a oxygenase in the chlorophyll catabolic pathway. *Plant Molecular Biol.* 67, 243–256. doi: 10.1007/s11103-008-9314-8
- Borges, A., Tsai, S. M., and Caldas, D. G. (2012). Validation of reference genes for RT-qPCR normalization in common bean during biotic and abiotic stresses. *Plant Cell Rep.* 31, 827–838. doi: 10.1007/s00299-011-1204-x
- Charoenchongsuk, N., Ikeda, K., Itai, A., Oikawa, A., and Murayama, H. (2015). Comparison of the expression of chlorophyll-degradation-related genes during ripening between stay-green and yellow-pear cultivars. *Sci. Hortic.* 181, 89–94. doi: 10.1016/j.scienta.2014.10.005
- Cheng, Y., Dong, Y., Yan, H., Ge, W., Shen, C., Guan, J., et al. (2012). Effects of 1-MCP on chlorophyll degradation pathway-associated genes expression and chloroplast ultrastructure during the peel yellowing of Chinese pear fruits in storage. *Food Chem.* 135, 415–422. doi: 10.1016/j.foodchem.2012.05.017
- Coleman, H. D., Yan, J., and Mansfield, S. D. (2009). Sucrose synthase affects carbon partitioning to increase cellulose production and altered cell wall ultrastructure. *PNAS* 106, 13118–13123. doi: 10.1073/pnas.0900188106
- Cui, B. L., Hu, Z. L., Zhang, Y. J., Hu, J. Y., Yin, W. C., Feng, Y., et al. (2016). Anthocyanins and flavonols are responsible for purple color of *Lablab purpureus* (L.) sweet pods. *Plant Physiol. Biochem.* 103, 183–190. doi: 10.1016/j.plaphy.2016.03.011
- Desprez, T., Juraniec, M., Crowell, E. F., Jouy, H., Pochylova, Z., Parcy, F., et al. (2007). Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. *PNAS* 104, 15572–15577. doi: 10.1073/pnas.0706569104
- Fang, H. X., Luo, F., Li, P. X., Zhou, Q., Zhou, X., Wei, B. D., et al. (2020). Potential of jasmonic acid (JA) in accelerating postharvest yellowing of broccoli by promoting its chlorophyll degradation. *Food Chem.* 309, 1–11. doi: 10.1016/j.spr.2008.6.8
- Gambi, F., Pilkington, S. M., McAtee, P. A., Donati, I., Schaffer, R. J., Montefiori, M., et al. (2018). Fruit of three kiwifruit (*Actinidia chinensis*) cultivars differ in their degreening response to temperature after harvest. *Postharvest Biol. Technol.* 141, 16–23. doi: 10.1016/j.postharvbio.2018.03.009
- Ginsburg, S., Schellenberg, M., and Matile, P. (1994). Cleavage of chlorophyll-Porphyrin. *Plant Physiol.* 105, 545–554. doi: 10.1104/pp.105.2.545
- Harpaz-Saad, S., Azoulay, T., Arazi, T., Ben-Yakov, E., Mett, A., Shibolet, Y. M., et al. (2007). Chlorophyllase is a rate-limiting enzyme in chlorophyll catabolism and is posttranslationally regulated. *Plant Cell* 19, 1007–1022. doi: 10.1105/tpc.107.050633
- Hörtensteiner, S., Vicentini, F., and Matile, P. (1995). Chlorophyll breakdown in senescent cotyledons of rape, *Brassica napus* L.: enzymatic cleavage of pheophorbide a in vitro. *New Phytologist* 129, 237–246. doi: 10.1111/j.1469-8137.1995.tb04293.x
- Hu, J. T., Chen, G. P., Zhang, Y. Y., Cui, B. L., Yin, W. C., Yu, X. H., et al. (2015). Anthocyanin composition and expression analysis of anthocyanin biosynthetic genes in kidney bean pod. *Plant Physiol. Biochem.* 97, 304–312. doi: 10.1016/j.plaphy.2015.10.019
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, 480–484. doi: 10.1093/nar/gkm882
- Kovinch, N., Kanyanja, G., Chanoca, A., Riedl, K., Otegui, M. S., and Grotewold, E. (2014). Not all anthocynins are born equal: distinct patterns induced by stress in *Arabidopsis*. *Planta* 240, 671–687.
- Kusaba, M., Ito, H., Morita, R., Iida, S., Sato, Y., Fujimoto, M., et al. (2007). Rice NON-YELLOW COLORING1 is involved in light-harvesting complex II and grana degradation during leaf senescence. *Plant Cell* 19, 1362–1375. doi: 10.1105/tpc.106.042911
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323
- Li, H. Y., Lv, Q. Y., Ma, C., Qu, J. T., Cai, F., Deng, J., et al. (2019). Metabolite profiling and transcriptome analyses provide insights into the flavonoid biosynthesis in the developing seed of Tartary Buckwheat (*Fagopyrum tataricum*). *J. Agric. Food Chem.* 67, 11262–11276.
- Li, Y., Yang, K., Yang, W., Chu, L. W., Chen, C. H., Zhao, B., et al. (2017). Identification of QTL and qualitative trait loci for agronomic traits using SNP markers in the adzuki bean. *Front. Plant Sci.* 8:840. doi: 10.3389/fpls.2017.00840
- Lv, J. Y., Zhang, M. Y., Bai, L., Han, X. Z., Ge, Y. H., Wang, W. H., et al. (2020). Effects of 1-methylcyclopropene (1-MCP) on the expression of genes involved in the chlorophyll degradation pathway of apple fruit during storage. *Food Chem.* 308, 1–7.
- Mesquita, F. R., Corrêa, A. D., Abreu, C. M., Lima, R. A., and Abreu, A. F. (2007). Linhagens de feijão (*Phaseolus vulgaris* L.): composição química e digestibilidade proteica. *Ciência e Agrotecnol.* 31, 1114–1121. doi: 10.1590/s1413-70542007000400026
- Motamayor, J. C., Mockaitis, K., Schmutz, J., Haiminen, N., Livingstone, D., Cornejo, O., et al. (2013). The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biol.* 14:r53.
- Nakai, T., Tonouchi, N., Konishi, T., Kojima, Y., Tsuchida, T., Yoshinaga, F., et al. (1999). Enhancement of cellulose production by expression of sucrose synthase in *Acetobacter xylinum*. *PNAS* 96, 14–18. doi: 10.1073/pnas.96.1.14
- Okonkwo, C. A., and Clayberg, C. D. (1984). Genetics of flower and pod color in *Phaseolus vulgaris*. *J. Heredity* 75, 440–444. doi: 10.1093/oxfordjournals.jhered.a109981
- Park, S. Y., Yu, J. W., Park, J. S., Li, J. J., Yoo, S. C., Lee, N. Y., et al. (2007). The senescence-induced staygreen protein regulates chlorophyll degradation. *Plant Cell* 19, 1649–1664. doi: 10.1105/tpc.106.044891
- Persson, S., Paredes, A., Carroll, A., Palsdottir, H., Doblin, M., Poindexter, P., et al. (2007). Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis*. *PNAS* 104, 15566–15571. doi: 10.1073/pnas.0706592104
- Petroni, K., and Tonelli, C. (2011). Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci.* 181, 219–229. doi: 10.1016/j.plantsci.2011.05.009
- Pruzinska, A., Tanner, G., Anders, I., Roca, M., and Hortensteiner, S. (2003). Chlorophyll breakdown: pheophorbide a oxygenase is a Rieske-type iron-sulfur protein, encoded by the accelerated cell death 1 gene. *PNAS* 100, 15259–15264. doi: 10.1073/pnas.2036571100
- Ribeiro, N. D., Londero, P. M., Filho, A. C., Jost, E., Poersch, N. L., and Mallmann, C. A. (2007). Composição de aminoácidos de cultivares de feijão e aplicações para o melhoramento genético. *Pesquisa Agropecuária Brasil.* 42, 1393–1399.
- Roca, M., and Minguez-Mosquera, M. I. (2006). Chlorophyll catabolism pathway in fruits of *Capsicum annuum* (L.): stay-green versus red fruits. *J. Agric. Food Chem.* 54, 4035–4040. doi: 10.1021/jf060213t
- Rodoní, S., Mühlecker, W., Anderl, M., Krautler, B., Moser, D., Thomas, H., et al. (1997). Chlorophyll breakdown in senescent chloroplasts. *Plant Physiol.* 115, 669–676. doi: 10.1104/pp.115.2.669
- Sakuraba, Y., Park, S. Y., Kim, Y. S., Wang, S. H., Yoo, S. C., Hortensteiner, S., et al. (2014). *Arabidopsis* STAY-GREEN2 is a negative regulator of chlorophyll degradation during leaf senescence. *Mol. Plant* 7, 1288–1302. doi: 10.1093/mp/ssu045
- Tanaka, Y., Sasaki, N., and Ohmiya, A. (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 54, 733–749. doi: 10.1111/j.1365-3113.2008.03447.x
- Taylor, N. G., Howells, R. M., Huttly, A. K., Vickers, K., and Turner, S. R. (2003). Interactions among three distinct CesA proteins essential for cellulose synthesis. *PNAS* 100, 1450–1455. doi: 10.1073/pnas.0337628100
- Tharanathan, R. N., and Mahadevamma, S. (2003). Grain legumes- a boon to human nutrition. *Trends Food Sci. Technol.* 14, 507–518. doi: 10.1016/j.tifs.2003.07.002
- Tsuchiya, T., Ohta, H., Okawa, K., Iwamatsu, A., Shimada, H., Masuda, T., et al. (1999). Cloning of chlorophyllase, the key enzyme in chlorophyll degradation: finding of a lipase motif and the induction by methyl jasmonate. *PNAS* 92, 15362–15367. doi: 10.1073/pnas.96.26.15362
- Verbancic, J., Lunn, J. E., Stitt, M., and Persson, S. (2018). Carbon supply and the regulation of cell wall synthesis. *Mol. Plant* 11, 75–94. doi: 10.1016/j.molp.2017.10.004
- Wang, H., Zhang, H., Yang, Y., Li, M. F., Zhang, Y. T., Liu, J. S., et al. (2019). The control of red color by a family of MYB transcription factors in octoploid

- strawberry (*Fragaria x ananassa*) fruits. *Plant Biotechnol. J.* 18, 1169–1184. doi: 10.1111/pbi.13282
- Wang, Z., Cui, Y., Vainstein, A., Chen, S., and Ma, H. (2017). Regulation of Fig (*Ficus carica* L.) fruit color: metabolomic and transcriptomic analyses of the flavonoid biosynthetic pathway. *Front. Plant Sci.* 8:1990. doi: 10.3389/fpls.2017.01990
- Wei, J., Wen, X., and Tang, L. (2017). Effect of methyl jasmonic acid on peach fruit ripening progress. *Sci. Hortic.* 220, 206–213. doi: 10.1016/j.scienta.2017.03.004
- Xu, P., Su, H., Jin, R., Mao, Y. X., Xu, A., Cheng, H. Y., et al. (2020). Shading effects on leaf color conversion and biosynthesis of the major secondary metabolites in the Albina Tea cultivar “Yujinxiang”. *J. Agric. Food Chem.* 68, 2528–2538. doi: 10.1021/acs.jafc.9b08212
- Yuste-Lisbona, F. J., Gonzalez, A. M., Capel, C., Garcia-Alcazar, M., Capel, J., Ron, A. M., et al. (2014). Genetic variation underlying pod size and color traits of common bean depends on quantitative trait loci with epistatic effects. *Mol. Breed.* 33, 939–952. doi: 10.1007/s11032-013-0008-9
- Zhuang, H. M., Lou, Q., Liu, H. F., Han, H. W., Wang, Q., Tang, Z. H., et al. (2019). Differential regulation of anthocyanins in green and purple turnips revealed by combined de novo transcriptome and metabolome analysis. *Int. J. Mol. Sci.* 20, 1–18.

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

Copyright © 2020 Hu, Zhu, Wu, Xu, Zhai, Guo, Gao, Yang, Zhu and Xia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Molecular Basis Underlying Common Cutworm Resistance of the Primitive Soybean Landrace Peking

OPEN ACCESS

Edited by:

Peerasak Srinives,
Kasetsart University, Thailand

Reviewed by:

Deyue Yu,
Nanjing Agricultural University, China
Xin Chen,
Jiangsu Academy of Agricultural
Sciences, China

*Correspondence:

Akito Kaga
kaga@affrc.go.jp

† Present address:

Ryu Nakata,
Department of Bioscience
and Biotechnology, Kyoto University
of Advanced Science, Kameoka,
Japan
Yutaka Okumoto,
Department of Agricultural Science
and Technology, Setsunan University,
Hirakata, Japan

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 10 July 2020

Accepted: 15 October 2020

Published: 13 November 2020

Citation:

Nakata R, Yano M, Hiraga S,
Teraishi M, Okumoto Y, Mori N and
Kaga A (2020) Molecular Basis
Underlying Common Cutworm
Resistance of the Primitive Soybean
Landrace Peking.
Front. Genet. 11:581917.
doi: 10.3389/fgene.2020.581917

Ryu Nakata^{1†}, Mariko Yano², Susumu Hiraga¹, Masayoshi Teraishi³, Yutaka Okumoto^{3†}, Naoki Mori² and Akito Kaga^{1*}

¹ Institute of Crop Science, National Agriculture and Food Research Organization, Tsukuba, Japan, ² Division of Applied Life Sciences, Graduate School of Agriculture, Kyoto University, Kyoto, Japan, ³ Division of Agronomy and Horticultural Science, Graduate School of Agriculture, Kyoto University, Kyoto, Japan

The common cutworm (CCW; *Spodoptera litura*) is one of the major insect pests of soybean in Asia and Oceania. Although quantitative trait loci related to CCW resistance have been introduced into leading soybean cultivars, these do not exhibit sufficient resistance against CCW. Thus, understanding the genetic and metabolic resistance mechanisms of CCW as well as integrating other new resistance genes are required. In this study, we focused on a primitive soybean landrace, Peking, which has retained resistances to various pests. We found a resistance to CCW in Peking by the detached-leaf feeding assay, and subsequently determined the genetic and metabolic basis of the resistance mechanism using chromosome segment substitution lines (CSSLs) of Peking. Several characteristic metabolites for Peking were identified by the metabolomic approach using liquid chromatography/mass spectrometry combined with a principle component analysis. The structure of seven metabolites were determined by nuclear magnetic resonance (NMR) analysis. The genomic segments of Peking on chromosome 06 (Chr06) and Chr20 had a clear association with these metabolites. Moreover, a line possessing a Peking genomic segment on Chr20 inhibited growth of the CCW. The genetic factors and the metabolites on Chr20 in Peking will be useful for understanding mechanisms underlying CCW resistance and breeding resistant soybean cultivars.

Keywords: soybean, common cutworm, resistance, chromosome segment substitution lines, Peking

INTRODUCTION

The soybean [*Glycine max* (L.) Merr.] is one of the most important leguminous crops in the world and is used for edible proteins, oils, fodder and in various processed foods. The production of soybean ranks fourth to rice, wheat and maize in terms of world crop production. The major production areas are the United States of America, Brazil, Argentina, China, Paraguay, India and Canada. The wild soybean [*G. soja* (Sieb. & Zucc.)], the direct ancestor of the soybean, is distributed mainly in eastern China, eastern Russia, Korea and Japan (Lu, 2004). The domestication of the

wild soybean to the soybean is believed to have started from the eleventh century B.C. in China and then spread to surrounding countries (Hymowitz, 1990). Undesirable agricultural traits of the wild soybean, such as seed dormancy, pod dehiscence and elongation of the twining stem, have been removed during the domestication process by farmers. Desirable agricultural traits, such as resistance to biotic and abiotic stress, high yield, and high seed quality, have been selected or introduced into leading soybean cultivars by breeders to ensure stable production under a changing climate and to meet market demand.

Insect pests are one of the factors limiting soybean production. More than 700 species of insect pests are found in soybean (Way, 1994). In Asia and Oceania, the common cutworm [CCW; *Spodoptera litura* (Lepidoptera: Noctuidae)] is one of the major insect pests of soybean. The CCW larvae feed on the leaves of the soybean and cause serious yield losses. The main strategy for controlling CCW is the use of insecticides. The development of CCW-resistant cultivars would reduce insecticide use and stabilize soybean production. In previous studies, two quantitative trait loci (QTLs) for CCW resistance, *CCW1* and *CCW2*, were identified on Chr07 from the CCW-resistant cultivar “Himeshirazu” (Komatsu et al., 2004, 2005). Both QTLs were introduced into a Japanese leading cultivar, “Fukuyutaka” to develop near isogenic lines (NILs), which was confirmed to exhibit significant resistance against CCW (Komatsu et al., 2008), including in a field assessment (Oki et al., 2015). However, the resistance level of NILs harboring *CCW1* and *CCW2* were lower than that of “Himeshirazu” (Komatsu et al., 2008; Oki et al., 2015).

In addition to *CCW1* and *CCW2*, *qRslx1* on Chr07 and *qRslx2* on Chr12 were identified as resistance QTLs from “Himeshirazu” and “Fukuyutaka,” respectively (Oki et al., 2012). These resistance QTLs affected CCW by different mechanisms; *CCW1* and *CCW2* are antibiosis QTLs (i.e., they have adverse effects on insect development and life history) (Komatsu et al., 2005), while *qRslx1* and *qRslx2* are antixenosis QTLs (i.e., they hinder insect behavior) (Oki et al., 2012). Based on these findings, multiple resistance genes would be required to achieve strong and stable resistance against CCW. To find other effective genes that can be introduced into soybean cultivars, the wild soybean has been studied as a source of new resistance genes. Two antixenosis QTLs, *qRslx3* on Chr07 and *qRslx4* on Chr02, were found in a *G. soja* accession [National Agriculture and Food Research Organization (NARO) Genebank accession JP110755]. Another resistant allele was also found in an accession JP267519 (Oki et al., 2019). Since the genetic diversity of the wild soybean is higher than that of the domesticated soybean (Li et al., 2014), it is expected that more effective CCW-resistance genes might be present in wild accessions.

The study of resistance genes often does not enable the explanation of the molecular basis of resistance. A metabolomic analysis is an effective strategy for filling the gap between genotype and phenotype (e.g., resistance). In particular, mass spectrometry (MS) coupled with gas-chromatography (GC) or liquid chromatography (LC) is a powerful technique to evaluate metabolomic diversity within broad genetic plant populations (Luo, 2015; Yandeu-Nelson et al., 2015; Fernie and Tohge, 2017).

Many crop phenotypes have been explained by the abundance of primary or secondary metabolites using a metabolomic approach aided with computational and statistical analyses (Hamany Djande et al., 2020).

Among the domestication-related traits, changes in the secondary metabolites are known to occur in 66% of crops (Meyer et al., 2012). For example, toxic or bitter compounds have decreased along with changes in the pigmentation of fruits or seeds. In maize and its wild ancestor, metabolic divergence and candidate genes were investigated (Xu et al., 2019). A large difference in the chemical responses between wild and domesticated wheat cultivars against aphid infection was reported (Batyrschina et al., 2020). A higher amount of organic acids and amino acids were found in the wild soybean than in the cultivated soybean, which are important for the plant to respond to salt stress (Zhang et al., 2016). Therefore, it is expected that chemical compounds for CCW resistance in the wild soybean have decreased or been lost in cultivated soybeans as a result of domestication.

In the present study, we focused on the primitive soybean landrace Peking, which is known to retain resistance to various pests, such as the soybean cyst nematode *Heterodera glycines* (Ross and Brim, 1957), Phytophthora stem and root rot (Jiang et al., 2017) and soybean mosaic virus (Hayes et al., 2000). In addition, Peking is resistant to another noctuid larvae, the corn earworm (*Helicoverpa zea*) (Joshi, 1980); therefore, CCW resistance is also expected. Here, we evaluated CCW resistance and investigated the genetic and metabolic mechanisms of CCW resistance in Peking by using chromosome segment substitution lines (CSSLs) (Watanabe et al., 2018). The CSSLs have partial genomic segments of Peking in a genetic background of the Japanese cultivar Enrei. Thus, an association analysis using the genotype and the metabolite abundance (i.e., phenotype) of CSSLs will enable us to find candidate genes contributing to the metabolite biosynthesis. Although the genetic factors for CCW resistance have been investigated by a genome-wide association analysis (Wang et al., 2015; Liu et al., 2016), the genetic and metabolic background of CCW resistance have not been comprehensively characterized. Understanding the resistance mechanism will help to breed elite cultivars with high CCW resistance.

MATERIALS AND METHODS

Plants and Insects

A total of 103 lines of Peking CSSLs with an Enrei genetic background (Watanabe et al., 2018) and their parents, Enrei and Peking, were used. Seeds were sown in plastic pots containing soil (Vegetable seedling soil type S; Yanmar Co., Ltd., Osaka, Japan) and vermiculite (Vermiculite GS; Nittai Co., Ltd., Osaka, Japan) in a ratio of 1:1 and placed in a greenhouse maintained at ca. 26 ± 8°C under natural light condition [ca. 12 h/12 h (light/dark)]. The CCW larvae were reared on an artificial diet (Insecta-LFS; Nihon Nosan Kogyo Ltd., Yokohama, Japan) under laboratory conditions of 26°C and a 16 h/8 h (light/dark) cycle.

The CCW developmental stages were synchronized at each molt by collecting new molting larvae.

Evaluation of Common Cutworm Resistance by Detached-Leaf Assay

The first and second trifoliate leaves of soybean plants in the V3–5 stages (Fehr and Caviness, 1977) were cut and placed in a petri dish with a ventilation mesh (Insect breeding dish; SPL Life Sciences Co., Ltd., Korea), which were previously lined with moistened filter paper. The 2nd instar CCW larvae (about 5 h after molting) were fasted for 4 h and subsequently placed into the petri dish at 26°C and a 16 h/8 h (light/dark) cycle. Ninety-six hours later, the larvae were frozen in liquid N₂ and stored at –80°C. After lyophilization, the dry weight of the larvae was measured. In this assay, two CCW larvae were put into one petri dish containing the first and second trifoliate leaves from one soybean plant and one moistened the lined filter paper; four petri dishes were replicated for each line and cultivar. All dishes were piled up to avoid excessive ventilation. Thus, the feeding assay was conducted with eight biological replicates of CCW larvae and four biological replicates of soybean plants.

Investigation of Characteristic Metabolites for Peking by Liquid Chromatography-Mass Spectrometry

A central leaflet of the first trifoliate leaf of soybean plants in the V3–5 stage was harvested and immersed in an 80% methanol (MeOH) solution (MeOH/H₂O, v/v) (0.2 mg fresh weight/μL) containing 7-hydroxyflavone (10 ng/μL) as an internal standard. After pulverization and centrifugation at 5,000 g for 5 min, the supernatants were filtered through a membrane filter (DISMIC-13HP, 0.45 μm pore size; Toyo Roshi Kaisha, Ltd., Tokyo, Japan). This liquid extraction was conducted with three biological replicates of the Enrei and Peking plants.

The filtrate was analyzed by the Prominence high performance liquid chromatography (HPLC) system coupled with LCMS-2020 (Shimadzu, Co., Kyoto, Japan). The LC separations were performed with an ODS column (Mightysil RP-18GP 50 × 2.0 mm I.D., 5 μm; Kanto Chemical Co., Inc., Tokyo, Japan) at 40°C with 0.2 mL/min flow rate. The solvent program was 5–50% (0–15 min), 50–99% (15–20 min), and 99% (20–25 min) of acetonitrile containing 0.08% acetic acid in H₂O containing 0.05% acetic acid. The MS was conducted with the following parameters: positive and negative ion mode; scan range, *m/z* 200–1,000; nebulizer gas flow, 1.5 L/min; drying gas flow, 15 L/min; ESI voltage, 4.5 kV in the positive ion mode and –4.5 kV in the negative ion mode; heat block temperature, 200°C; DL temperature, 250°C.

The metabolites in Peking associated with CCW resistance were investigated by comparing the metabolites between Peking and Enrei using a principal component analysis (PCA). The LC-MS data (*m/z* and retention times) was exported using an alignment software (Profiling Solution 1.1 Build 104; Shimadzu, Co., Kyoto, Japan) with the following parameters: retention time, 2–22 min; mass range from 200 to 1,000 Da; ion *m/z* tolerance, 25 mDa; ion retention time tolerance, 1.5 min; ion intensity

threshold, 10,000 counts; detecting 20% isomer valley; allowing some ions without isotope peaks. The obtained metabolite array list was analyzed by PCA using SIMCA P (version 13.0.0.0; Umetrics, Umeå, Sweden) after processing by column centering and Pareto scaling.

Purification of Metabolites A–G From the Peking Leaf Extract

Whole leaves from V5 stage plants of Peking (46 g) were harvested and immersed in 1 L of 80% MeOH solution (MeOH/H₂O, v/v) at 4°C. After filtration, the filtrate was evaporated under a vacuum to yield a crude extract (4.6 g). This extract was separated with hexane and H₂O; then, a H₂O layer (4.5 g extract) was separated with ethyl acetate and H₂O. The obtained H₂O layer (3.6 g extract) was passed through an ODS cartridge (Sep-pak C18 (10 g; Waters, Milford, MA, United States) by MeOH elution to remove highly hydrophobic compounds. The eluate (i.e., the through fraction, 3.5 g extract) was collected and separated with a flash chromatograph (CombiFlash Rf + UV; Teledyne ISCO, Lincoln, NE, United States) equipped with a RediSep Rf Gold Reversed-phase C18 column (50 g; Teledyne ISCO, Lincoln, NE, United States). The separation was performed with a 40 L/min flow rate by a stepwise gradient. The solvent program was 30% (0–15 min, Fr. I), 40% (15–30 min, Fr. II), 50% (30–45 min, Fr. III), and 100% (45–70 min, Fr. IV) of MeOH solution (MeOH/H₂O, v/v). Then, Fr. II (143 g) was chromatographed by HPLC using the Prominence HPLC system. The LC separations were performed with an ODS column (Mightysil RP-18GP 250 × 10 mm I.D., 5 μm; Kanto Chemical Co., Inc., Tokyo, Japan) at 50°C with a 5 mL/min flow rate. The column was eluted in the isocratic mode with a mobile phase containing 31% MeOH (MeOH/H₂O, v/v) to yield Fr. II-i (*t_R*, 9.6–14.6 min), Fr. II-ii (*t_R*, 14.5–15.9 min), Fr. II-iii (*t_R*, 17.0–20.0 min), Fr. II-iv (*t_R*, 22.4–24.3 min), Fr. II-v (*t_R*, 24.6–26.6 min), Fr. II-vi (*t_R*, 26.8–28.8 min); Fr. II-ii, Fr. II-iv, Fr. II-v, and Fr. II-vi contained metabolites C, E, F, and G in high purity, respectively. Then, Fr. II-i and Fr. II-iii were re-chromatographed using the same apparatus with an ODS column (Mightysil RP-18GP 250 × 4.6 mm I.D., 5 μm; Kanto Chemical Co., Inc., Tokyo, Japan). Then, Fr. II-i was separated in the isocratic mode with a mobile phase containing 16% MeOH (MeOH/H₂O, v/v) at 60°C with 1 mL/min flow rate to yield Fr. II-i-1 (*t_R*, 58.4–66.2 min) and Fr. II-i-2 (*t_R*, 64.6–69.4 min); Fr. II-iii was separated in the isocratic mode with a mobile phase containing 16% MeOH (MeOH/H₂O, v/v) at 40°C with 1 mL/min flow rate to yield Fr. II-iii-1 (*t_R*, 16.2–17.6 min). The obtained fraction, Fr. II-i-1, Fr. II-i-2, and Fr. II-iii-1, contained metabolites A, B, and D in high purity, respectively. The purification scheme was summarized in **Supplementary Figure S1**.

Structure Determination of Metabolites A–G

The nuclear magnetic resonance (NMR) experiments were performed on a Bruker Avance III 400 spectrometer and

Avance III 500 NMR spectrometer (Bruker Corp., Billerica, MA, United States). Each metabolite was dissolved in CD₃OD (>99.8% D; Euriso-Top, Saint-Aubin, France) containing 0.1% tetramethylsilane as an internal standard. The ¹H NMR, ¹³C NMR, ¹H-¹H COSY, HSQC, and HMBC spectra were measured. The molecular formula and MS/MS fragment of each metabolite were analyzed by LC-Orbitrap MS under the same conditions as described in the next section.

Liquid Chromatography-Orbitrap Mass Spectrometry Analysis of Metabolites A–G in Chromosome Segment Substitution Lines

A central leaflet of the 1st trifoliate leaf of each line of the CSSLs in the V3 stage was extracted with an 80% MeOH solution (MeOH/H₂O, v/v) (0.2 mg fresh weight/mL) containing 7-hydroxyflavone (10 μM) as an internal standard. After pulverization and centrifugation at 5,000 g for 5 min, the supernatants were filtered through a membrane filter (GL chromat disk 4P, 0.45 μm pore size; GL Sciences Inc., Tokyo, Japan). The leaves of Enrei and Peking were also extracted in the same way. The liquid extraction was conducted with three biological replicates of all plant lines.

The filtrate was analyzed by an Ultimate 3000 SD HPLC system coupled with LTQ Orbitrap discovery (Thermo Fisher Scientific Inc., MA, United States). The LC separations were performed with a PFP column (Ascentis Express 90 Å F5 50 × 2.1 mm I.D., 2.7 μm; Merck KGaA, Darmstadt, Germany) at 50°C with 0.2 mL/min flow rate. The solvent program was 0% (0–1 min), 0–20% (1–21 min), and 99.5% (21–23.5 min) of acetonitrile containing 0.1% formic acid in H₂O containing 0.1% formic acid. The LTQ Orbitrap MS was operated in the ESI positive ion mode with the following parameters: a capillary temperature of 380°C, sheath gas flow rate of 5 (arbitrary unit), aux gas flow rate of 6 (arbitrary unit), sweep gas flow rate of 0 (arbitrary unit), source voltage of 4.5 kV, capillary voltage of 30 V, and tube lens voltage of 80 V. The MS full scan was acquired in 30,000 resolution at *m/z* 100–1,500 in the FT detector mode. The MS/MS scan was operated by data dependent acquisition (DDA) in 30,000 resolution at *m/z* 100–1,500 in the FT detector mode. For the dynamic exclusion, a repeat count was set at one; the repeat duration and exclusion duration were set at 30 and 20 s, respectively. The most intense ion was selected for the collision induced dissociation (CID) with the following parameters: default charge state, 1; isolation width of *m/z*, 2.0; normalized collision energy, 35; acquisition Q, 0.250; acquisition time, 30 ms; minimum signal required, 1,000. The exact mass of diisooctyl phthalate ([M + H]⁺ = 391.28429) was used for the lock mass.

All collected raw data were imported into the data processing software Compound Discoverer v2.1 (Thermo Fisher Scientific Inc., MA, United States). After retention time alignment and peak picking, the peak area of metabolites A–G and the internal standard were measured.

Investigation of the Genomic Segments of Peking Associated With the Accumulation of Metabolites A–G and Common Cutworm Resistance

The peak area of each metabolite was divided by that of the internal standard (IS) to give the peak area ratio. In each CSSL and their parents, three peak area ratios derived from three biological replicates were averaged out for each metabolite to represent the amount. The obtained value was used to find the genomic segments that affected the amount of each metabolite by statistical analysis. The genomic segments of the 103 CSSLs were characterized by 321 simple sequence repeat markers that cover all chromosomes evenly (Watanabe et al., 2018). By using the genotype data, the effect of the genotype of each marker on the amounts of each metabolite was evaluated by a one-way ANOVA. The genotype of each marker (i.e., Peking-homozygous, heterozygous, and Enrei-homozygous) was used as a factor, and Type II sum of squares was used for the calculation. Then, the amounts of each metabolite were compared between the lines possessing Peking- and Enrei-genomic segments on each marker.

The CCW resistance of the lines possessing Peking-genomic segments associated with the amounts of each metabolite was evaluated as described in section “Evaluation of Common Cutworm Resistance by Detached-Leaf Assay.”

Statistics

All statistical analyses were performed with R software (ver. 3.6.1). Boxplots depicted the first quartile minus 1.5 × interquartile range (IQR) (lower whiskers) and the third quartile plus 1.5 × IQR (upper whiskers), the IQR (box) and the median (horizontal line); outliers are indicated by dots outside the whiskers. The quartile was calculated according to Definition 7 in Hyndman and Fan (1996).

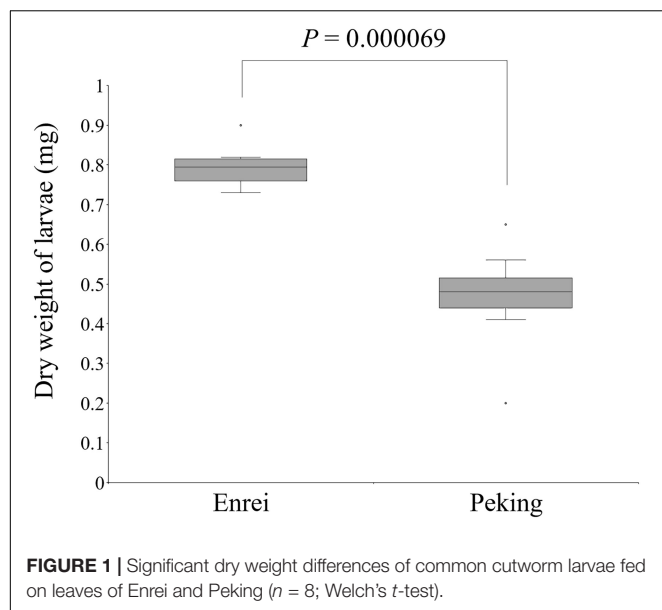
RESULTS

Common Cutworm Resistance of Peking

The CCW resistance of Enrei and Peking was evaluated by the detached-leaf feeding assay. The weights of the CCW larvae feeding on Peking were significantly lower than those feeding on Enrei (Welch's *t*-test, *P* < 0.01; **Figure 1**), suggesting that Peking has factor(s) that inhibit the growth of CCW.

Investigation of Characteristic Metabolites for Peking by Liquid-Chromatography-Mass Spectrometry

The characteristic metabolites for the leaf of Peking were screened by a comparative metabolic analysis using LC/MS followed by a PCA. The total metabolites were extracted from the leaves of Peking and Enrei, and analyzed by LC-MS. All detected ions were aligned along retention times and *m/z* values, then subjected to the PCA. Score plots exhibited a clear difference of the metabolic profiles between Peking and Enrei (**Figure 2A**).



The principal component (PC) scores of PC1 and PC2 explained 47.5 and 17.1% of the variation, respectively. The PC1 mainly explained the separation between Peking and Enrei; thus, we focused on the PC1 loading value of each component. The components possessing negative loading values indicated that the amounts in Peking were much higher than in Enrei. To find characteristic metabolites for Peking, three components possessing large negative loading values were picked based on PC1 (**Figure 2B**); these exhibited much larger negative loading values than the other components and represented negative ions at m/z 609, 593, and 623 for components 1, 2, and 3, respectively. Further analyses of these ions by LC/MS resolved that a total 5 chromatographic peaks were intermingled: two peaks of m/z 609 ions, two peaks of m/z 593 ions, and one peak of m/z 623 ions (**Figure 2C**). This is because the poorly separated peaks with the same m/z were picked as single components by the alignment software. A total of three ions at m/z 609 were further identified in the process of purification (see the next section). Finally, a total of seven characteristic metabolites for Peking were identified and assigned as metabolites A–G.

Purification and Structural Determination of Metabolites A–G

Metabolites A–G were purified from 46 g of Peking leaves. The purification scheme and yields were summarized in **Supplementary Figure S1**. Their structures were determined by NMR analysis and LC/Orbitrap-MS spectrometry.

Metabolite A was identified as kaempferol 3-*O*-glucosyl-(1→6)-galactoside by 1D and 2D-NMR (**Figure 3A**). The chemical shifts in ^1H and ^{13}C -NMR, and the correlation observed in ^1H - ^1H COSY and HMBC are summarized in **Supplementary Table S1**. The position at which the sugars conjugated were determined by the HMBC correlation between H-1'' and C-3, and also between H-1''' and C-6''. The spectrum of ^1H -NMR was also in accordance with values in the literature (Tanaka et al., 2001).

Then, the molecular formula was confirmed by MS spectrometry as $\text{C}_{27}\text{H}_{31}\text{O}_{16}$ (error: -1.25 ppm, **Supplementary Table S2**). In the MS/MS analysis, the diglycoside moiety was also signified by two glycoside-specific fragments, $[\text{M} + \text{H} - \text{C}_6\text{H}_{10}\text{O}_5]^+$ and $[\text{M} + \text{H} - \text{C}_{12}\text{H}_{20}\text{O}_{10}]^+$ ions.

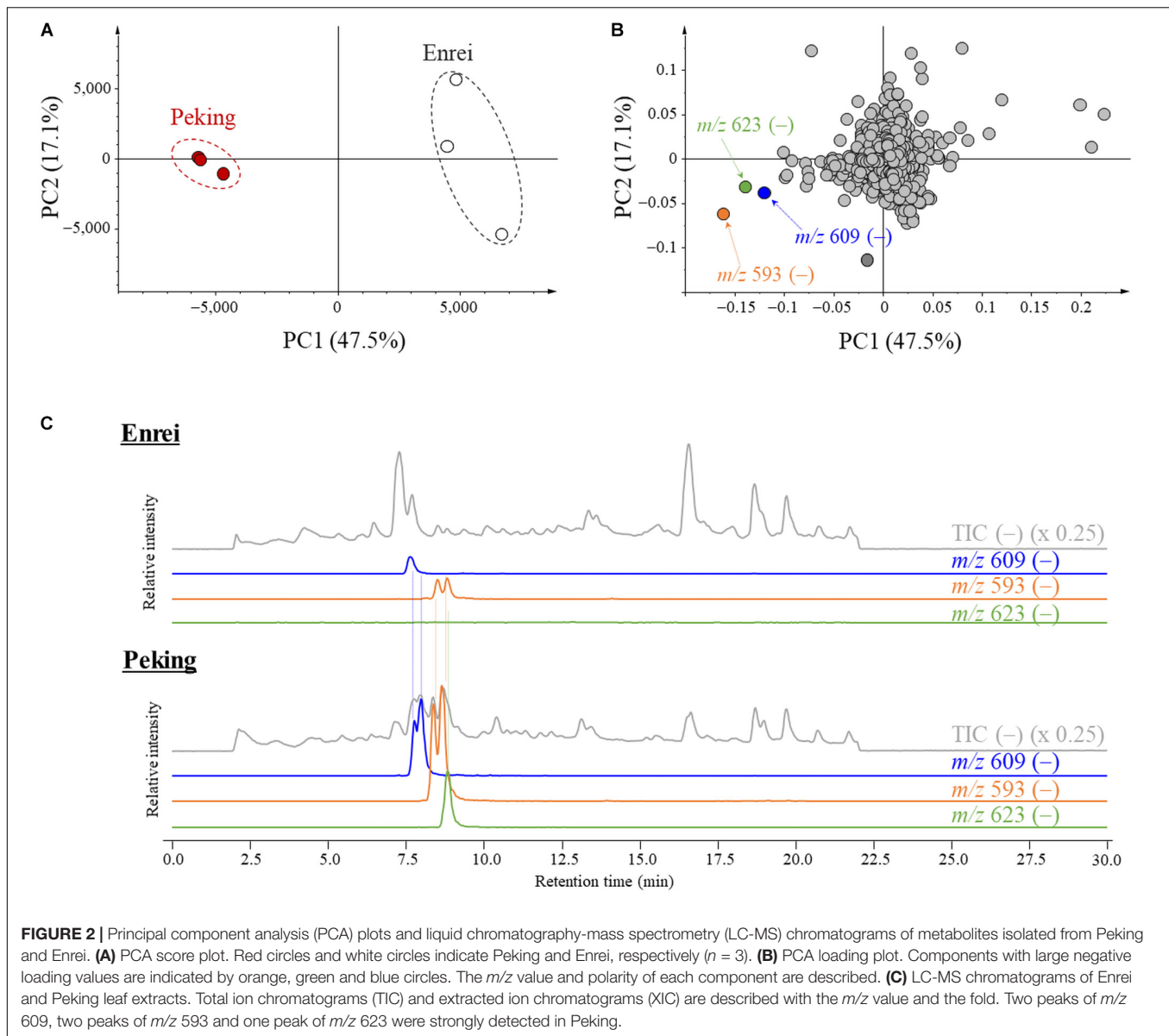
Metabolites B–G were identified as quercetin 3-*O*-rutinoside, kaempferol 3-*O*-gentiobioside, kaempferol 3-*O*-robinobioside, kaempferol 3-*O*-rutinoside, isorhamnetin 3-*O*-robinobioside and isorhamnetin 3-*O*-rutinoside, respectively (**Figures 3B–G**). The observed chemical shifts accorded with those of values reported in the literature (Demirezer et al., 2006; Güvenalp et al., 2006; Nogueira and Lopes, 2012; Duong et al., 2017; Viet Thanh et al., 2018; **Supplementary Table S3**). Each molecular formula and the MS/MS fragments are summarized in **Supplementary Table S2**.

Genomic Segments of Peking Associated With the Accumulation of Metabolites A–G and Common Cutworm Resistance

The genomic segments of Peking associated with the accumulation of metabolites A–G were screened using the CSSLs. The amounts of each metabolite in all 103 CSSLs were quantified by LC-Orbitrap MS equipped with a PFP column, which enabled the separation of metabolites A–G in a short run time (**Supplementary Figure S2**). Next, the association of 321 marker genotypes with the amount of metabolites A–G was tested by a one-way ANOVA. The obtained P -values, converted to $-\text{Log}P$, were plotted with respect to each marker number (#) (**Figure 4**). A large $-\text{Log}P$ value in these plots indicated that the genomic segments of Peking tagged by the marker affected the amount of each metabolite. The marker numbers (#) with the largest $-\text{Log}P$ values were #318 (metabolite A), #89 (B), #89 (C), #92 (D), #92 (E), #89 (F), and #89 (G) (indicated by red colored arrows in **Figure 4**). The marker #89 and #92 were located on Chr06, while the marker #318 was on Chr20.

The amounts of metabolites A–G in three CSSLs (B0341, B0331, and B0309), Enrei and Peking were compared (**Figure 5**). The CSSLs B0341, B0331 and B0309 had either one of the Peking-genomic segments tagged by the marker #89, #92, and #318, respectively. The amount of metabolite A in B0309 was higher than that in the other two lines and Enrei. The CSSL B0341 contained a higher amount of metabolites B, F, and G than the other two lines and Enrei, while B0331 contained a higher amount of metabolites D and E. In addition, the amount of metabolites C in B0309 was slightly higher than that in the other two lines and Enrei. According to these results, the amount of metabolite A, B, D, E, F, and G was clearly increased by the presence of the Peking-genomic segment, which exhibited the largest $-\text{Log}P$ values in the ANOVA for each metabolite (**Figure 4**). On the other hand, the amount of metabolite C was not increased by this (i.e., the marker #89) but increased by the marker #318, which exhibited a relatively large $-\text{Log}P$ value in the ANOVA (see the blue colored arrow in **Figure 4C**).

The effect of the Peking-genomic segments tagged by the markers #89, 92, and 318 on CCW growth was examined (**Figure 6**). Among the three CSSLs (B0341, B0331, and B0309),



Enrei and Peking, the weight of larvae fed on B0309 was smaller but not significantly different from that of larvae fed on Enrei and the other CSSLs. Although B0309 did not show the same level of CCW resistance of Peking, the Peking-genomic segments tagged by the marker #318 have genetic factor (s) that inhibit the growth of CCW. In the vicinity of the marker #318 (Chr20: 40,999,771 bp), the biosynthetic genes or pseudogenes of the biosynthesis of metabolite A and C were not found on the Glyma.Wm.82.a2.v1 assembly using SoyBase (last accessed on 24 June 2020)¹. However, a glycosyltransferase gene, Glyma.20g167900, was identified on Chr20: 40,568,843–40,578,286 bp. This gene is annotated to have quercetin 3-O-glucosyltransferase activity (GO:0080043) and quercetin 7-O-glucosyltransferase activity (GO:0080044). In

the vicinity of marker #89 (Chr06: 18,737,344 bp), flavonoid 3'-monooxygenase (Glyma.06g202300) is located on Chr06: 18,731,105–18,738,025 bp. In the vicinity of marker #92 (Chr06: 48,016,322 bp), the biosynthetic genes of these metabolites were not identified; however, a glycosyltransferase gene, Glyma.06g285700, is located on Chr06: 47,431,347–47,433,567 bp, 584,975 bp apart from the marker #92. This gene is annotated to have quercetin 3-O-glucosyltransferase activity (GO:0080043) and quercetin 7-O-glucosyltransferase activity (GO:0080044).

DISCUSSION

In the present study, we revealed the genetic and metabolic background of CCW resistance of Peking using CSSLs. Seven

¹<https://soybase.org/>

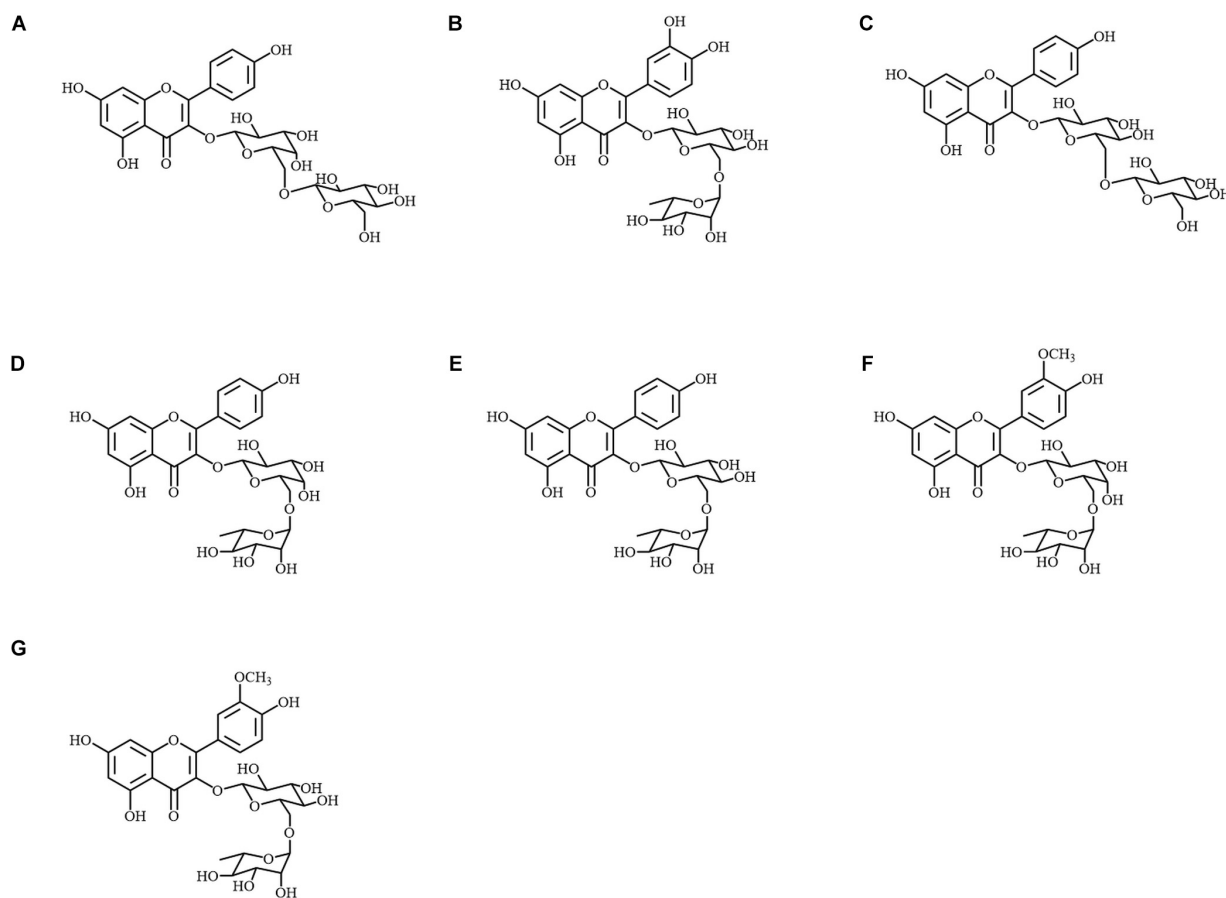


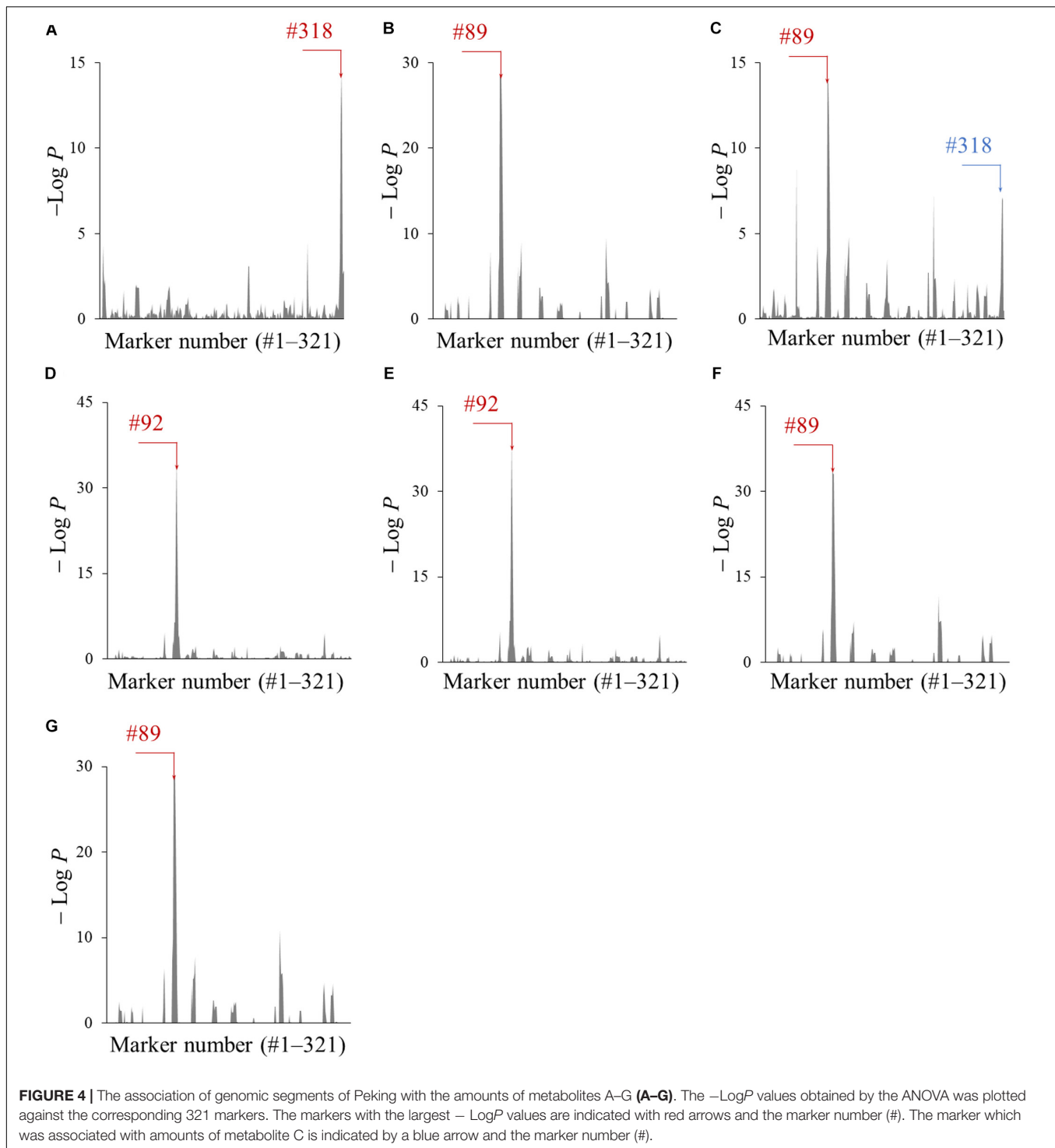
FIGURE 3 | Chemical structures of metabolites A–G isolated from Peking. **(A)** Kaempferol 3-O-glucosyl-(1→6)-galactoside; **(B)** quercetin 3-O-rutinoside; **(C)** kaempferol 3-O-gentiobioside; **(D)** kaempferol 3-O-robinobioside; **(E)** kaempferol 3-O-rutinoside; **(F)** isorhamnetin 3-O-robinobioside; **(G)** isorhamnetin 3-O-rutinoside.

characteristic metabolites for Peking expected to associate with CCW resistance were identified by metabolomic analyses using LC-MS followed by a PCA. Two Peking genomic segments on Chr06 and Chr20 were associated with the accumulation of these metabolites. Subsequently, the Peking genomic segment on Chr20 was associated with CCW resistance.

The CSSL B0309 possessing the Peking Chr20 genomic segment tagged by marker #318 inhibited the growth of CCW (**Figure 6**) and accumulated metabolite A (kaempferol 3-O-glucosyl-(1→6)-galactoside) and C (kaempferol 3-O-gentiobioside) (**Figure 5**). These results suggest that metabolite A and C are the potential resistance factors against CCW. The biological activity of these metabolites on insects have not been reported; however, Varghese et al. (2013) revealed α -glucosidase inhibitory activity of metabolite C. α -Glucosidase is one of the important digestive enzymes in the insect midgut (Terra et al., 2012). The inhibition of α -glucosidase reduces the survival and development of insects, including CCW larvae (Kaur et al., 2019). To demonstrate that metabolite A and C are resistance factors in Peking, the evaluation of the growth inhibition of CCW larvae and their biological activity

using purified substances is ongoing. Although genes related to the biosynthesis and accumulation of metabolite A and C have not been fully identified, the biosynthesis of kaempferol (i.e., an aglycone moiety) following glycosylation is deduced. In the vicinity of the marker #318, two genes annotated as quercetin 3-O-glucosyltransferase activity and quercetin 7-O-glucosyltransferase. Further functional studies of this gene and its relationship with CCW resistance is required. The comprehensive gene expression study based on RNA-seq analysis may facilitate narrow down candidate genes in the vicinity of the marker #318. In addition, genes in other segments are also required for reproducing the same level of CCW resistance as in Peking, because this segment alone did not confer the high accumulation level of metabolites A and C that was observed in Peking (**Figure 5**) and did not produce strong CCW resistance comparable to that of Peking (**Figure 6**).

The metabolites B (quercetin 3-O-rutinoside), F (isorhamnetin 3-O-robinobioside), and G (isorhamnetin 3-O-rutinoside) were accumulated in B0341 possessing the marker #89 (**Figure 5**). In the vicinity of marker #89, flavonoid 3'-monooxygenase is located on Chr06. Flavonoid



3'-monooxygenase is responsible for the biosynthesis of the aglycone of metabolites B, F, and G. This enzyme converts kaempferol to quercetin by oxidation at position 3', which is an aglycone of metabolite B; subsequently, quercetin is converted to isorhamnetin by methyltransferase, which is an aglycone of metabolite G. The accumulation level of metabolite B in B0341 was clearly higher than that in Peking, whereas that

of metabolites F and G were lower than in Peking. The over-accumulation of metabolite B in B0341 would be caused by a lack of other Peking-genomic segments, which would precede the metabolism of metabolite B to others in Peking. In contrast, the low accumulation level of metabolite F and G would be due to the requirements of other Peking-genomic segments to reproduce the high accumulation levels, as observed in Peking.

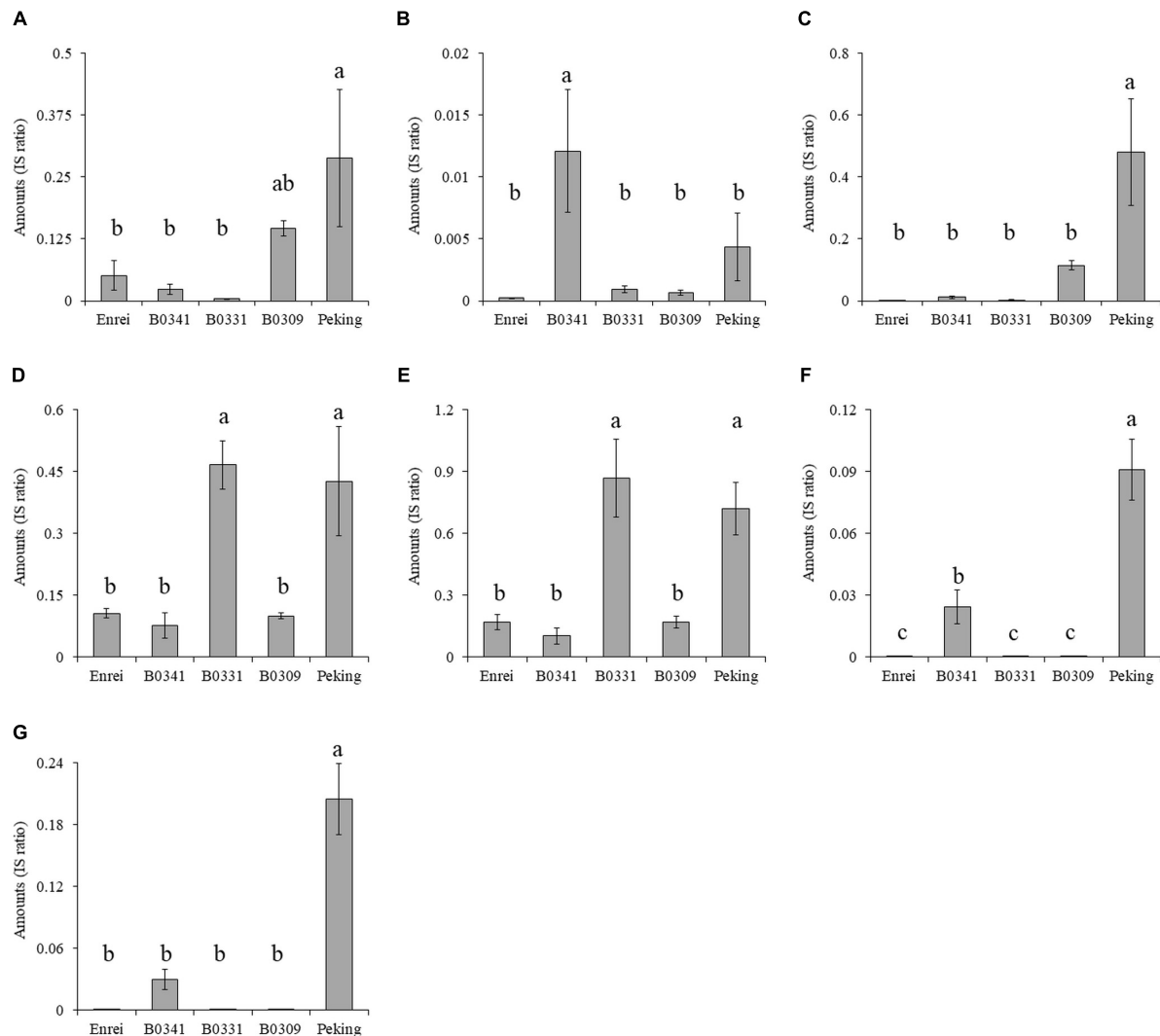


FIGURE 5 | The amounts of metabolites (A–G) in three chromosome segment substitution lines (B0341, B0331 and B0309), Enrei and Peking. Different letters indicate significant differences between lines (Tukey's test, $n = 3$, $P < 0.05$). B0341, B0331 and B0309 possess the Peking-genomic segment tagged by the markers #89, #92 and #318, respectively.

The accumulation level of metabolites D (kaempferol 3-*O*-robinobioside) and E (kaempferol 3-*O*-rutinoside) in B0331 were the same as in Peking (Figure 5). This suggests that the Chr06 Peking-genomic segment tagged by the marker #92 contained gene(s) that enable the accumulation of these metabolites at the same levels as observed in Peking. The biosynthetic genes of these metabolites were not identified in the vicinity of marker #92, while two genes (quercetin 3-*O*-glucosyltransferase activity and quercetin 7-*O*-glucosyltransferase activity) were located 585 kbp apart from the marker #92.

Although B0309 inhibited the larval growth of CCW, the inhibitory effect was not at the same level as Peking (Figure 6). This indicates that other Peking-genomic segments would be required to reproduce the same resistance level of Peking. In previous studies, two antibiosis QTLs (Komatsu et al., 2005) and one antixenosis QTL (Oki et al., 2012) for CCW from a

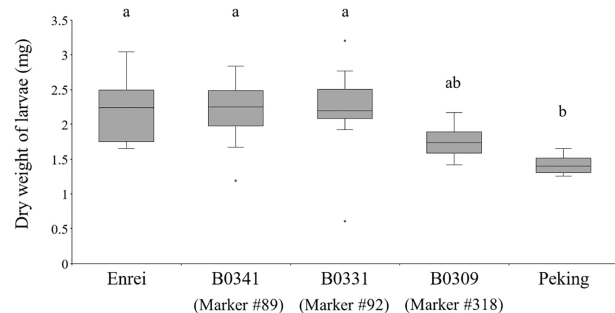


FIGURE 6 | The dry weight of common cutworm larvae fed on leaves of the three chromosome segment substitution lines (B0341, B0331, and B0309), Enrei and Peking. Different letters indicate significant differences between lines (Tukey's test, $n = 8$, $P < 0.05$).

resistant cultivar, “Himeshirazu,” were reported. Among them, two antibiosis QTLs, *CCW1*, and *CCW2*, did not reproduce CCW resistance at the same level as “Himeshirazu” (Komatsu et al., 2004). Although there are limitations for evaluating CCW resistance in terms of the replication scale and uniformity of plant and CCW growth, we are planning to evaluate abundance of characteristic secondary metabolites for Peking and the larval growth of CCW for all the CSSLs to advance our understanding of the resistance mechanisms of Peking.

In the present study, we uncovered CCW-resistant genetic factor(s) on Chr20 derived from the primitive soybean landrace Peking. In addition, metabolite A (kaempferol 3-O-glucosyl-(1→6)-galactoside) and C (kaempferol 3-O-gentiobioside) are potential resistance compounds. These would be novel resistant factors because *CCW1* and *CCW2* in “Himeshirazu” were identified on Chr07 (Komatsu et al., 2005). On the other hand, a direct evidence for CCW resistance of metabolite A and C is required for proving the contribution of these metabolites to CCW resistance in Peking. Thus, we are planning to evaluate CCW resistant activity of these metabolites by feeding assay using purified substances. In the present soybean breeding program in Japan, efforts of pyramiding several CCW-resistant genes from wild soybeans have continued (Oki et al., 2017). Although the QTLs related to CCW resistance found in “Himeshirazu” (*CCW1* and *CCW2*) have been introduced into a leading cultivar, it has not exhibited strong enough resistance to CCW. Therefore, the genetic factors on Chr20 will contribute to understanding CCW resistance mechanisms and can be used to improve the resistance of current resistant cultivars in a breeding program. Since break down linkage between genetic factors for resistant and undesirable characters and integration of resistant genetic factors into genetic background of a leading cultivar are required, multigenic nature of resistance will impede breeding within short period even if we found good wild or un-adapted genetic resources for the resistance. Similarly, the primitive landrace Peking may have undesirable characters beside blackish and flat small seed. CSSLs used in the present study are useful to examine undesirable characters and will be good materials for the development of a durable CCW-resistant cultivars. Especially, a series of CSSLs retain all resistance factors of Peking because CSSLs cover all chromosomes in contrast to breeding materials which lose many minor resistant factors during selection in the breeding. If we can develop methods that efficiently evaluate metabolites that collectively affect the insect resistance as shown in the present study and evaluate precisely morphological traits such as shape of trichome to the insect resistance, an elite

cultivar with more durable insect resistance can be developed by inter-crossing CSSLs without undesirable characters in future.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**. Metabolomics data have been deposited to the EMBL-EBI MetaboLights database (Haug et al., 2020) with the identifier MTBLS1960. The complete dataset can be accessed here <https://www.ebi.ac.uk/metabolights/MTBLS1960>.

AUTHOR CONTRIBUTIONS

RN, MY, MT, YO, NM, and AK: study design and data analysis. RN and MY: phenotype data collection. RN, MY, and SH: metabolomics data collection. MY and NM: structural determination of metabolites. RN, NM, and AK: manuscript drafting. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by JSPS KAKENHI grant nos. JP18KT0042 and JP19J01010.

ACKNOWLEDGMENTS

We would like to thank Kyouka Ise of Kyoto University for support of the CCW rearing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.581917/full#supplementary-material>

The purification scheme and LC-MS chromatograms (XIC) of authentic samples of the metabolites are provided in **Supplementary Figures S1, S2**. The NMR spectra of metabolite A is presented as **Supplementary Table S1**. The molecular formulae and MS/MS fragments obtained by LC/Orbitrap-MS analyses are summarized in **Supplementary Table S2**. The chemical shifts of the metabolites are provided in **Supplementary Table S3**.

REFERENCES

- Batyrshina, Z. S., Yaakov, B., Shavit, R., Singh, A., and Tzin, V. (2020). Comparative transcriptomic and metabolic analysis of wild and domesticated wheat genotypes reveals differences in chemical and physical defense responses against aphids. *BMC Plant Biol.* 20:19. doi: 10.1186/s12870-019-2214-z
- Demirezer, L. Ö., Gürbüz, F., Güvenalp, Z., Ströck, K., and Zeeck, A. (2006). Iridoids, flavonoids and monoterpene glycosides from *Galium verum* subsp. *verum*. *Turk. J. Chem.* 30, 525–534.
- Duong, P. Q., Duyen, N. T., Quyen, P. T., Tung, N. Q., Son, V. H., and Hung, V. D. (2017). Isolation and identification of phenolic compounds from the leaf extract of *Cassia alata* L. *Vietnam J. Chem.* 55, 589–594.
- Fehr, W. R., and Caviness, C. E. (1977). Stages of soybean development. *Special Rep.* 87, 1–11.
- Fernie, A. R., and Tohge, T. (2017). The genetics of plant metabolism. *Annu. Rev. Genet.* 51, 287–310. doi: 10.1146/annurev-genet-120116-024640
- Güvenalp, Z., Özbek, H., Ünsalar, T., Kazaz, C., and Demirezer, L. Ö. (2006). Iridoid, flavonoid, and phenylethanoid glycosides from *Wiedemannia orientalis*. *Turk. J. Chem.* 30, 391–400.

- Hamany Djande, C. Y., Pretorius, C., Tugizimana, F., Piater, L. A., and Dubery, I. A. (2020). Metabolomics: a tool for cultivar phenotyping and investigation of grain crops. *Agronomy* 10:831. doi: 10.3390/agronomy10060831
- Haug, K., Cochrane, K., Nainala, V. C., Williams, M., Chang, J., Jayaseelan, K. V., et al. (2020). MetaboLights: a resource evolving in response to the needs of its scientific community. *Nucleic Acids Res.* 48, D440–D444. doi: 10.1093/nar/gkz1019
- Hayes, A. J., Ma, G., Buss, G. R., and Maroof, M. A. S. (2000). Molecular marker mapping of Rsv4, a gene conferring resistance to all known strains of soybean mosaic virus. *Crop. Sci.* 40, 1434–1437. doi: 10.2135/cropsci2000.4051434x
- Hymowitz, T. (1990). "Soybeans: the success story," in *Advances in New Crops*, eds J. Janick and J. E. Simon (Portland: Timber Press), 159–163.
- Hyndman, R. J., and Fan, Y. (1996). Sample quantiles in statistical packages. *Am. Stat.* 50, 361–365. doi: 10.1080/00031305.1996.10473566
- Jiang, C. J., Sugano, S., Kaga, A., Lee, S. S., Sugimoto, T., Takahashi, M., et al. (2017). Evaluation of resistance to phytophthora sojae in soybean mini core collections using an improved assay system. *Phytopathology* 107, 216–223. doi: 10.1094/PHYTO-06-16-0233-R
- Joshi, J. M. (1980). Research notes: evaluation of soybean germplasm for resistance to corn earworm—III. *Soybean Genet. Newslett.* 6:37.
- Kaur, J., Sharma, A., Sharma, M., Manhas, R. K., Kaur, S., and Kaur, A. (2019). Effect of α -glycosidase inhibitors from endophytic fungus *alternaria destruens* on survival and development of insect pest *Spodoptera litura* fab. and fungal phytopathogens. *Sci. Rep.* 9:11400. doi: 10.1038/s41598-019-47838-6
- Komatsu, K., Okuda, S., Takahashi, M., and Matsunaga, R. (2004). Antibiotic effect of insect-resistant soybean on common cutworm (*Spodoptera litura*) and its inheritance. *Breed. Sci.* 54, 27–32. doi: 10.1270/jsbbs.54.27
- Komatsu, K., Okuda, S., Takahashi, M., Matsunaga, R., and Nakazawa, Y. (2005). QTL mapping of antibiosis resistance to common cutworm (*Spodoptera litura* Fabricius) in soybean. *Crop Sci.* 45, 2044–2048. doi: 10.2135/cropsci2004.0744
- Komatsu, K., Takahashi, M., and Nakazawa, Y. (2008). Antibiosis resistance of QTL introgressive soybean lines to common cutworm (*Spodoptera litura* fabricius). *Crop Sci.* 48, 527–532. doi: 10.2135/cropsci2006.12.0770
- Li, Y. H., Zhou, G., Ma, J., Jiang, W., Jin, W., Jin, L. G., et al. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* 32, 1045–1052. doi: 10.1038/nbt.2979
- Liu, H., Che, Z., Zeng, X., Zhang, G., Wang, H., and Yu, D. (2016). Identification of single nucleotide polymorphisms in soybean associated with resistance to common cutworm (*Spodoptera litura* Fabricius). *Euphytica* 209, 49–62. doi: 10.1007/s10681-016-1631-4
- Lu, B. R. (2004). Conserving biodiversity of soybean gene pool in the biotechnology era. *Plant Spec. Biol.* 19, 115–125. doi: 10.1111/j.1442-1984.2004.00108.x
- Luo, J. (2015). Metabolite-based genome-wide association studies in plants. *Curr. Opin. Plant Biol.* 24, 31–38. doi: 10.1016/j.pbi.2015.01.006
- Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x
- Nogueira, C. R., and Lopes, L. M. X. (2012). Nitro derivatives and other constituents of *Aristolochia melastoma*. *Quim. Nova* 35, 780–782. doi: 10.1590/S0100-40422012000400024
- Oki, N., Kaga, A., Shimizu, T., Takahashi, M., Kono, Y., and Takahashi, M. (2017). QTL mapping of antixenosis resistance to common cutworm (*Spodoptera litura* Fabricius) in wild soybean (*Glycine soja*). *PLoS One* 12:e0189440. doi: 10.1371/journal.pone.0189440
- Oki, N., Komatsu, K., Sayama, T., Ishimoto, M., Takahashi, M., and Takahashi, M. (2012). Genetic analysis of antixenosis resistance to common cutworm (*Spodoptera litura* Fabricius) and its relationship with pubescence characteristics in soybean (*Glycine max* (L.) Merr.). *Breed. Sci.* 61, 608–617. doi: 10.1270/jsbbs.61.608
- Oki, N., Komatsu, K., Takahashi, M., Takahashi, M., Kono, Y., and Ishimoto, M. (2015). Field assessment of resistance QTL to common cutworm in soybean. *Crop Sci.* 55, 624–630. doi: 10.2135/cropsci2014.05.0364
- Oki, N., Takagi, K., Ishimoto, M., Takahashi, M., and Takahashi, M. (2019). Evaluation of the resistance effect of QTLs derived from wild soybean (*Glycine soja*) to common cutworm (*Spodoptera litura* Fabricius). *Breed. Sci.* 69, 529–535. doi: 10.1270/jsbbs.18157
- Ross, J., and Brim, C. A. (1957). Resistance of soybean to the soybean-cyst nematode as determined by a double-row method. *Plant Dis. Rep.* 41, 923–924.
- Tanaka, M., Fujimori, T., Uchida, I., Yamaguchi, S., and Takeda, K. (2001). A malonylated anthocyanin and flavonols in blue *Meconopsis* flowers. *Phytochemistry* 56, 373–376. doi: 10.1016/S0031-9422(00)00357-5
- Terra, W. R., Ferreira, C., and Gilbert, L. I. (2012). "Biochemistry and molecular biology of digestion," in *Insect Molecular Biology and Biochemistry*, (San Diego: Academic Press), 365–418.
- Varghese, G. K., Bose, L. V., and Habtemariam, S. (2013). Antidiabetic components of *Cassia alata* leaves: identification through α -glucosidase inhibition studies. *Pharm. Biol.* 51, 345–349. doi: 10.3109/13880209.2012.729066
- Viet Thanh, N. T., Thu Hien, D. T., Minh, T. T., Nhiem, N. X., Yen, P. H., and Van Kiem, P. (2018). Flavonol glycosides from *Phoebe poilanei* Kosterm. *Vietnam J. Chem.* 56, 711–716. doi: 10.1002/vjch.201800075
- Wang, H., Yan, H., Du, H., Chao, M., Gao, Z., and Yu, D. (2015). Mapping quantitative trait loci associated with soybean resistance to common cutworm and soybean compensatory growth after defoliation using SNP marker-based genome-wide association analysis. *Mol. Breed.* 35:168. doi: 10.1007/s11032-015-0360-z
- Watanabe, S., Shimizu, T., Machita, K., Tsubokura, Y., Xia, Z., Yamada, T., et al. (2018). Development of a high-density linkage map and chromosome segment substitution lines for Japanese soybean cultivar Enrei. *DNA Res.* 25, 123–136. doi: 10.1093/dnares/dsx043
- Way, M. O. (1994). "Status of soybean insect pests in the United States," in *Handbook of Soybean Insect Pests*, eds L. G. Hingley and D. J. Boethel (Landham: Entomological Society of America), 15–16.
- Xu, G., Cao, J., Wang, X., Chen, Q., Jin, W., Li, Z., et al. (2019). Evolutionary metabolomics identifies substantial metabolic divergence between maize and its wild ancestor. *Teosinte. Plant Cell* 31, 1990–2009. doi: 10.1105/tpc.19.00111
- Yandeau-Nelson, M. D., Lauter, N., and Zabolina, O. A. (2015). Advances in metabolomic applications in plant genetics and breeding. *CAB Rev.* 10, 1–15. doi: 10.1079/PAVSNNR201510040
- Zhang, J., Yang, D., Li, M., and Shi, L. (2016). Metabolic profiles reveal changes in wild and cultivated soybean seedling leaves under salt stress. *PLoS One* 11:e0159622. doi: 10.1371/journal.pone.0159622

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

Copyright © 2020 Nakata, Yano, Hiraga, Teraishi, Okumoto, Mori and Kaga. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Impact of Domestication on Aboveground and Belowground Trait Responses to Nitrogen Fertilization in Wild and Cultivated Genotypes of Chickpea (*Cicer* sp.)

OPEN ACCESS

Edited by:

Gaofeng Zhou,

Department of Primary Industries and
Regional Development of Western
Australia (DPIRD), Australia

Reviewed by:

Cengiz Toker,

Akdeniz University, Turkey

Jiayin Pang,

University of Western Australia,
Australia

*Correspondence:

Eric von Wettberg

eric.bishop-von-wettberg@uvm.edu;

ebishopv@uvm.edu

Specialty section:

This article was submitted to

Evolutionary and Population

Genetics,

a section of the journal

Frontiers in Genetics

Received: 25 June 2020

Accepted: 06 November 2020

Published: 02 December 2020

Citation:

Marques E, Krieg CP,

Dacosta-Calheiros E, Bueno E,

Sessa E, Penmetsa RV and von

Wettberg E (2020) The Impact of

Domestication on Aboveground and

Belowground Trait Responses to

Nitrogen Fertilization in Wild and

Cultivated Genotypes of Chickpea

(*Cicer* sp.).

Front. Genet. 11:576338.

doi: 10.3389/fgene.2020.576338

Edward Marques^{1,2}, Christopher P. Krieg³, Emmanuel Dacosta-Calheiros², Erika Bueno², Emily Sessa³, R. Varma Penmetsa⁴ and Eric von Wettberg^{1,2*}

¹Department of Plant and Soil Science and Gund Institute for the Environment, University of Vermont, Burlington, VT, United States, ²Department of Biological Sciences, Florida International University, Miami, FL, United States, ³Department of Biology, University of Florida, Gainesville, FL, United States, ⁴Department of Plant Sciences, University of California, Davis, CA, United States

Despite the importance of crop responses to low fertility conditions, few studies have examined the extent to which domestication may have limited crop responses to low-fertility environments in aboveground and belowground traits. Moreover, studies that have addressed this topic have used a limited number of wild accessions, therefore overlooking the genotypic and phenotypic diversity of wild relatives. To examine how domestication has affected the response of aboveground and belowground agronomic traits, we measured root and leaf functional traits in an extensive set of wild and domesticated chickpea accessions grown in low and high nitrogen soil environments. Unlike previous studies, the wild accessions used in this study broadly capture the genetic and phenotypic diversity of domesticated chickpea's (*Cicer arietinum*) closest compatible wild relative (*C. reticulatum*). Our results suggest that the domestication of chickpea led to greater capacities for plasticity in morphological and biomass related traits but may have lowered the capacity to modify physiological traits related to gas exchange. Wild chickpea displayed greater phenotypic plasticity for physiological traits including stomatal conductance, canopy level photosynthesis, leaf level photosynthesis, and leaf C/N ratio. In contrast to domesticated chickpea, wild chickpea displayed phenotypes consistent with water loss prevention, by exhibiting lower specific leaf area, stomatal conductance and maintaining efficient water-use. In addition to these general patterns, our results indicate that the domestication dampened the variation in response type to higher nitrogen environments for belowground and aboveground traits, which suggests reduced genetic diversity in current crop germplasm collections.

Keywords: resource use efficiency, functional traits, cicer, roots, phenotypic plasticity 3

INTRODUCTION

The practice of artificial selection was a critical innovation in human history that allowed for the rapid directional modification of traits in plants and animals. In crops like cereals and legumes, humans primarily selected for traits such as fruit indehiscence, reduced seed dormancy, and yield (e.g., Gross and Olsen, 2010; Meyer et al., 2012; Olsen and Wendel, 2013; Smýkal et al., 2018). However, the selection to modify one trait can often lead to a modification in other traits due to trait covariation and underlying genetic linkage (Lande and Arnold, 1983; Price and Langen, 1992). The evolution of traits through the inadvertent selection of correlated traits and genetic linkage are well documented in a variety of species (Rauw et al., 1998; Hoekstra et al., 2001; Kingsolver et al., 2001; Hereford et al., 2004; Kingsolver and Pfennig, 2007). In crops, similar patterns of correlated selection combined with population bottlenecks during domestication may have unintentionally altered non-target traits, potentially canalizing crop responses to different environmental conditions (Flatt, 2005; Morrell et al., 2011; Meyer et al., 2012; Gaut et al., 2018; Smýkal et al., 2018; Lye and Purugganan, 2019). Understanding the degree to which domestication has canalized or otherwise altered plant traits, and the ability of plants to respond to low fertility environments can aid agricultural programs to combat food insecurity in a changing global climate.

Although many comparative studies have demonstrated how artificial selection can lead to marked decreases in genomic and phenotypic variation in domesticated plants compared to wild relatives (e.g., Morrell et al., 2011; Olsen and Wendel, 2013; Gaut et al., 2018; Hufford et al., 2019; Lye and Purugganan, 2019), the majority of comparative research of phenotypes has focused on the impacts of domestication on aboveground agronomic traits such as seed size or shattering (e.g., Milla et al., 2015; Ogutcen et al., 2018; Smýkal et al., 2018). Relatively few studies have examined the potentially canalizing effects of domestication on belowground functional traits such as root architecture and root-soil-nutrient dynamics (e.g., Bulgarelli et al., 2015; Milla et al., 2015; Pérez-Jaramillo et al., 2016). Even fewer have taken a whole-plant approach to understand the impact of domestication on aboveground and belowground traits in tandem; thus, limiting our understanding of how domestication may have impacted plant function. Furthermore, most studies assessing the effects of domestication on crops have utilized very small numbers of genotypes of wild relatives (e.g., Warschewsky et al., 2014), limiting the power and potential to extrapolate from these comparisons.

Many crop wild relatives are found in environments with limited water availability and more nutrient-poor soils compared to their domesticated counterparts that occur in agricultural farms (e.g., Grossman and Rice, 2012; McKey et al., 2012; Milla et al., 2015). However, these wild habitats are heterogeneous, and likely to maintain phenotypic plasticity, in contrast to trait canalization (*sensu* Flatt, 2005). As a result, a shift to fertile environments during domestication (i.e., as humans often initially cultivated richer valley soils, and learned to till soils and fertilize crops with animal waste) may have

relaxed selective pressures on plant functional traits that impact resource acquisition like carbon, nitrogen, and water uptake from nutrient-poor soils or canalized responses under high fertility conditions (Grossman and Rice, 2012; Martin and Isaac, 2015). Canalization of nutrient uptake traits under nutrient-rich environments in cultivated crop lineages could lead to poorer performance than ancestral wild populations in nutrient-limiting environments from erosion of alleles for these traits in cultivated gene pools. This would result in a reduced capacity to grow in low nutrient conditions, such as those typical in many small holder farming systems in the developing world, for farmers restricted to marginal soils and to some organic production systems.

The impacts of domestication on belowground traits may be particularly pronounced for crops with complex soil interactions such as legumes. A recent study suggests that the domestication of common bean (*Phaseolus vulgaris* L.), for which domestication and post-domestication selection by humans has focused largely on the bean itself (i.e., pod shattering, yield, size, and flavor), has also resulted in shifts in traits critical to soil interactions and nutrient dynamics including root microbiome composition, increased specific root length (SPL), and decreased root density (Pérez-Jaramillo et al., 2017). Despite the impact of domestication on agronomic traits, a broad set of root functional traits remain unexplored for most of the world's most economically important crop species. For example, chickpea (*Cicer arietinum* L.) is the second most important grain legume globally and the leading legume in South Asia (FAO et al., 2017). The need for such studies in economically important crop species such as chickpea is more urgent than ever, with reductions in rainfall and soil fertility predicted to result in decreased yields in several food-insecure areas like India, Ethiopia, and Turkey, where chickpea is a key source of nutritional security and a cash crop (Singh et al., 2014; Ahmed et al., 2016). Therefore, understanding the degree to which domestication has impacted plant traits, and the ability of plants and traits to respond to new environments, is critical to adapting agricultural programs in a changing climate.

To understand how domestication affected aboveground and belowground agronomic traits, resource-use efficiency, and adaptive capacity in crops, we assembled a uniquely large collection of wild chickpeas from southeastern Turkey, providing sufficient numbers of genetically distinct wild genotypes to examine differentiation in aboveground and belowground phenotypes between cultivated crops and their wild relatives (von Wettberg et al., 2018). We grew wild and domesticated chickpea accessions in low and high nitrogen concentrations and measured root and leaf functional traits. We hypothesized that if domestication for typical agronomic traits has resulted in inadvertent selection in other functional traits due to cultivation in higher fertility environments that are typical of agriculture, then (1) wild accessions will have traits consistent with greater performance and resource use efficiency in low nutrient conditions compared to domesticated accessions and (2) domesticated accessions will exhibit lower phenotypic plasticity in root and leaf functional traits.

MATERIALS AND METHODS

Plant Germplasm Used

Twenty-seven genetically diverse accessions of chickpea were used in this study (Table 1). Six accessions: CDC Frontier, ICC16207, Gokce, Dwelley, Myles, and UC15 are cultivars originating from four countries: United States, Canada, Turkey, and India. These accessions were selected because they represent both chickpea market types, Desi and Kabuli (Penmetsa et al., 2016) and are widely grown in their native countries. The remaining 21 accessions are wild chickpea lines systematically collected from different regions of Turkey, the native range of wild chickpea (von Wettberg et al., 2018). These accessions were selected to maximize genetic and native environmental differences in the material to capture as much wild diversity as possible.

Experimental Design

All accessions were grown in a shade house at Fairchild Tropical Botanic Garden in Coral Gables, Florida, from December 2016 to March 2017. Average day and night temperature during this period ranged between 27 and 16°C, and average monthly rainfall was 5 cm (+/−0.75) (USclimatedata.com, 2020). Seeds of each accession were planted in 11-L pots containing 8 L of a mixture of sand and coconut coir. This mixture was used as a planting media to minimize the nitrogen present before preparation. Plants were watered every 48 h by an automatic sprinkler system.

TABLE 1 | Germplasm used in the study

Germplasm	Species	Geographical origin	History	Market type
CDC Frontier	<i>Cicer arietinum</i>	Canada	Domesticated	Kabuli
ICC16207	<i>Cicer arietinum</i>	India	Domesticated	Desi
Gokce	<i>Cicer arietinum</i>	Syria	Domesticated	Kabuli
Dwelley	<i>Cicer arietinum</i>	United States	Domesticated	Kabuli
Myles	<i>Cicer arietinum</i>	United States	Domesticated	Kabuli
UC 15	<i>Cicer arietinum</i>	United States	Domesticated	Kabuli
Bari1 092	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Bari2 072	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Bari3 072n2	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Bari3 100	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Bari3 106	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Besev 075	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Besev 079	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
CudiA 152	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
CudiB 022C	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Derei 070	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Derei 072	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Egill 065	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Egill 073	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Kalka 064	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Kayat 077	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Kesen 075	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Oyali 084	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Oyali 111	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Sarik 067	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Savur 063	<i>Cicer reticulatum</i>	Turkey	Wild	Wild
Sirna 060	<i>Cicer reticulatum</i>	Turkey	Wild	Wild

Eight replicates of each accession were subjected to two different nitrogen treatments: 1 ppm (2.362 mg N source/L planting media) and 100 ppm (238506.2 mg/L). ESN Polymer Coated Urea (Agrium U.S. Inc.), a slow-release nitrogen pellet, was used as the nitrogen source. These treatments were chosen to represent generally nitrogen poor conditions in the wild, and the nitrogen rich conditions that may be found in an agricultural field setting, respectively (Keen, 2020). To make sure other nutrients were not limiting for chickpea growth, all pots received 2.40 mg/L phosphorus (P) as $\text{Al}(\text{PO}_3)_3$, 470.8 mg/L calcium (Ca) as $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$, 507.8 mg/L magnesium (Mg) as $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 2.598 mg/L copper (Cu) as $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 5.401 mg/L zinc as $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 22.96 mg/L manganese (Mn) as $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 2.499 mg/L boron (B) as $\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$ [//], and 0.119 mg/L molybdenum (Mo) as $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$. Plants were grown in the absence of rhizobial symbionts, as evidence suggests that wild and cultivated chickpea differ in symbiont preference (Greenlon et al., unpublished; Greenlon et al., 2019), and symbionts differ in their tolerance of different soils (Alford et al., unpublished; Greenlon et al., 2019). All pots were randomly arranged in a grid in the shade house.

Gas-Exchange

We measure instantaneous rates of gas exchange to estimate key aboveground traits related to carbon gain, water conservation, and its impacts on nutrient turnover. Gas-exchange measurements were performed on mature leaflets for 6–8 individuals per genotype using the LI-6400 infrared gas analyzer (Li-6400, Li-Cor Inc., NE, United States). Chamber conditions were set to 1,300 μmol PAR and CO_2 concentration of 400 ppm. The block temperature was set to 28°C achieving an average leaf temperature of 28.91°C (+/− 1.35°C) and a cuvette vapor pressure deficit (VPD) of 1.38 kPa (+/− 0.37 kPa). After gas exchange rates had stabilized (≥ 6 min), net photosynthetic rates (A_N) and stomatal conductance (g_s) were recorded. The leaf area was corrected using digital photographs of the leaf material that was inside the chamber using ImageJ (Wayne Rasband/NIH, Bethesda, MD, United States). Gas-exchange measurements were taken between 800 and 1,300 h.

Stable Isotope Chemistry

We measured stable isotopes of carbon and nitrogen to gain an integrated lifetime estimate of water relations and nutrient movement to complement instantaneous water use measurements. The leaflets used for gas-exchange were cut and digitally photographed in the field (for later analysis of specific leaf area) and then placed into coin envelopes and stored in a drying oven at 75°C for at least 72 h before being weighed. Leaflet area was calculated in ImageJ (Wayne Rasband/NIH, Bethesda, MD, United States), and specific leaf area (SLA) was calculated from the ratio of fresh area (cm^2) and dry mass (g). The dried samples were then run through a Carlo Erba NC2500 elemental analyzer (CE Instruments Ltd., England, United Kingdom) in tandem with a Thermo Delta V Stable Isotope Mass Spectrometer (Thermo Fisher Scientific Inc., Waltham, MA, United States) at the Cornell University Isotope

Lab (COIL) to measure elemental chemistry, i.e., $\delta^{13}\text{C}$ and %N. Leaf chlorophyll investments were estimated using a Photosynthesis MultispeQ V1.0 (East Lansing, MI). Carbon isotope chemistry was used to estimate water-use-efficiency ($\delta^{13}\text{C}$), and total leaf carbon and nitrogen content were used to calculate C/N ratios, photosynthetic nitrogen use efficiency (PNUE), and nitrogen investment into chlorophyll (i.e., estimated by the ratio of leaf nitrogen and chlorophyll index).

Root and Canopy Morphology

We harvested whole plants to examine plant allocation to above and belowground tissues. Aboveground and belowground plant biomass was harvested 12 weeks after sowing. Aboveground biomass was defined as all living biomass above the soil. A subset of wild and domesticated replicates of each accession for each treatment (1 and 100 ppm) was randomly selected for leaf area measurements. All leaves of selected plants were removed, laminated, and scanned at 1200 dpi using an Epson Perfection V700 scanner (Epson America, Long Beach CA). Lamination prevented folding of leaves during scanning and allowed more measurements to be taken by slowing down wilting. The total canopy area was used to scale leaf level gas exchange measurements to canopy level photosynthesis. For the remaining plants, aboveground biomass was placed in a drying oven for 24 h, after which the dry mass was recorded using an analytical balance (Mettler Toledo ME103TE, Columbus, OH, United States). All belowground biomass was carefully separated from the soil, cleaned with deionized water, and scanned for image analysis. The samples were then dried and weighed as described above. The image analysis system, WinRHIZO (version *Arabidopsis*) was used to calculate root length (i.e., RL), average root diameter (i.e., RD), root surface area (i.e., RSA), root volume (i.e., RV), and leaf area (i.e., LA) from root and leaf scans (Regent Instruments, Quebec City, QC, Canada). Specific root length and root density were calculated by dividing root length (cm) by belowground biomass (g) and belowground biomass (g) by root volume (cm^3), respectively.

Data Analysis

A nested linear mixed model (LMM) was used to test for significant differences for all measurements between nitrogen treatments (1 and 100 ppm), history (wild or domesticated), and accessions (i.e., ecotype). Treatment was used as a fixed factor, while accession, a random factor, was nested into history, a fixed factor, making the combined term random. Furthermore, a Tukey's HSD test with a Bonferroni correction was used for pairwise comparisons. Furthermore, to test if ecotype had a significant effect, we calculated the value of p of ecotype through the *Ranova* function in the *Lmertest* package in R (Kuznetsova et al., 2017).

To understand if responses to soil nitrogen are dependent on domestication history, we examined the history by treatment interactions of our LMM for all traits (Bates et al., 2015; Harrel, 2015). To better understand the intensity of trait

responses, we further calculated the relative distance plasticity index (RDPI) for traits with significant history by treatment interactions (Valladares et al., 2006). Specifically, we calculated RDPI and tested for differences between domesticated and wild chickpea for six variables: specific root length, root density, water-use efficiency, aboveground biomass, canopy level photosynthesis, leaf level photosynthesis, and stomatal conductance using the “*Plasticity*” R package (Ameztegui, 2017). Lastly, we correlated three independent variables: specific root length (x^λ -transformed), root density (log-transformed), and water-use efficiency (x^λ -transformed), against aboveground biomass (log-transformed) to test whether plant plasticity may affect plant fitness. For each correlation, we calculated the Pearson correlation coefficient and determined its significance using a *Psych* R package (Revelle, 2019). We used plant biomass as a fitness indicator rather than seed set because we harvested before flowering to capture intact root systems. Additionally, seed set and plant biomass are often correlated, making it an appropriate fitness indicator (e.g., von Wettberg et al., 2008). All statistical analyses were performed in R (R Core Team, 2019).

RESULTS

Adaptation to the Low Nitrogen Environment

Domesticated and wild chickpea displayed similar adaptations to the lower nitrogen environment (1 ppm) with only a handful of traits being significantly different between these groups. All belowground morphological measurements between domesticated and wild chickpea were non-significant; belowground biomass ($t_{46} = 1.763$, $p = 0.507$; **Figure 1B**), root length ($t_{49} = 1.870$, $p = 0.404$), root density ($t_{59} = -1.940$, $p = 0.343$; **Figure 1D**), average root diameter ($t_{53} = 1.460$, $p = 0.902$), root volume ($t_{54} = 1.128$, $p = 1.000$), SRL ($t_{87} = 1.145$, $p = 1.000$; **Figure 1C**), and root surface area ($t_{43} = 1.903$, $p = 0.383$).

This trend continued in regard to aboveground traits, we found very few differences between domesticated and wild chickpea in low nitrogen conditions with respect to aboveground biomass ($t_{56} = 1.054$, $p = 1.000$; **Figure 1A**), water-use efficiency ($t_{42} = -2.397$, $p = 0.127$; **Figure 2A**), %N ($t_{63} = 1.172$, $p = 1.000$), %C ($t_{85} = 0.294$, $p = 1.000$), chlorophyll content index ($t_{52} = 1.247$, $p = 1.000$), nitrogen investment into chlorophyll ($t_{48} = -0.581$, $p = 1.000$), stomatal conductance ($t_{144} = -3.236$, $p = 0.009$), canopy photosynthetic rates ($t_{27} = 0.237$, $p = 1.000$; **Figure 2D**), photosynthetic nitrogen use efficiency (per area $t_{32} = 0.106$, $p = 1.000$; per mass $t_{35} = 0.697$, $p = 1.000$; **Figure 2B**), and leaf level photosynthetic rate (photosynthetic rate per area $t_{27} = 0.549$, $p = 1.000$; leaf level photosynthetic rate per mass $t_{30} = 0.871$, $p = 1.000$). However, several other aboveground functional traits varied significantly between wild and domesticated chickpea within the low nitrogen environment, with wild chickpea displaying lower specific leaf area ($t_{41} = 5.608$, $p < 0.001$; **Figure 2C**) and stomatal conductance ($t_{144} = -3.236$, $p = 0.009$), and higher carbon to nitrogen ratio ($t_{292} = -3.909$, $p = 0.001$) than domesticated chickpea.

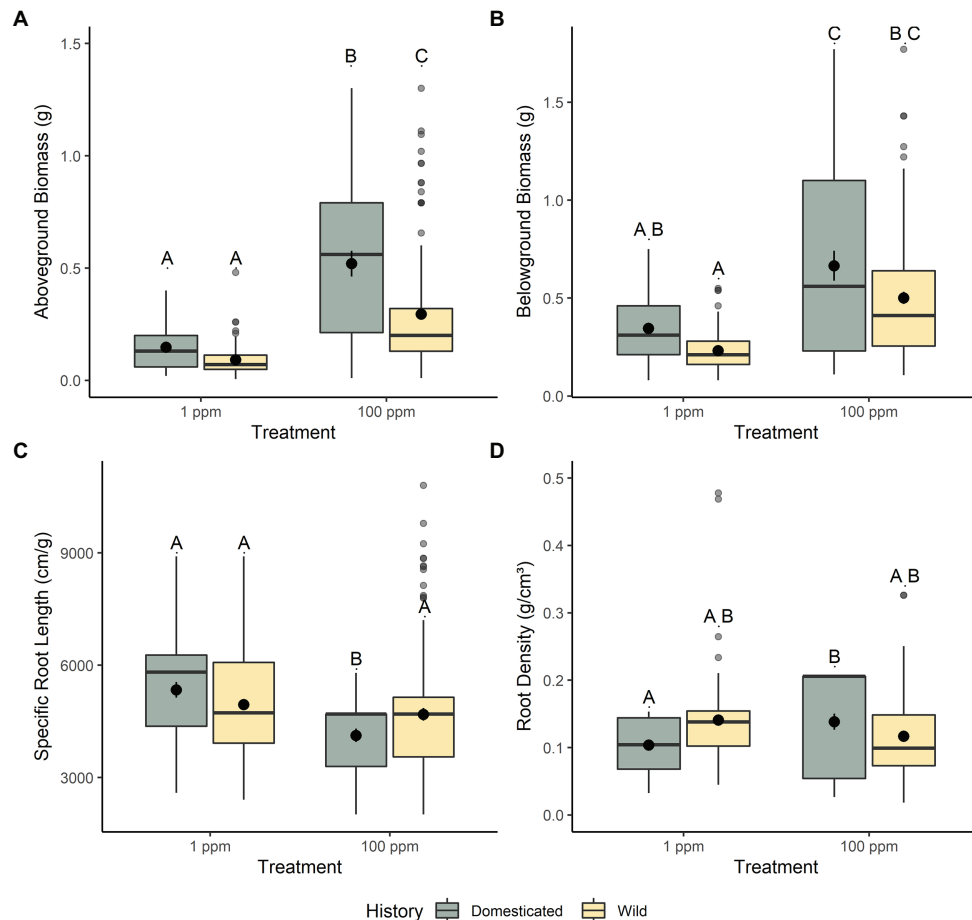


FIGURE 1 | The response of chickpea morphology to increased nitrogen availability. **(A)** Aboveground biomass, **(B)** belowground biomass, **(C)** specific root length (SRL), and **(D)** root density. Domesticated (yellow) and wild (green) chickpea accessions are grouped. Different letters indicate statistically significant differences, $p < 0.05$ (Tukey's HSD test).

Adaptation to the High Nitrogen Environment

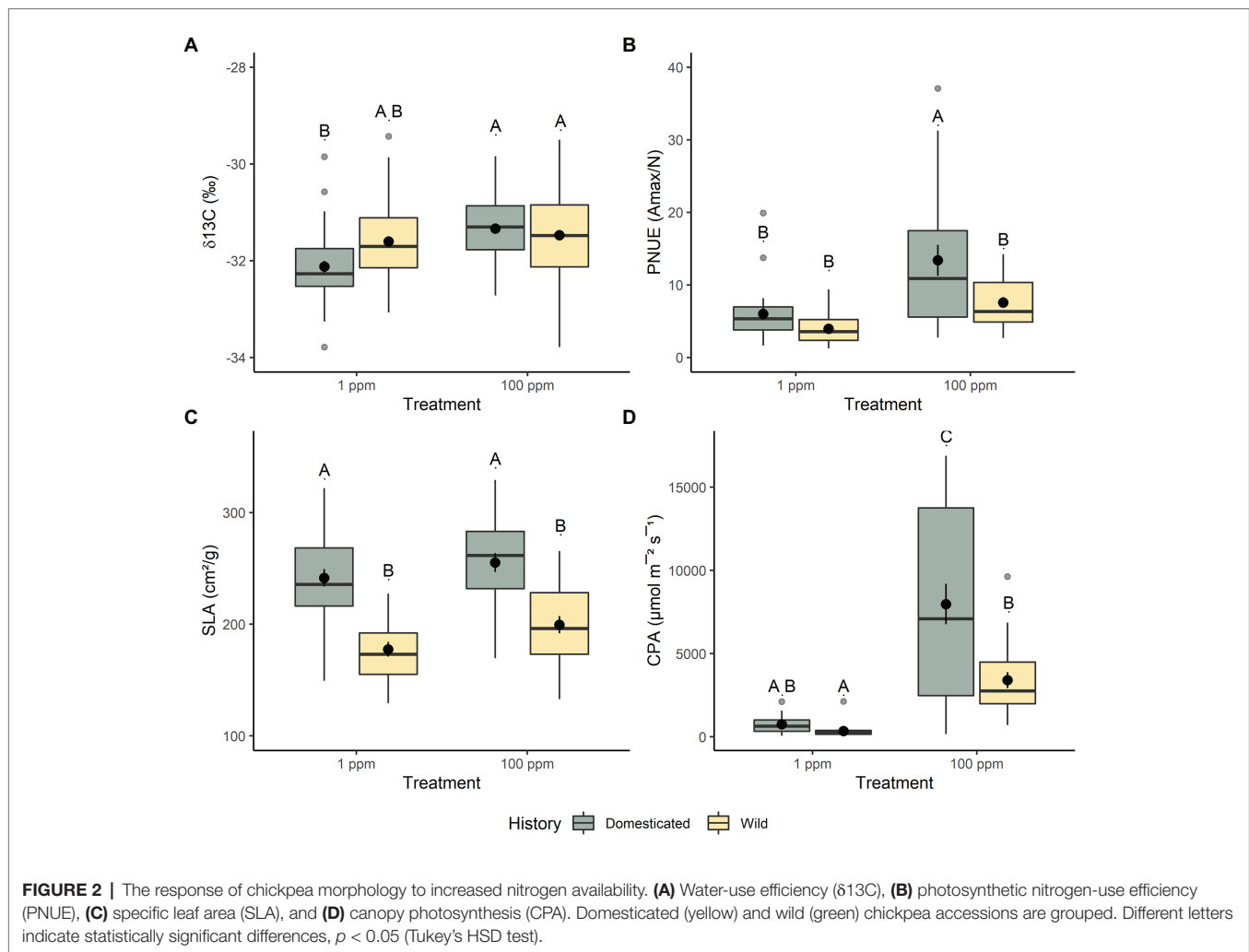
Domesticated and wild chickpea displayed differential adaptation to the higher nitrogen environment (100 ppm), with several belowground and aboveground traits being significantly different between these groups. Two belowground morphological traits between domesticated and wild chickpea were significantly different from each other, indicating differences in belowground adaptation to high nitrogen conditions. Domesticated chickpea exhibited higher root volume ($t_{51} = 3.062$, $p = 0.021$) and lower SRL ($t_{77} = -2.924$, $p = 0.0271$; **Figure 1C**) than wild chickpea. Similar to the low nitrogen environment, domesticated and wild chickpea exhibited comparable belowground biomass ($t_{45} = 2.215$, $p = 0.192$; **Figure 1B**), root length ($t_{47} = 1.267$, $p = 1.000$), root density ($t_{55} = 2.566$, $p = 0.0782$; **Figure 1D**), average root diameter ($t_{50} = 2.517$, $p = 0.090$), and root surface area ($t_{43} = 2.425$, $p = 0.118$).

Additionally, several aboveground functional traits were significantly different between domesticated and wild chickpea within the higher nitrogen level (100 ppm), with domestic chickpea exhibiting greater aboveground biomass ($t_{53} = 4.361$, $p = <0.001$;

Figure 1A), leaf level photosynthetic rate (maximum photosynthetic rate per area $t_{26} = 3.127$, $p = 0.026$; maximum photosynthetic rate per mass $t_{29} = 4.539$, $p = <0.001$), canopy photosynthetic rates ($t_{26} = 3.605$, $p = 0.008$; **Figure 2D**), photosynthetic nitrogen use efficiency per mass ($t_{35} = 2.961$, $p = 0.033$; **Figure 2B**), and specific leaf area ($t_{40} = 4.949$, $p = <0.001$; **Figure 2C**). Domesticated and wild chickpea did not differ in %N ($t_{59} = 1.343$, $p = 1.000$), %C ($t_{76} = 1.083$, $p = 1.000$), chlorophyll content index ($t_{49} = 0.787$, $p = 1.000$), nitrogen investment into chlorophyll ($t_{46} = 0.352$, $p = 1.000$), photosynthetic nitrogen use efficiency per area ($t_{32} = 1.976$, $p = 3.423$), water-use efficiency ($t_{40} = 1.187$, $p = 1.000$; **Figure 2A**), leaf C/N ratio ($t_{59} = 0.114$, $p = 1.000$), and stomatal conductance ($t_{107} = -1.070$, $p = 1.000$) within the higher soil nitrogen treatment (100 ppm).

Phenotypic Plasticity: Phenotypic Response to Nitrogen Availability

Significant interactions between nitrogen level and domestication history revealed differences between wild and domesticated chickpea phenotypic response to nitrogen availability for several



belowground and aboveground traits. For belowground traits, root density ($f_{1,279} = 14.849$, $p < 0.001$; **Figure 1D**) and SRL ($f_{1,296} = 8.719$, $p = 0.003$; **Figure 1C**) were significantly different between the two groups. As nitrogen levels increased, domesticated chickpea ($t_{279} = -3.553$, $p = 0.003$) significantly increased root density, while root density for wild chickpea ($t_{279} = 1.545$, $p = 0.741$) remained relatively consistent, illustrating a canalized response (**Figure 1D**). Additionally, domesticated chickpea exhibited significantly reduced SRL as nitrogen level increased ($t_{284} = 3.466$, $p = 0.004$), while SRL for wild chickpea remained broadly consistent between nitrogen environments ($t_{283} = 0.020$, $p = 1.000$) (**Figure 1C**). These results were corroborated by our analyses of RDPI, an index that quantifies phenotypic plasticity (Valladares et al., 2006). RDPI revealed that domesticated chickpea had significantly higher plasticity for SRL ($t_{2122} = 4.273$, $p < 0.001$) and root density ($t_{2068} = 19.059$, $p < 0.001$) than wild chickpea.

Moreover, significant interactions between nitrogen level and domestication history were also present for several aboveground traits, including aboveground biomass ($f_{275} = 8.600$, $p = 0.004$; **Figure 1A**), water use efficiency ($f_{273} = 16.901$, $p < 0.001$; **Figure 2A**), C/N ratio ($t_{276} = 11.471$, $p < 0.001$), canopy level

photosynthesis ($f_{76} = 11.179$, $p = 0.001$; **Figure 2D**), leaf level photosynthesis (per area, $f_{76} = 6.462$, $p = 0.013$; per mass, $f_{76} = 11.321$, $p = 0.001$), PNUE (per area $t_{29} = 8.599$, $p < 0.001$; per mass $f_{33} = 8.941$, $p < 0.001$; **Figure 2B**), and stomatal conductance ($f_{76} = 4.137$, $p = 0.045$). As nitrogen levels increased, both domesticated and wild chickpea increased aboveground biomass, whole canopy photosynthesis, leaf level photosynthesis, PNUE, and stomatal conductance. However, despite both groups increasing aboveground biomass and PNUE in response to higher nitrogen presence, domesticated chickpea exhibited higher plasticity for aboveground biomass ($t_{2232} = 11.411$, $p < 0.001$) and PNUE (per area $t_{37} = 3.99$, $p < 0.001$; per mass $t_{941} = 3.806$, $p < 0.001$) with significantly higher RDPI. On the other hand, wild chickpea had significantly higher RDPI for gas exchange traits including stomatal conductance ($t_{905} = -4.144$, $p < 0.001$), leaf level photosynthesis (photosynthetic rate per area $t_{946} = -1.608$, $p = 0.109$; photosynthetic rate per mass $t_{944} = -2.883$, $p = 0.004$), and canopy photosynthesis ($t_{938} = -4.169$, $p < 0.001$).

As nitrogen levels increased, both domesticated ($t_{279} = 8.021$, $p < 0.001$) and wild chickpea ($t_{278} = 19.302$, $p < 0.001$) significantly decreased their C/N ratio. However, despite both groups decreasing C/N ratio, wild chickpea displayed a higher

phenotypic response to nitrogen presence by exhibiting a greater RDPI ($t_{2105} = -9.943$, $p = <0.001$) than domesticated chickpea. Additionally, as nitrogen levels increased, wild chickpea exhibited a canalized response with no significant change in water-use efficiency ($t_{276} = -1.272$, $p = 1.000$), whereas domesticated chickpea increased water-use efficiency ($t_{276} = -5.612$, $p = <0.001$). RDPI analyses also concluded that domesticated chickpea had a significantly higher plasticity for water-use efficiency ($t_{1959} = 9.794$, $p = <0.001$), but this was driven by a decreased efficiency in the low soil nitrogen environment. Surprisingly, both wild ($t_{83} = -1.282$, $p = 0.576$) and domesticated ($t_{83} = -1.937$, $p = 0.221$) chickpea displayed a canalized response in regard to SLA.

Substantial Variation by Ecotype

In addition to domestication history having a significant effect on chickpea's capacity for plasticity in response to nitrogen availability, our results revealed that ecotypes also greatly varied in their response to nitrogen availability. This was determined by an overall significant random term in the linear mixed model (**Figure 3**). We found significant variation among ecotypes for the following belowground traits: belowground biomass [$\chi^2(1) = 17.092$, $p = <0.001$], root density [$\chi^2(1) = 4.396$, $p = 0.036$; **Figure 3B**], root length [$\chi^2(1) = 9.987$, $p = 0.002$], average root diameter density [$\chi^2(1) = 8.619$, $p = 0.003$], root volume [$\chi^2(1) = 7.598$, $p = 0.005$], root surface area [$\chi^2(1) = 23.037$, $p = <0.001$], and SRL [$\chi^2(1) = 5.298$, $p = 0.022$; **Figure 3A**]. Additionally, a significant response variation for ecotype was found for several aboveground traits including, aboveground biomass [$\chi^2(1) = 5.506$, $p = 0.019$], whole canopy photosynthesis [$\chi^2(1) = 3.957$, $p = 0.047$; **Figure 3D**], leaf level photosynthetic rate [per area $\chi^2(1) = 4.203$, $p = 0.040$], water-use efficiency [$\chi^2(1) = 26.821$, $p = <0.001$; **Figure 3C**], chlorophyll content

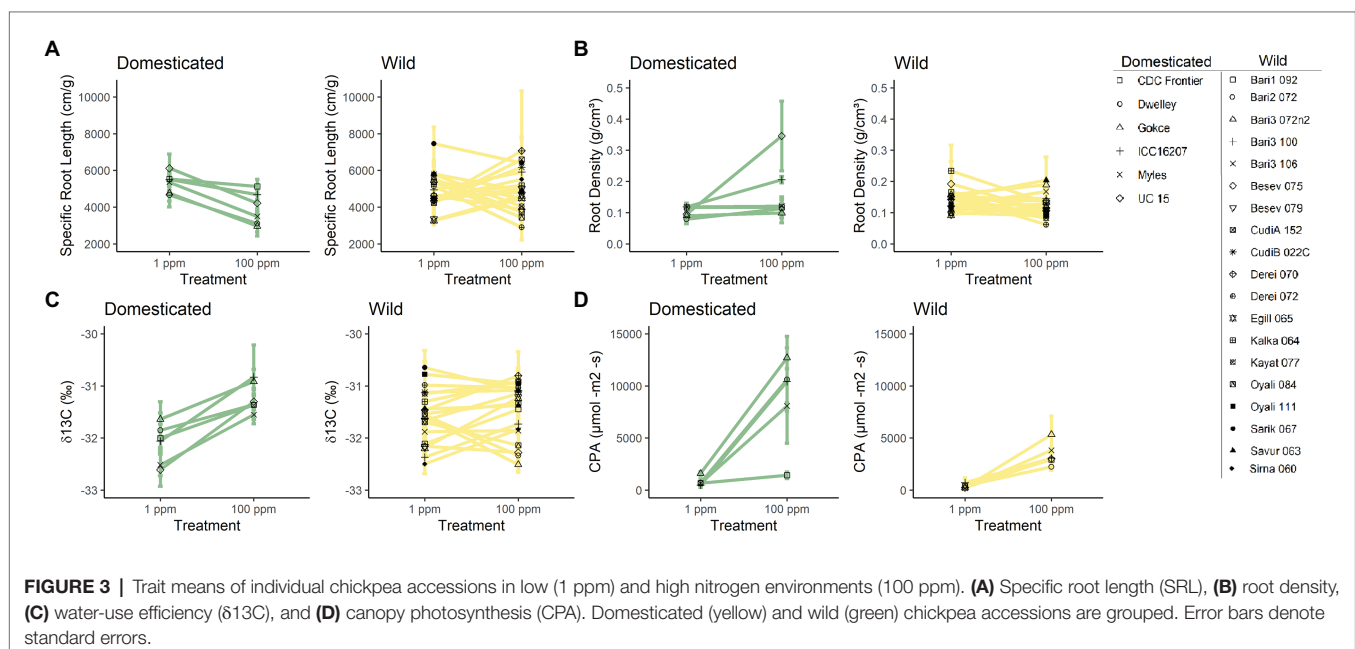
index [$\chi^2(1) = 9.428$, $p = 0.002$], and nitrogen investment into chlorophyll [$\chi^2(1) = 12.109$, $p = <0.001$].

Phenotypic Plasticity: Relationships to Indicators of Plant Fitness

Plant size (i.e., aboveground or belowground biomass) and measures of plant plasticity were negatively correlated across treatments and history (**Supplementary Figure S1**). There was a significant overall negative correlation between plasticity in SRL and aboveground plant biomass ($t_{294} = -4.838$, $p = <0.001$, $r = -0.272$; **Supplementary Figure S1b**) and between water-use efficiency and aboveground plant biomass ($t_{294} = -2.432$, $p = 0.016$, $r = -0.141$; **Supplementary Figure S1a**). Root density and aboveground biomass were not significantly correlated overall ($t_{294} = -1.298$, $p = 0.196$, $r = -0.075$). The negative correlation for SRL and aboveground plant biomass held true for both wild ($t_{212} = -2.321$, $p = 0.021$, $r = -0.157$) and domesticated chickpea ($t_{80} = -5.563$, $p = <0.001$, $r = -0.528$) across treatments. However, for water-use efficiency and aboveground biomass, domesticated and wild chickpea differed in their correlations; with wild chickpea ($t_{212} = -0.513$, $p = 0.609$, $r = -0.035$; **Supplementary Figure S1c**) having a non-significant correlation while domesticated chickpea having a strong negative correlation ($t_{80} = -4.168$, $p = <0.001$, $r = -0.422$; **Supplementary Figure S1d**) across chickpea.

DISCUSSION

The overall objectives in our study were to determine (1) if wild chickpea performed better under lower nitrogen conditions and (2) if domesticated chickpea had reduced phenotypic plasticity. Our results revealed that wild and domesticated



chickpea had similar phenotypes in the low nitrogen environment for both belowground and aboveground traits, indicating that the domestication has not affected the chickpea's response to low nitrogen conditions in the absence of rhizobia. However, wild chickpea did display phenotypes consistent with water conservation, by exhibiting lower specific leaf area and stomatal conductance, adaptations consistent with one of our core hypotheses. At higher nitrogen concentrations, wild and domesticated chickpea differed in respect to many below and aboveground phenotypes, with domesticated chickpea consistently displaying phenotypes adapted to high resource conditions. This is not surprising as thousands of years of cultivation and breeding has most likely adapted domesticated chickpea, like most crops, to high resource conditions (e.g., Bogaard et al., 2013; Brown, 2018). Furthermore, significant two-way interactions between nitrogen concentration and history (wild vs. cultivated) for SRL, root density, aboveground biomass, water-use-efficiency, C/N ratio, canopy level photosynthesis, leaf level photosynthesis, and stomatal conductance demonstrated that wild and domesticated chickpea exhibited differences in their responses to nitrogen levels. Our RDPI results, surprisingly, suggest that the wild chickpea had a canalized response for SRL, root density, and greater water-use efficiency and limited phenotypic plasticity for aboveground biomass but greater phenotypic plasticity for stomatal conductance, canopy photosynthesis, leaf level photosynthesis, and C/N ratio. Both cultivated and wild chickpea exhibited a canalized response for SLA, but SLA was consistently higher for domesticated chickpea at both nitrogen treatments, indicating this trait may be adaptive for nutrient rich environments.

The lower plasticity of some traits in wild chickpea is primarily explained by the substantial ecotypic variation within the wild germplasm, which lowered the average phenotypic response to increased nitrogen. The substantial ecotypic variation in wild chickpea is not surprising as ecotypes originate from different environmental conditions (von Wettberg et al., 2018). However, the lack of plasticity for some ecotypes, yet similar performance and resource use efficiency in low nutrient conditions is surprising, as these are potential mechanisms to increase plant survival in natural environments (reviewed in Ghalambor et al., 2007; Hauvermale and Sanad, 2019). Specifically, root plasticity is beneficial for wild plants due to heterogeneous nutrient distribution and limiting nutrients found in natural habitats when compared to agroecosystems (Paz-González et al., 2000; Bennett et al., 2005). For instance, the distribution of inorganic nitrogen was found to be homogenous in agricultural top-soil, (Jackson and Bloom, 1990), while nutrient distribution varied significantly in natural sagebrush steppe-habitat (Jackson and Caldwell, 1993) and tropical forests (John et al., 2007). Furthermore, domesticated crops such as barley (Grossman and Rice, 2012), cassava (Ménard et al., 2013), and soybeans (Kiers et al., 2007) have undergone a reduction in phenotypic plasticity, which is believed to be due to a reduction in genetic diversity driven by agronomic selection (Sadras, 2007) or continuous selection in a more homogenous agricultural environment.

As expected, average aboveground and belowground biomass increased with higher nitrogen levels for both domesticated and wild chickpea. However, for several wild ecotypes, aboveground and belowground biomass decreased or remained relatively the same in higher nitrogen conditions, indicating limited phenotypic plasticity for these traits to nitrogen availability. These results are surprising, as they contrast against previous results comparing plasticity in aboveground biomass to nutrient availability in domesticated and wild: chard (*Beta vulgaris* L.), cabbage (*Brassica oleracea* DC.), sunflower (*Helianthus annuus* L.), tomato (*Solanum lycopersicum* L.), durum wheat (*Triticum durum* Desf.), maize (*Zea mays* L.), and pea (*Pisum sativum* L.; Matesanz and Milla, 2018). Differences between our results and previous findings could stem from the number of accessions used in each study. The limited number of accessions used in previous studies likely was not sufficient to adequately capture the phenotypic variation or plasticity present in each crop or wild relative (Krieg et al., 2017).

As nitrogen availability increased, domesticated chickpea ecotypes reacted uniformly with decreased SRL and increased root density. These results indicate that the domesticated chickpea has increased root diameter and decreased root length, a physiological response to higher nitrogen presence that has been shown in other studies (Callaway et al., 2003; von Wettberg and Weiner, 2003). Low SRL and high root density are optimum root phenotypes for plants in nutrient-rich soil environments, as these phenotypes are believed to be most efficient when nutrients are abundant (Reich, 2014; Kong et al., 2019). Conversely, on average, SRL and root density remained relatively unchanged for wild chickpea in both nitrogen treatments and were not significantly different. However, when taking into account how individual wild ecotypes reacted to increased nitrogen availability, we observed variation among ecotypes in phenotypically plastic responses to nitrogen. With respect to SRL and root density, wild ecotypes decreased, increased, or remained constant in response to nitrogen availability. These results were surprising as wild bean ecotypes on average have greater SRL and root density than domesticated ecotypes (Pérez-Jaramillo et al., 2017). Greater SRL has been hypothesized to provide higher efficiency of nutrient search and uptake, a beneficial phenotype for nutrient "foraging" in nutrient heterogeneous environments.

Leaf measurements such as leaf level photosynthesis and stomatal conductance, and canopy level photosynthesis were consistently greater for domesticated chickpea at both nitrogen levels; however, domestication history was not statistically significant while nitrogen level was (**Supplementary Table S1**). This is likely due to the cultivated varieties having been selected under higher fertility agricultural conditions when compared to those experienced by wild ecotypes. Domesticated chickpea showed a similar general increase in water-use-efficiency (less negative $\delta^{13}\text{C}$), %N in leaves, chlorophyll content, and PNUE to increased nitrogen availability, an expected response to nutrient-rich environments (Matesanz and Milla, 2018). When focusing on ecotypic variation, wild chickpea ecotypes did not respond consistently to increased nitrogen level in regard to water-use-efficiency ($\delta^{13}\text{C}$). However, at lower nitrogen

conditions, wild ecotypes displayed similar nitrogen and water use phenotypes, as might be expected from adaptation to low-nitrogen conditions of the native range of wild chickpea in Southeastern Turkey. One of the few measurements that was primarily influenced by domestication history was SLA, which was not affected by increased nitrogen presence. Domesticated chickpea ecotypes had consistently higher SLA than wild chickpea (**Figure 2; Supplementary Table S1**), perhaps as an indirect consequence of selection in domesticated chickpea for early growth and plant maturity, as evidenced by early phenology of domesticated chickpea (Ortega et al., 2019). Additionally, domesticated chickpea had a slight increase in leaf %C relative to wild ecotypes, but this was not significant.

An important factor impacting our results is that we performed our study in the absence of symbiotic rhizobia. During root morphology measurement, nodules were rarely found. Sterile soil was an experimental necessity, as wild and cultivated chickpea differ in their preferred rhizobia and have substantial interactions with soil substrate (Cook et al., unpublished; Greenlon et al., 2019), adding multiple rhizobial strains would make the experiment too complicated to dissect a signal of response to nitrogen fertility. An experiment without rhizobia is a realistic scenario for cases when a crop is grown in new soil or in a soil that has not had chickpea for over several years prior. We suspect that wild relatives with short dispersal distances may have a greater chance to encounter nearby co-adapted symbionts than their cultivated relatives (Greenlon et al., 2019). When moved beyond their native range or grown in soils lacking a compatible symbiont, wild chickpea may consequently perform more poorly under nutrient limiting conditions. However, when wild chickpea does occur in agricultural conditions, they may on average experience fertility much higher than in uncultivated habitats. It is also possible that selection may not have been sufficient to canalize responses to low nutrient availability, particularly if there is a very limited cost to plasticity for root responses to low nutrient availability. The only existing data of which we are aware of is that of Grossman and Rice (2012), who showed a loss of root plasticity in cultivated barley ecotypes. An earlier study by Kiers et al. (2007) showed that bred soybean varieties had a reduced capacity to enforce sanctions on low-performing rhizobia, although they did not examine other root traits. Conversely, in other crops, the impacts of domestication on crop functional traits remain difficult to predict, especially for belowground traits that have not been systematically studied.

CONCLUSION

The potentially widespread loss of phenotypic plasticity of crops to low fertility environments as a consequence of domestication could be a concern for farmers working on degraded or marginal soils without access to expensive inputs, as well as many organic production systems. Here, we find evidence that wild and domesticated chickpea display similarly efficient responses to low nitrogen conditions, but that domestication may have led to greater capacities for plasticity in morphological

and biomass related traits but may have lowered the capacity to modify physiological traits related to gas exchange and efficient water use. Moreover, when focusing on the ecotype or variety level, we found significantly more response variation for agronomic important traits, including specific root length and water-use efficiency in wild chickpea than domesticated chickpea, indicating that the wild chickpea is a repository for novel responses to nitrogen conditions. Under Green Revolution agroecological conditions, it is not uncommon for there to be such high levels of added nitrogen in the soil that it results in reduced levels of nodulation in legumes (e.g., Kiers et al., 2007). However, if such excess nitrogen is not present, the loss of phenotypic plasticity is a concern for the performance of crops in more challenging conditions. For a crop like chickpea, which is still largely produced by small-holder farmers as a low or minimal fertilizer input crop in South Asia and East Africa and that serves a critical food security role in many diets, lost phenotypic plasticity may reduce resilience against climate change. The genetic bottlenecks that arise from domestication, post-domestication divergence, and the intensive breeding for agronomic traits may have additional, inadvertent effects on unselected belowground traits (e.g., Morrell et al., 2013; Gaut et al., 2018). These inadvertent effects are one of several reasons why large collections of wild relatives with a greater range of adaptive traits or plasticity than in the cultigen are needed in breeding programs to increase the resilience of our crops within a changing global climate (Warschefsky et al., 2014; Coyne et al., 2020).

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

EW, CK, and VP conceived and designed the experiment. ED-C, EB, EM, and CK performed the experiment. EM and CK analyzed the data. EM, CK, and EW wrote the manuscript. ES provided editorial advice. ES provided editorial advice and photosynthetic analysis resources. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by a cooperative agreement from the United States Agency for International Development under the Feed the Future Program AID-OAA-A-14-00008 to D. R. Cook and Co-PIs VP and EW; a grant from the US National Science Foundation Plant Genome Program under Award IOS-1339346 to D. R. Cook, and CoPIs VP and EW; US NIFA grant 2018-67013-27619 to VP and EW; Northeast SARE graduate research grant GNE-179 to Edward Marques; National Science Foundation Graduate Research Fellowship

Program grant DGE-1315138 to CK. EW is further supported by the USDA Hatch program through the Vermont State Agricultural Experimental Station.

ACKNOWLEDGMENTS

We thank Camilo Soares, Joseph Rahm, Annie Manzor, Rebecca Valls, Christina Currais, Katja Diaz-Granados, Carlos Cano, Manny Briceno, Allan Ponce, Lilian Rodicio, Daniel Gitlin, Ellen Garcia, Giovanna Carlson, Beck Morrow, Shelby Rosten, Seth Epling, Yadira Reynaldo, and Poornima Narayana for help

with the greenhouse work and root morphology measurements, and Emily Warschewsky and R Ford Dennison for helpful discussion and Abdullah Kahraman for providing seeds. We would also like to thank Aitor Ameztegui for his help with RDPI calculations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.576338/full#supplementary-material>

REFERENCES

- Ahmed, A. M., Tana, T., Singh, P., and Molla, A. (2016). Modeling climate change impact on chickpea production and adaptation options in the semi-arid North-Eastern Ethiopia. *JAIED* 110, 377–395. doi: 10.12895/jaeid.20162.510
- Ameztegui, A. (2017). Plasticity: an R package to determine several plasticity indices. GitHub repository.
- Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bennett, E. M., Carpenter, S. R., and Clayton, M. K. (2005). Soil phosphorus variability: scale-dependence in an urbanizing agricultural landscape. *Landsc. Ecol.* 20, 389–400. doi: 10.1007/s10980-004-3158-7
- Bogaard, A., Fraser, R., Heaton, T. H., Wallace, M., Vaiglova, P., Charles, M., et al. (2013). Crop manuring and intensive land management by Europe's first farmers. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12589–12594. doi: 10.1073/pnas.1305918110
- Brown, T. A. (2018). The role of humans in a protracted transition from hunting-gathering to plant domestication in the Fertile Crescent. *Front. Plant Sci.* 9:1287. doi: 10.3389/fpls.2018.01287
- Bulgarelli, D., Garrido-Oter, R., Münch, P. C., Weiman, A., Dröge, J., Pan, Y., et al. (2015). Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host Microbe* 17, 392–403. doi: 10.1016/j.chom.2015.01.011
- Callaway, R. M., Pennings, S. C., and Richards, C. L. (2003). Phenotypic plasticity and interactions among plants. *Ecology* 84, 1115–1128. doi: 10.1890/0012-9658(2003)084[1115:PPAIAIP]2.0.CO;2
- Coyne, C. J., Kumar, S., von Wettberg, E. J. B., Marques, E., Berger, J. D., Redden, R. J., et al. (2020). Potential and limits of exploitation of crop wild relatives for pea, lentil, and chickpea improvement. *Legume Science* 2:e36. doi: 10.1002/leg3.36
- FAO, IFAD, UNICEF, WFP, and WHO (2017). The state of food security and nutrition in the world. Building resilience for peace and food security.
- Flatt, T. (2005). The Evolutionary Genetics of Canalization. *Q. Rev. Biol.* 80, 287–316. doi: 10.1086/432265
- Gaut, B. S., Seymour, D. K., Liu, Q., and Zhou, Y. (2018). Demography and its effects on genomic variation in crop domestication. *Nat. Plants* 4, 512–520. doi: 10.1038/s41477-018-0210-1
- Ghalambor, C. K., McKAY, J. K., Carroll, S. P., and Reznick, D. N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Funct. Ecol.* 21, 394–407. doi: 10.1111/j.1365-2435.2007.01283.x
- Greenlon, A., Chang, P. L., Damte, Z. M., Muleta, A., Carrasquilla-Garcia, N., Kim, D., et al. (2019). Global-level population genomics reveals differential effects of geography and phylogeny on horizontal gene transfer in soil bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 116, 15200–15209. doi: 10.1073/pnas.1900056116
- Gross, B. L., and Olsen, K. M. (2010). Genetic perspectives on crop domestication. *Trends Plant Sci.* 15, 529–537. doi: 10.1016/j.tplants.2010.05.008
- Grossman, J. D., and Rice, K. J. (2012). Evolution of root plasticity responses to variation in soil nutrient distribution and concentration. *Evol. Appl.* 5, 850–857. doi: 10.1111/j.1752-4571.2012.00263.x
- Harrel, F. (2015). Package “Hmisc.” Available at: <https://cran.r-project.org/web/packages/Hmisc> (Accessed November 20, 2020).
- Hauvermale, A. L., and Sanad, M. N. (2019). Phenological Plasticity of Wild and Cultivated Plants [Online First], IntechOpen. Available at: <https://www.intechopen.com/online-first/phenological-plasticity-of-wild-and-cultivated-plants> (Accessed November 20, 2020).
- Hereford, J., Hansen, T. F., and Houle, D. (2004). Comparing strengths of directional selection: how strong is strong? *Evolution* 58, 2133–2143. doi: 10.1111/j.0014-3820.2004.tb01592.x
- Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hoang, A., Hill, C. E., et al. (2001). Strength and tempo of directional selection in the wild. *Proc. Natl. Acad. Sci. U. S. A.* 98, 9157–9160. doi: 10.1073/pnas.161281098
- Hufford, M. B., Teran, J. C. B. M. Y., and Gepts, P. (2019). Crop biodiversity: an unfinished magnum opus of nature. *Annu. Rev. Plant Biol.* 70, 727–751. doi: 10.1146/annurev-arplant-042817-040240
- Jackson, L. E., and Bloom, A. J. (1990). Root distribution in relation to soil nitrogen availability in field-grown tomatoes. *Plant Soil* 128, 115–126. doi: 10.1007/BF00011100
- Jackson, R. B., and Caldwell, M. M. (1993). Geostatistical patterns of soil heterogeneity around individual perennial plants. *J. Ecol.* 81, 683–692. doi: 10.2307/2261666
- John, R., Dalling, J. W., Harms, K. E., Yavitt, J. B., Stallard, R. F., Mirabello, M., et al. (2007). Soil nutrients influence spatial distributions of tropical tree species. *Proc. Natl. Acad. Sci. U. S. A.* 104, 864–869. doi: 10.1073/pnas.0604666104
- Keen, C. (2020). Growing chickpea in North Dakota. North Dakota State Extension Bulletin. Available at: <https://www.ag.ndsu.edu/publications/crops/soilfertilityrecommendations-for-field-pea-lentil-and-chickpea-in-northdakota#section-3> (Accessed November 20, 2020).
- Kiers, E. T., Hutton, M. G., and Denison, R. F. (2007). Human selection and the relaxation of legume defences against ineffective rhizobia. *Proc. Biol. Sci.* 274, 3119–3126. doi: 10.1098/rspb.2007.1187
- Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., et al. (2001). The strength of phenotypic selection in natural populations. *Am. Nat.* 157, 245–261. doi: 10.1086/319193
- Kingsolver, J. G., and Pfennig, D. W. (2007). Patterns and power of phenotypic selection in nature. *Bioscience* 57, 561–572. doi: 10.1641/B570706
- Kong, D., Wang, J., Wu, H., Valverde-Barrantes, O. J., Wang, R., Zeng, H., et al. (2019). Nonlinearity of root trait relationships and the root economics spectrum. *Nat. Commun.* 10:2203. doi: 10.1038/s41467-019-10245-6
- Krieg, C. P., Kassa, M. T., and von Wettberg, E. J. B. (2017). “Germplasm Characterization and Trait Discovery” in *The Pigeonpea Genome. Compendium of Plant Genomes*. eds. R. Varshney, R. Saxena and S. Jackson (Berlin: Springer), 65–79.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 22–24. doi: 10.18637/jss.v082.i13
- Lande, R., and Arnold, S. J. (1983). The Measurement of Selection on Correlated Characters. *Evolution* 37, 1210–1226. doi: 10.1111/j.1558-5646.1983.tb00236.x
- Lye, Z. N., and Purugganan, M. D. (2019). Copy number variation in domestication. *Trends Plant Sci.* 24, 352–365. doi: 10.1016/j.tplants.2019.01.003
- Martin, A. R., and Isaac, M. E. (2015). REVIEW: Plant functional traits in agroecosystems: a blueprint for research. *J. Appl. Ecol.* 52, 1425–1435. doi: 10.1111/1365-2664.12526

- Matesanz, S., and Milla, R. (2018). Differential plasticity to water and nutrients between crops and their wild progenitors. *Environ. Exp. Bot.* 145, 54–63. doi: 10.1016/j.envexpbot.2017.10.014
- McKey, D. B., Elias, M., and Pujol, B. (2012). “Ecological Approaches to Crop Domestication, 17” in *Biodiversity in agriculture: domestication, evolution, and sustainability*. eds. J. R. Harlan, P. Gepts, T. R. Famula, R. L. Bettinger, S. B. Brush and A. B. Damania, et al. (Cambridge University Press), 377–406.
- Ménard, L., McKey, D., Mühlen, G. S., Clair, B., and Rowe, N. P. (2013). The evolutionary fate of phenotypic plasticity and functional traits under domestication in manioc: changes in stem biomechanics and the appearance of stem brittleness. *PLoS One* 8:e74727. doi: 10.1371/journal.pone.0074727
- Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x
- Milla, R., Osborne, C. P., Turcotte, M. M., and Violle, C. (2015). Plant domestication through an ecological lens. *Trends Ecol. Evol.* 30, 463–469. doi: 10.1016/j.tree.2015.06.006
- Morrell, P. L., Buckler, E. S., and Ross-Ibarra, J. (2011). Crop genomics: advances and applications. *Nat. Rev. Genet.* 13, 85–96. doi: 10.1038/nrg3097
- Morrell, P. L., Gonzales, A. M., Meyer, K. K. T., and Clegg, M. T. (2013). Resequencing data indicate a modest effect of domestication on diversity in barley: a cultigen with multiple origins. *J. Hered.* 105, 253–264. doi: 10.1093/jhered/est083
- Ogutcen, E., Pandey, A., Khan, M. K., Marques, E., Penmetsa, R. V., Kahraman, A., et al. (2018). Pod shattering: A homologous series of variation underlying domestication and an avenue for crop improvement. *Agronomy* 8:137. doi: 10.20944/preprints201806.0162.v1
- Olsen, K. M., and Wendel, J. F. (2013). A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu. Rev. Plant Biol.* 64, 47–70. doi: 10.1146/annurev-arplant-050312-120048
- Ortega, R., Hecht, V. F. G., Freeman, J. S., Carrasquilla-Garcia, N., and Rouf, R. M., Penmetsa, R. V., et al. (2019). Altered expression of an FT cluster underlies a major locus controlling domestication-related changes to chickpea phenology and growth habit. *Front. Plant Sci.* 10:824. doi: 10.3389/fpls.2019.00824
- Paz-González, A., Vieira, S. R., and Taboada Castro, M. (2000). The effect of cultivation on the spatial variability of selected properties of an umbric horizon. *Geoderma* 97, 273–292. doi: 10.1016/S0016-7061(00)00066-5
- Penmetsa, R. V., Carrasquilla-Garcia, N., Bergmann, E. M., Vance, L., Castro, B., Kassa, M. T., et al. (2016). Multiple post-domestication origins of kabuli chickpea through allelic variation in a diversification-associated transcription factor. *New Phytol.* 211, 1440–1451. doi: 10.1111/nph.14010
- Pérez-Jaramillo, J. E., Carrión, V. J., Bosse, M., Ferrão, L. F. V., de Hollander, M., Garcia, A. A. F., et al. (2017). Linking rhizosphere microbiome composition of wild and domesticated *Phaseolus vulgaris* to genotypic and root phenotypic traits. *ISME J.* 11, 2244–2257. doi: 10.1038/ismej.2017.85
- Pérez-Jaramillo, J. E., Mendes, R., and Raaijmakers, J. M. (2016). Impact of plant domestication on rhizosphere microbiome assembly and functions. *Plant Mol. Biol.* 90, 635–644. doi: 10.1007/s11103-015-0337-7
- Price, T., and Langen, T. (1992). Evolution of correlated characters. *Trends Ecol. Evol.* 7, 307–310. doi: 10.1016/0169-5347(92)90229-5
- R Core Team (2019). R: a language and environment for statistical computing (version 3.1.2). Vienna, Austria: R Foundation for Statistical Computing; 2014.
- Rauw, W. M., Kanis, E., Noordhuizen-Stassen, E. N., and Grommers, F. J. (1998). Undesirable side effects of selection for high production efficiency in farm animals: a review. *Livest. Prod. Sci.* 56, 15–33. doi: 10.1016/S0301-6226(98)00147-X
- Reich, P. B. (2014). The world-wide “fast-slow” plant economics spectrum: a traits manifesto. *J. Ecol.* 102, 275–301. doi: 10.1111/1365-2745.12211
- Revelle, W. (2019). psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 1.9.12. Available at: <https://CRAN.R-project.org/package=psych> (Accessed November 20, 2020).
- Sadras, V. O. (2007). Evolutionary aspects of the trade-off between seed size and number in crops. *Field Crop Res.* 100, 125–138. doi: 10.1016/j.fcr.2006.07.004
- Singh, P., Nedumaran, S., Boote, K. J., Gaur, P. M., Srinivas, K., and Bantilan, M. C. S. (2014). Climate change impacts and potential benefits of drought and heat tolerance in chickpea in South Asia and East Africa. *Eur. J. Agron.* 52, 123–137. doi: 10.1016/j.eja.2013.09.018
- Smykal, P., Nelson, M., Berger, J., and von Wettberg, E. (2018). The impact of genetic changes during crop domestication. *Agronomy* 8:119. doi: 10.3390/agronomy8070119
- USClimateData.com (2020). Your Climate Data LLC, Santiago Chile. Available at: <https://www.usclimatedata.com/climate/miami/florida/united-states/usfl0316> (Accessed April 20, 2020)
- Valladares, F., Sanchez-gomez, D., and Zavala, M. A. (2006). Quantitative estimation of phenotypic plasticity: bridging the gap between the evolutionary concept and its ecological applications. *J. Ecol.* 94, 1103–1116. doi: 10.1111/j.1365-2745.2006.01176.x
- von Wettberg, E. J. B., Chang, P. L., Başdemir, F., Carrasquilla-Garcia, N., Korbu, L. B., Moenga, S. M., et al. (2018). Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nat. Commun.* 9:649. doi: 10.1038/s41467-018-02867-z
- von Wettberg, E. J., Remington, D. L., and Schmitt, J. (2008). Partitioning adaptation differentiation across a patchy landscape: shade avoidance traits in *Impatiens capensis*. *Evolution* 62, 654–667. doi: 10.1111/j.1558-5646.2007.00309.x
- von Wettberg, E. J., and Weiner, J. (2003). Larger *Triticum aestivum* plants do not preempt nutrient-rich patches in a glasshouse experiment. *Plant Ecol.* 169, 85–92. doi: 10.1023/A:1026253007056
- Warschefsky, E., Penmetsa, R. V., Cook, D. R., and von Wettberg, E. J. B. (2014). Back to the wilds: tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *Am. J. Bot.* 101, 1791–1800. doi: 10.3732/ajb.1400116

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

Copyright © 2020 Marques, Krieg, Dacosta-Calheiros, Bueno, Sessa, Penmetsa and von Wettberg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Wild and Domestic Differences in Plant Development and Responses to Water Deficit in *Cicer*

Jens Berger^{1*}, Raju Pushpavalli¹, Christiane Ludwig¹, Sylvia Parsons¹, Fatma Basdemir² and Kelly Whisson¹

¹Commonwealth Scientific and Industrial Research Organization (CSIRO) Agriculture and Food, Wembley, WA, Australia,

²Ceylanpınar Agriculture Vocational School, University of Harran, Sanlıurfa, Turkey

OPEN ACCESS

Edited by:

Gaofeng Zhou,
Department of Primary Industries and
Regional Development of Western
Australia (DPIRD), Australia

Reviewed by:

Carlos Lara-Romero,
Rey Juan Carlos University, Spain
Jaroslav Klápšte,
New Zealand Forest Research
Institute Limited (Scion), New Zealand

*Correspondence:

Jens Berger
jens.berger@csiro.au

Specialty section:

This article was submitted to
Evolutionary and Population
Genetics,
a section of the journal
Frontiers in Genetics

Received: 18 September 2020

Accepted: 10 November 2020

Published: 04 December 2020

Citation:

Berger J, Pushpavalli R, Ludwig C,
Parsons S, Basdemir F and
Whisson K (2020) Wild and Domestic
Differences in Plant Development and
Responses to Water Deficit in *Cicer*.
Front. Genet. 11:607819.
doi: 10.3389/fgene.2020.607819

There is growing interest in widening the genetic diversity of domestic crops using wild relatives to break linkage drag and/or introduce new adaptive traits, particularly in narrow crops such as chickpea. To this end, it is important to understand wild and domestic adaptive differences to develop greater insight into how wild traits can be exploited for crop improvement. Here, we study wild and domestic *Cicer* development and water-use over the lifecycle, measuring responses to reproductive water deficit, a key Mediterranean selection pressure, using mini-lysimeters (33 L round pots) in common gardens under contrasting water regimes. Wild and domestic *Cicer* were consistently separated by later phenology, greater water extraction and lower water use efficiency (WUE) and harvest index in the former, and much greater yield-responsiveness in the latter. Throughout the lifecycle, there was greater vegetative investment in wild, and greater reproductive investment in domestic *Cicer*, reflected in root and harvest indices, rates of leaf area, and pod growth. Domestic WUE was consistently greater than wild, suggesting differences in water-use regulation and partitioning. Large wild-domestic differences revealed in this study are indicative of evolution under contrasting selection pressures. *Cicer* domestication has selected for early phenology, greater early vigor, and reproductive efficiency, attributes well-suited to a time-delimited production system, where the crop is protected from grazing, disease, and competition, circumstances that do not pertain in the wild. Wild *Cicer* attributes are more competitive: higher peak rates of leaf area growth, greater *ad libitum* water-use, and extraction under terminal drought associated with greater vegetative dry matter allocation, leading to a lower reproductive capacity and efficiency than in domestic chickpea. These traits strengthen competitive capacity throughout the growing season and are likely to facilitate recovery from grazing, two significant selection pressures faced by wild, rather than domesticated *Cicer*. While increased water extraction may be useful for improving chickpea drought tolerance, this trait must be evaluated independently of the other associated wild traits. To this end, the wild-domestic populations have been developed.

Keywords: crop wild relative, partitioning, water-deficit, water-use, domestication

INTRODUCTION

Given genetic bottlenecks in most crop species, there is widespread interest in exploiting crop wild relatives (CWRs) that typically harbor much greater diversity (Tanksley and McCouch, 1997; Dempewolf et al., 2014; Gepts, 2014). Advocates of base broadening suggest that the greater genetic diversity of CWR will improve complex traits like yield by breaking linkage drag through the introgression of novel, previously unexploited alleles (Tanksley and McCouch, 1997), and/or be reflected in greater adaptive diversity (Dempewolf et al., 2017; Coyne et al., 2020). Indeed, ideas such as exploiting the climatic resilience of wild populations have gained considerable traction recently (McCouch et al., 2013; Dempewolf et al., 2014). Supporting this, there is an ever-increasing understanding of plant domestication in terms of history, geography, trait selection over time, and the genomic implications of these (Purugganan and Fuller, 2009; Meyer et al., 2012; Meyer and Purugganan, 2013; Abbo et al., 2014; Gepts, 2014; Larson et al., 2014). However, this work largely focuses on changes in the domesticate, rather than the attributes of the wild progenitor. As a result, when we want to turn to the wild relatives for crop improvement, we lack a framework to guide us. Genomic approaches maximizing genetic diversity to break up linkage drag advocated by Tanksley and McCouch (1997) are being deployed in breeding and pre-breeding programs in diverse ways from nested association mapping (von Wettberg et al., 2018) to MAGIC populations through to genomic selection (Bazakos et al., 2017). While these offer a methodology for increasing the genetic diversity of our breeding programs, they do not guide us in the identification of new adaptive traits and strategies that are missing in the cultigen, and which may or may not be present in the CWR (see discussion in Dempewolf et al., 2017). Thus, the big picture questions such as “where among wild populations do we find the adaptive traits with which to improve our crops and why?” remain important priorities (Thormann et al., 2014).

Clearly domestication has changed the adaptive capacity of crop compared to wild progenitor. In addition to the selection of domestication syndrome traits (lower seed dispersal and dormancy, larger seed size and more erect habit, and modified phenology), humans have altered many aspects of the cropping environment compared to that in which the progenitors evolved (Milla et al., 2015). Domesticated crops grow in managed (mostly) fields under reduced competition relative to the wild state (Gepts, 2014), their lifecycle is regulated (Milla et al., 2015), and they are not typically found as weedy escapes in the habitats, where their wild relatives occur. *Ipso facto* wild and domestic adaptive capacities do differ. But this does not mean that wild relatives will automatically contain all the adaptive traits lacking in the domesticate. Both wild and domestic plants are shaped by evolutionary responses to selection pressure which should be taken into account when evaluating their adaptive potential (Milla et al., 2015). There is a disconnect here between the agricultural and ecological approaches to this problem. While the former often samples widely (see examples in Singh et al., 1990, 1995, 1998), when screening for a shopping list

of required traits ignores questions such as how representative is our collection and what selection pressures were imposed by the environment of origin, no greater insights into the biology of the species are made. While ecological approaches do address these issues, their experimental material may lack the depth to provide certainty. Recently, an argument has been advanced that under domestication crops evolved into resource-acquisitive, fast growing plants (competitors *sensu* Grime, 2006) as a result of their cultivation in well-managed, resource rich environments relative to those of their wild progenitors (Milla et al., 2015). By extension, wild progenitors should be relatively slower growing plants with traits tending toward the stress tolerator spectrum *sensu* Grime (2006). This is an interesting idea which has been tested quite widely across species, but with little depth (typically one wild and domesticated accession per species), and only in the early vegetative phase (Milla et al., 2014; Matesanz and Milla, 2018), where the strong selection for early vigor in domesticated crops (Evans and Dunstone, 1970; Berger et al., 2017) might be expected to influence the results.

These concepts are particularly tractable in wild and domestic *Cicer*. The genetic narrow base of chickpea as an adaptive constraint has been long recognized (Abbo et al., 2003). Early work at the International Center for Agricultural Research in the Dry Areas (ICARDA) demonstrated that the wild *Cicer* had a wider, potentially useful range of responses to pests, diseases, and stresses than domestic chickpea, particularly for ascochyta blight, leaf miner, bruchids, cyst nematode, and vegetative cold (Singh et al., 1998). However, at that time the world collection of wild, *Cicer* was far too narrow to adequately characterize the adaptive potential of any species, including those that can readily cross with domestic chickpea, with only 18 independent accessions of *Cicer reticulatum*, the wild progenitor, and even less for its close relative, *Cicer echinospermum* (Berger et al., 2003). This has changed recently with extensive new collection across the habitat range of both these species in Southeastern Anatolia (von Wettberg et al., 2018) that is driving renewed interest in trait discovery in these CWRs (Reen et al., 2019; Newman et al., 2020).

In this paper, we investigate wild and domestic *Cicer* responses to reproductive water deficit (terminal drought), one of the principal selection pressures exerted by the Mediterranean climate, using mini-lysimeters in common gardens under contrasting reproductive water regimes. This approach allows us to compare wild and domestic responses to contrasting resource (water) availability as well as the underlying water-use patterns. This is followed up by a plant above- and below-ground development and water-use study to describe wild-domestic differences in greater depth across the lifecycle. We were interested to discover to what extent the wild and domestic *Cicer* would segregate along stress tolerator-competitor continuum of Grime (2006). Secondly, we were interested to explain a consistent field observation made during the collection that the annual wild *Cicer* species tend to have a longer lifecycle, reproducing and maturing considerably later than most of their sympatric annual plant competitors such as *Lens*, *Pisum*, and many *Vicia* species. To validate the evolution of “acquisitiveness” in domestication hypothesis, the following expectations should be met:

1. Lower growth rates, water-use, and above- and below-ground productivity in wild compared to domestic.
2. Improved stress tolerance in wild compared to domestic.
3. Lower response to resource (water) availability in wild compared to domestic.

MATERIALS AND METHODS

Rainout Shelter Water Deficit Studies-Overview (Exp 1–4)

A series of reproductive phase water deficit studies were run in the Commonwealth Scientific and Industrial Research Organization (CSIRO) Floreat rainout shelter from 2016 to 2019 comparing wild and domestic *Cicer* species collected from contrasting environments (Table 1). CSIRO Floreat is located in Perth, Western Australia (31.95°S, 115.79°E), a Mediterranean-type climate. The experiments were conducted in the standard Mediterranean late autumn to late spring growing season. The two larger 2016–2017 experiments (see Table 1 for accession details) were randomized complete block designs (RCBDs, $n = 3$) randomized within water regime [terminal drought (TD) and well-watered (WW)]. Because of space limitations, the two water regime treatments were allocated individually in two contiguous areas along the rainout shelter bay separated by 2 m (see Figure 1 for layout). This was done so that the rainout roof would shelter only the TD treatment

when rain was detected. Rainout shelter closure was an automatic sensor driven process whereby the roof covers the TD treatment during rainfall and then withdraws. The 2018–2019 experiments were smaller split plot designs ($n = 4$) with water regime as main plots, accessions as sub-plots, and all located in the same parcel of the rainout shelter bay (see Figure 1 for layout).

Experimental plots were 33 L round pots (430 × 340 mm) containing ca. 38 kg of Gingin loam. Five evenly spaced plants were planted in each pot (one in the middle and one in each quadrant).

Germplasm

The 2016 and 2017 trials evaluated a wide range of wild germplasm (*C. echinospermum* and *C. reticulatum*) collected from a range of sites against domestic chickpea check varieties (Table 1). The 2018 and 2019 experiments evaluated a subset ($n = 27$) of this material, choosing contrasting accessions based on the previous results.

Water Regimes

Reproductive water deficits were set-up by with-holding water in the terminal drought (TD) treatment when pod set was underway, defined by the first appearance of enlarged, but unfilled pods. To this end, phenological observations (dates of first flowering, podding, and pod enlargement) were recorded

TABLE 1 | Accession numbers evaluated for water deficit response in the Commonwealth Scientific and Industrial Research Organization (CSIRO) rainout shelter from 2016 to 2019, categorized by species and collection site (accession counts within species presented in bold type).

Sp/collection site	Dry-down 2016	Dry-down 2017	Dry-down 2018	Dry-down 2019
<i>Cicer arietinum</i>	3	3	4	4
Domestic (Aust desi)	1	1	1	1
Domestic (Aust kabuli)	1	1		
Domestic (Indian desi)	1	1	3	3
<i>Cicer echinospermum</i>	30	39	11	11
Cermik	3	5	1	1
Destek	5	11	1	1
Gunasan	2	1	1	1
Karabahce	8	11	2	2
Ortanca	2	1	1	1
Siv-Diyar	10	10	5	5
<i>Cicer reticulatum</i>	102	82	12	12
Baristepe 1	8	8	1	1
Baristepe 2	5	3		
Baristepe 3	8	6	1	1
Beslever	7	6	2	2
Cudi	9	4	1	1
Cudi 2	9	4	2	2
Dereici	10	7	1	1
Egil	6	4	2	2
Kalkan	6	3	1	1
Kayatepe	7	6		
Kesentas	9	7		
Oyali	6	7	1	1
Sarikaya	9	2		
Savur	1	1		
Sirnak	2	14		
Grand total	135	124	27	27

Bold values are species summaries and grand total.

FIGURE 1 | Plot layout of the 2016–2017 randomized complete block design (RCBD; $n = 3$) and 2018–2019 split-plot ($n = 4$) rainout shelter trials at CSIRO Floreat listing block identity for each plot [red for terminal drought (TD) and blue for well-watered (WW)].

In the 2016, trial all plots were planted on 9th June. Large phenological differences among accessions meant that 10 separate dry-down groups were required to initiate the water deficit treatment from the onset of pod filling (**Figure 2A**; **Table 2**). The combined dry-down period in the 2016 trial was characterized by gradually rising temperatures punctuated by temperature spikes at approximately 10-day intervals such that

TABLE 2 | Mean temperatures and rates of change for dry-down (DD) groups evaluated in the 2016–2019 water deficit trials at CSIRO Floreat.

Dry-down group/trial year	Mean temperature (°C)			
	2016	2017	2018	2019
1	16.4	17.4	17.7	17.8
2	16.8	18.4	18.0	
3	17.4	20.6	19.2	
4	18.2	22.4		
5	19.3			
6	19.4			
7	19.0			
8	20.7			
9	21.8			
10	21.3			
DD period temp change (°C/day)	0.09	0.18	0.13	−0.02
Vegetative phase rainfall (mm)	463.2	540.4	444.6	461.8
Veg phase mean temp (°C)	13.3	15.0	14.9	15.2

Vegetative phase rainfall and temperature calculated up to the start of the dry-down phase.

later groups experienced more terminal drought stress than earlier groups (**Figure 2A**; **Table 2**).

To ensure consistent terminal drought stress later trials used a combination of vernalization (4 weeks at 4°C) and staggered sowing dates to make the onset date of water deficit as uniform as possible. To this end, the 2017 and 2018 trials were sown over six staggered occasions according to phenology from 7th June to 12th July and 18th June to 9th July, respectively. Despite our attempts to synchronize reproductive phases, in the 2017 trial, the onset of water deficit was still staggered across four groups starting from 2nd to 30th October. Over this period, there was a linear temperature increase punctuated by high temperature spikes driving sequentially increasing terminal drought stress in later dry-down groups (**Figure 2B**; **Table 2**). The 2018 staggered sowing was more effective with the onset of water deficit staggered across three groups at weekly intervals from 15th to 29th October, with smaller temperature differences driven by a weaker linear temperature rise (**Figure 2C**; **Table 2**). Nevertheless, isolated hot days (maxima > 30.0°C) in early-mid November 2018 added stress to the end of the dry-down period in groups 2 and particularly 3 (**Figure 2C**).

In the 2019, trial sowing dates were widened even more to ensure reproductive synchrony. Fourteen staggered sowing dates over 37 days were used in each of three staggered starting dates (3rd June, 10th June, and 17th June) to grow three times as many pots as required for the experiment. In early October, plants with similar pod development were selected for all treatments and moved into the final split plot configuration such that the 2019 trial contained only a single, synchronous dry-down group exposed to a similar temperature range as previous early groups (**Figure 2D**; **Table 2**). Unlike previous water deficit treatments, the 2019 TD treatment was irrigated on day 2 (receiving 50% of the water used calculated individually for each pot) to extend the dry down period because of a temperature spike (**Figure 2D**). Thereafter, no water was added to the TD treatment, as in the 2017–2018 experiments.

Observations

Phenology was measured as described previously. Plant maturity (defined as 95% of pods ripe) was recorded in 2016 and 2018.

Plants were bagged prior to pod maturity to prevent seed loss during shattering. Subsequently, plants were counted and harvested, total biomass, seed weight, and number recorded. Seed size and harvest index were calculated from this data. A TD stress index was calculated based on the percentage of the well-watered value (Bouslama and Schapaugh Jr., 1984).

Reproductive water-use was measured by weighing pots at 2-day intervals after the onset of the terminal drought. Final pot weights were used to calculate plant available water (PAW) at each weighing. This data were used to calculate reproductive WUE, PAW% and transpiration per unit time, expressed in days, thermal time or cumulative evapotranspiration. In 2018 and 2019, water-use was measured in both water regimes.

Wild and Domestic Developmental Differences Over the Lifecycle (Exp 5)

To better understand behavior of wild and domestic *Cicer* under contrasting water supply, it was necessary to develop an understanding of developmental differences between the groups. To this end, we designed a glasshouse pot trial measuring above and below-ground dry matter partitioning and their effects on water extraction and WUE. To describe the effects of domestication independently of phenology, we studied plant development throughout the lifecycle in balanced early and late phenology subsets of wild and domesticated *Cicer* (**Table 3**). These were grown as single plants in 120 × 16 cm split pots filled to 100 cm with Gingin loam using an RCB (n = 3), with extra replication (n = 3 per harvest) allowing for destructive harvests at 35, 70, and 105 days and a final harvest at physiological maturity (155–170 days). The experiment was sown on June 20 2017 to coincide with the normal Mediterranean winter growing season after 4 weeks of vernalization at 4°C.

At sowing, the soil was filled to field capacity and water-use monitored at approximately 2-day intervals throughout the growing season. This was done by applying a known amount of water and calculating daily water-use by subtracting the overflow emerging from a tube at the base of each pot. To minimize evaporation, the soil surface layer was covered by plastic beads to a depth of 5 cm. Cumulative water-use curves were generated by fitting logistic functions to the daily summed water-use (see Statistical Analysis). Water inputs were matched to water use to avoid over-filling. Water-use data were used to calculate WUE of above ground biomass at each destructive harvest and reproductive WUE using pod weights at final harvest. Reproductive WUE was calculated on cumulative water-use between the start of podding and final harvest (as in the rainout shelter experiments), biomass WUE was based on water-use since the start of the experiment and each destructive harvest. Phenology observations were made on all pots throughout the growing season as described in the rainout shelter experiments.

Water inputs were stopped for those pots slated for destructive harvest 1 week before each harvest date to reveal water extraction

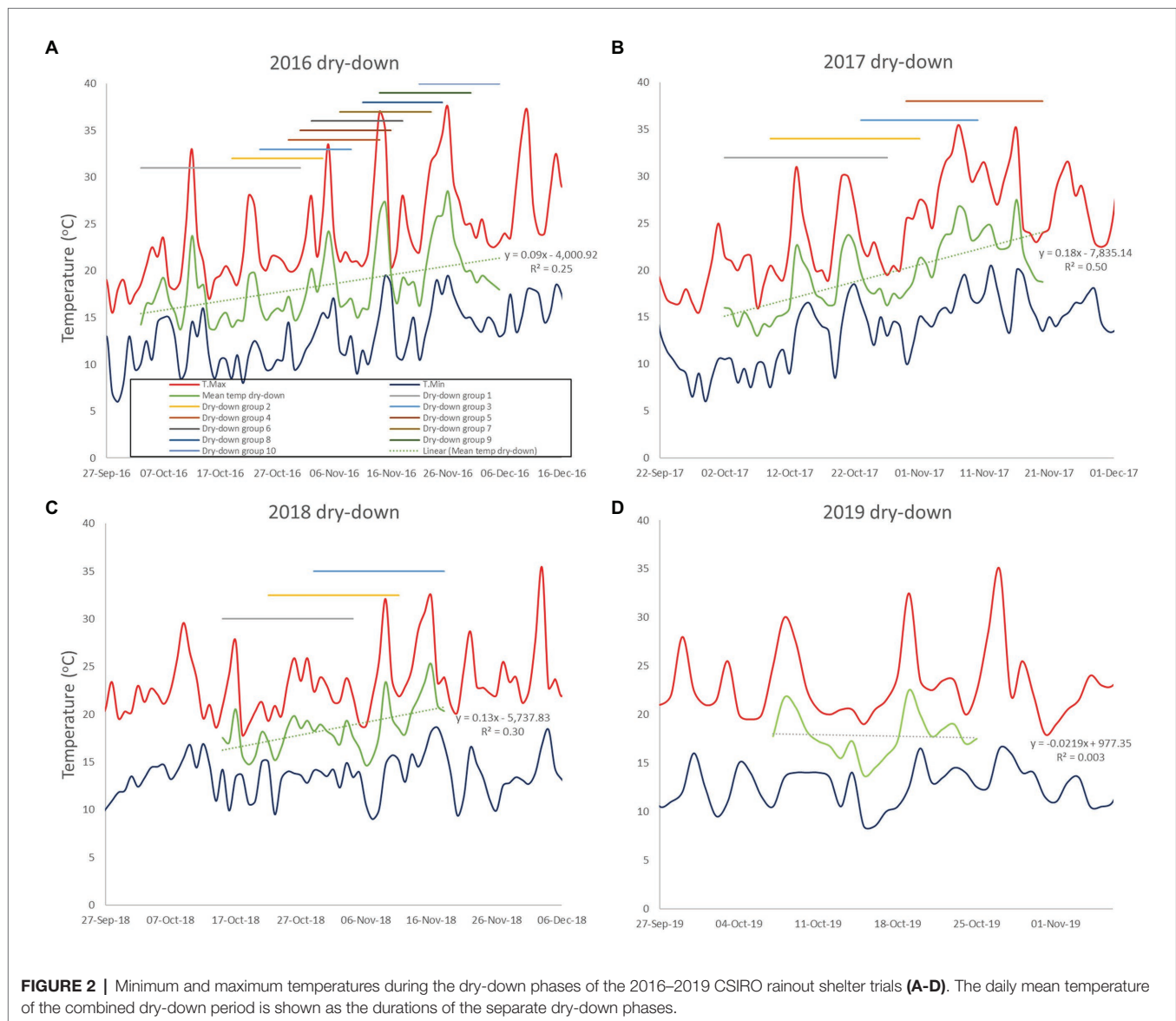


TABLE 3 | Early and late phenology subsets of domestic (*C. arietinum*) and wild *Cicer* (*C. echinospermum* and *C. reticulatum*) selected for studying developmental differences across the lifecycle.

Species	Early	Late
<i>Cicer arietinum</i>	ICCV 93929 (Indian desi cv.)	Almaz (Australian kabuli cv.)
<i>Cicer echinospermum</i>	S2Drd_061	Ortan_066, Cermi_063
<i>Cicer reticulatum</i>	Besev_066	Sirna_060, Sirna_063

profiles along the soil column. The above ground biomass was removed, processed into vegetative and reproductive tissue, and leaf area measured (only until day 105 because leaves were shed by physiological maturity). The split pots were carefully divided in two without disturbing the soil column, which was then separated into 20 cm segments (0–20, 20–40, 60–80, and 80–100 cm). Soil in each segment was sub-sampled, fresh and dry weights recorded (after 48 h oven drying at

60°C) to calculate the relative water content [(fresh-dry weight)/dry weight]. Roots were carefully washed out of each soil segment, sieved, and dried. The data from each segment were used to explore relationships with depth (see Statistical Analysis). Root weights from each segment were also summed to provide a total root biomass, which was used to calculate root index (percent of total biomass attributable to roots), shoot to root ratio, vegetative, and total biomass (sum of above and below ground biomass). Below ground data were only available for the first three harvest dates because by physiological maturity the roots had started to decay.

Statistical Analysis

Genstat (V20) was used for all statistical analyses. Nested ANOVA and regression models were used to partition variance between species, between collection sites within species and finally between accessions within collection sites within species.

Wild and domestic differences were analyzed using orthogonal contrasts. The same nested approach was used in the lifecycle development study, except that accessions were nested within phenology categories within species, rather than collection sites within species.

Replications were fitted within water regime in the RCBDs used in 2016–2017, where the water regime treatments were established on separate contiguous areas of the rainout shelter. In the 2018–2019, split plot designs water regime was treated as the main plot, accessions as the subplot. The lifecycle development study was analyzed as a standard RCBD using reps as blocks.

Linear and non-linear {exponential, $Y = a + br^X$; logistic, $Y = a + c/[1 + e^{-b(X - m)}]$ } regression was used to model changes in PAWC and biomass over time, and root growth and water extraction over depth in the development study. These regression models were chosen to best fit the trend in the data, producing residual plots with normally, and independently distributed errors. Changes in PAWC and leaf area over time, and root growth over time and depth were exponential. Water-use, vegetative matter, and total biomass followed logistic patterns over time, while WUE over the growing season was well-modeled by quadratic regression.

Residual plots were used throughout to identify outliers and confirm that errors were normally and independently distributed.

Correlation-based principal components analysis (PCA) was used to integrate the results using two-way accession by water regime means and curve parameters generated by the analyses described above.

RESULTS

Cicer Species Responses to Terminal Water Deficit: Productivity, Phenology, and Water-Use

Nested ANOVA demonstrated consistent patterns among traits across years. In those phenological observations taken prior to the onset of water deficit treatments, the largest differences occurred between species, followed by collection sites within species, and finally accessions within collection sites ($p < 0.001$ for all). Observations made after the imposition of water deficit tended to follow a similar pattern, albeit with significant interactions with water regime.

Domestic chickpea was characterized by a consistently earlier phenology than wild *C. reticulatum* and *C. echinospermum* (Table 4). Moving from a common (2016) to a staggered sowing date (2017 onward) to try to synchronize the reproductive phase shortened the vegetative phase, particularly in the 2019 trial, based on the widest combination of sowing dates (Table 4).

The reduction of the vegetative phase had ramifications on biomass production. Total biomass decreased as the vegetative phase was shrunk from 2016 to 2019 (Table 4; Figure 3). Despite these scale changes, water deficit consistently reduced above ground biomass production in all trials, albeit with some differences between species. In 2016 and 2019, wild *Cicer* was more responsive to the WW treatment than domestic chickpea ($p < 0.001$), in 2017, there was no difference ($p_{\text{diff}} = 0.390$),

while in 2018, *C. echinospermum* was uncharacteristically unresponsive (Figure 3). These interactions notwithstanding, domestic chickpea tended to accumulate more above ground biomass than wild *Cicer* on most occasions (Figure 3, TD: 2016–2018; WW: 2017–2018).

Under terminal drought, wild and domestic *Cicer* had consistently similar reproductive investment (Figure 3: harvest index and seed weight). Indeed, only the 2019, TD treatment plants exhibited a significant harvest index difference in favor of domestic chickpea ($p < 0.05$), while seed weights were similar. However, there were clear domestic/wild differences in the reproductive response to the additional irrigation provided by the WW treatment. In 3 of the 4 trial years (2016–2018) harvest index and seed weight increased more in domesticated than in wild *Cicer* (Figure 3; $p_{\text{diff}} = 0.013$ – <0.001) in response to WW treatment irrigation.

The reduction of vegetative phase length and total biomass production over years had implications on plant reproductive water-use, measured gravimetrically during the dry-down cycle. This is clearly indicated by reduced water extraction from 2016 to 2019 in all species (Figure 4; Table 5). Nevertheless, there were remarkably consistent, contrasting water-use patterns between domestic and wild *Cicer* across all years (Figure 4). Thus, domestic chickpea tended to extract less water in the dry down cycle than either wild species, as indicated by significant γ intercept differences (Figure 4; $p < 0.001$) in all years, captured by the exponential curve parameter B (see Table 5 for raw water uptake values). Surprisingly, the exponential rate of water-use (parameter R) was higher in chickpea than in wild *Cicer* from 2017 to 2018 ($p < 0.05$). These curve parameter differences are clearly evident across years in Figure 3, with chickpea varieties forming a tight cluster at the lower end of the water extraction range (except for ICCV 93929 as a high extracting outlier in 2016 only). Conversely, the two wild species were characterized by a wider range of water-use curves among accessions, corresponding to higher mean values than in domestic chickpea (Figure 4).

In 2018 and 2019, water-use was also measured in the WW treatment. ANOVA was dominated by very large wild and domestic differences ($p_{\text{diff}} = 0.005$ – <0.001), without water regime interaction ($p_{\text{diff}} = 0.2110.298$). This is important because it indicates that while the wild *Cicer* species extracted more water under terminal drought, they also consistently used more water under the WW treatment (Table 5).

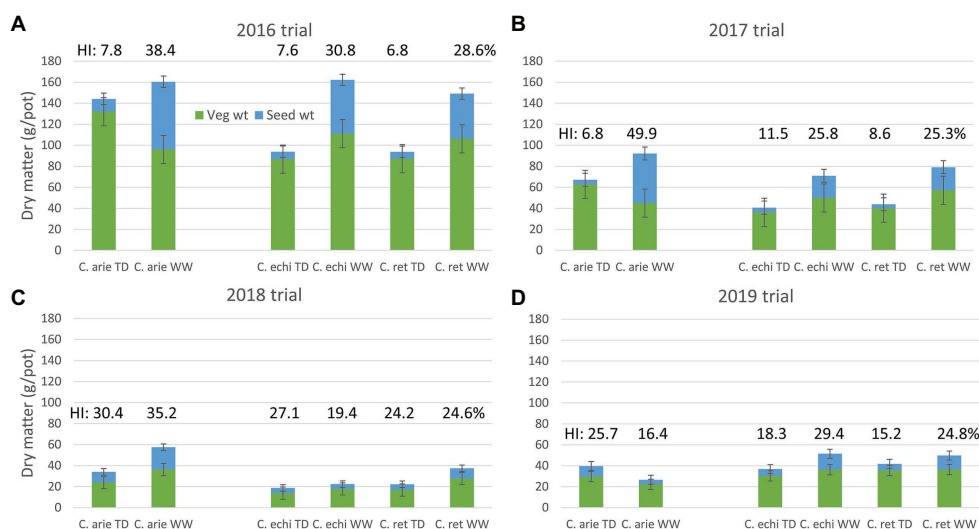
Integrating the Results With Multivariate Analysis

Principal components analysis integrated the observations made in the TD treatment, capturing 55–66% of variance in two components in the ordinations performed for the 2016–2019 trials (Figure 5). Wild and domestic *Cicer* were consistently separated by phenology, water extraction, and WUE (Figure 5; Table 5; all years), and to a lesser extent rate of use (2017–2019). Productivity traits (seed weight, biomass, etc.) were consistently closely associated with harvest index and negatively correlated with water extraction and WUE. Thus, high yielding plants had high harvest index and WUE but extracted less water during the reproductive

TABLE 4 | Domestic (*C. arietinum*) vs. wild (*C. echinospermum* and *C. reticulatum*) phenology (flowering and podding means) in the 2016–2019 dry-down experiments.

Exp yr	<i>Cicer arietinum</i>	<i>Cicer echinospermum</i>	<i>Cicer reticulatum</i>	LSD*	Pval: wild vs. dom
Flowering					
2016	90	117	121	1	<0.001
2017	67	93	92	4	<0.001
2018	78	97	92	2	<0.001
2019	68	77	78	2	<0.001
Podding					
2016	109	127	130	1	<0.001
2017	82	103	102	3	<0.001
2018	88	105	101	2	<0.001
2019	79	86	88	5	<0.001

*Least significant difference (LSD).

**FIGURE 3** | Domestic (*C. arie*, *C. arietinum*) vs. wild *Cicer* (*C. echi*, *C. echinospermum*; *C. ret*, *C. reticulatum*) productivity (total biomass, vegetative, and seed weight) under well-watered (WW) and terminal drought (TD) treatments in the 2016–2019 rainout shelter trials (A–D). Error bars represent 1 LSD, the proportion of vegetative to seed weight represents harvest index (HI), values included above columns.

phase dry-down than low yielding plants. In 2 of the 3 years where TD stress indices were calculated (2016–2017) plants with high TD productivity achieved a high proportion of their WW productivity, that is to say they were less responsive to the more benign WW treatment than poorly yielding plants [TD stress indices were not calculated in 2019 because domestic chickpea performed so poorly in the WW treatment (Figure 3D) rendering the wild comparison meaningless].

Principal components analysis confirmed the specific distinctions described earlier that wild *Cicer* tends to have later phenology, lower productivity, and greater water extraction coupled with lower maximum use rates than domestic chickpea. However, PCA also demonstrated considerable within species variation in these traits, particularly in 2016–2017, when a much larger wild cohort was investigated. In both these years, some *C. reticulatum* and *C. echinospermum* accessions had

similar, or greater productivity, WUE and harvest index than the most productive domestic chickpea (Figures 5A,B). The ordinations for the WW treatment were remarkably similar (data not presented). Wild and domestic *Cicer* were separated by phenology, productivity, harvest index, and WUE (measured in 2018 only). As in the TD treatment, seed yield was positively correlated with harvest index and WUE, and negatively correlated with phenology. With very few exceptions, domestic chickpea was more productive than wild *Cicer* in the WW treatment, associated with greater reproductive investment and WUE.

Wild and Domestic Developmental Differences

To independently test the effects of phenology and “wildness” on above and below-ground dry matter vegetative and reproductive partitioning, and their effects on water extraction

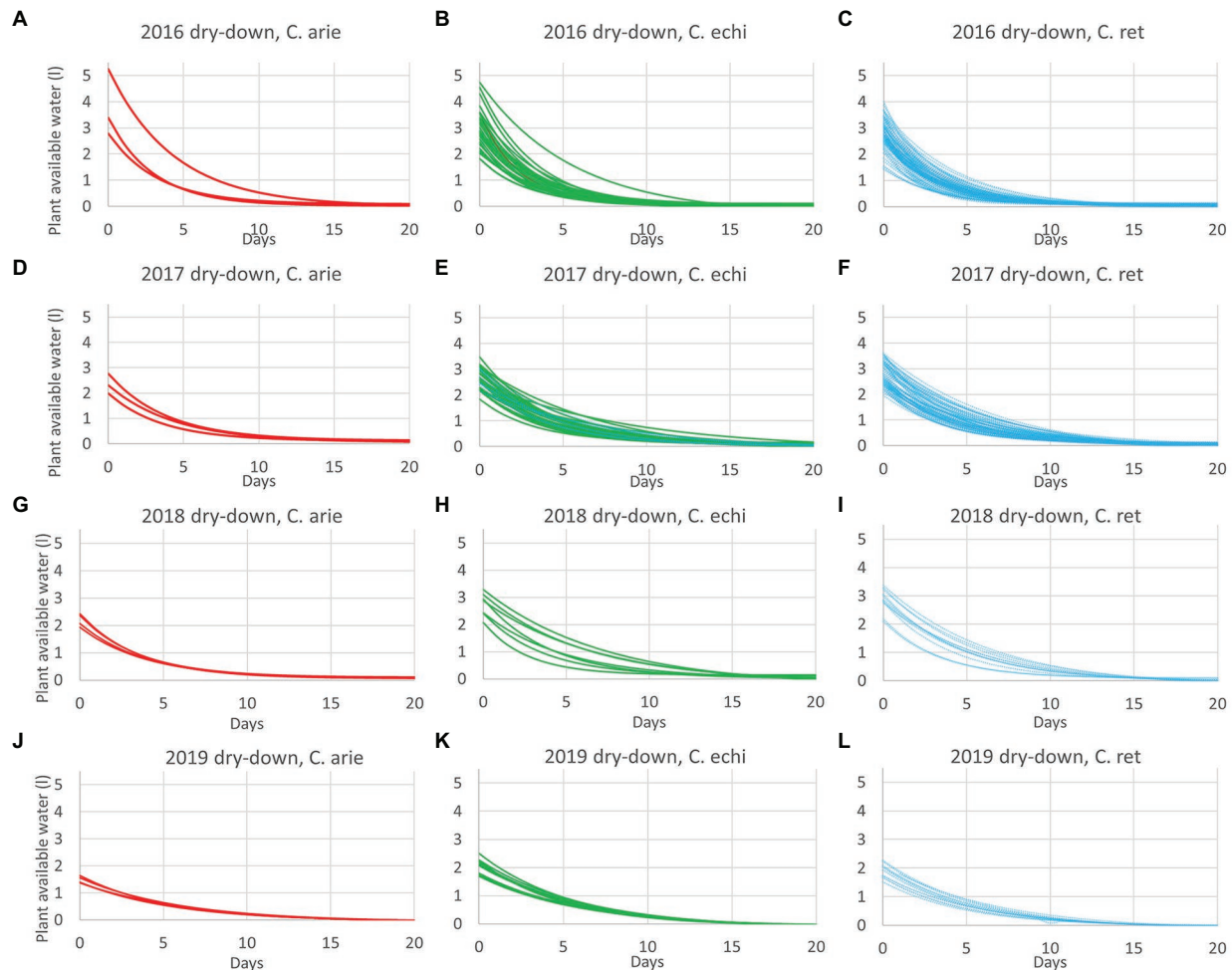
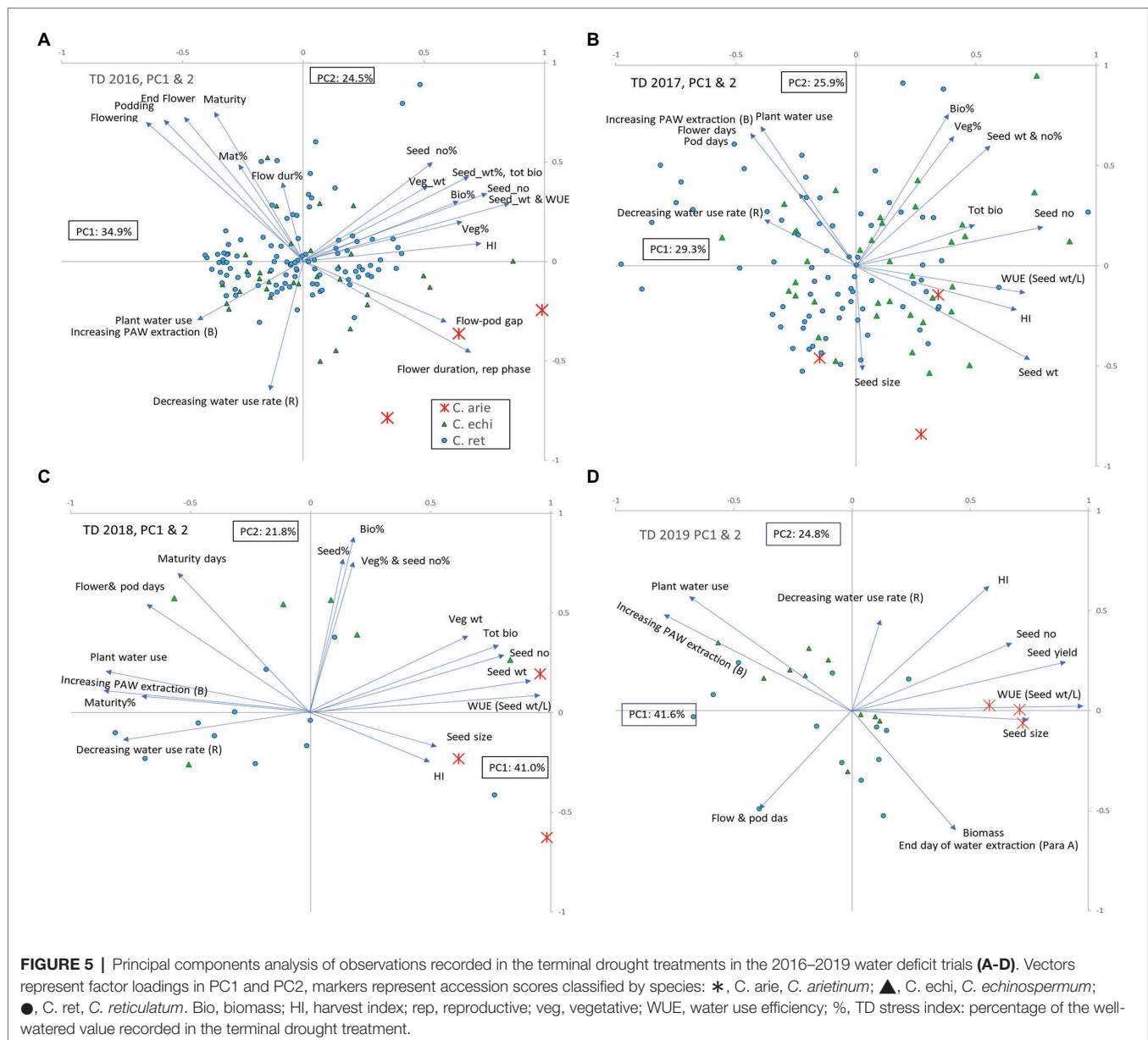


FIGURE 4 | Domestic (*C. arie*, *C. arietinum*, —) vs. wild *Cicer* (*C. echi*, *C. echinospermum*, —; *C. ret.*, *C. reticulatum*, —) water-use after the onset of terminal drought in dry-down experiments from 2016 to 2019 (A–L). Exponential curves fitted for accessions within species captured 89.6–96.9% of variance.

TABLE 5 | Domestic (*C. arietinum*) vs. wild (*C. echinospermum* and *C. reticulatum*) plant water uptake and reproductive water use efficiency (WUE) in the 2016–2019 dry-down experiments.

Treat/yr	<i>Cicer arietinum</i>	<i>Cicer echinospermum</i>	<i>Cicer reticulatum</i>	LSD	Pval: wild vs. dom
Plant water uptake (L)					
2016 TD	3.76	3.02	2.78	0.33	<0.001
2017 TD	2.45	2.66	2.79	0.30	0.065
2018 TD	2.24	2.79	2.77	0.29	<0.001
2018 WW	7.20	7.91	9.23	1.35	0.033
2019 TD	2.10	2.70	2.40	0.22	<0.001
2019 WW	6.80	7.70	7.50	0.62	0.006
Reproductive water use efficiency (seed weight g/L)					
2016 TD	4.0	2.8	2.6	0.8	0.002
2017 TD	1.9	2.2	1.5	0.6	0.947
2018 TD	4.9	2.2	2.2	1.1	<0.001
2018 WW	3.0	0.9	1.1	0.5	<0.001
2019 TD	4.9	2.5	2.7	0.7	<0.001
2019 WW	0.6	2.0	1.9	0.4	<0.001

TD, terminal drought; WW, well-watered reproductive phase; LSD, least significant difference.



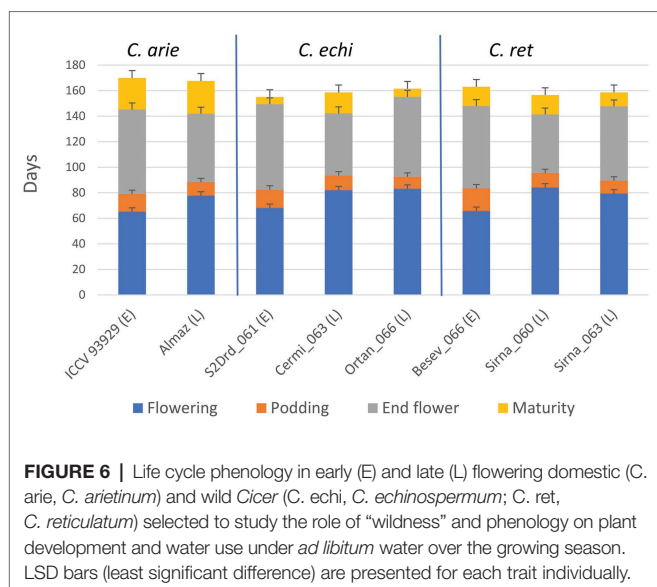
and WUE, we studied plant development throughout the lifecycle in balanced early and late phenology subsets of wild and domestic *Cicer* grown in the glasshouse.

This experimental approach set vegetative phase limits within and between species. **Figure 6** shows consistent phenology category differences between species for flowering and podding, not for end of flowering and particularly maturity. Thus, our phenology categories reliably determined the length of the vegetative phase and the onset of reproduction (e.g., short and early vs. longer and later) with no, or very minor species differences (**Figure 6**). However, domestic chickpea matured approximately 10 days later than wild *Cicer* (**Figure 6**; $p < 0.001$), with important flow-on effects on the length of the reproductive phase. While the reproductive phase was consistently 9–22 days longer in the early compared to late

phenology groups, the domestic reproductive phase was 9–20 and 12–15 days longer than wild in the early and late categories, respectively (**Figure 6**).

These species and phenology category differences played out in plant development. Polynomial contrasts in nested ANOVA showed strong linear and quadratic interactions between species and phenology categories within species for most measured traits ($p < 0.001$). Because of developmental lag phases and ceiling values in most of these plant structures, non-linear logistic and occasionally exponential regression was generally more appropriate (**Figure 7**), confirming the significant interactions indicate by ANOVA.

The early harvests (35 and 70 days) were dominated by wild vs. domestic *Cicer* differences. Domestic chickpea had significantly greater early vigor than wild *Cicer* ($p < 0.001$),



producing far greater leaf area, root, and shoot mass at 35 and 70 days, accounting for almost all significant differences (Figure 7). The only exception was at 35 days where the late kabuli variety Almaz had much greater biomass than the early desi ICCV 93929, a difference that disappeared by day 70. Thus, the early-mid vegetative phase growth rates were far higher in domestic chickpea than wild *Cicer*, particularly for leaf area (Figure 7).

These wild-domestic early vigor differences were reflected in root development and water extraction down the soil profile. Root mass and water extraction decreased curvi-linearly with depth in all species (Figure 8). At 35 days, domestic chickpea (particularly the kabuli cultivar, Almaz) had far greater rates of root weight decline and water extraction over depth than wild *Cicer* ($p < 0.001$), driven by massive differences in the 0–20 cm soil layer, disappearing by 40–60 cm ($p_{\text{diff}} = 0.334$), the limit of root exploration (Figure 8A). Interestingly, water extraction of the two chickpea cultivars was similar, despite their differences in surface root production. Domestic water extraction was greater than wild in the upper soil layer ($p < 0.001$) but not at greater depths (Figure 8B; $p_{\text{diff}} = 0.150$). At 70 days, wild and domestic differences still dominated, albeit these were becoming smaller in terms of root distribution (Figure 8C). Root decline rates over depth were similar across all accessions, regardless of species or phenology category, except for Almaz ($p_{\text{diff}} < 0.001$) with its high surface layer root mass. Nevertheless, there were significant intercept differences between the desi cultivar ICCV 93929 and most of the wild *Cicer* ($p_{\text{diff}} = 0.127$ – 0.001), reflecting root mass differences at most depths (Figure 8C). These wild-domestic differences had a large impact on water extraction down the soil profile (Figure 8D). Domestic chickpea depleted water from the top three soil layers, leaving progressively more water in the remaining two soil layers, captured by a similar upward trending quadratic curve in both the early desi and late kabuli varieties. Conversely, the wild *Cicer* water extraction curve

was much more linear over depth, leaving significantly more residual water in all layers ($p < 0.001$).

Day 70 marked the end of the lag phase when all species showed rapid growth rates for above and below ground biomass, and new species differences and phenology category by species interactions were emerged. Thus, the early specific differences in root growth rates disappeared during the rapid growth phase, while phenology category by species interactions emerged, indicated by higher growth rates in the late compared to the early *Cicer arietinum* ($p_{\text{diff}} = 0.015$) and *C. echinospermum* ($p_{\text{diff}} = 0.027$), but not *C. reticulatum* (Figure 7, $p_{\text{diff}} = 0.733$). This pattern was clearly evident in the root distribution over depth, with steeper declines in late compared to early *C. arietinum* ($p_{\text{diff}} = 0.065$) and *C. echinospermum* ($p_{\text{diff}} = 0.002$), but not *C. reticulatum* (Figure 8E, $p_{\text{diff}} = 0.259$). Accordingly, there were no specific differences in water extraction by day 70, with a common, relatively flat curvi-linear response (Figure 8F, $p_{\text{diff}} = 0.467$ – 0.708), leaving only minor variety within phenology category differences. Above-ground vegetative biomass growth curves were clearly logistic across the range of harvest dates (Figure 7). As with root growth, there were no consistent specific differences.

In terms of leaf area growth species differences trumped phenology interactions, but now the wild *Cicer* had a higher rate of leaf area production than domestic chickpea ($P_{\text{diff}} = 0.048$), such that by 105 days there were no specific differences in leaf area (Figure 7, $P_{\text{diff}} = 0.192$).

Phenology and “wildness” were both important in the relative above- and below-ground dry matter partitioning, indicated by strong species by phenology category interaction ($p < 0.001$). While later types of all species invested more heavily in roots, as indicated by higher root indices and lower shoot:root ratios (Figure 9), the contrast was much stronger in wild than domesticated *Cicer*, particularly *C. reticulatum*. Thus, late wild *Cicer* have a much greater root index than late domestic chickpea, while there were no specific differences among the early group. Interestingly, this pattern was evident already at 35 days after sowing, well before the start of flowering, suggesting that these differences are not explained by differences in the vegetative phase length.

Pod growth rates also showed strong species by phenology category interaction ($p < 0.001$). Chickpea rates were considerably higher than wild, while late types of all species tended to fill pods at higher rates than early types, reaching similar final pod weights at maturity despite a later podding onset (Figure 7). Pod growth rate differences between early and late types were larger in *C. reticulatum* than in *C. arietinum* and *C. echinospermum*, accounting for the significant interaction. Ultimately at maturity, reproductive investment (harvest index) was far greater in domestic than wild *Cicer* ($p < 0.001$), and greater in early compared to late *C. arietinum* and *C. echinospermum*, but not *C. reticulatum*. The combination of rapid pod growth rates and high reproductive investment was responsible for higher rates of aerial biomass production in domestic vs. wild *Cicer* over the growing season ($p < 0.001$), with no differences between phenology categories within species (Figure 7). The final areal biomass values at maturity largely

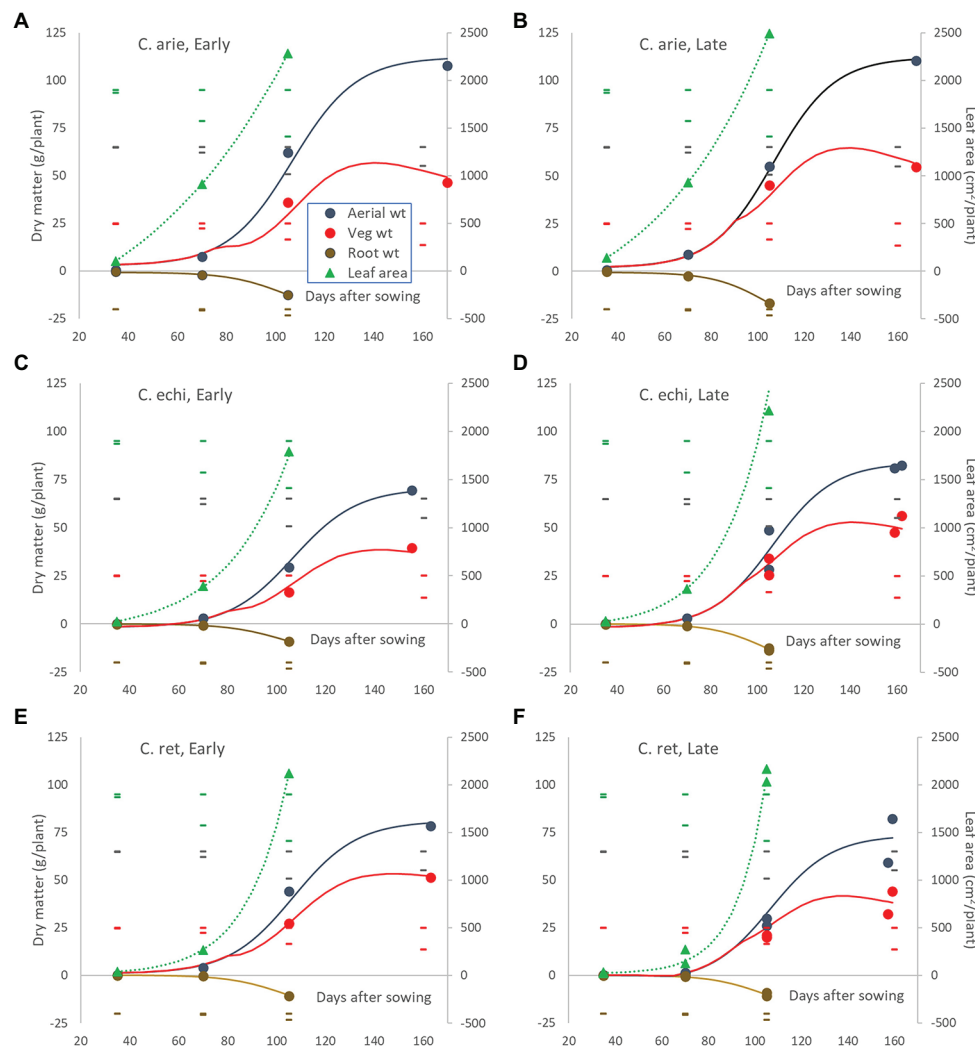


FIGURE 7 | Leaf area and above- and below-ground dry matter partitioning over the growing season in early and late flowering domestic (A,B) and wild *Cicer* (C–F). Color coded markers, fitted curves and daily LSD bars (least significant difference) are presented for leaf area (—), root weight (—), vegetative matter (—), and total aerial biomass (—). Point values represent accession means from destructive harvests at 35, 70, 105, and 155–170 days after sowing. Logistic and exponential curves fitted for species/phenology categories captured 92.6–96.3% of variance, accounting for all differences between accessions. The area between the vegetative matter and aerial biomass curves represents pod weight. Root weights are presented as negative values to facilitate visual root-shoot comparisons over time.

reflected these rate differences: domestic larger than wild ($p < 0.001$), late *C. echinospermum* larger than early, whereas the opposite was the case for *C. reticulatum* ($P_{\text{diff}} = 0.05$).

Plant water-use throughout the growing season followed a logistic curve largely, but not totally mirroring biomass production (Figure 10A). Daily water-use rates remained flat for the first 70 days with minor wild vs. domestic differences (12 vs. 13 ml/day, $P_{\text{diff}} = 0.014$), turning sharply at 90 days, and peaking at 111 days. Moreover, the species/phenology characteristics described earlier also played out in the water-use curves: later *C. arietinum* and *C. echinospermum* used more water [parameter C (curve maximum value), $p < 0.001$] at higher rates (logistic growth rate k , $p < 0.001$) than early types, while the opposite was the case for *C. reticulatum* (Figure 10A; $p < 0.001$).

Nevertheless, the wild and domestic curves were remarkably similar, given differences in their aerial biomass. This is reflected in large wild vs. domestic differences in WUE across the life cycle (Figure 10B). While all species became more efficient in their water-use for aerial biomass production over time, the rate of increase was much larger in domestic vs. wild *Cicer* (Figure 10B; $p < 0.001$). Domestic WUE was larger than wild ($p < 0.001$) at every point sampled throughout the lifecycle. Moreover, there were important differences in the shape of the response. In domestic chickpea (and early *C. reticulatum*), the rise in WUE over time was curvi-linear, peaking approximately 2/3 of the way through the life-cycle, whereas the response in the remaining wild *Cicer* groups was much linear, peaking at maturity (Figure 10B). Nevertheless,

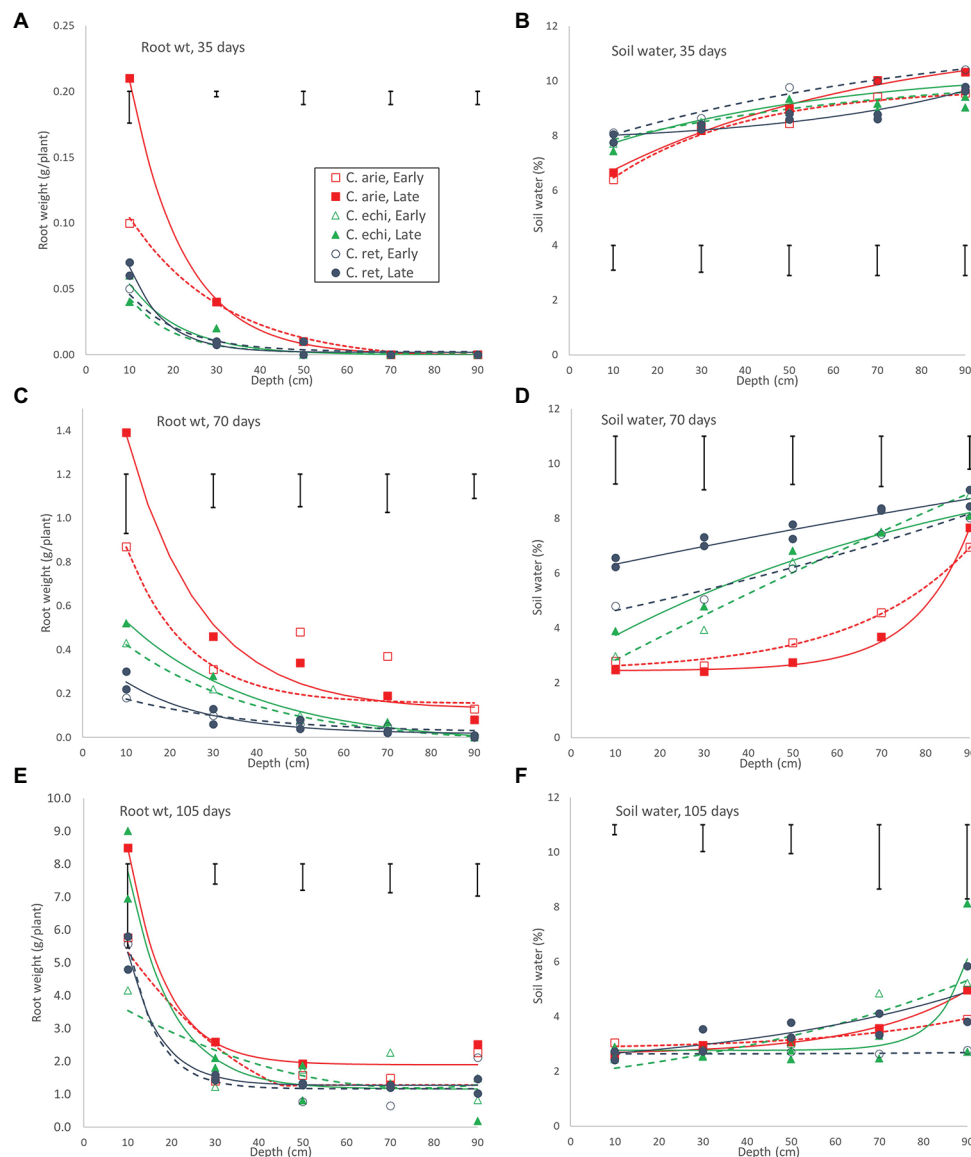


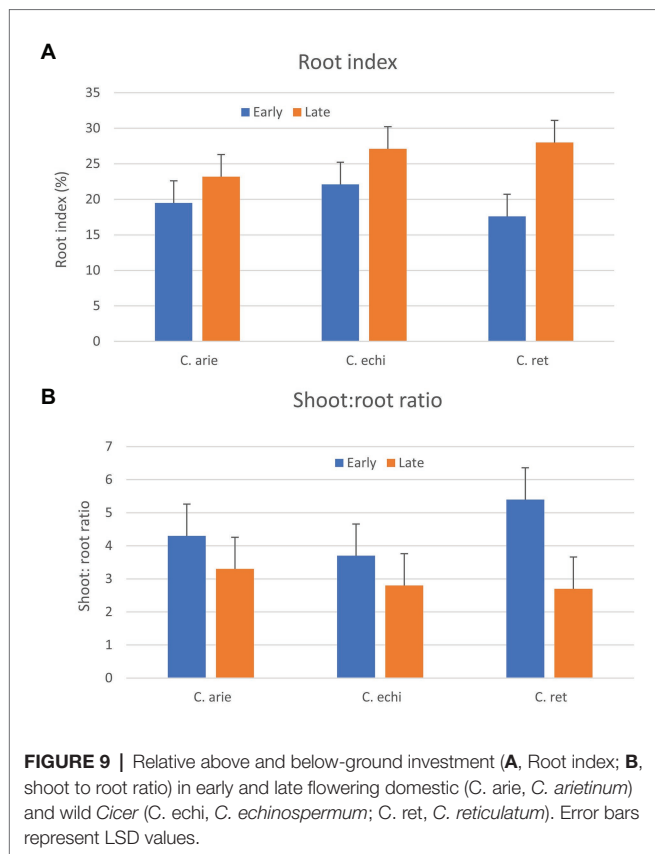
FIGURE 8 | Root development and water extraction over depth over the growing season in early and late flowering domestic (*C. arie*, *C. arietinum*) and wild *Cicer* (*C. echi*, *C. echinospermum*; *C. ret*, *C. reticulatum*). Point values represent accession means from destructive harvests at 35 (A,B), 70 (C,D), and 105 days after sowing (E,F). Error bars represent LSD values for individual harvests. Exponential curves fitted for species/phenology categories captured 86.8–97.4% and 37.6–83.1% of variance for root weight and water extraction, respectively.

at maturity the domestic WUE remained higher than wild (Figure 9B; $p < 0.001$), accounting for all phenology/species category or variety differences ($p_{\text{diff}} = 0.901$ and 0.561).

DISCUSSION

Our work shows that domestication has ramifications throughout the entire lifecycle in this *Cicer* example, but that domestication as a promotor of acquisitiveness is not particularly helpful to distinguish wild and domesticate. While domestic chickpea was

indeed much more responsive to resource-rich conditions, and wild and domestic differences minimized under terminal drought (similar to Matesanz and Milla, 2018), these differences were not driven by greater resource acquisition in domestic chickpea. On the contrary, wild *Cicer* was able to extract more water under water deficit, but also used more water when it was freely available. Nor were there consistent wild-domestic growth rate differences across all plant organs over time, as would be predicted if wild and domestic occupied different ends of the stress tolerator-competitor continuum (Grime, 2006). While early vegetative growth and water extraction was much more



rapid in domestic compared to wild *Cicer*, this is likely to be a function of greater early vigor, presumably a by-product of selecting large seed sizes. This is a common phenomenon in both domesticated cereals (Evans and Dunstone, 1970) and grain legumes (Berger et al., 2017), and underlined in this example by the early wild and domestic growth differences, and between desi and the much larger seeded kabuli type within domesticated chickpea. The fact that these growth rate differences disappeared over the growing season highlights the need to study development over the entire lifecycle. Indeed, in the late vegetative phase, as growth rates became exponential, rates of leaf area expansion were considerably greater in wild compared to domestic *Cicer*.

Clearly, the slow-wild/fast-domestic dichotomy is not supported in this *Cicer* example. Unpacking the stress tolerator-competitor continuum (Grime, 2006) in an agricultural context indicates why this may be. The triangle of Grime (2006) suggests that acquisitive traits such as rapid growth rates above and below ground are selected for in resource-rich environments, where there is strong competition for these resources (the “use it or lose it” scenario, see examples in Grime, 1988, 2006). Conversely, this strategy is risky and maladaptive in resource-poor environments, where periods of stress have to be tolerated. These environments select for slow, resource efficient growth, and stress tolerating physiology (Grime, 2006) that is strongly expressed in extremophiles such as desert cacti and succulents (Ehleringer and Mooney, 1983), and is less

relevant to annual plants and agriculture (Berger et al., 2016). The lifecycles of annual plants and most agricultural crops balance stress escape (the ruderal strategy, third apex in the triangle of Grime, 2006) against acquisitiveness. Although well-managed fertile fields are likely to represent a more resource-rich environment for crops than the natural systems in which their wild progenitors evolved, this is unlikely to have selected for greater acquisitiveness (aka competitive capacity) because of the disparate selection pressures imposed by the two systems. Whereas wild plants are selected as individuals, crops are grown and selected as populations rather than as single plants. This has selected for crops that are poor competitors as individual plants, lifting the productivity of the community as a whole, rather than maximizing the fitness of the individual (Donald, 1963, 1981; Reynolds et al., 1994). This is exactly what we see in our results, where the wild *Cicer* consistently extract and use more water (which would otherwise be lost to their competitors) than their domesticated counterparts. In fact, far from being parsimonious and efficient stress tolerators, the wild *Cicer* seems to be profligate competitors compared to domestic chickpea. Indeed, the poor competitor aspect of chickpea water-use is being exploited in the development of cultivars for short season, stored soil moisture systems in the semi-arid tropics. ICRISAT is promoting the use of cultivars that use less water in the vegetative phase, leaving more residual soil water for seed filling (Zaman-Allah et al., 2011), a scenario that is difficult to envisage occurring in the natural world of inter-plant competition.

Instead of greater acquisitiveness *Cicer* domestication appears to have selected for greater reproductive efficiency, whether measured in terms of dry matter allocation or WUE. This is particularly apparent under high input conditions, where domestic chickpea is far more responsive than wild in terms of pod growth rates and harvest index. Conversely, these studies provide evidence of greater vegetative investment by the wild *Cicer* in terms of later phenology, higher rates of leaf area production, and greater relative investment in roots, particularly in late types. This combination facilitates greater water extraction under terminal drought, but also allows greater water-use when it is freely available. Here, there are interesting parallels with the Old World lupins, where there are similar wild-domestic differences in reproductive and vegetative investment, and flow-on effects on water use (Berger and Ludwig, 2014; Berger et al., 2020). The cereals may be more conservative in terms of relative dry matter partitioning (Wacker et al., 2002; see Wang et al., 2017 for counter-indication), but do show a similar trend in tillering along the domestication series. Thus, post anthesis tillering decreases in domestic tetra and hexaploid wheat compared to the diploid wild progenitors (Evans and Dunstone, 1970). Interestingly, similar phenology-reproductive investment trade-offs were seen in wild emmer × durum RIL populations, highlighted by the remarkable similarity in the ordinations presented by Peleg et al. (2009) and those in this manuscript. These patterns align well with competitor-ruderal continuum of Grime (2006). By selecting for earlier phenology to fit crops into a time delimited production system, domesticated

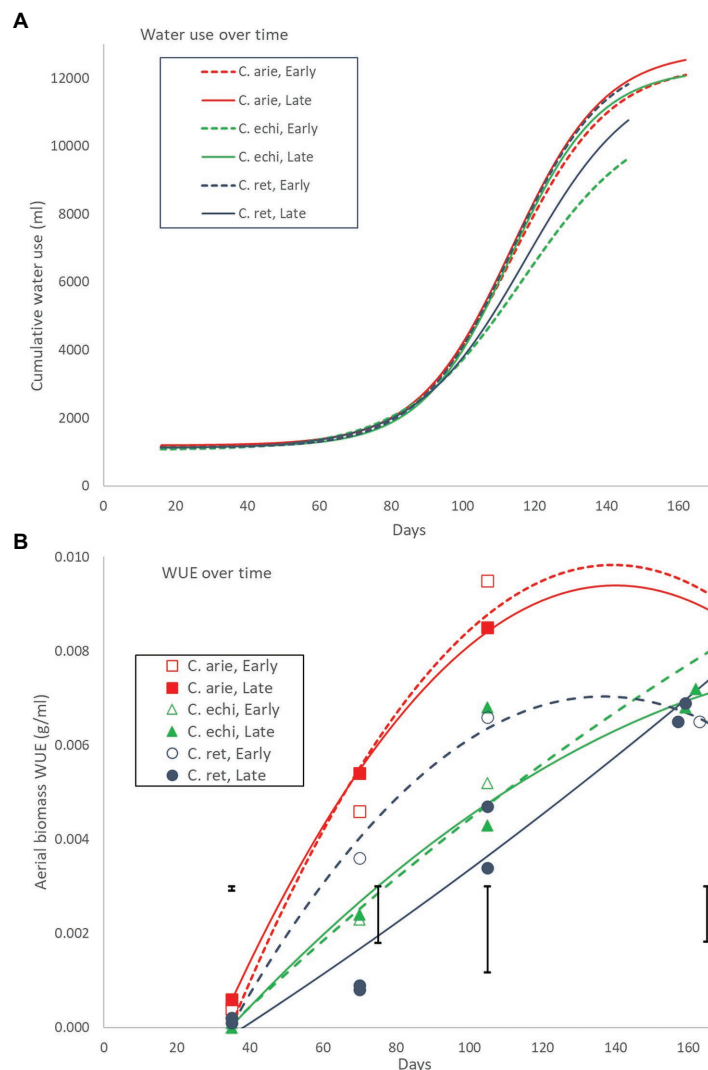


FIGURE 10 | Water-use **(A)** and WUE **(B)** over the growing season in early and late flowering domestic (*C. arie*, *C. arietinum*) and wild *Cicer* (*C. echi*, *C. echinospermum*; *C. ret*, *C. reticulatum*). The non-linear logistic and linear quadratic curves in **(A)** and **(B)** account for 98.0 and 92.1% of variance, with significant species/phenology category differences for all parameters. Point values in **(B)** represent accession means from destructive harvests at 35, 70, 105, and 155–170 days after sowing. Error bars represent LSD values for individual harvests.

crops took on ruderal attributes such as increased harvest index, and seedling establishment was improved by greater early vigor, associated with large grain size (Berger et al., 2017). This is a productive and efficient, but risky reproductive strategy that works in agriculture where the crop is protected from grazing, disease, and competition, circumstances that do not pertain in the wild. Life is less certain for wild *Cicer*, and flexibility appears to be more important than reproductive efficiency over the long term. Our field observations support this idea: we have seen wild accessions re-growing from the base after *Ascochyta* has killed the aerial plant parts, or after grazing and also after very late rains when the plant had seemingly matured, leaving only shattered pods, and dry straw.

Apart from differences in vegetative and reproductive dry matter allocation, domestic water-use was also more efficient than wild. In cereals, this has been attributed to greater harvest index (Wang et al., 2017). While our later season results confirm this idea, the fact that this trend was already apparent in the vegetative phase when calculated over aerial (Figure 10) or total biomass (data not presented) indicates that these WUE differences are not entirely attributable to vegetative vs. reproductive partitioning. Given previous reports of wild-domestic differences in stomatal conductance and photosynthesis (Evans and Dunstone, 1970; Matesanz and Milla, 2018), this underlines the possibility of differential water-use regulation in the genus *Cicer* that warrant further investigation.

CONCLUSION

This study has demonstrated large wild-domestic differences in vigor, vegetative and reproductive investment, water extraction, and WUE in the genus *Cicer* indicative of evolution under contrasting selection pressures. Dry matter allocation in wild *Cicer* is more vegetative than in domestic, which appears to be responsible for greater water extraction under terminal drought, and also greater water-use when it is freely available, but leads to a lower reproductive capacity and efficiency. While increased water extraction may be useful for chickpea improvement in water limiting environments, the wild trait combination should be disassembled as much as possible to evaluate its potential independently. To this end, wild × domestic populations have already been developed. It will be fascinating to see to what extent it is possible to recombine wild and domestic trait assemblages, whether water extraction capacity can be evaluated without simultaneously introducing low harvest index, or whether this returns a similar wild-domestic cline as observed in wild emmer × durum populations (Peleg et al., 2009).

REFERENCES

- Abbo, S., Berger, J., and Turner, N. C. (2003). Viewpoint: evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. *Funct. Plant Biol.* 30, 1081–1087. doi: 10.1071/FP03084
- Abbo, S., Pinhasi Van-Oss, R., Gopher, A., Saranga, Y., Ofner, I., and Peleg, Z. (2014). Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends Plant Sci.* 19, 351–360. doi: 10.1016/j.tplants.2013.12.002
- Bazakos, C., Hanemian, M., Trontin, C., Jiménez-Gómez, J. M., and Loudet, O. (2017). New strategies and tools in quantitative genetics: how to go from the phenotype to the genotype. *Annu. Rev. Plant Biol.* 68, 435–455. doi: 10.1146/annurev-arplant-042916-040820
- Berger, J., Abbo, S., and Turner, N. C. (2003). Ecogeography of annual wild *Cicer* species: the poor state of the world collection. *Crop Sci.* 43, 1076–1090. doi: 10.1146/annurev-arplant-042916-040820
- Berger, J. D., and Ludwig, C. (2014). Contrasting adaptive strategies to terminal drought stress gradients in Mediterranean legumes: phenology, productivity and water relations in wild and domesticated *Lupinus luteus* L. *J. Exp. Bot.* 65, 6219–6229. doi: 10.1093/jxb/eru006
- Berger, J. D., Ludwig, C., and Whisson, K. (2020). Changing water use and adaptive strategies along rainfall gradients in Mediterranean lupins. *Plant Biol.* 22, 298–308. doi: 10.1111/plb.13076
- Berger, J., Palta, J., and Vadez, V. (2016). Review: an integrated framework for crop adaptation to dry environments: responses to transient and terminal drought. *Plant Sci.* 253, 58–67. doi: 10.1016/j.plantsci.2016.09.007
- Berger, J. D., Shrestha, D., and Ludwig, C. (2017). Reproductive strategies in mediterranean legumes: trade-offs between phenology, seed size and vigor within and between wild and domesticated *Lupinus* species collected along aridity gradients. *Front. Plant Sci.* 8:548. doi: 10.3389/fpls.2017.00548
- Bouslama, M., and Schapaugh, W. T. Jr. (1984). Stress tolerance in soybeans. I. Evaluation of three screening techniques for heat and drought tolerance. *Crop Sci.* 24, 933–937. doi: 10.2135/cropsci1984.0011183X002400050026x
- Coyne, C. J., Kumar, S., Von Wettberg, E. J. B., Marques, E., Berger, J. D., Redden, R. J., et al. (2020). Potential and limits of exploitation of crop wild relatives for pea, lentil, and chickpea improvement. *Legum. Sci.* 2:e36. doi: 10.1002/leg3.36
- Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Sci.* 57, 1070–1082. doi: 10.2135/cropsci2016.10.0885
- Dempewolf, H., Eastwood, R. J., Guarino, L., Khoury, C. K., Müller, J. V., and Toll, J. (2014). Adapting agriculture to climate change: a global initiative

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

JB conceived the research, helped to run the experiments, analyzed the data, and wrote the manuscript. RP, CL, SP, FB, and KW implemented the experimental program, discussed the results, and provided feedback on the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

GRDC CSP00185 and CSIRO have co-invested in funding this research.

- to collect, conserve, and use crop wild relatives. *Agroecol. Sustain. Food Syst.* 38, 369–377. doi: 10.1080/21683565.2013.870629
- Donald, C. M. (1963). “Competition among crop and pasture plants” in *Advances in agronomy*. ed. A. G. Norman (New York: Academic Press), 1–118.
- Donald, C. M. (1981). “Competitive plants, communal plants, and yield in wheat crops” in *Wheat science – Today and tomorrow*. eds. L. T. Evans and W. J. Peacock (New York: Cambridge University Press), 223–247.
- Ehleringer, J., and Mooney, H. (1983). “Productivity of desert and Mediterranean-climate plants” in *Physiological plant ecology IV*. eds. O. L. Lange, P. S. Nobel, C. B. Osmond and H. Ziegler (Berlin, Heidelberg: Springer), 205–231.
- Evans, L., and Dunstone, R. (1970). Some physiological aspects of evolution in wheat. *Aust. J. Biol. Sci.* 23, 725–742. doi: 10.1071/bi9700725
- Gepts, P. (2014). “Domestication of plants,” in *Encyclopedia of agriculture and food systems*. ed. N. K. Van Alfen (Oxford: Academic Press), 474–486.
- Grime, J. P. (1988). “The C-S-R model of primary plant strategies — origins, implications and tests” in *Plant evolutionary biology*. eds. L. D. Gottlieb and S. K. Jain (Dordrecht: Springer Netherlands), 371–393.
- Grime, J. P. (2006). *Plant strategies, vegetation processes and ecosystem properties*. Chichester: Wiley & Sons.
- Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., et al. (2014). Current perspectives and the future of domestication studies. *Proc. Natl. Acad. Sci. U. S. A.* 111, 6139–6146. doi: 10.1073/pnas.1323964111
- Matesanz, S., and Milla, R. (2018). Differential plasticity to water and nutrients between crops and their wild progenitors. *Environ. Exp. Bot.* 145, 54–63. doi: 10.1016/j.envexpbot.2017.10.014
- McCouch, S., Baute, G. J., Bradeen, J., Bramel, P., Bretting, P. K., Buckler, E., et al. (2013). Agriculture: feeding the future. *Nature* 499, 23–24. doi: 10.1038/499023a
- Meyer, R. S., Duval, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x
- Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14:840. doi: 10.1038/nrg3605
- Milla, R., Morente-López, J., Alonso-Rodrigo, J. M., Martín-Robles, N., and Chapin, F. S. (2014). Shifts and disruptions in resource-use trait syndromes during the evolution of herbaceous crops. *Proc. Biol. Sci.* 281:20141429. doi: 10.1098/rspb.2014.1429
- Milla, R., Osborne, C. P., Turcotte, M. M., and Violle, C. (2015). Plant domestication through an ecological lens. *Trends Ecol. Evol.* 30, 463–469. doi: 10.1016/j.tree.2015.06.006

- Newman, T. E., Jacques, S., Grime, C., Kamphuis, F. L., Lee, R. C., Berger, J., et al. (2020). Identification of novel sources of resistance to ascochyta blight in a collection of wild *Cicer* accessions. *Phytopathology* doi: 10.1094/phyto-04-20-0137-r (in press).
- Peleg, Z., Fahima, T., Abbo, S., Yakir, D., Korol, A. B., et al. (2009). Genomic dissection of drought resistance in durum wheat x wild emmer wheat recombinant inbred line population. *Plant Cell Environ.* 32, 758–779. doi: 10.1111/j.1365-3040.2009.01956.x
- Purugganan, M. D., and Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature* 457:843. doi: 10.1038/nature07895
- Reen, R. A., Mumford, M. H., and Thompson, J. P. (2019). Novel sources of resistance to root-lesion nematode (*Pratylenchus thornei*) in a new collection of wild *Cicer* species (*C. reticulatum* and *C. echinospermum*) to improve resistance in cultivated chickpea (*C. arietinum*). *Phytopathology* 109, 1270–1279. doi: 10.1094/PHYTO-02-19-0047-R
- Reynolds, M. P., Acevedo, E., Sayre, K. D., and Fischer, R. A. (1994). Yield potential in modern wheat varieties: its association with a less competitive ideotype. *Field Crop Res.* 37, 149–160. doi: 10.1016/0378-4290(94)90094-9
- Singh, K. B., Malhotra, R. S., and Saxena, M. C. (1990). Sources for tolerance to cold in *Cicer* species. *Crop Sci.* 30, 1136–1138. doi: 10.2135/cropsci1990.0011183X003000050036x
- Singh, K. B., Malhotra, R. S., and Saxena, M. C. (1995). Additional sources of tolerance to cold in cultivated and wild *Cicer* species. *Crop Sci.* 35, 1491–1497. doi: 10.1126/science.277.5329.1063
- Singh, K. B., Robertson, L. D., and Ocampo, B. (1998). Diversity for abiotic and biotic stress resistance in the wild annual *Cicer* species. *Genet. Resour. Crop. Evol.* 45, 9–17. doi: 10.1023/A:1008620002136
- Tanksley, S. D., and McCouch, S. R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277, 1063–1066. doi: 10.1126/science.277.5329.1063
- Thormann, I., Parra-Quijano, M., Endresen, D. T. F., Rubio-Teso, M. L., Lriondo, M. J., and Maxted, N. (2014). “Predictive characterization of crop wild relatives and landraces” in *Technical guidelines version 1*. (Rome, Italy: Bioversity international).
- von Wettberg, E. J. B., Chang, P. L., Başdemir, F., Carrasquilla-Garcia, N., Korbu, L. B., Moenga, S. M., et al. (2018). Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nat. Commun.* 9:649. doi: 10.1038/s41467-018-02867-z
- Wacker, L., Jacomet, S., and Körner, C. (2002). Trends in biomass fractionation in wheat and barley from wild ancestors to modern cultivars. *Plant Biol.* 4, 258–265. doi: 10.1055/s-2002-25735
- Wang, J. Y., Turner, N. C., Liu, Y. X., Siddique, K. H. M., and Xiong, Y. C. (2017). Effects of drought stress on morphological, physiological and biochemical characteristics of wheat species differing in ploidy level. *Funct. Plant Biol.* 44, 219–234. doi: 10.1071/FP16082
- Zaman-Allah, M., Jenkinson, D. M., and Vadez, V. (2011). Chickpea genotypes contrasting for seed yield under terminal drought stress in the field differ for traits related to the control of water use. *Funct. Plant Biol.* 38, 270–281. doi: 10.1071/FP10244

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

Copyright © 2020 Berger, Pushpavalli, Ludwig, Parsons, Basdemir and Whisson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Targeting Low-Phytate Soybean Genotypes Without Compromising Desirable Phosphorus-Acquisition Traits

Mireadili Kuerban¹, Wenfeng Jiao¹, Jiayin Pang^{2,3}, Jingying Jing⁴, Li-Juan Qiu⁵, Wenli Ding¹, Wen-Feng Cong^{1*}, Fusuo Zhang¹ and Hans Lambers^{1,2,6}

¹ College of Resources and Environmental Sciences, National Academy of Agriculture Green Development, Key Laboratory of Plant-Soil Interactions, Ministry of Education, China Agricultural University, Beijing, China, ² The UWA Institute of Agriculture, The University of Western Australia, Perth, WA, Australia, ³ School of Agriculture and Environment, The University of Western Australia, Perth, WA, Australia, ⁴ College of Grassland Science and Technology, China Agricultural University, Beijing, China, ⁵ National Key Facility for Gene Resources and Genetic Improvement, Key Laboratory of Crop Germplasm Utilization, Ministry of Agriculture, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, ⁶ School of Biological Sciences, The University of Western Australia, Perth, WA, Australia

OPEN ACCESS

Edited by:

Rogério Margis,
Federal University of Rio Grande do
Sul, Brazil

Reviewed by:

Renu Pandey,
Indian Agricultural Research Institute
(ICAR), India
Monica Jolly,
Indian Agricultural Research Institute
(ICAR), India

*Correspondence:

Wen-Feng Cong
wenfeng.cong@cau.edu.cn

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 20 June 2020

Accepted: 23 November 2020

Published: 14 December 2020

Citation:

Kuerban M, Jiao W, Pang J,
Jing J, Qiu L-J, Ding W, Cong W-F,
Zhang F and Lambers H (2020)
Targeting Low-Phytate Soybean
Genotypes Without Compromising
Desirable Phosphorus-Acquisition
Traits. *Front. Genet.* 11:574547.
doi: 10.3389/fgene.2020.574547

Phytate-phosphorus (P) in food and feed is not efficiently utilized by humans and non-ruminant livestock, potentially contributing to high losses of P to the environment. Crops with high P-acquisition efficiency can access soil P effectively. It remains elusive whether crop genotypes with high P-acquisition efficiency can also have low seed phytate concentrations. A core collection of 256 soybean [*Glycine max* (L.) Merr.] genotypes from China with diverse genetic background were grown in the same environment and seeds were sampled to screen for seed phytate-P concentration. Some of these genotypes were also grown in a low-P soil in the glasshouse to measure root morphological and physiological traits related to P acquisition. Large genotypic variation was found in seed phytate-P concentration (0.69–5.49 mg P g⁻¹ dry weight), total root length, root surface area, rhizosheath carboxylates, and acid phosphatase activity in rhizosheath soil. Geographically, seed phytate-P concentration was the highest for the genotypes from Hainan Province, whereas it was the lowest for the genotypes from Inner Mongolia. Seed phytate-P concentration showed no correlation with any desirable root traits associated with enhanced P acquisition. Two genotypes (Siliyuan and Diliuhuangdou-2) with both low phytate concentrations and highly desirable P-acquisition traits were identified. This is the first study to show that some soybean genotypes have extremely low seed phytate concentrations, combined with important root traits for efficient P acquisition, offering material for breeding genotypes with low seed phytate-P concentrations.

Keywords: soybean germplasm, phytate, phosphorus-acquisition efficiency, root morphology, carboxylates

INTRODUCTION

Phosphorus (P) deficiency is a major limiting factor for crop production worldwide owing to low soil P availability. It is readily sorbed by oxides and hydroxides of aluminium (Al) and iron (Fe) in acid soils, and precipitated by calcium (Ca) in alkaline soils (Hinsinger, 2001; Raghothama and Karthikeyan, 2005), leading to 70–90% of P applied as fertilizer becoming unavailable to most crop

plants (Holford, 1997). The high rates of P-fertilizer application in crop production result not only in gradually diminishing phosphate rock reserves, which is associated with decreasing P-fertilizer quality, but also in environmental issues associated with off-site effects of P fertilizers (Cordell et al., 2009; Ghaffar et al., 2017). Therefore, it is essential to enhance P-acquisition efficiency by breeding highly P-efficient crop genotypes (Cong et al., 2020).

Soybean [*Glycine max* (L.) Merr.] as a widely cultivated grain legume, is an important source of protein and vegetable oil for human consumption, and also widely used in animal feed globally (Yuan et al., 2007; Liang et al., 2010). In addition, soybean has a pivotal ecological function in cropping system, e.g., nitrogen fixation (Salvagiotti et al., 2008), soil carbon sequestration (Cong et al., 2015), enhancing soil P availability (Xia et al., 2013), and decreasing soil-borne diseases (Gao et al., 2014) for themselves and the following crops. Approximately 60–80% of total P in soybean seed accumulates as phytate (myo-inositol 1, 2, 3, 4, 5, 6-hexakisphosphate, PA, or IP6) (Raboy et al., 1984). A high phytate concentration in seed is an undesirable trait, because it renders zinc (Zn) and other micro-nutrients unavailable for humans and livestock, contributing to malnutrition, especially Zn and iron (Fe) deficiency (Raboy, 2001; Perera et al., 2019). Moreover, phytate cannot be efficiently utilized by humans and non-ruminant animals, contributing to high losses of P to the environment. Thus, breeding for low-phytate soybean genotypes with high P-acquisition efficiency is a highly desirable sustainable strategy. It is one of the vital and preliminary steps for improving the desired traits in breeding programs that explore the genetic variation in germplasm (Perera et al., 2019). Hence, there is a need to study the genetic variation in seed phytate-P using a large set of soybean genotypes with diverse genetic backgrounds.

Identifying P-efficient crop genotypes is a sustainable and effective way to tighten the P cycle, thereby reducing P-fertilizer input, and mitigating the risks of pollution of ground- and surface water (Cong et al., 2020). Genotypic differences in nutrient-use efficiency are closely related to differences in efficiency of nutrient acquisition by roots (Marschner, 1998). Root traits such as root size, morphology, physiology, and mycorrhizal associations play a dominant role, both in P acquisition and in the exploration of a large soil volume, particularly under low P availability in soil (Marschner, 1998). Root morphological and physiological traits may respond to P deficiency (Zhou et al., 2016). Fernandez and Rubio (2015) reported higher specific root length and smaller average root diameter in soybean where P-acquisition efficiency increased under P deficiency. Root physiological traits, mainly root exudates such as carboxylates and phosphatase enzymes, may also increase plant P-acquisition efficiency (Dinkelaker et al., 1989; Pang et al., 2018). Earlier work showed genetic variability of root architecture in a core collection of soybean germplasm, with a shallow root architecture associated with higher P efficiency (Ao et al., 2010). However, knowledge of genotypic variation in root size, morphological and physiological traits is largely unknown for soybean. Therefore, there is an urgent need to investigate the variation in root size including root dry weight and rhizosheath soil dry weight, root morphological traits including root surface area and total root length, and root physiological traits including

the total amount of carboxylates in the rhizosheath and acid phosphatase activity in a large set of soybean genotypes with diverse genetic backgrounds.

Previous studies compared the seed phytate-P concentration of soybean genotypes (Horner et al., 2005; Maharjan et al., 2019), and mainly focused on variation in seed phytate-P concentration among genotypes and its relationship with nutrients such as Fe, Zn, and calcium (Ca), and protein within a relatively small set of genotypes. Thus, there is a chance they missed identifying soybean genotypes with low seed phytate-P concentration. Wang et al. (2007) have identified a core collection of soybean germplasm, on average representing 81.5% of genetic variation. This collection provides us with a great opportunity to identify genotype with low seed phytate-P concentration and its relationship with desirable root traits related to P acquisition in soil with a low P availability. Identification of genotypes with highly desirable P-acquisition traits with low phytate-P concentrations has many benefits. First, lowering the seed phytate-P concentration would enhance the bio-availability of micro-nutrients such as Zn and Fe for humans and farm animals (Raboy, 2001), and thus improve the nutritional quality of soymeal. Increased nutritional quality of soymeal would go hand in hand with decreasing environmental pollution, especially eutrophication of aquatic environments (Sharpley et al., 2000; Brinch et al., 2002). Second, genotypes with desirable P-acquisition traits and low unavailable seed P concentrations require less P fertilization, and therefore reduce the cost for farmers and risk of P losses from the field (Cong et al., 2020).

The aims of the present study were (1) to investigate the variation in seed phytate-P concentration among 256 soybean genotypes and their geographically distributed differences; (2) to study the variation in root dry weight and rhizosheath soil dry weight, root morphological traits including total root length and surface area, as well as root physiological traits such as the amount of carboxylates and acid phosphatase activity in the rhizosheath among a subset of genotypes originally from the North China Plain (the main food producing region) under low-P condition; (3) to examine if soybean genotypes with desirable low phytate-P concentrations can simultaneously express desirable root traits associated with high P-acquisition efficiency, using a subset of genotypes. We hypothesized that (1) there is a large variation in seed phytate-P concentration and root size, morphological and physiological traits among soybean genotypes; (2) seed phytate-P concentration does not correlate with any of the measured desirable root traits related to P acquisition.

MATERIALS AND METHODS

Plant Material and Growing Conditions

We grew 256 genotypes (**Supplementary Table 1**) from divergent provinces of China in a single field in Hainan Province (108°37'–111°03' E, 18°10'–20°10' N), China to measure seed phytate-P concentrations. Annual mean temperature is 22°C with a minimum and maximum temperature of 17 and 27°C, respectively. Annual rainfall ranges from 1000 to 2600 mm with an average annual rainfall of 1639 mm. Annual sunshine

ranges from 1750 to 2650 h, and the total solar radiation is 4600–5800 MJ m⁻² year⁻¹. The soil is classified latosol. Information on these genotypes including Province of origin is described in **Supplementary Table 1**.

Of the 256 genotypes, 43 from five provinces in the Huang-Huai-Hai plain and Middle-Lower Yangtze plain were used to study six root traits that are associated with P acquisition (**Supplementary Table 1**). Those 43 genotypes are representative, because they include both ones that show the lowest and the highest phytate concentration, and the coefficient of variation (CV) is similar to that of the data set comprising 256 genotypes.

Pot experiment was conducted to determine root traits in a glasshouse at China Agricultural University, Beijing, with a day temperature of 25–30°C and a night temperature of 18–22°C. River sand and a soil with a low P availability were collected from Changping in the Beijing area. River sand was washed, and then both river sand and soil were air-dried and sieved (2-mm mesh size) to remove coarse fragments and macroarthropods prior to potting. Soil and sand were sterilized by gamma irradiation (>25 K Gray gamma irradiation before potting). Each pot (85 cm × 85 mm × 180 mm) was filled with 1.2 kg mixture of sterilized washed river sand and soil with a low P availability (in a ratio of 3:7, w/w). The pot experiment followed a completed randomized design with soybean genotype as the main factor. Each of 43 genotypes was replicated in four pots. Seeds were surface sterilized with 10% (v/v) hydrogen peroxide for 3 min, then washed with deionized water and germinated in the dark on moist filter paper at 25/20°C (day/night). In each pot, we planted four seeds at 20 mm depth and inoculated with *Rhizobium leguminosarum* bv. NM353 (provided by the Culture Collection of China Agricultural University). Seedlings were thinned to one plant per pot at 7 days after sowing. Each pot was watered with deionized water to 60% field capacity by weighing every other day.

Determination of Seed Phytate Concentration

Grains of each soybean genotype were collected at maturity, dried and ground to a fine powder. The powder was then passed through a 0.5 mm sieve. The phytate concentration was determined using the trichloroacetic acid method (Perera et al., 2019). Briefly, 4 g of seed powder was added to 50 ml of 30 g l⁻¹ trichloroacetic acid and kept for 2 h for digestion, and then shaken for 30 min, and filtered using dry filter paper. Ten ml of supernatant was treated with 4 ml of ferric chloride solution (FeCl₃) in a boiling water bath for 45 min. After cooling, samples were centrifuged at 1,710 × g for 10 min. The supernatant was removed, and 20–25 ml of 0.18 M trichloroacetic acid solution was added to wash the samples; after that, the samples were boiled in a water bath for 10 min twice. The precipitate was treated with 20 ml of water and 3 ml of 1.5 M NaOH in a boiling water bath for 30 min. After centrifugation, 5–10 ml of supernatant was treated with 3 ml of a mixture of di-acids (2 nitric acid + 1 perchloric acid) for digestion on an electric furnace at low temperature until there were white fumes. After cooling, samples were washed with 30 ml water several times; then, 3 ml of nitric acid solution

and 10 ml of chromogenic agent were added. After 20 min, the absorbance was measured spectrophotometrically at 420 nm.

Plant Harvest and Measurements

Plants grown in the glasshouse were harvested 45 days after sowing. Shoots were separated from the roots. In order to collect rhizosheath exudates, the method of Pearse et al. (2007) was modified. Briefly, at harvest, each pot was squeezed gently to allow dislodgement of the soil column and loosening of soil around the roots. The roots were shaken lightly to remove excess bulk soil; the soil and sand remaining attached to the roots was defined as rhizosheath soil (Pang et al., 2017). The root system was then transferred into a beaker containing 50 ml of 0.2 mM CaCl₂ to avoid cell damage and gently shaken for 60 s to remove as much of the rhizosheath soil as possible. For the determination of acid phosphomonoesterase (APase) activity in the rhizosheath soil, a 0.5 ml subsample of the rhizosheath extract was transferred into a 2 ml centrifuge tube (Alvey et al., 2001). This solution was incubated at 30°C for 30 min after adding 0.4 ml of 200 mM sodium acetate and 0.1 ml of 150 mM pNPP. After incubation, 0.5 ml of 0.5 M NaOH was added and the solution was filtered. Then, the absorbance of this solution was measured spectrophotometrically at 405 nm, representing rhizosphere APase (Nuruzzaman et al., 2006; Lyu et al., 2016). For the analysis of carboxylates, 10 ml of supernatant was taken and three drops of concentrated phosphoric acid and a microbial inhibitor, Micropur (Sicheres Trinkwasser, Graz, Austria) were added, and filtered through a 0.22 µm syringe filter into high performance liquid chromatography (HPLC) vials according to Shen et al. (2003). HPLC samples were frozen at –20°C until analysis. The rhizosheath soil in the rhizosphere exudate solution after extraction of APase and carboxylates was kept in the greenhouse to allow evaporation of excess solution. The rhizosheath soil was oven-dried at 105°C for 72 h, and the dry weight was recorded.

After collection of rhizosheath exudates, roots were washed free of remaining soil, spread out on a transparent plastic tray, and root images were obtained by Epson Perfection V700 dual lens scanning system at a resolution of 600 dpi (dots per inch). Root images were analyzed for total root length and total root surface area using WinRHIZO software (Pro 2009b, Regent Instruments Inc., Quebec City, Canada). After analyzing root images, roots were oven-dried at 70°C for 72 h until constant weight to measure biomass.

Classification of Soybean Genotypes

Soybean genotypes were classified into four categories as proposed by Bilal et al. (2018). These categories include (i) a low-phytate group with desirable root traits, (ii) a low-phytate group with undesirable root traits, (iii) a high-phytate group with desirable root traits, and (iv) a high phytate-P group with undesirable root traits. Low phytate-P refers to genotypes having seed phytate-P concentration lower than the mean seed phytate-P concentration, and desirable root traits means genotypes having more desirable root traits than the mean root traits and *vice versa*.

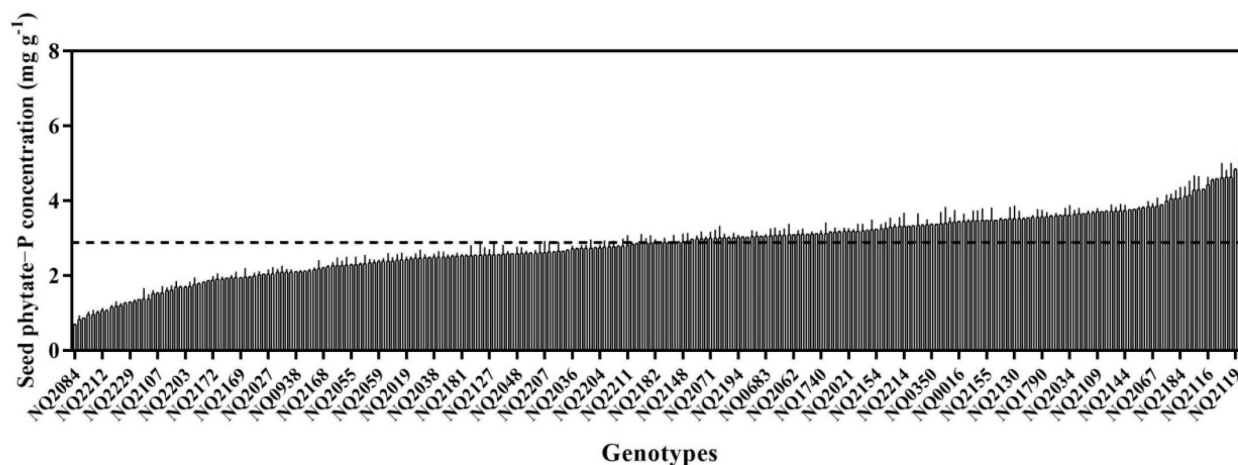


FIGURE 1 | Seed phytate-P concentrations of 256 soybean genotypes, expressed in units of phosphorus (P). Data are means + SE ($N = 2$). The dotted line represents the mean value of 256 soybean genotypes seed phytate-P concentration.

Statistical Analyses

Analysis of variance was used to analyse the differences in seed phytate-P concentration between genotypes used in different agroecological regions using SPSS19. Tukey's *post hoc* test was further used to examine the differences in case the effect was significant. Principal component analysis (PCA) of all plant traits comprising seed phytate-P concentration and root traits based on Pearson's correlation matrix was conducted by Genstat (version 18.2, Genstat Procedure Library Release PL26.2, VSN International, Hemel Hempstead, United Kingdom, 2016) to investigate their correlations.

RESULTS

Seed Phytate-P Concentration

Seed phytate-P concentrations from plants grown in the same environment varied greatly, with approximately a seven-fold difference among 256 varieties, ranging from 0.69 mg P g⁻¹ DW to 5.49 mg P g⁻¹ DW, with a mean value of 2.88 mg P g⁻¹ DW (Figure 1). Over half of the genotypes had a higher phytate concentration than the mean value. Two genotypes (Pixianalayinguang and Heikewudou) had the highest seed phytate-P concentration (>5 mg g⁻¹ DW), whereas five varieties (Dabaimaodou, Longchuanhuangniumao, Liushiribaidou, Siliyuan, and Lvcaodou) had very low values (<1 mg g⁻¹ DW) (Figure 1).

Geographic Background of Genotypes With Specific Seed Phytate-P Concentrations

There was a large variation in seed phytate-P concentration among genotypes from 26 provinces ($P < 0.05$). Averaged for all provinces, the mean seed phytate concentration in the 26 provinces ranged from 2.31 mg P g⁻¹ DW in Gansu to 5.08 mg P g⁻¹ DW in Hainan Province, with a mean value of

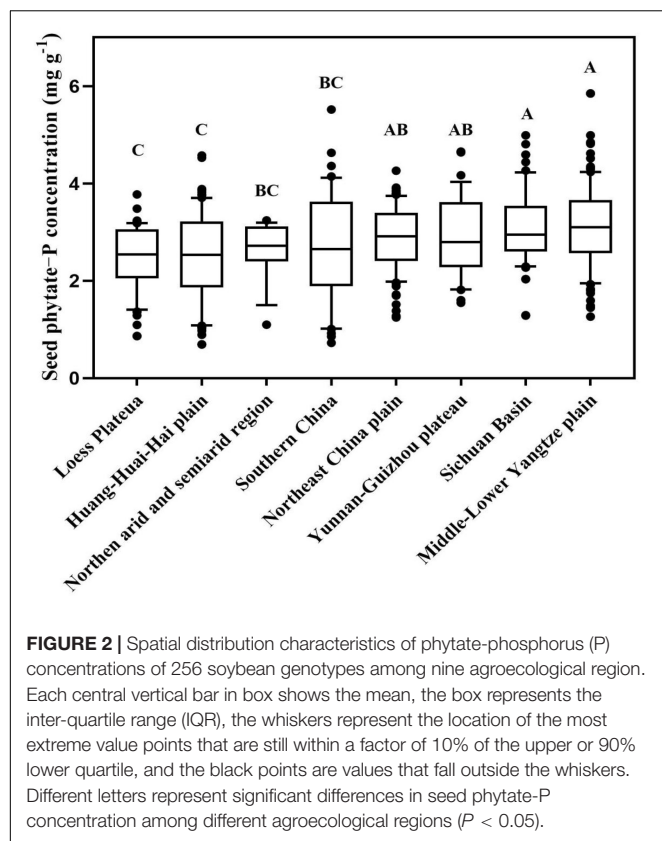
2.94 mg P g⁻¹ DW across China (Supplementary Figure 1 and Supplementary Table 2). The seed phytate-P concentration for genotypes from Hainan Province was 24~54% greater than that in other provinces ($P < 0.05$). The coefficient of variation of seed phytate-P concentration in each Province ranged from 0 to 52%. Among nine agroecological region, seed phytate-P concentration ranged from 2.50 mg P g⁻¹ DW on the Loess Plateau to 3.11 mg P g⁻¹ DW on the Middle-Lower Yangtze Plain ($P < 0.05$, Figure 2).

Root Traits

There was a large variation among 43 genotypes in root dry weight (DW) and rhizosheath soil dry weight ($P < 0.001$ for both) (Figures 3A,B). Root dry weight varied 2.5-fold, ranging from 0.18 g DW plant⁻¹ in Pingdingheito 0.45 g DW plant⁻¹ in Dadunxiaohaidou (Figure 3A). Rhizosheath soil dry weight varied 2.7-fold, from 2.0 g DW plant⁻¹ in Binhaidahuangkezijia to 5.4 g DW plant⁻¹ in Siliyuan (Figure 3B).

An eight-fold difference in total root length was found among 43 soybean varieties, ranging from 0.77 to 6.88 m plant⁻¹ ($P < 0.001$, Figure 3C). Similarly, root surface area also varied greatly among 43 soybean genotypes, ranging from 1.8×10^{-2} m² plant⁻¹ in Xichuanjiwohuang to 4.4×10^{-2} m² plant⁻¹ in Heidou ($P < 0.001$, Figure 3D).

There was a large variation among the 43 genotypes in root physiological traits including the total amount of carboxylates and acid phosphatase activity ($P < 0.001$ for both) (Figures 3E,F). We found a 35-fold difference in the activity of acid phosphatase in the rhizosheath among 43 genotypes, ranging from 0.19 nkat g⁻¹ soil DW in Baiqidawandou to 6.63 nkat g⁻¹ soil DW in Xinyangyangyandou ($P < 0.001$, Figure 3E). Similarly, the amount of carboxylates in the rhizosheath soil relative to root dry weight also differed seven-fold among genotypes, ranging from 14.9 μmol g⁻¹ root DW in Pixianalayanguang to 110.8 μmol g⁻¹ root DW in Datunxiaohaidou ($P < 0.001$, Figure 3F). The composition of carboxylates showed a large variation among 43 genotypes ($P < 0.001$, Figure 3G). All 43 genotypes showed



a predominant combination of malonate and malate in the rhizosphere soil, whereas other carboxylates including fumarate and *trans*-aconitate only accounted for 0.25–67.5% of total carboxylates (Supplementary Figure 2). Malonate accounted for the largest proportion of the total carboxylates in the rhizosphere soil (19–85%, $P < 0.001$) (Supplementary Figure 2). There was a large variation in the proportion of malate (10–43%, $P < 0.001$) and succinate (0–33%, $P < 0.001$, Supplementary Figure 2).

Correlations Among Traits

Based on nine plant traits of 43 genotypes, PCA explained 74% of the variation in the first, second, and third components (Figures 4A,B and Supplementary Table 3). The first component (PC1) represented 44% of the variance and primarily comprised root DW, total root length and root surface area. The second component represented 16% of variability and accounted mainly for shoot DW, root: shoot ratio, acid phosphatase activity and rhizosphere soil DW. The third component (PC3) represented 14% of variability, and primarily accounted for seed phytate-P concentration, acid phosphatase activity and shoot DW (Figures 4A–C and Supplementary Table 3).

Seed phytate-P concentrations showed no significant correlation with either root size, root morphological traits, physiological traits, or shoot DW (Table 1). Root DW had a significant positive correlation with all plant traits, except acid phosphatase activity, whereas acid phosphatase activity was only significantly correlated with rhizosphere soil DW. Total root

length, root surface area and the total amount of carboxylates showed a significant positive correlation with all other traits, except seed phytate-P concentration and acid phosphatase activity (Table 1).

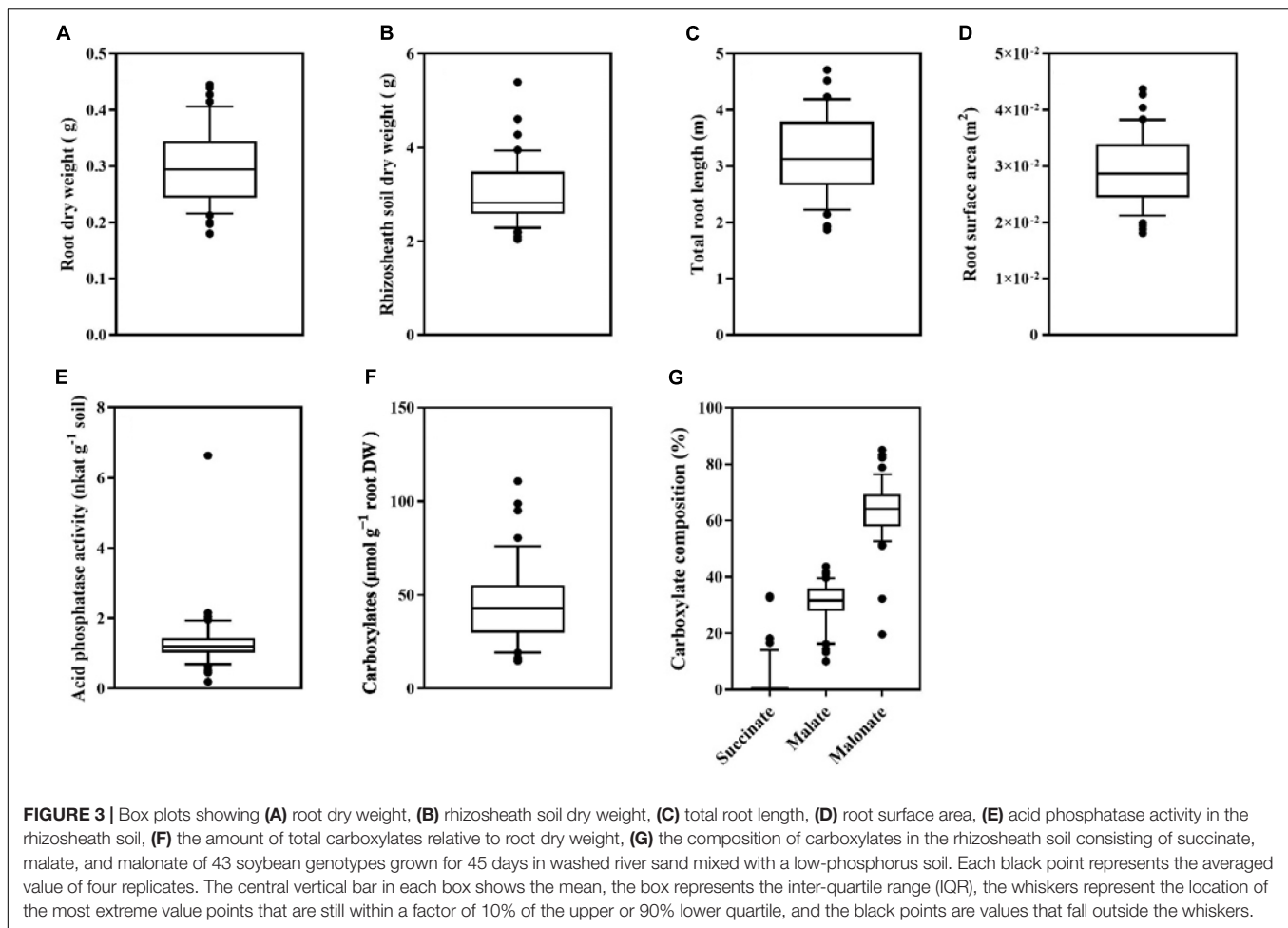
Classification of Soybean Genotypes

There were 19 (out of 43) soybean genotypes that had a lower seed phytate-P concentration and at least one desirable P-acquisition trait (Table 2). The soybean genotypes showed a large variation with respect to each trait (Supplementary Figure 3 and Table 2). The maximum and minimum comprehensive score was gained by genotype Diliuhuangdou-2, Siliyuan and Jidou no 7 (5 out of 6), and genotype Lvcaodou, Zheng84240-B1, Miyangxiaozihuang, Binhaidahuangkezijia, Baiqidawandou and Zheng8516 (1 out of 6), respectively. Three soybean genotypes showed a higher score (5 out of 6) of these, two genotypes had seed phytate-P concentration below $1.53 \text{ mg g}^{-1} \text{ DW}$ including Siliyuan and Diliuhuangdou-2. The high-score genotypes all exhibited efficient root morphological traits (Table 2).

DISCUSSION

A vital first step for improving trait in breeding programs is to explore the genetic variation in germplasm (Perera et al., 2019). The soybean core collections with different sample size were established based on the genetic diversity existed among its originated country of China (Wang et al., 2006; Qiu et al., 2013), in which the mini core collection had the rich diversity with the least sample size (Song et al., 2010). Thus, they had priority for discovering new traits or genes. The present study shows a large variation in seed phytate-P concentration among 256 soybean genotypes (Figure 1) because almost all mini core collection was included. Since the present genotypes were grown at the same time in the same field, we minimized environmental variables, growing location, irrigation condition, fertilizer applications soil type, and planting time (Boehm et al., 2017). Therefore, variation in seed phytate-P concentration in the present study can be primarily attributed to the broad genotypic variability of 256 soybean genotypes. This agrees with results of Raboy et al. (1984), who found extensive variation in soybean seed phytate-P concentration among 38 genotypes that were grown at the same time in the same field. Seed phytate P concentration ranged from 13.9 to $23.0 \text{ mg g}^{-1} \text{ DW}$ with a mean of $17.6 \text{ mg g}^{-1} \text{ DW}$. Likewise, Horner et al. (2005) found a large variation in phytate-P concentration among 86 soybean genotypes that were grown at the same time at three locations, ranging from 7.7 to $22.2 \text{ mg g}^{-1} \text{ DW}$ with a mean of $14.5 \text{ mg g}^{-1} \text{ DW}$. These results provide compelling evidence that genotypic variation in phytate-P concentration exists within soybean germplasm, despite a previous study showing that soybean has higher concentrations of seed phytate-P compared with our study (Wilcox et al., 2000).

Importantly, we found that more than half of the studied genotypes had a lower seed phytate P concentration than that of Gm-lpa-ZC-2 (range from 5.6 to $9.7 \text{ mg g}^{-1} \text{ DW}$) and Gm-lpa-TW75-1 (range from 3.0 to $6.4 \text{ mg g}^{-1} \text{ DW}$), although both Gm-lpa-ZC-2 and Gm-lpa-TW75-1 were previously identified as



low-phytate soybean mutants (Yuan et al., 2007, 2009). This offers the exciting possibility of identifying genotypes with even lower phytate-P concentrations than known before. This information is highly valuable for breeding and identifying soybean genotypes with low phytate-P concentrations, because seed phytate-P not only reduces the availability of micro-nutrients, especially Zn, Fe, and Cu, for humans and livestock, contributing to malnutrition, but also cannot be efficiently utilized by humans and non-ruminant animals, contributing to high losses of P to the environment (Raboy, 2001).

As we hypothesized, we found that under low-P conditions, root size (root dry weight), rhizosphere soil dry weight, morphological (total root length and root surface area) and physiological (carboxylate exudation and acid phosphate activity in rhizosphere soil) traits showed significant variations among 43 soybean genotypes (Figure 3), confirming a large genotypic variability for these traits. Root traits are critically important by determining soil exploration and therefore nutrient acquisition (Lynch, 2007). They affect a plant's acquisition of P including through the release of root exudates such as carboxylates and enzymes that alter the rhizosphere soil properties and increase the amount of available nutrients, distribution of roots, and morphological characteristic of the root system such as

root surface area and specific root length. All of these traits influence the soil volume that is explored by the root system for acquisition of nutrients (Lambers et al., 2006; Richardson et al., 2011; White et al., 2013; Krishnapriya and Pandey, 2016; Zhou et al., 2016; Lynch, 2019; Wang et al., 2019; White, 2019). Therefore, this information provides a scientific basis for breeding and identifying soybean genotypes with greater P-acquisition efficiency.

The absence of correlations between desirable root traits and seed phytate-P concentration gives valuable insight for selection in breeding programs. In this study, principal component analysis explained 74% of variation in the first, second, and third components (Figures 4A–C and Supplementary Table 3). The first (PC1) and second (PC2) components primarily comprised all of the root traits measured, whereas the third component (PC3) comprised seed phytate-P concentration. Our study shows that seed phytate-P concentration had no significant correlation with any root trait measured, while there was a significant correlation among all of these root traits measured (Table 1 and Supplementary Figure 3). This suggests that there is potential to breed low seed phytate-P soybean genotypes without compromising desirable P-acquisition traits. Classifying the genotypes based on the seed phytate-P concentration and

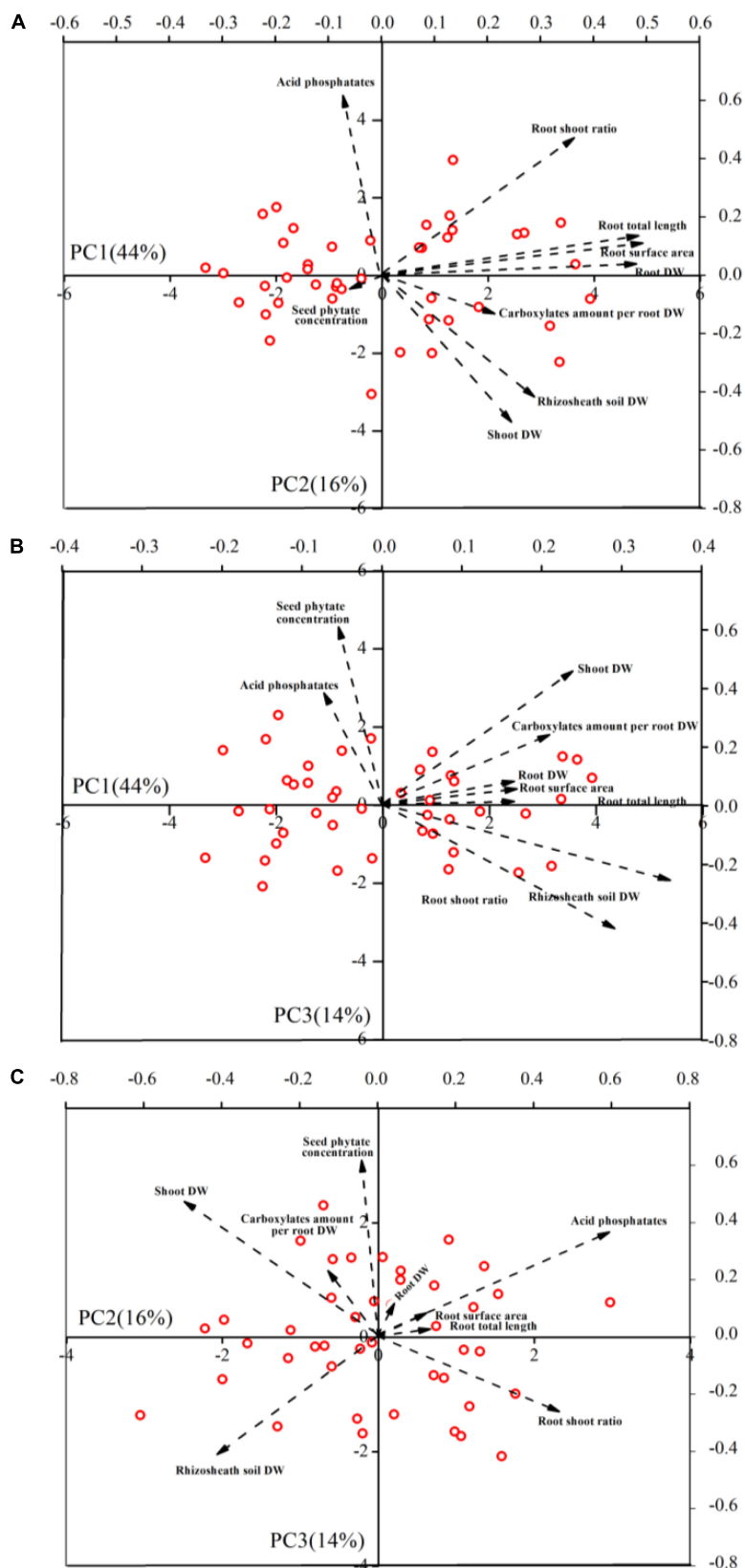


FIGURE 4 | Principal component analyses of 9 plant traits for 43 soybean genotypes. Biplot vectors are trait factor loadings, whereas the position of each genotype is shown. Panels refers to PC1 vs. PC2 (A), PC1 vs. PC3 (B), and PC2 vs. PC3 (C), respectively.

root traits is a useful tool in selecting potential genotypes to be used in biofortification of well-adapted genotypes with enhanced P acquisition. In the present study, we identified 19 soybean genotypes belonging to a low-phytate group with desirable root traits based on the plant performance score (Table 2 and Supplementary Figure 3), while some genotypes only had a low score (1 out of 6). The low-phytate soybean genotypes with desirable root traits identified in the present study are recommended to be used as genetic resource for breeding programs, whereas the low-phytate soybean genotypes with undesirable traits could be used as experimental materials in future genomic studies on low seed phytate. Although there were three genotypes that had a higher score (5 out of 6), we selected two soybean genotypes including Siliyuan and Diliuhuangdou-2 with both highly desirable P-acquisition traits

and low phytate-P concentrations (Table 2), because these had far lower seed phytate-P concentration than the other two genotypes. In addition, seed phytate-P concentrations of these genotypes were lower than that of Gm-lpa-ZC-2 and Gm-lpa-TW75-1, which were previously identified as low phytate-P soybean mutants (Yuan et al., 2007, 2009), suggesting the possibility of identifying and selecting further low-phytate soybean genotypes with highly desirable P-acquisition traits. Importantly, these two genotypes have crop yields of more than 3 t ha⁻¹ (unpublished), holding promise for application in agriculture. Further work on agronomic and physiological traits as well as seed quality parameters of the low-phytate P genotypes is needed. For example, it is important to examine whether seed P content of the low-phytate P genotype is sufficient for seed establishment.

TABLE 1 | Pearson's correlation matrix for 9 plant traits in 43 soybean genotypes.

	Phytate-P	Root DW	Shoot DW	RhiS DW	Root/shoot ratio	TRL	RSA	APase
Root DW	-0.04							
Shoot DW	0.08	0.54***						
RhiS DW	-0.25	0.40**	0.36**					
Root/shoot ratio	-0.11	0.71***	-0.21	0.16				
TRL	-0.05	0.83***	0.34*	0.47**	0.68***			
RSA	-0.04	0.90***	0.41**	0.43**	0.70***	0.97***		
APase	0.10	-0.12	-0.11	-0.38**	-0.06	-0.06	-0.06	
Carb_root	-0.10	0.31*	0.28	0.21	0.11	0.30	0.33*	-0.01

Root DW, root dry weight; Shoot DW, shoot dry weight; RhiS DW, rhizosphere soil DW; TRL, total root length; RSA, root surface area; APase, acid phosphatase activity in the rhizosphere soil; Carb_root, the amount of total carboxylates relative to root dry weight. Significant correlations are shown in bold; *P < 0.05; **P < 0.01; ***P < 0.001.

TABLE 2 | Comprehensive plant performance scores of low phytate-P soybean genotypes.

Genotype name	Phytate-P Conc mg g ⁻¹	Root size		Root morphological trait		Root physiological trait	
		Root DW	RhiS DW	TRL	RSA	APase	Carb_root
Lvcaodou	0.69					✓	
Siliyuan	0.95	✓	✓	✓	✓		✓
Pingdinghei	1.01		✓	✓	✓	✓	
Huaheihu	1.07	✓		✓	✓		
Gaozuoxuan no. 1	1.16	✓	✓	✓	✓		
Diliuhuangdou-2	1.53	✓	✓	✓	✓	✓	
Pixiannianzhuangliuyuxian	2.09		✓			✓	✓
Zhechengxiaohuangdou	2.11					✓	✓
Jidou no. 7	2.27	✓	✓	✓	✓		✓
Chadou	2.28	✓		✓	✓		
Bo'aiahongpizaojiaozi	2.48					✓	✓
Tongshanqingdadou	2.49	✓	✓		✓		✓
Miyangxiaozihuang	2.50					✓	
Zheng8516	2.55		✓				
Pixiansilicao	2.56	✓		✓	✓	✓	
Yangtianxiaohuangdou	2.58		✓	✓	✓		✓
Binhaidahuangkezijia	2.77					✓	
Miyangniumaohuang	2.79	✓	✓				✓
Baiqidawandou	2.87		✓				

Root DW, root dry weight; RhiS DW, rhizosphere soil dry weight; TRL, total root length; RSA, root surface area; APase, acid phosphate activity in the rhizosphere soil; Carb_root, the total amount of total carboxylates relative to root DW. Ticks mean that genotypes have lower seed phytate-P concentration than the mean seed phytate-P concentration of 43 soybean genotypes, and higher amount or larger root traits than the average value. Genotype with bold is the recommended genotypes.

In summary, selecting low phytate-P soybean genotypes with desirable P-acquisition traits will tighten the P cycle in crop production systems, animal production systems as well as human consumption, and lead to reduction of P-fertilizer cost, increase of nutritional quality of soymeal, and decrease of P-related environmental pollution.

CONCLUSION

We show substantial genotypic variation in seed phytate-P concentration and a range of root traits related to P acquisition. Some of the present genotypes showed even lower seed phytate-P concentrations than mutants that were considered the best available in terms of low seed phytate-P. Most importantly, seed phytate-P concentration was not correlated with any desirable root traits measured in this study.

DATA AVAILABILITY STATEMENT

The generated datasets available by request to the corresponding author.

AUTHOR CONTRIBUTIONS

MK, HL, JP, and W-FC designed the study. MK and WJ performed the experiments and collected the data. MK analyzed the data. MK, HL, JJ, L-JQ, JP, WD, W-FC, and FZ interpreted the data and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported financially by the Program of Advanced Discipline Construction in Beijing [Agriculture Green

Development, Chinese Universities Scientific Fund (Green, high quality and efficient crop production technology, and application model innovation project 2019TC121)], the Program of Introducing Talents of Discipline to Universities (Plant-soil interactions innovative research platform 1031-00100701), and funds provided by the Deputy Vice Chancellor Research of the University of Western Australia to HL. In addition, we thank Yongzhe Gu for providing soybean seeds as well as related information which were funded by Ministry of Science and Technology, Platform of National Crop Germplasm Resources of China (NICGR-2019-04), Ministry of Agriculture and Rural Affairs: Crop Germplasm Resources Protection Platform (2019NWB036-05).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.574547/full#supplementary-material>

Supplementary Figure 1 | Spatial distribution characteristics of phytate phosphorus (P) concentrations of 256 soybean genotypes. Abbreviations used along the x-axis refer to the provinces, spelled out toward the right of the figure panel.

Supplementary Figure 2 | The proportion of carboxylates in the rhizosphere soil consisting of malate, citrate, fumarate, succinate, and *trans*-aconitate of 43 soybean genotypes.

Supplementary Figure 3 | Plant performance scores of 43 soybean genotypes grown in washed river sand with a low supply of phosphorus (P).

Supplementary Table 1 | Detailed information on the 256 soybean genotypes used in the present study to measure seed phytate phosphorus concentration, and to study six root traits, associated with phosphorus acquisition.

Supplementary Table 2 | Descriptive statistics of phytate-phosphorus (P) concentration of 256 soybean genotypes commonly cultivated in China.

Supplementary Table 3 | Variable scores of nine plants traits in 43 soybean genotypes and proportion of variation of each component.

REFERENCES

- Alvey, S., Bagayoko, M., Neumann, G., and Buerkert, A. (2001). Cereal/legume rotations affect chemical properties and biological activities in two West African soils. *Plant Soil* 231, 45–54. doi: 10.1023/A:1010386800937
- Ao, J., Fu, J., Tian, J., Yan, X., and Liao, H. (2010). Genetic variability for root morph-architecture traits and root growth dynamics as related to phosphorus efficiency in soybean. *Funct. Plant Biol.* 37, 304–312. doi: 10.1071/FP09215
- Bilal, H. M., Aziz, T., Maqsood, M. A., Farooq, M., and Yan, G. J. (2018). Categorization of wheat genotypes for phosphorus efficiency. *PLoS One* 13:e0205471. doi: 10.1371/journal.pone.0205471
- Boehm, J. D., Walker, F. R., Bhandari, H. S., Kopsell, D., and Pantalone, V. R. (2017). Seed inorganic phosphorus stability and agronomic performance of two low-phytate soybean lines evaluated across six southeastern US environments. *Crop Sci.* 57, 2555–2563. doi: 10.2135/cropsci2017.02.0107
- Brinch, P. H., Sorensen, L. D., and Holm, P. B. (2002). Engineering crop plants: getting a handle on phosphate. *Trend Plant Sci.* 7, 118–125. doi: 10.1016/S1360-1385(01)02222-1
- Cong, W.-F., Hoffland, E., Li, L., Six, J., Sun, J., Bao, X., et al. (2015). Intercropping enhances soil carbon and nitrogen. *Global Change Biol.* 21, 1715–1726. doi: 10.1111/gcb.12738
- Cong, W.-F., Suriyagoda, L., and Lambers, H. (2020). Tightening the phosphorus cycle through phosphorus-efficient crop genotypes. *Trend Plant Sci.* 10, 967–975. doi: 10.1016/j.tplants.2020.04.013
- Cordell, D., Drangert, J. O., and White, S. (2009). The story of phosphorus: global food security and food for thought. *Global Environ. Change* 19, 292–305. doi: 10.1016/j.gloenvcha.2008.10.009
- Dinkelaker, B., Römhild, V., and Marschner, H. (1989). Citric-acid excretion and precipitation of calcium citrate in the rhizosphere of White Lupin (*Lupinus albus* L.). *Plant Cell Environ.* 12, 285–292. doi: 10.1111/j.1365-3040.1989.tb01942.x
- Fernandez, M. C., and Rubio, G. (2015). Root morphological traits related to phosphorus-uptake efficiency of soybean, sunflower, and maize. *J. Plant Nutr. Soil Sci.* 178, 807–815. doi: 10.1002/jpln.201500155
- Gao, X., Wu, M., Xu, R., Wang, X., Pan, R., Kim, H., et al. (2014). Root interactions in a maize/soybean intercropping system control soybean soil-borne disease, red crown rot. *PLoS One* 9:e95031. doi: 10.1371/journal.pone.0095031
- Ghaffar, S., Stevenson, R. J., and Khan, Z. (2017). Effect of phosphorus stress on *Microcystis aeruginosa* growth and phosphorus uptake. *PLoS One* 12:e0174349. doi: 10.1371/journal.pone.0174349
- Hinsinger, P. (2001). Bioavailability of soil inorganic P in the rhizosphere as affected by root-induced chemical changes: a review. *Plant Soil* 237, 173–195. doi: 10.1023/A:1013351617532

- Holford, I. C. R. (1997). Soil phosphorus: its measurement, and its uptake by plants. *Aus. J. Soil Res.* 35, 227–239. doi: 10.1071/S96047
- Horner, H. T., Cervantes, M. T., Healy, R., Reddy, M. B., Deardorff, B. L., Bailey, T. B., et al. (2005). Oxalate and phytate concentrations in seeds of soybean genotypes [*Glycine max* (L.) Merr.]. *J. Agric. Food Chem.* 53, 7870–7877. doi: 10.1021/jf051193i
- Krishnapriya, V., and Pandey, R. (2016). Root exudation index: screening organic acid exudation and phosphorus acquisition efficiency in soybean genotypes. *Crop Past. Sci.* 67, 1096–1109. doi: 10.1071/CP15329
- Lambers, H., Shane, M. W., Cramer, M. D., Pearce, S. J., and Veneklaas, E. J. (2006). Root structure and functioning for efficient acquisition of phosphorus: matching morphological and physiological traits. *Ann. Bot.* 98, 693–713. doi: 10.1093/aob/mcl114
- Liang, Q., Cheng, X., Mei, M., Yan, X., and Liao, H. (2010). QTL analysis of root traits as related to phosphorus efficiency in soybean. *Ann. Bot.* 106, 223–234. doi: 10.1093/aob/mcq097
- Lynch, J. P. (2007). Roots of the second green revolution. *Aus. J. Bot.* 55, 493–512. doi: 10.1071/BT06118
- Lynch, J. P. (2019). Root phenotypes for improved nutrient capture: an underexploited opportunity for global agriculture. *New Phytol.* 223, 548–564. doi: 10.1111/nph.15738
- Lyu, Y., Tang, H. L., Li, H., Zhang, F., Rengel, Z., Whalley, W. R., et al. (2016). Major crop species show differential balance between root morphological and physiological responses to variable phosphorus supply. *Front. Plant Sci.* 7:1939. doi: 10.3389/fpls.2016.01939
- Maharjan, P., Penny, J., Partington, D. L., and Panozzo, J. F. (2019). Genotype and environment effects on the chemical composition and rheological properties of field peas. *J. Sci. Food Agric.* 99, 5409–5416. doi: 10.1002/jsfa.9801
- Marschner, H. (1998). Role of root growth, arbuscular mycorrhiza, and root exudates for the efficiency in nutrient acquisition. *Field Crops Res.* 56, 203–207. doi: 10.1016/S0378-4290(97)00131-7
- Nuruzzaman, M., Lambers, H., Bolland, M. D. A., and Veneklaas, E. J. (2006). Distribution of carboxylates and acid phosphatase and depletion of different phosphorus fractions in the rhizosphere of a cereal and three grain legumes. *Plant Soil* 281, 109–120. doi: 10.1007/s11104-005-3936-2
- Pang, J., Ryan, M. H., Siddique, K. H. M., and Simpson, R. J. (2017). Unwrapping the rhizosheath. *Plant Soil* 418, 129–139. doi: 10.1007/s11104-017-3358-y
- Pang, J., Zhao, H., Bansal, R., Bohuon, E., Lambers, H., Ryan, M. H., et al. (2018). Leaf transpiration plays a role in phosphorus acquisition among a large set of chickpea genotypes. *Plant Cell Environ.* 41, 2069–2079. doi: 10.1111/pce.13139
- Pearce, S. J., Veneklaas, E. J., Cawthray, G., Bolland, M. D. A., and Lambers, H. (2007). Carboxylate composition of root exudates does not relate consistently to a crop species' ability to use phosphorus from aluminium, iron or calcium phosphate sources. *New Phytol.* 173, 181–190. doi: 10.1111/j.1469-8137.2006.01897.x
- Perera, I., Fukushima, A., Arai, M., Yamada, K., Nagasaka, S., Seneweera, S., et al. (2019). Identification of low phytic acid and high Zn bioavailable rice (*Oryza sativa* L.) from 69 accessions of the world rice core collection. *J. Cereal Sci.* 85, 206–213. doi: 10.1016/j.jcs.2018.12.010
- Qiu, L.-J., Xing, L., Guo, Y., Wang, J., Jackson, S. A., and Chang, R. (2013). A platform for soybean molecular breeding: the utilization of core collections for food security. *Plant Mol. Biol.* 83, 41–50. doi: 10.1008/S11103-013-0076-6
- Raboy, V. (2001). Seeds for a better future: 'low phytate' grains help to overcome malnutrition and reduce pollution. *Trends Plant Sci.* 6, 458–462. doi: 10.1016/S1360-1385(01)02104-5
- Raboy, V., Dickinson, D. B., and Below, F. E. (1984). Variation in seed total phosphorus, phytic acid, zinc, calcium, magnesium, and protein among lines of *Glycine max* and *Glycine soja*. *Crop Sci.* 24, 431–434. doi: 10.2135/cropsci1984.0011183X002400030001x
- Raghothama, K. G., and Karthikeyan, A. S. (2005). Phosphate acquisition. *Plant Soil* 274, 37–49. doi: 10.1007/s11104-004-2005-6
- Richardson, A. E., Lynch, J. P., Ryan, P. R., Delhaize, E., Smith, F. A., Smith, S. E., et al. (2011). Plant and microbial strategies to improve the phosphorus efficiency of agriculture. *Plant Soil* 349, 121–156. doi: 10.1007/s11104-011-0950-4
- Salvagiotti, F., Cassman, K. G., Specht, J. E., Walters, D. T., Weiss, A., and Dobermann, A. (2008). Nitrogen uptake, fixation and response to fertilizer N in soybeans: a review. *Field Crops Res.* 108, 1–13. doi: 10.1016/j.fcr.2008.03.001
- Sharpley, A., Foy, B., and Withers, P. (2000). Practical and innovative measures for the control of agricultural phosphorus losses to water: an overview. *J. Environ. Qual.* 29, 1–9. doi: 10.2134/jeq2000.00472425002900010001x
- Shen, J., Rengel, Z., Tang, C., and Zhang, F. (2003). Role of phosphorus nutrition in development of cluster roots and release of carboxylates in soil-grown *Lupinus albus*. *Plant Soil* 248, 199–206. doi: 10.1023/A:1022375229625
- Song, X., Li, Y., Chang, R., Guo, P., and Qiu, L.-J. (2010). Population structure and genetic diversity of mini core collection of cultivated soybean (*Glycine max*) in China. *Sci. Agric. Sin.* 43, 2209–2219. doi: 10.3864/j.issn.0578-1752.2010.11.00
- Wang, L., Guan, R., Li, Y., Lin, F., and Qiu, L.-J. (2007). Genetic diversity of Chinese spring soybean germplasm revealed by SSR markers. *Plant Breed.* 127, 56–61. doi: 10.1111/j.1439-0523.2007.01429.x
- Wang, L., Guan, Y., Guan, R., Li, Y., Ma, Y., Dong, Z., et al. (2006). Establishment of Chinese soybean (*Glycine max*) core collection with agronomic traits and SSR markers. *Euphytica* 151, 215–223. doi: 10.1007/s10681-006-9142-3
- Wang, W., Ding, G., White, P. J., Wang, X., Jin, K., Xu, F., et al. (2019). Mapping and cloning of quantitative trait loci for phosphorus efficiency in crops: opportunities and challenges. *Plant Soil* 439, 91–112. doi: 10.1007/s11104-018-3706-6
- White, P. J. (2019). Root traits benefitting crop production in environments with limited water and nutrient availability. *Ann. Bot.* 124, 883–890. doi: 10.1093/aob/mcz162
- White, P. J., George, T. S., Dupuy, L. X., Karley, A. J., and Valentine, T. (2013). Root traits for infertile soils. *Front. Plant Sci.* 4:193. doi: 10.3389/fpls.2013.00193
- Wilcox, J. R., Premachandra, G. S., Young, K. A., and Raboy, V. (2000). Isolation of high seed inorganic P, low-phytate soybean mutants. *Crop Sci.* 40, 1601–1605. doi: 10.2135/cropsci2000.4061601x
- Xia, H., Wang, Z., Zhao, J., Sun, J., Bao, X., Christie, P., et al. (2013). Contribution of interspecific interactions and phosphorus application to sustainable and productive intercropping systems. *Field Crops Res.* 154, 53–64. doi: 10.1016/j.fcr.2013.07.011
- Yuan, F., Zhao, H., Ren, X., Zhu, S., Fu, X., and Shu, Q. (2007). Generation and characterization of two novel low phytate mutations in soybean (*Glycine max* L. Merr.). *Theor. Appl. Genet.* 115, 945–957. doi: 10.1007/s00122-007-0621-2
- Yuan, F., Zhu, D., Deng, B., Fu, X., Dong, D., Zhu, S., et al. (2009). Effects of two low phytic acid mutations on seed quality and nutritional traits in soybean (*Glycine max* L. Merr.). *J. Agric. Food Chem.* 57, 3632–3638. doi: 10.1021/jf803862a
- Zhou, T., Du, Y., Ahmed, S., Liu, T., Ren, M., Liu, W., et al. (2016). Genotypic differences in phosphorus efficiency and the performance of physiological characteristics in response to low phosphorus stress of soybean in southwest of China. *Front. Plant Sci.* 7:1776. doi: 10.3389/fpls.2016.01776

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

Copyright © 2020 Kuerban, Jiao, Pang, Jing, Qiu, Ding, Cong, Zhang and Lambers. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Identification of a Locus Controlling Compound Raceme Inflorescence in Mungbean [*Vigna radiata* (L.) R. Wilczek]

Eunsoo Lee¹, Xuefei Yang^{1,2}, Jungmin Ha³, Moon Young Kim^{1,2}, Keum Yong Park^{1,2} and Suk-Ha Lee^{1,2*}

¹ Department of Agriculture, Forestry and Bioresources and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea, ² Plant Genomics and Breeding Institute, Seoul National University, Seoul, South Korea, ³ Department of Plant Science, Gangneung-Wonju National University, Gangneung, South Korea

OPEN ACCESS

Edited by:

Peerasak Srinives,
Kasetsart University, Thailand

Reviewed by:

Norihiko Tomooka,
National Agriculture and Food
Research Organization, Japan
Prakit Somta,
Kasetsart University, Thailand
Lixia Wang,
Chinese Academy of Agricultural
Sciences, China

*Correspondence:

Suk-Ha Lee
sukhalee@snu.ac.kr

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 16 December 2020

Accepted: 16 February 2021

Published: 08 March 2021

Citation:

Lee E, Yang X, Ha J, Kim MY, Park KY
and Lee S-H (2021) Identification of a
Locus Controlling Compound Raceme
Inflorescence in Mungbean [*Vigna*
radiata (L.) R. Wilczek].
Front. Genet. 12:642518.
doi: 10.3389/fgene.2021.642518

Mungbean [*Vigna radiata* (L.) R. Wilczek] produces a compound raceme inflorescence that branches into secondary inflorescences, which produce flowers. This architecture results in the less-domesticated traits of asynchronous pod maturity and multiple harvest times. This study identified the genetic factors responsible for the compound raceme of mungbean, providing a unique biological opportunity to improve simultaneous flowering. Using a recombinant inbred line (RIL) population derived from VC1973A, an elite cultivar with a compound raceme type, and IT208075, a natural mutant with a simple raceme type, a single locus that determined the inflorescence type was identified based on 1:1 segregation ratio in the F₈ generation, and designated *Comraceme*. Linkage map analysis showed *Comraceme* was located on chromosome 4 within a marker interval spanning 520 kb and containing 64 genes. RILs carrying heterozygous fragments around *Comraceme* produced compound racemes, indicating this form was dominant to the simple raceme type. Quantitative trait loci related to plant architecture and inflorescence have been identified in genomic regions of soybean syntenic to *Comraceme*. In IT208075, 15 genes were present as distinct variants not observed in other landrace varieties or wild mungbean. These genes included *Vradi04g00002481*, a development-related gene encoding a B3 transcriptional factor. The upstream region of *Vradi04g00002481* differed between lines producing the simple and compound types of raceme. Expression of *Vradi04g00002481* was significantly lower at the early vegetative stage and higher at the early reproductive stage, in IT208075 than in VC1973A. *Vradi04g00002481* was therefore likely to determine inflorescence type in mungbean. Although further study is required to determine the functional mechanism, this finding provides valuable genetic information for understanding the architecture of the compound raceme in mungbean.

Keywords: B3 transcription factor, compound raceme, inflorescence, legume, mungbean, synchronous maturity

INTRODUCTION

Mungbean [*Vigna radiata* (L.) R. Wilczek] is a diploid ($2n = 2x = 22$) warm-season legume that belongs to the tribe Phaseoleae within the subfamily of Papilionoideae and is cultivated widely across Asia (Shanmugasundaram et al., 2009). Mungbean is not only a highly nutritious crop, due to its high concentration of folate and iron, but also highly efficient during cultivation, when factors such as self-pollination, a short life-cycle, and nitrogen fixation are considered (Kim et al., 2015). Despite the economic importance of mungbean, several agronomical traits, including early flowering, synchronous pod maturity, yield, and suitability of plant type for mechanized harvest, remain to be improved. All these traits are regarded as part of a domestication syndrome. Linkage mapping of domestication-related traits of mungbean has been less intensive than in other leguminous crops including soybean (*Glycine max*) and chickpea (*Cicer arietinum*). Indeed, most mapping results reported in mungbean involve resistance to insects, for example, to bruchid beetles (Chen et al., 2013; Chotechung et al., 2016; Schafleitner et al., 2016; Kaewwongwal et al., 2017, 2020) or to diseases, including yellow mosaic virus (Chen et al., 2013; Kitsanachandee et al., 2013; Alam et al., 2014; Mathivathana et al., 2019), powdery mildew (Chankaew et al., 2013; Yundaeng et al., 2020) and *Cercospora* leaf spot (Chankaew et al., 2011; Yundaeng et al., 2021). Although quantitative trait loci (QTLs) associated with flowering and traits related to domestication have been identified, including 34 quantitative traits and 4 qualitative traits (Isemura et al., 2012; Hwang et al., 2017), studies of yield-related traits are still at an early stage as there is a lack of the background information required to develop varieties suited to modern cultivation.

Inflorescence architecture, which determines where the flowers form, determines the reproductive success and survival of plants, and is a result of evolution by natural selection (Prusinkiewicz et al., 2007); during domestication, inflorescence architecture is a key factor affecting crop yield and harvesting method. Inflorescence architecture is classified according to the types of branches produced (for e.g., raceme, panicle, and cyme) and the patterns of flowering (indeterminate and determinate inflorescence) (Benlloch et al., 2007). The raceme is an archetypal example of indeterminate inflorescence in which the axis elongates continuously through indefinite growth of the apical meristem and flowers are laterally generated along the axis. The raceme may be divided into simple and compound forms depending on whether or not the primary inflorescence meristem directly produces the flowers (Han et al., 2014). The *Arabidopsis thaliana* inflorescence is a simple raceme, as the flowers are borne on the main stem of the primary inflorescence (Ungerer et al., 2002; Han et al., 2014). Leguminous plants typically produce compound racemes on which secondary (I_2) or higher order inflorescences develop from the primary (I_1) inflorescence and bear flowers (Figure 1). As the axis of primary inflorescence is elongated, secondary inflorescences are continuously generated laterally and bear flowers in succession (Figure 1). In contrast to the simple raceme, individual inflorescences of the mungbean compound raceme contain flowers of many different ages, with the oldest at the bottom and younger ones above (Benlloch

et al., 2015). This compound raceme inflorescence architecture is shared by most cultivated mungbean varieties and their wild relative, *Vigna radiata* var. *sublobata* (IBPGR, 1985; Bernardo et al., 2018). The cultivated mungbean therefore retains more less-domesticated and agronomically undesirable traits associated with inflorescence development, such as asynchronous pod maturity and multiple harvesting times, than does the cultivated soybean. Unraveling the genetic factors that control raceme inflorescence development is therefore important, both for understanding the evolution of plant form (Benlloch et al., 2015) and improving key architectural traits in mungbean.

We report here a natural mungbean mutant that showed a simple raceme inflorescence (Korean gene bank accession no. IT208075). This accession was used as a parental line to develop a recombinant inbred line (RIL) population of VC1973A (compound raceme) $\times$ IT208075 (simple raceme). We investigated the inheritance pattern of the compound raceme inflorescence and used a single nucleotide polymorphism (SNP)-based genetic linkage map to identify a locus controlling raceme inflorescence type. In addition, we identified a candidate gene located in a locus associated with the compound raceme inflorescence.

MATERIALS AND METHODS

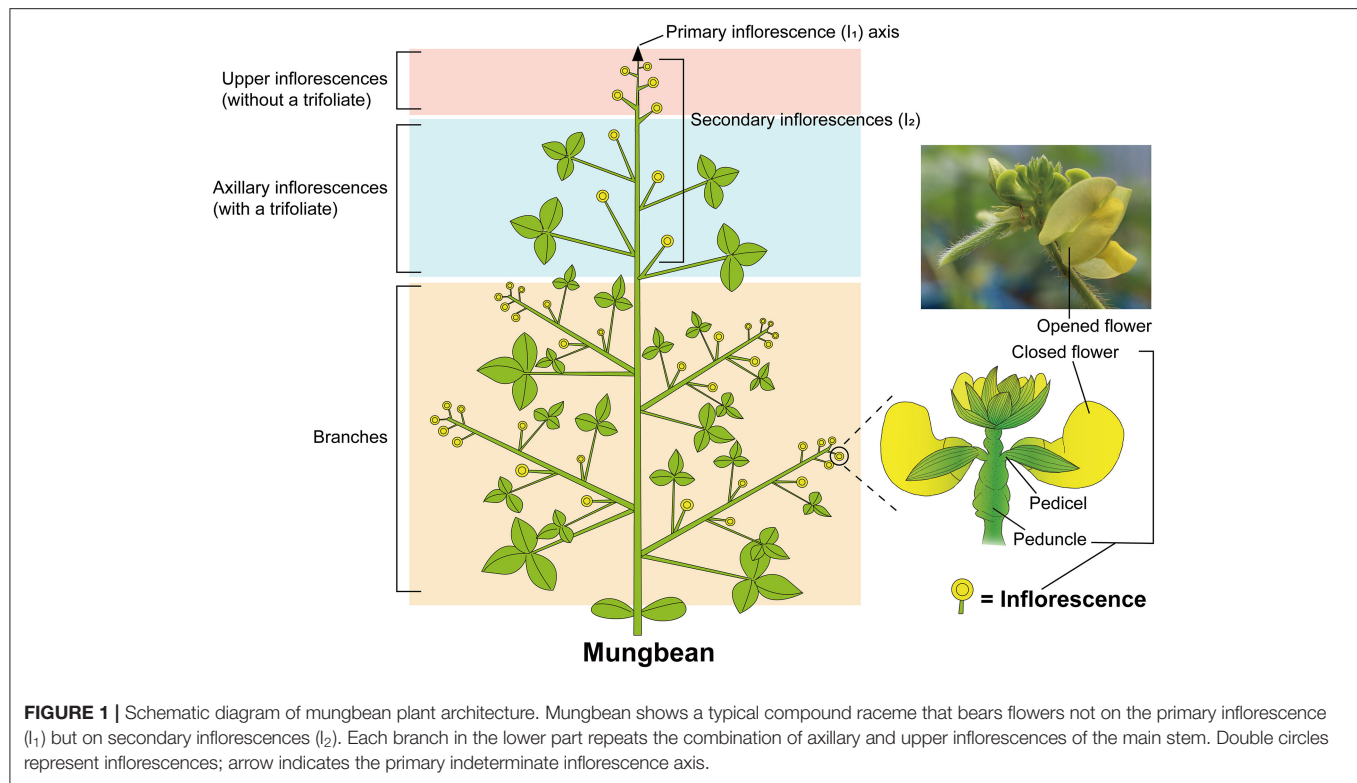
Plant Materials and Trait Phenotyping

A RIL population consisting of 235 F_7 lines was developed from a cross between VC1973A (compound raceme) and IT208075 (simple raceme) by advancing the F_2 to the F_6 generation using the single seed descent method. The maternal line, VC1973A, was an elite cultivar developed at the World Vegetable Center (previously known as the Asian Vegetable Research and Development Center, AVRDC) in 1982, which was used to build the mungbean reference genome (Kang et al., 2014). The paternal line, IT208075, was a local variety from Vietnam named as Binh khe D.X. and was obtained from the National Agrobiodiversity Center, Republic of Korea (<http://genebank.rda.go.kr/>).

The 235 RILs and the mapping parents, VC1973A and IT208075, were planted in field trials over 2 growing seasons (planting dates: June 25, 2018 and June 20, 2019) at Seoul National University Experiment Farm ($37^{\circ}16'14.6''$ N, $126^{\circ}59'19.9''$ E). Individual plants were set out in rows with 15 cm intervals between each plant, and the distance between rows was set at 70 cm. The plants were grown under natural photoperiods of 11.5 to 14.5 h per day. The inflorescence type of each RIL and parental line was scored as a compound or simple raceme depending on the presence or absence of a secondary inflorescence on the upper part of the main stem during the pod maturity period; 3 replicates were scored for each line. The number of secondary inflorescences on the upper part without a trifoliate was recorded at the time of harvest and used as a quantitative trait.

DNA Extraction, Resequencing, and Genotyping-by-Sequencing

Genomic DNA was extracted from healthy young leaves of IT208075 and the 235 $F_{7:8}$ RILs using the GenoAll[®] Exgene[™]



Plant SV kit (GeneAll Biotechnology, Seoul, Korea). DNA quality was assessed by the 260/280 nm ratio using a Nanodrop 3000 spectrometer (Thermo Scientific, Wilmington, DE, USA). DNA was quantified using the Invitrogen Quant-iT PicoGreen[®] dsDNA Assay kit (Life Technologies, Burlington, ON, Canada), and its concentration was adjusted to 20 ng/ μ L. IT208075 was re-sequenced with 9 $\times$ sequencing depth on an Illumina HiSeq2000 platform (Illumina Inc., San Diego, CA, USA). The Illumina reads were deposited in the database of the National Center for Biotechnology Information (NCBI) under BioProject accession number PRJNA698712. The genotyping-by-sequencing (GBS) method was used to genotype the RILs (Elshire et al., 2011). After digestion with the restriction enzyme *ApeKI*, a GBS library was constructed following the procedure described previously (Hwang et al., 2017; Yoon et al., 2019); the barcode adapters used in this study are listed in **Supplementary Table 1**. Three separate libraries were constructed using pooling amplified DNA samples from 84 or 85 RILs for each library. Single-end sequencing of the GBS libraries was performed on 3 lanes of an Illumina HiSeq2000 instrument (Illumina Inc.).

Sequence Analysis, Genetic Map Construction, and QTL Analysis

Paired-end reads of IT208075 and GBS reads were mapped onto the mungbean reference genome (cv. VC1973A, PRJNA560716) using BWA v0.7.15 (Li, 2013) after trimming low quality sequence, barcode adapter, and *ApeKI* overhang sequences from raw reads. SNPs were called by SAMtools v1.3 (Li et al., 2009) and VCFtools (Danecek et al., 2011) with the following filtering criteria: mapping quality ≥ 30 , read depth ≥ 3 , heterozygosity

$\leq 10\%$, and missing genotypes $\leq 70\%$. We constructed two separate genetic linkage maps for qualitative/quantitative analysis using JoinMap v4.1 (Van Ooijen, 2006). To anchor the locus for the qualitative trait of inflorescence type to the genetic map, the phenotypic difference was used as a molecular marker for linkage grouping along with the identified SNPs, where the order of the markers was determined using a maximum likelihood (ML) method. To confirm the position of the marker associated with inflorescence type on the genetic map, an additional genetic map was constructed with only SNP markers using regression mapping methods for QTL analysis. In both linkage mappings, the Kosambi mapping function was employed to translate the recombination frequency into a map distance in centimorgans (cM). QTL analysis was performed using IciMapping v4.1 software (Meng et al., 2015) through the ICIM-ADD mapping method, and the LOD threshold was determined by 1,000 permutation tests at a significance level of 0.05. Allelic associations with phenotypic differences were explored using RILs mapped to the genomic region surrounding the locus associated with a compound raceme.

Comparative Analysis of the Locus for Compound Raceme

Soybean synteny blocks were identified by searching the soybean reference genome (Wm82.a2.v1, <https://phytozome.jgi.doe.gov/>) through MCScanX using BLASTp with default parameters to provide a comparative analysis of the 3 Mb genomic region on chromosome 4 surrounding the locus associated with a compound raceme (Vr04:26,001,525 to Vr04:28,997,867 bp). QTLs, markers, and gene information for soybean were obtained

from the SoyBase database (<http://www.soybase.org>). The plant transcription factor database PlantTFDB (<http://planttfdb.gao-lab.org/>) was used to obtain information about transcription factor genes (Jin et al., 2017). TAIR version 10 (<http://www.arabidopsis.org>) was searched using BLASTp to find orthologous genes of *A. thaliana*.

Identification of Sequence Variation in Genes Within the Compound Raceme Locus

We investigated sequence variation between VC1973A and IT208075 in genes located in the locus associated with the compound raceme to identify likely candidate genes responsible for this phenotype. The variant sequences were compared with previously reported sequences of two other mungbean lines producing compound racemes, TC1966 (wild mungbean, *Vigna radiata* var. *sublobata*) and V2984 (a Korean landrace), to identify sequence variation specific to IT208075, which has a simple raceme (NCBI accession code JJMO00000000, Kang et al., 2014). We used Sanger sequencing to validate the sequence variation in two promising genes, *VrDet1* and *Vradi04g00002481*. PCR products, amplified with sequence-specific primers (Supplementary Table 2), were purified from 1% agarose gels using the AccuPrep® Gel Purification Kit (Bioneer, Daejeon, Korea), cloned into *Escherichia coli* using the pGEM-T Easy Vector (Promega, Madison, WI, USA), according to the manufacturer's instructions, and sequenced using an ABI 3730XL DNA analyzer (Applied Biosystems, Foster City, CA, USA).

Motif Discovery and qRT-PCR Analysis of Selected Genes

To determine the domain conserved between mungbean and *A. thaliana* orthologs, sequences were aligned using the MEME suite motif analysis tool (Bailey et al., 2009). Differences in expression levels of selected genes between the mapping parents were compared using newly emerged shoot apical tissue (SAT) collected at each developmental stage. Unifoliate opening marked the first developmental stage. The SAT of each subsequent developmental stage was sampled when the second uppermost trifoliate leaflet of the node had fully opened; as VC1973A had 8 nodes and IT208075 had 4 nodes, they consisted of 9 and 5 developmental stages, respectively. Three biological samples were analyzed at each developmental stage, and one biological sample consisted of SAT from 5 individual plants exhibiting the same developmental stage. Total RNA was extracted from each sample using a Ribospin™ Plant kit (GeneAll Biotechnology, Seoul, Korea). cDNA was synthesized using a Bio-Rad iScript™ cDNA Synthesis Kit (Bio-Rad, Hercules, CA, USA). Primer sequences for qRT-PCR were designed using Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) and are listed in Supplementary Table 2. qRT-PCR was performed with 2 technical replicates using the Bio-Rad iQ™ SYBR Green Supermix Kit in a LightCycler® 480 (Roche Diagnostics, Laval, QC, Canada). The amplification conditions were 95°C for 5 min, followed by 40 cycles of 95°C for 10 s, 60°C for 15 s, and 72°C for 15 s. *eIF5A* (*Vradi05g00001056*),

which encodes eukaryotic initiation factor 5A, was used as a reference gene for normalizing expression levels of target genes, and relative expression of each gene was calculated according to the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). Statistical significance was determined using the Student's *t*-test.

RESULTS

A Simple Raceme Inflorescence Was Observed in IT208075

The maternal line, VC1973A, had a typical compound raceme. It showed indeterminate flowering, and thus an individual inflorescence produced different cohorts of flowers, with the oldest ones located at the bottom and younger ones developing consecutively toward the top (Figure 2A). Consequently, the pods matured at different times, and the same inflorescence simultaneously carried mature pods and new flowers. We identified a landrace variety, IT208075, which did not produce secondary inflorescences on the upper part of its main stem. Instead, a single primary inflorescence was transferred from the axis of the main stem and bore flowers in an acropetal succession (Figure 2B). This architecture is defined as a simple raceme inflorescence, as typified by the model plant *A. thaliana*.

Inheritance of the Compound Raceme Inflorescence

A phenotypic investigation was conducted in a RIL population derived from a cross between VC1973A and IT208075 to identify the pattern of genetic inheritance underlying the inflorescence architecture of mungbean. The mean numbers of upper inflorescences produced by VC1973A were 7.0 and 4.7 in 2018 and 2019, respectively, whereas IT208075 produced a single primary inflorescence (Table 1). Within the population of 235 F_{8:9} RILs, 122 produced a compound raceme, 99 produced a simple raceme, and 14 lines showed a segregating phenotype (Supplementary Table 3). A Chi-square test indicated that these observations were consistent with segregation according to a 1:1 Mendelian ratio ($X^2 = 2.4$, $P = 0.12$), suggesting that a single gene, designated *Comraceme*, controlled the compound raceme phenotype in VC1973A.

Resequencing and GBS Analysis

In total, 4.3 Gb raw data from 31 million reads were generated from IT208075 (Supplementary Table 4). After quality trimming, 23 million high quality reads were mapped against the mungbean reference genome, which resulted in 91.2% genome coverage and 7.7× mapping depth. To genotype the RIL population, a total of 672 million GBS reads of 235 RILs were mapped to the reference genome. The number of GBS reads mapped to the reference genome ranged from 1,202,609 to 6,565,325 (mean: 2,680,760 reads). The percentage coverage of the genome ranged from 1.3 to 3.4% (mean: 2.2%). A total of 4,177 SNPs was used to construct genetic maps.

Mapping of the *Comraceme* Locus

To locate *Comraceme*, the locus associated with the qualitative raceme trait, we constructed a genetic map containing

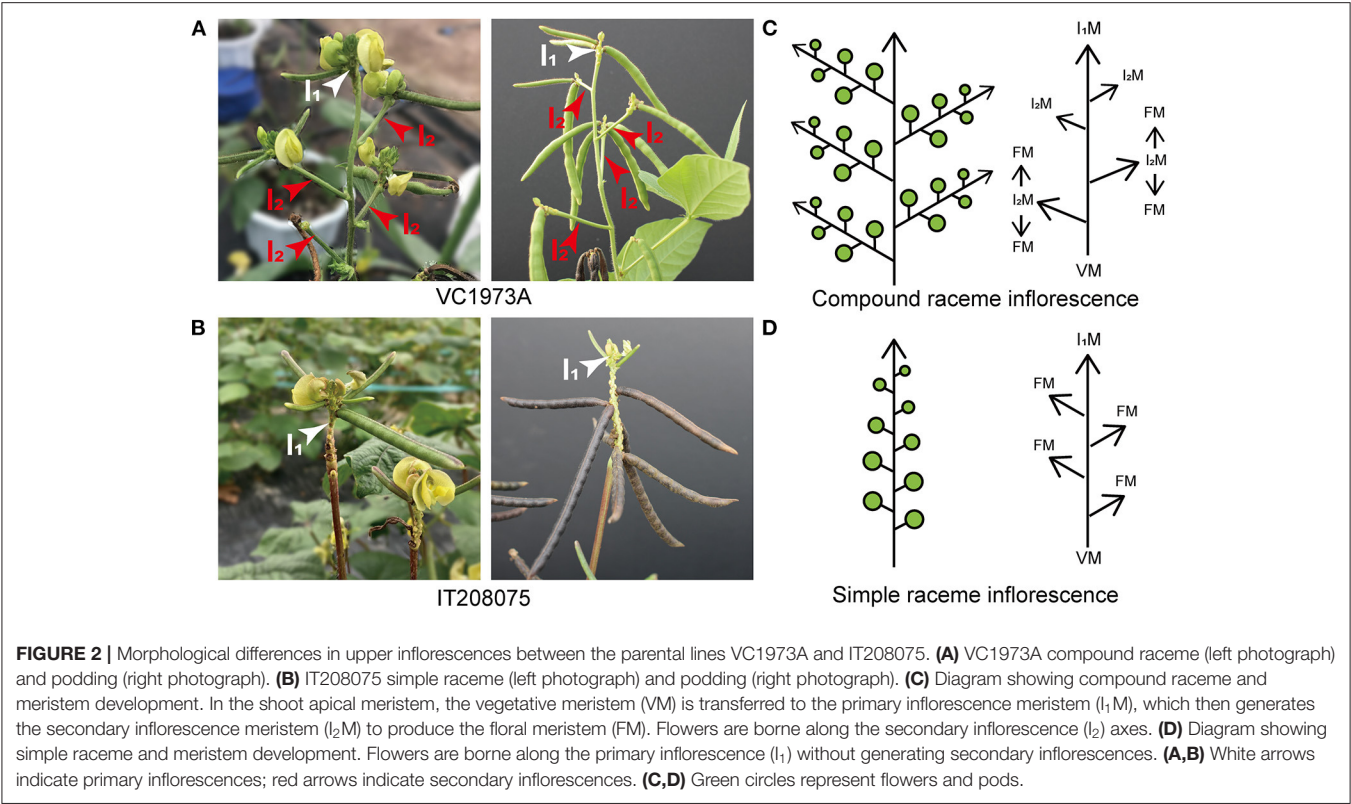


TABLE 1 | Phenotypic variation and heritability (H^2) of the number of upper inflorescences in VC1973A and IT208075 and their derived RIL population.

Year	Parental lines $\pm$ SD			RIL population				
	VC1973A	IT208075	<i>P</i> -value	Average	SD	Min	Max	H^{2a}
2018	7.0 $\pm$ 1.0	1.0	0.0091**	3.6	2.8	1.0	8.3	0.94
2019	4.7 $\pm$ 0.3	1.0	0.0081**	2.5	1.5	1.0	6.0	0.98

^aBroad-sense heritability.
**Significance at the 0.01 probability level; SD, standard deviation.

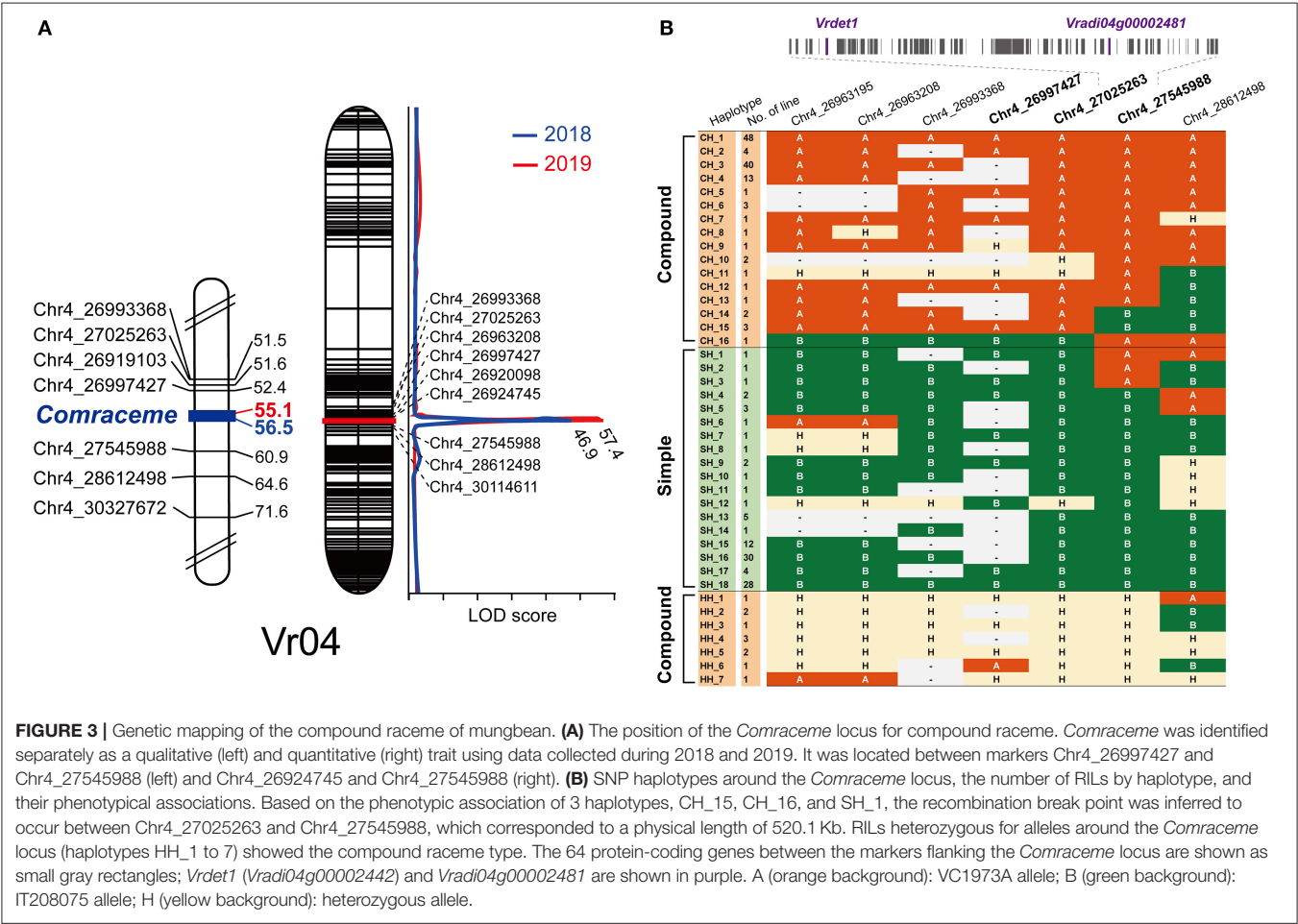
phenotypic data of compound/single raceme type as a molecular marker among the RIL population. Of the 4,177 SNP markers available, the final genetic map contained 1,799 SNPs over 11 chromosomes (Vr01 to Vr11) and spanned 1,353.5 cM with a mean marker interval of 0.82 cM by the ML method (Supplementary Table 5). The number of SNPs mapped to each chromosome ranged between 54 (Vr07) and 357 (Vr02) with a mean of 164. *Comraceme* was located in the interval between markers Chr4_26997427 and Chr4_27545988, which spanned 545.6 Kb of Vr04 (Table 2, Figure 3A). To validate *Comraceme*, composite interval mapping was conducted using the number of the upper inflorescences as the quantitative trait. RILs producing a simple raceme were recorded as having one inflorescence. The additional genetic map spanning 450.1 cM was constructed using the regression mapping method (Supplementary Figure 1, Supplementary Table 6). The mean interval between markers was 0.33 cM, although it ranged from 0.06 cM on Vr10 to 0.77 cM on Vr11 (Supplementary Table 5).

A major QTL was detected between Chr4_26924745 and Chr4_27545988 that exceeded the significance level on Vr04 for the number of upper inflorescences in both the 2018 and 2019 datasets. This QTL had LOD scores of 46.9 (2018 data) and 57.4 (2019 data), which explained 62.0 and 67.6% of the phenotypic variation, respectively (Table 2, Figure 3A). Both qualitative and quantitative analyses of the compound raceme inflorescence trait identified the marker Chr4_27545988 as a boundary of the interval containing the *Comraceme* locus (Table 2). We compared the phenotypes of RILs with different haplotypes around the *Comraceme* locus to delimit the position of *Comraceme*. The order of the marker Chr4_27025263 in the genetic maps differed from its physical position in the reference genome between Chr4_26997427 and Chr4_27545988, probably because of insufficient genotype data (Figures 3A,B). Although haplotypes CH_15 and CH_16 had different alleles at Chr4_27025263 and Chr4_27545988, both produced the compound raceme inflorescence phenotype. By contrast,

TABLE 2 | Genetic mapping for the compound raceme inflorescence in the RIL population derived from VC1973A × IT208075.

Phenotyping	Year	Chr	Pos ^a (cM)	Marker interval	LOD ^b	PVE ^c (%)	Add ^d	No. of genes ^e
Qualitative	2018	4	55.1	chr4_26997427:chr4_27545988				67
	2019	4	56.5	chr4_26997427:chr4_27545988				67
Quantitative	2018	4	15.0	chr4_26924745:chr4_27545988	46.9	62.0	−2.1	73
	2019	4	15.0	chr4_26924745:chr4_27545988	57.4	67.6	−1.2	73

^aMapped position of a qualitative trait as a marker in the genetic linkage map following ML mapping method, and genetic position of a QTL peak in the genetic linkage map constructed following regression mapping method.
^bMaximum-likelihood logarithm of odds (LOD) score for the individual QTL.
^cPercent of phenotypic variance explained by the QTL.
^dAllelic additive effect.
^eNumber of protein-coding genes within marker intervals.



haplotypes CH_16 and SH_1 had the same alleles at both marker positions but produced compound and simple raceme inflorescences, respectively. This suggested the recombination break occurred between Chr4_27025263 and Chr4_27545988, corresponding to a physical distance of 520.1 Kb. In addition, 11 RILs carrying a heterozygous fragment around the *Comraceme* locus produced a compound raceme, indicating that a single, dominant gene controlled *Comraceme* and was responsible for the compound raceme phenotype.

Syntenic Analysis of the *Comraceme* Locus and the Soybean Genome

As soybean is the closest model legume species to mungbean, we analyzed the syntenic relationship between the genomic region containing *Comraceme* and the soybean genome. A comparison between the 3 Mb mungbean genomic region containing the 520 Kb segment between Chr4_27025263 and Chr4_27545988 and the soybean genome identified 3 colinear blocks, located on the soybean chromosomes Gm03, Gm10,

and Gm 19. The syntenic regions located on Gm03 and Gm19 contained soybean QTLs associated with inflorescence-related agronomic traits, including plant height, branching, first flower, flower number, pod number, and seed yield (Lark et al., 1995; Mansur et al., 1996; Orf et al., 1999; Zhang et al., 2010; Kim et al., 2012; Kuroda et al., 2013; Shim et al., 2017). The soybean QTL *Plant height* 33–1 and *Seed yield* 27–4 were located on Gm03, whereas *Branching* 5–4, *First flower* 3–3 and 6–3, *Flower number* 1–10, *Plant height* 3–1 and 6–1, and *Pod number* 8–1 were all located on Gm19 (Figure 4). In addition, the syntenic blocks on Gm03 and Gm19 contained *GmTFL1a* and *GmDt1* (*GmTFL1b*), respectively; both are orthologs of *Arabidopsis* *TERMINAL FLOWER 1* (*TFL1*). *VrDet1*, the ortholog of *GmDt1*, was also located in the *Comraceme* locus. *GmDt1* and *VrDet1* modulate determinate stem growth habit in soybean and mungbean, respectively (Liu et al., 2010; Tian et al., 2010; Li et al., 2018). No QTL associated with inflorescence-related traits were located in the syntenic region of Gm10, although this region did contain *GmCEN2*, a homolog of *AtTFL1*.

Identification of a Gene Candidate for *Comraceme*

The *Comraceme* locus between Chr4_27025263 and Chr4_27545988 contained 64 protein-coding genes, 55 of which have functionally annotated *A. thaliana* orthologs (Supplementary Table 7). After comparing all the variant sequences found in the *Comraceme* genomic region with those known from *Vigna* varieties with a compound raceme (the landrace V2984 and the wild mungbean TC1966), we identified 15 genes with variants present only in the IT208075 paternal line, which produced a simple raceme inflorescence. Among these genes, 2 genes, *VrDet1* and *Vradi04g00002481* were likely to affect inflorescence type based on previous studies demonstrating their roles in growth and development (Supplementary Table 8). *VrDet1* modulates the stem growth habit of mungbean (Li et al., 2018). The first intron region of *VrDet1* showed variation of simple repeat sequences (SSRs) [(AC)_n(AT)_n] between lines: TC1966 had a 4 bp insertion of SSRs (“AC” and “AT”), IT208075 had a 2bp deletion (“AC”), whereas VC1973A and V2984 had the same sequence (Supplementary Table 8, Supplementary Figure 2A). The mapping parents also differed with respect to two insertion-deletion mutations (indels) in the upstream and second intron regions of *VrDet1* but these mutations were not detected in TC1966 and V2984 owing to unmapped gaps (Supplementary Figure 2A). *Vradi04g00002481* is a homolog of an *A. thaliana* gene, *ABNORMAL SHOOT 2* (*ABS2*, also named *NGATHA-Like 1*, AT2G36080), which causes pleiotropic growth defects (Shao et al., 2012). IT208075 showed specific sequence variation of two-base insertion (“TC”) in the upstream region of *Vradi04g00002481*, whereas TC1966, V2984, and VC1973A had the same sequence (Figure 5A). The predicted *Vradi04g00002481* protein consisted of 224 residues and included a B3 DNA-binding domain that showed a high similarity with the motif of *ABS2* (Figure 5B).

Expression Analysis During Mungbean Development

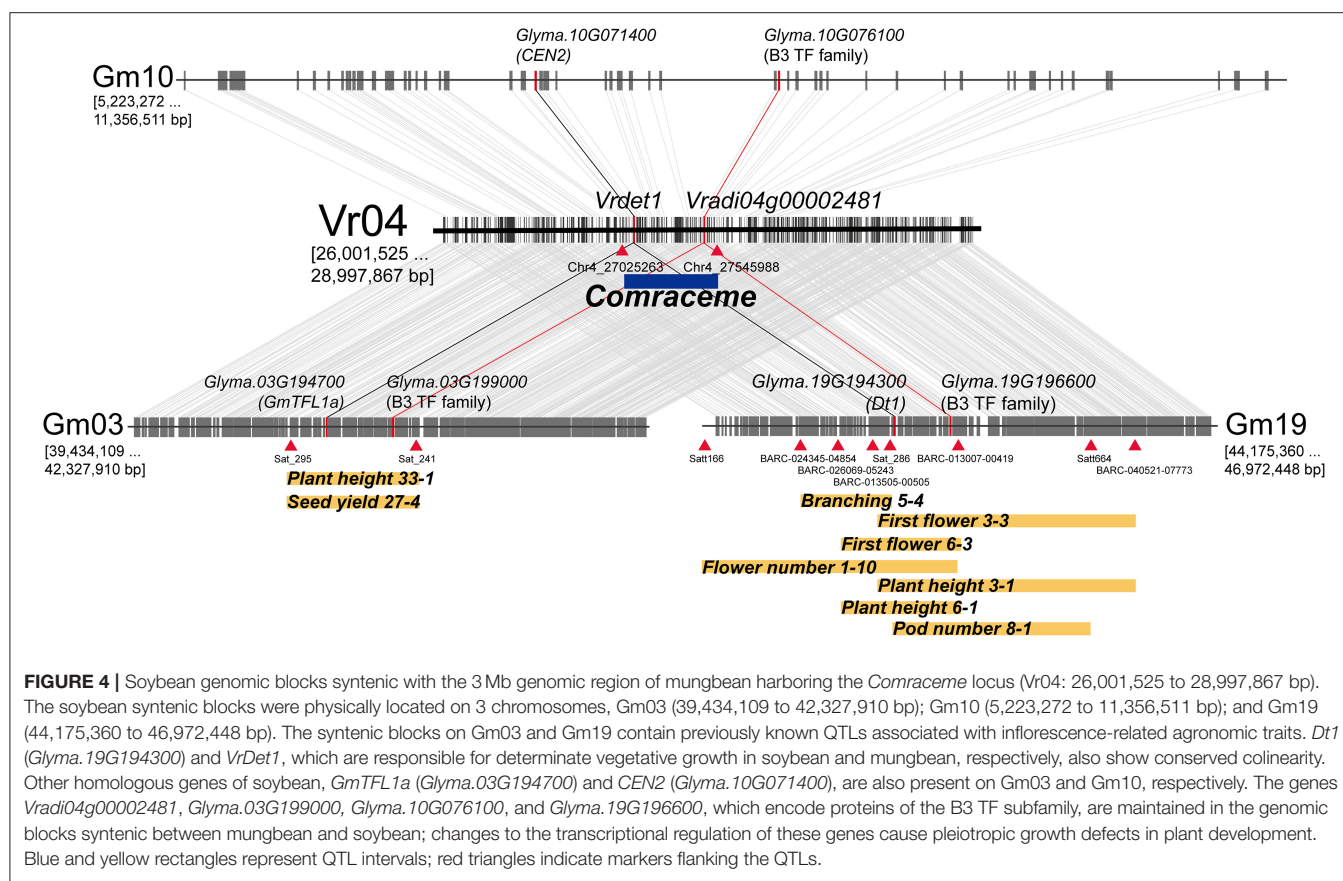
To determine whether transcription levels differed in the selected genes *VrDet1* and *Vradi04g00002481*, we used qRT-PCR to examine their expression level in the SAT at each developmental stage. We used the differences in plant architecture between the parental lines VC1973A and IT208075 to select 3 developmental stages in which the same organs were generated: the beginning stage when the trifoliate emerged (stage 1 for both accessions) after the unifoliate was fully opened; the terminal stage of the vegetative state when the final trifoliate emerged on the main stem (stage 8 for VC1973A and stage 4 for IT208075); and the early reproductive stage when the apical shoot was transformed to generate the peduncle for inflorescence (stage 9 for VC1973A and stage 5 for IT208075).

There were no significant differences in expression levels of *VrDet1* between the parental lines at any of the 3 developmental stages (Supplementary Figure 2B). By contrast, significant differences in the levels of *Vradi04g00002481* expression were observed: *Vradi04g00002481* expression in VC1973A was significantly higher than in IT208075 at the beginning stage (stage 1), but expression in IT208075 was significantly higher than in VC1973A at the other 2 stages (Figure 5C). These results indicated that transcription of *Vradi04g00002481* was more likely to affect the inflorescence architecture of mungbean than *VrDet1*.

DISCUSSION

The complex processes controlling inflorescence architecture formation significantly affect plant reproduction. A compound raceme inflorescence is defined by the presence of secondary inflorescences to produce flowers and is often observed in legume species in the Papilionoideae subfamily (Berbel et al., 2012; Han et al., 2014; Benlloch et al., 2015). Plants with a compound raceme generate peduncles from the axis of the main stem toward the direction of elongation; the peduncles become stalks, each supporting a raceme inflorescence and thus enabling constant blooming. This architecture, depicted in a simplified form in Figure 2C, is observed in the maternal line, VC1973A. The number of flowers per plant depends on the duration of activation of the apical meristem in the primary inflorescence (*I*₁M) as this controls the indeterminate initiation of the secondary inflorescence meristems (*I*₂M), which produce the floral meristems that generate flowers (Tucker, 2003; Benlloch et al., 2007). Plants with a simple raceme type, such as *Arabidopsis*, do not generate peduncles and produce flowers only from the axis of the main stem as the apical meristem develops into the floral meristem (Shannon and Meeks-Wagner, 1993; Bradley et al., 1997; Ungerer et al., 2002). This architecture is observed in the paternal line, IT208075, and is depicted in Figure 2D.

Relatively few studies of inflorescence architecture have been reported from mungbean compared with other major crops. Although inheritance of the inflorescence type, the number of clusters per node, and sterility have been investigated (Sen and Ghosh, 1959; Singh and Singh, 1970; Selvi et al., 2003; Mehendi



et al., 2019), the genetic basis underlying inflorescence formation remains unexplained. In the current study, the compound and simple inflorescence types appeared in a 1:1 ratio among the RIL population, as would be expected for a qualitative trait controlled by a single gene. To map the gene controlling the qualitative trait, we treated the phenotypic data of RILs as a molecular marker following the alleles of parental lines (Isemura et al., 2012; Mei et al., 2017; Pereira et al., 2018). This showed that the *Comraceme* locus was on Vr04 and determined the inflorescence type. The number of upper inflorescences varied among the subset of the RIL population, even though they showed a compound inflorescence type. Analysis of the number of upper inflorescences as a quantitative trait identified a major QTL on Vr04 that shared a marker (Chr4_27545988) with *Comraceme*. Furthermore, we observed that a few RILs with heterozygous haplotypes around the *Comraceme* locus produced a compound raceme, and concluded that the type of inflorescence was controlled by a single dominant allele.

Comraceme was located in a 520.1 kb genomic region that contained 64 protein-coding genes, 15 of which had conserved nucleotide variants in IT208075 that were not observed in *Vigna* varieties with a compound raceme. We selected two of these genes, *VrDet1* and *Vradi04g00002481*, as potential candidates for *Comraceme* and analyzed their expression level in the SAT between the parental lines. *VrDet1* is an ortholog of the inflorescence meristem identity gene

TFL1, which together with two other floral meristem identity genes, *LEAFY* (*LFY*) and *APETAL1* (*API*), participates in the transition between the vegetative and reproductive stages in *Arabidopsis* (Shannon and Meeks-Wagner, 1993; Liljegren et al., 1999). In mungbean, the determinate growth habit is determined by a mutation in the promoter region of *VrDet1* that regulates gene expression during the vegetative stage (Li et al., 2018). Although we identified a variant in the upstream region of *VrDet1* (Supplementary Figure 2A), there were no significant differences between the parental lines in *VrDet1* expression levels in the SATs throughout the vegetative and reproductive developmental stages (Supplementary Figure 2B). *Vradi04g00002481* is a homolog of *Arabidopsis* *ABS2* (AT2G36080) and belongs to a B3-plant specific transcription factor regulating plant growth and development. In *Arabidopsis*, *ABS2* encodes a transcription factor in the RAV subfamily, and its over-expression produces abnormal shoot and floral petal development as a pleiotropic effect (Shao et al., 2012). Transgenic *Medicago truncatula* plants over-expressing *MtRAV* exhibit increased bolting and bifurcation of sub-branches up to tertiary level (Wang et al., 2020). The soybean *E1* gene is one of the major flowering genes and has a genetic region resembling the B3 domain (Xia et al., 2012; Benlloch et al., 2015). In this context, it is notable that VC1973A had many more nodes and branches, produced more inflorescences, and showed a later flowering time than IT208075. *Vradi04g00002481* had one

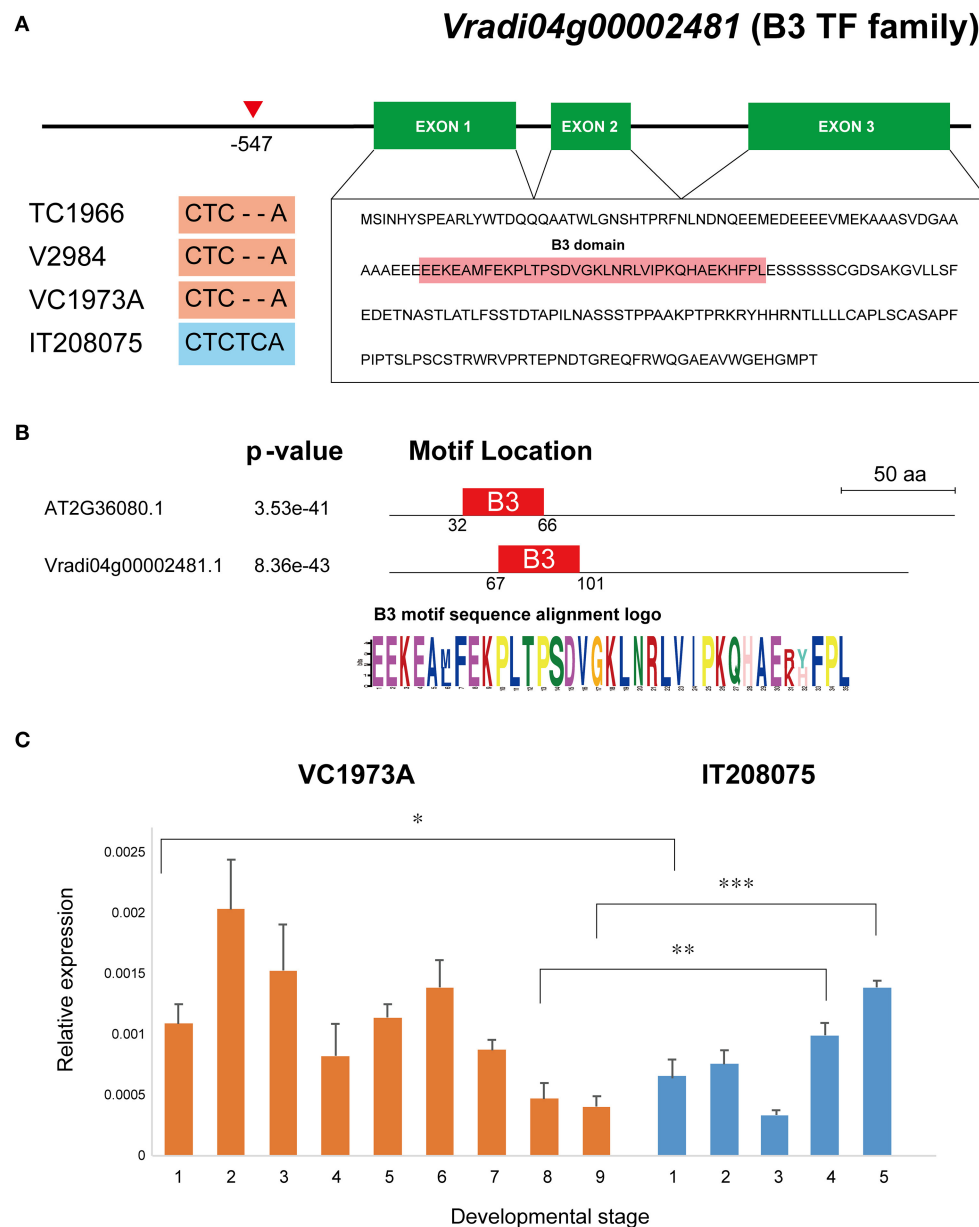


FIGURE 5 | Gene structure and expression level of *Vradi04g00002481*, which encodes a B3 TF family protein. **(A)** Two-base insertion specific to IT208075 located -547 bp upstream of the coding region of the gene. Three mungbean lines producing compound racemes, VC1973A, V2984, and TC1966, showed a two-base deletion at this point, indicated by a red triangle. *Vradi04g00002481* consists of 3 exons encoding 224 amino acids and contains a B3 domain. **(B)** B3 motif sequence alignment between AT2G36080 and *Vradi04g00002481*. **(C)** Expression levels of *Vradi04g00002481* in shoot apical tissue from lines VC1973A and IT208075 at each developmental stage. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

variant in the promoter region only from IT208075 (Figure 5A) and the significant differences in levels of *Vradi04g00002481* expression were observed between VC1973A and IT208075 (Figure 5C). These results suggested that *Vradi04g00002481* functioned as a modulator that determined inflorescence architecture and was a key factor affecting flowering.

Many agronomically important traits, such as seed dormancy, 100-seed weight, plant type, and phenology-related traits, have

been improved for the benefit of human beings during mungbean domestication (Isemura et al., 2012). In general, traits related to high yield and mechanical cultivation have been the main target of crop breeding programs. Several traits have been selected from wild and exotic species to produce modern cultivated varieties with a suitable plant architecture and high crop productivity. Such traits include vine to erect growth habit (Wang et al., 2019), stem termination (Tian et al., 2010), and pod shattering

(Dong et al., 2014). In this study, we identified IT208075, a simple raceme genotype having a single primary inflorescence. This genotype is expected to show shorter flowering duration than VC1973A of a compound raceme, probably leading to higher synchronicity of pod maturity. IT208075 will be a useful breeding material for mungbean improvement, even though the correlation between the simple raceme inflorescence and other agricultural traits such as flowering duration, synchronous pod maturity and seed yield is yet to be elucidated. Furthermore, the linkage mapping and QTL analysis of a RIL population derived from a cross of VC1973A and IT208075 suggested that *Vradi04g00002481* was the candidate gene responsible for the formation of compound raceme inflorescence architecture. Although further study is required to determine the functional role played by this gene in modulating inflorescence type, these results broaden our understanding of the genetic background of inflorescence architecture in mungbean.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found at: National Center for Biotechnology Information (NCBI) BioProject, <https://www.ncbi.nlm.nih.gov/bioproject/>, PRJNA698712.

REFERENCES

- Alam, A. K. M. M., Somta, P., and Srinives, P. (2014). Identification and confirmation of quantitative trait loci controlling resistance to mungbean yellow mosaic disease in mungbean [*Vigna radiata* (L.) wilczek]. *Mol. Breed.* 34, 1497–1506. doi: 10.1007/s11032-014-0133-0
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME suite: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335
- Benlloch, R., Berbel, A., Ali, L., Gohari, G., Millán, T., and Madueño, F. (2015). Genetic control of inflorescence architecture in legumes. *Front. Plant Sci.* 6:543. doi: 10.3389/fpls.2015.00543
- Benlloch, R., Berbel, A., Serrano-Mislata, A., and Madueno, F. (2007). Floral initiation and inflorescence architecture: a comparative view. *Ann. Bot.* 100, 659–676. doi: 10.1093/aob/mcm146
- Berbel, A., Ferrándiz, C., Hecht, V., Dalmais, M., Lund, O. S., Sussmilch, F. C., et al. (2012). VEGETATIVE1 is essential for development of the compound inflorescence in pea. *Nat. Commun.* 3:797. doi: 10.1038/ncomms1801
- Bernardo, K. A. S., Freire Filho, F. R., Ribeiro, V. Q., Vieira, P. F. M. J., Lopes, Â. C. A., and Oliveira, R. M. (2018). Incorporation of compound inflorescences and selection of high-yielding progenies in cowpea. *Pesqui. Agropecuária Bras.* 53, 1150–1157. doi: 10.1590/s0100-204x2018001000008
- Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R., and Coen, E. (1997). Inflorescence commitment and architecture in arabidopsis. *Science* 275, 80–83. doi: 10.1126/science.275.5296.80
- Chankaew, S., Somta, P., Isemura, T., Tomooka, N., Kaga, A., Vaughan, D. A., et al. (2013). Quantitative trait locus mapping reveals conservation of major and minor loci for powdery mildew resistance in four sources of resistance in mungbean [*Vigna radiata* (L.) Wilczek]. *Mol. Breed.* 32, 121–130. doi: 10.1007/s11032-013-9856-6
- Chankaew, S., Somta, P., Sorajapinun, W., and Srinives, P. (2011). Quantitative trait loci mapping of cercospora leaf spot resistance in mungbean, *Vigna radiata* (L.) Wilczek. *Mol. Breed.* 28, 255–264. doi: 10.1007/s11032-010-9478-1

AUTHOR CONTRIBUTIONS

EL and S-HL conceived and designed the research. EL and KYP developed the RIL population. EL and XY conducted the experiments. EL analyzed the data. EL and MYK wrote the manuscript. MYK, JH, and S-HL revised the manuscript. S-HL supervised the project. All authors read and approved the manuscript.

FUNDING

This research was supported by a grant from the Next Generation BioGreen 21 Program (Code No. PJ01326101), Rural Development Administration, Republic of Korea. This work was also supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through the Agri-Bio Industry Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (114054-3), Republic of Korea.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.642518/full#supplementary-material>

- Chen, H.-M., Ku, H.-M., Schafleitner, R., Bains, T. S., George Kuo, C., Liu, C.-A., et al. (2013). The major quantitative trait locus for mungbean yellow mosaic Indian virus resistance is tightly linked in repulsion phase to the major bruchid resistance locus in a cross between mungbean [*Vigna radiata* (L.) wilczek] and its wild relative *Vigna radiata* ssp. sublobata. *Euphytica* 192, 205–216. doi: 10.1007/s10681-012-0831-9
- Chotechung, S., Somta, P., Chen, J., Yimram, T., Chen, X., and Srinives, P. (2016). A gene encoding a polygalacturonase-inhibiting protein (PGIP) is a candidate gene for bruchid (*Coleoptera: Bruchidae*) resistance in mungbean (*Vigna radiata*). *Theor. Appl. Genet.* 129, 1673–1683. doi: 10.1007/s00122-016-2731-1
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- 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
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. doi: 10.1371/journal.pone.0019379
- Han, Y., Yang, H., and Jiao, Y. (2014). Regulation of inflorescence architecture by cytokinins. *Front. Plant Sci.* 5:669. doi: 10.3389/fpls.2014.00669
- Hwang, W. J., Ha, J., Lee, T., Jeong, H., Kim, M. Y., Kim, S. K., et al. (2017). A candidate flowering gene in mungbean is homologous to a soybean phytochrome A gene. *Euphytica* 213:79. doi: 10.1007/s10681-017-1866-8
- IBPGR (1985). *Descriptors for Vigna mungo and V. radiata (revised)*. Rome: IBPGR Secretariat.
- Isemura, T., Kaga, A., Tabata, S., Somta, P., Srinives, P., Shimizu, T., et al. (2012). Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). *PLoS ONE* 7:e41304. doi: 10.1371/journal.pone.0041304
- Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. doi: 10.1093/nar/gkw982

- Kaewwongwal, A., Chen, J., Somta, P., Kongjaimun, A., Yimram, T., Chen, X., et al. (2017). Novel alleles of two tightly linked genes encoding polygalacturonase-inhibiting proteins (VrPGIP1 and VrPGIP2) associated with the Br locus that confer bruchid (*Callosobruchus* spp.) resistance to mungbean (*Vigna radiata*) accession V2709. *Front. Plant Sci.* 8:1692. doi: 10.3389/fpls.2017.01692
- Kaewwongwal, A., Liu, C., Somta, P., Chen, J., Tian, J., Yuan, X., et al. (2020). A second VrPGIP1 allele is associated with bruchid resistance (*Callosobruchus* spp.) in wild mungbean (*Vigna radiata* var. sublobata) accession ACC41. *Mol. Genet. Genomics* 295, 275–286. doi: 10.1007/s00438-019-01619-y
- Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B.-K., et al. (2014). Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat. Commun.* 5:5443. doi: 10.1038/ncomms6443
- Kim, K.-S., Diers, B. W., Hyten, D. L., Mian, M. R., Shannon, J. G., and Nelson, R. L. (2012). Identification of positive yield QTL alleles from exotic soybean germplasm in two backcross populations. *Theor. Appl. Genet.* 125, 1353–1369. doi: 10.1007/s00122-012-1944-1
- Kim, S. K., Nair, R. M., Lee, J., and Lee, S.-H. (2015). Genomic resources in mungbean for future breeding programs. *Front. Plant Sci.* 6:626. doi: 10.3389/fpls.2015.00626
- Kitsanachandee, R., Somta, P., Chatchawankanphanich, O., Akhtar, K. P., Shah, T. M., Nair, R. M., et al. (2013). Detection of quantitative trait loci for mungbean yellow mosaic India virus (MYMIV) resistance in mungbean (*Vigna radiata* (L.) wilczek) in India and Pakistan. *Breed. Sci.* 63, 367–373. doi: 10.1270/jsbbs.63.367
- Kuroda, Y., Kaga, A., Tomooka, N., Yano, H., Takada, Y., Kato, S., et al. (2013). QTL affecting fitness of hybrids between wild and cultivated soybeans in experimental fields. *Ecol. Evol.* 3, 2150–2168. doi: 10.1002/ece3.606
- Lark, K. G., Chase, K., Adler, F., Mansur, L. M., and Orf, J. H. (1995). Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. *Proc. Natl. Acad. Sci. U.S.A.* 92, 4656–4660. doi: 10.1073/pnas.92.10.4656
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio*. Available online at: <http://arxiv.org/abs/1303.3997> (accessed June 8, 2020).
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, S., Ding, Y., Zhang, D., Wang, X., Tang, X., Dai, D., et al. (2018). Parallel domestication with a broad mutational spectrum of determinate stem growth habit in leguminous crops. *Plant J.* 96, 761–771. doi: 10.1111/tpj.14066
- Liljgren, S. J., Gustafson-Brown, C., Pinyopich, A., Ditta, G. S., and Yanofsky, M. F. (1999). Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. *Plant Cell* 11, 1007–1018. doi: 10.1105/tpc.11.6.1007
- Liu, B., Watanabe, S., Uchiyama, T., Kong, F., Kanazawa, A., Xia, Z., et al. (2010). The soybean stem growth habit gene Dt1 is an ortholog of arabidopsis TERMINAL FLOWER1. *Plant Physiol.* 153, 198–210. doi: 10.1104/pp.109.150607
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔCT} method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Mansur, L. M., Orf, J. H., Chase, K., Jarvik, T., Cregan, P. B., and Lark, K. G. (1996). Genetic mapping of agronomic traits using recombinant inbred lines of soybean. *Crop Sci.* 36, 1327–1336. doi: 10.2135/cropsci1996.0011183X003600050042x
- Mathivathana, M. K., Murukarthick, J., Karthikeyan, A., Jang, W., Dhasarathan, M., Jagadeeshselvam, N., et al. (2019). Detection of QTLs associated with mungbean yellow mosaic virus (MYMV) resistance using the interspecific cross of *Vigna radiata* × *Vigna umbellata*. *J. Appl. Genet.* 60, 255–268. doi: 10.1007/s13353-019-00506-x
- Mehandi, S., Quatadah, S., Mishra, S. P., Singh, I., Praveen, N., and Dwivedi, N. (2019). “Mungbean (*Vigna radiata* L. wilczek): retrospect and prospects,” in *Legume Crops - Characterization and Breeding for Improved Food Security*, ed. El-Esawi, M. A. (London: IntechOpen), 49–66. doi: 10.5772/intechopen.85657
- Mei, H., Liu, Y., Du, Z., Wu, K., Cui, C., Jiang, X., et al. (2017). High-density genetic map construction and gene mapping of basal branching habit and flowers per leaf axil in sesame. *Front. Plant Sci.* 8:636. doi: 10.3389/fpls.2017.00636
- Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001
- Orf, J. H., Chase, K., Jarvik, T., Mansur, L. M., Cregan, P. B., Adler, F. R., et al. (1999). Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. *Crop Sci.* 39, 1642–1651. doi: 10.2135/cropsci1999.3961642x
- Pereira, L., Ruggieri, V., Pérez, S., Alexiou, K. G., Fernández, M., Jahrman, T., et al. (2018). QTL mapping of melon fruit quality traits using a high-density GBS-based genetic map. *BMC Plant Biol.* 18:324. doi: 10.1186/s12870-018-1537-5
- Prusinkiewicz, P., Erasmus, Y., Lane, B., Harder, L. D., and Coen, E. (2007). Evolution and development of inflorescence architectures. *Science* 316, 1452–1456. doi: 10.1126/science.1140429
- Schaffleitner, R., Huang, S., Chu, S., Yen, J., Lin, C., Yan, M., et al. (2016). Identification of single nucleotide polymorphism markers associated with resistance to bruchids (*Callosobruchus* spp.) in wild mungbean (*Vigna radiata* var. sublobata) and cultivated *V. radiata* through genotyping by sequencing and quantitative trait locus analysis. *BMC Plant Biol.* 16:159. doi: 10.1186/s12870-016-0847-8
- Selvi, R., Muthiah, A. R., Maheswaran, M., and Shanmugasundaram, P. (2003). Genetic diversity analysis in the genus *Vigna* based on morphological traits and isozyme markers. *Sabara J. Breed. Genet.* 35, 103–112.
- Sen, N. K., and Ghosh, A. K. (1959). Genetic studies in green gram. *Indian J. Genet.* 19, 210–227.
- Shanmugasundaram, S., Keatinge, J. D. H., and Hughes, J. D. (2009). “Counting on beans: mungbean improvement in Asia,” in *Millions Fed: Proven Successes in Agricultural Development*, eds D. J. Spielman and R. Pandya-Lorch (Washington, DC: International Food Policy Research Institute), 103–108.
- Shannon, S., and Meeks-Wagner, D. R. (1993). Genetic interactions that regulate inflorescence development in arabidopsis. *Plant Cell* 5, 639–655. doi: 10.2307/3869807
- Shao, J., Liu, X., Wang, R., Zhang, G., and Yu, F. (2012). The over-expression of an arabidopsis B3 transcription factor, ABS2/NGAL1, leads to the loss of flower petals. *PLoS ONE* 7:e49861. doi: 10.1371/journal.pone.0049861
- Shim, S., Kim, M. Y., Ha, J., Lee, Y.-H., and Lee, S.-H. (2017). Identification of QTLs for branching in soybean (*Glycine max* (L.) merrill). *Euphytica* 213, 225. doi: 10.1007/s10681-017-2016-z
- Singh, T. P., and Singh, K. B. (1970). Inheritance of clusters per node in mungbean (*Phaseolus aureus* Roxb.). *Curr. Sci.* 39:265.
- Tian, Z., Wang, X., Lee, R., Li, Y., Specht, J. E., Nelson, R. L., et al. (2010). Artificial selection for determinate growth habit in soybean. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8563–8568. doi: 10.1073/pnas.1000088107
- Tucker, S. C. (2003). Floral development in legumes. *Plant Physiol.* 131, 911–926. doi: 10.1104/pp.102.017459
- Ungerer, M. C., Halldorsdottir, S. S., Modliszewski, J. L., Mackay, T. F. C., and Purugganan, M. D. (2002). Quantitative trait loci for inflorescence development in arabidopsis thaliana. *Genetics* 160, 1133–1151. doi: 10.1093/genetics/160.3.1133
- Van Ooijen, J. (2006). *JoinMap® 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Kyazma BV Wageningen.
- Wang, R., Liu, L., Kong, J., Xu, Z., Akhter Bhat, J., and Zhao, T. (2019). QTL architecture of vine growth habit and gibberellin oxidase gene diversity in wild soybean (*Glycine soja*). *Sci. Rep.* 9:7393. doi: 10.1038/s41598-019-43887-z
- Wang, S., Guo, T., Wang, Z., Kang, J., Yang, Q., Shen, Y., et al. (2020). Expression of three related to ABI3/VP1 genes in medicago truncatula caused increased stress resistance and branch increase in arabidopsis thaliana. *Front. Plant Sci.* 11:611. doi: 10.3389/fpls.2020.00611
- Xia, Z., Watanabe, S., Yamada, T., Tsukokura, Y., Nakashima, H., Zhai, H., et al. (2012). Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2155–E2164. doi: 10.1073/pnas.1117982109

- Yoon, M. Y., Kim, M. Y., Ha, J., Lee, T., Kim, K. D., and Lee, S.-H. (2019). QTL analysis of resistance to high-intensity UV-B irradiation in soybean (*Glycine max* [L.] Merr.). *Int. J. Mol. Sci.* 20:3287. doi: 10.3390/ijms20133287
- Yundaeng, C., Somta, P., Chen, J., Yuan, X., Chankaew, S., and Chen, X. (2021). Fine mapping of QTL conferring *Cercospora* leaf spot disease resistance in mungbean revealed TAF5 as candidate gene for the resistance. *Theor. Appl. Genet.* 134, 701–714. doi: 10.1007/s00122-020-03724-8
- Yundaeng, C., Somta, P., Chen, J., Yuan, X., Chankaew, S., Srinives, P., et al. (2020). Candidate gene mapping reveals VrMLO12 (MLO Clade II) is associated with powdery mildew resistance in mungbean (*Vigna radiata* [L.] wilczek). *Plant Sci.* 298:110594. doi: 10.1016/j.plantsci.2020.110594
- Zhang, D., Cheng, H., Wang, H., Zhang, H., Liu, C., and Yu, D. (2010). Identification of genomic regions determining flower and pod numbers development in soybean (*Glycine max* L.). *J. Genet. Genomics* 37, 545–556. doi: 10.1016/S1673-8527(09)60074-6
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2021 Lee, Yang, Ha, Kim, Park and Lee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Single-Molecule Real-Time and Illumina-Based RNA Sequencing Data Identified Vernalization-Responsive Candidate Genes in Faba Bean (*Vicia faba* L.)

Xingxing Yuan^{1,2}, Qiong Wang², Bin Yan^{1,2}, Jiong Zhang^{1,2}, Chenchen Xue², Jingbin Chen², Yun Lin², Xiaoyan Zhang², Wenbiao Shen^{1*} and Xin Chen²

¹ College of Life Sciences, Nanjing Agricultural University, Nanjing, China, ² Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China

OPEN ACCESS

Edited by:

Longjiang Fan,
Zhejiang University, China

Reviewed by:

Yordan Muhovski,
Walloon Agricultural Research Centre,
Belgium

Ahmed Sallam,
Assiut University, Egypt

Hussein M. Migdadi,
King Saud University, Saudi Arabia

*Correspondence:

Wenbiao Shen
wbshenh@njau.edu.cn

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 20 January 2021

Accepted: 07 June 2021

Published: 05 July 2021

Citation:

Yuan X, Wang Q, Yan B, Zhang J,
Xue C, Chen J, Lin Y, Zhang X,
Shen W and Chen X (2021)
Single-Molecule Real-Time
and Illumina-Based RNA Sequencing
Data Identified
Vernalization-Responsive Candidate
Genes in Faba Bean (*Vicia faba* L.).
Front. Genet. 12:656137.
doi: 10.3389/fgene.2021.656137

Faba bean (*Vicia faba* L.) is one of the most widely grown cool season legume crops in the world. Winter faba bean normally has a vernalization requirement, which promotes an earlier flowering and pod setting than unvernallized plants. However, the molecular mechanisms of vernalization in faba bean are largely unknown. Discovering vernalization-related candidate genes is of great importance for faba bean breeding. In this study, the whole transcriptome of faba bean buds was profiled by using next-generation sequencing (NGS) and single-molecule, real-time (SMRT) full-length transcriptome sequencing technology. A total of 29,203 high-quality non-redundant transcripts, 21,098 complete coding sequences (CDS), 1,045 long non-coding RNAs (lncRNAs), and 12,939 simple sequence repeats (SSRs) were identified. Furthermore, 4,044 differentially expressed genes (DEGs) were identified through pairwise comparisons. By Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, these differentially expressed transcripts were found to be enriched in binding and transcription factor activity, electron carrier activity, rhythmic process, and receptor activity. Finally, 50 putative vernalization-related genes that played important roles in the vernalization of faba bean were identified; we also found that the levels of vernalization-responsive transcripts showed significantly higher expression levels in cold-treated buds. The expression of *VfSOC1*, one of the candidate genes, was sensitive to vernalization. Ectopic expression of *VfSOC1* in *Arabidopsis* brought earlier flowering. In conclusion, the abundant vernalization-related transcripts identified in this study will provide a basis for future researches on the vernalization and faba bean breeding and established a reference full-length transcriptome for future studies on faba bean.

Keywords: faba bean, transcriptome, SMRT sequencing, vernalization, *VfSOC1*

INTRODUCTION

Vicia faba (L.), commonly known as faba bean or broad bean, is an annual or biennial herb of the family Fabaceae. It is a cross-pollinating diploid plant species ($2n = 12$) (Fuchs et al., 1998) that has a “giant genome” with an estimated genome size of 13 Gb (Johnston et al., 1999). The nuclear genome size of faba bean is one of the largest yet described among crop legumes, which is approximately 25 times greater than that of the model legume *Medicago truncatula* and 2.5 times greater than that of *Pisum sativum* (Rispaill et al., 2010). Up to now, the genomic information of faba bean is blank, and there is no genome sequence for this species. Faba bean is one of the most effective nitrogen-fixing legumes for soil improvement abilities and is widely used for animal feed and human food, especially as a staple dietary protein source in North African and Middle Eastern cultures (O’Sullivan and Angra, 2016). Faba bean is also consumed as a vegetable crop. The immature faba bean is a high-quality vegetable, supplying well-balanced protein and carbohydrate together with numerous antioxidants and essential vitamins (Heather and Fawzy, 2010). China is the largest faba bean producer (in both sowing area and production) in the world, followed currently by Ethiopia, Australia, France, and Egypt (O’Sullivan and Angra, 2016). In the southern parts of China, faba bean is sown in areas between 21 and 35°N latitude from October to November and harvested from March to May (Ye et al., 2003). Winter faba bean comprised 85.5% of the total area and 78.2% of the total production of faba bean in China (Ye et al., 2003). It is the fourth most widely grown cool season legume after pea (*P. sativum*), chickpea (*Cicer arietinum*), and lentil (*Lens culinaris*) (Alharbi and Adhikari, 2020).

Vernalization plays an important role in the regulation of flowering time in higher plants for flower transformation, which is an evolutionarily adaptive mechanism preventing plant growth transition from the vegetative to the reproductive phase before winter and allowing flowering in favorable conditions in spring (Mahfoozi et al., 2006). The vernalization process has been thoroughly studied in *Arabidopsis* and in some winter crops such as wheat in the past years; however, the molecular regulatory mechanisms in other plants including faba bean are still largely unknown (Cao et al., 2017). In previous studies, researchers have identified some quantitative trait nucleotides (QTNs) or quantitative trait loci (QTLs) related to winter hardiness or days after flowering (DAF), but due to the limitation of genomic information, candidate genes were not found (Sallam et al., 2016; Catt et al., 2017).

In *Arabidopsis*, many reports indicated that flowering-time genes regulated vernalization. For example, FLOWERING LOCUS C (*FLC*) determines the extent of the vernalization response during flowering (Michaels and Amasino, 1999), FRIGIDA (*FRI*) activates *FLC* expression and drives *Arabidopsis* floral transition along the vernalization pathway (Sallam et al., 2016), VERNALIZATION1 (*VRN1*) encodes a protein that binds *FLC* in a non-sequence-specific manner and functions in the stable repression of *FLC* (Levy et al., 2002), and *VRN2* encodes a nuclear-localized zinc finger protein, maintains *FLC* repression, and serves as a mechanism for cellular memory

(Gendall et al., 2001). Lately, *SOC1*, a MADS box transcription factor regulated by two antagonistic flowering regulators, CONSTANS (*CO*) and *FLC*, has acted as a floral activator and repressor (Lee and Lee, 2010). Recently, *TAF15b* (Eom et al., 2018), *PEP1* (Lazaro et al., 2018), *TRAF1a/1b* (Qi et al., 2017), and *VIN3* (Hepworth et al., 2018) were found to participate in vernalization directly or indirectly.

Due to the reading length, GC (guanine and cytosine) sensitivity, and other factors, the UniGene assembled by the traditional second-generation sequencing has incomplete transcripts. The third-generation sequencing—single-molecule, real-time (SMRT) sequencing—has good quality sequencing results, with an average reading length of 10–15 kb compared with the 1 kb sequenced by the second-generation sequencing technology, which can easily span the complete transcripts, and the full-length transcripts of messenger RNAs (mRNAs) can be obtained without assembly (Hepworth et al., 2018). In the present study, we used the SMRT sequencing technology to identify candidate vernalization genes and found that *VfSOC1* could promote flowering and response to vernalization. Furthermore, 50 putative flowering-related genes were identified and could be associated with vernalization. Thus, this study provides resources and directions for researchers to study the vernalization genes in faba bean.

MATERIALS AND METHODS

Plant Materials and Treatment Conditions

Faba bean (*V. faba* L.) cv. Tongcanxian-6 seedlings were generated. Pacific Biosciences (PacBio, Menlo Park, CA, United States) isoform sequencing (Iso-Seq) libraries of faba bean were constructed using buds grown for 5 days to 1 cm (sample 0), buds and leaves with cold treatment (4°C) and dark culture for 14 days (sample 1), cold treated and dark cultured for 12 days then transferred to incubators (25°C/16 h and 20°C/8 h) for 2 days (sample 2), and cultured in incubators (25°C/16 h and 20°C/8 h) for 14 days (sample 3). For samples 1, 2, 3, and 0, and their T correspondence, the corresponding relations are as follows: T01–T03 represent three biological repetitions of sample 0; T04–T06 represent three biological repetitions of sample 1; T07–T09 represent three biological repetitions of sample 2; and T10–T12 represent three biological repetitions of sample 3. Finally 12 samples were sequenced. After treatment, the leaves were collected for RNA extraction. Faba bean total RNA was isolated using the Trizol reagent (Invitrogen, Foster City, CA, United States). Of the total RNA (equally mixed with RNAs from T01 to T12), 5 µg was reverse transcribed into complementary DAN (cDNA) using a SMARTer PCR cDNA synthesis kit (TaKaRa, Shiga, Japan), mRNA was purified from total RNA using poly-T oligo-attached magnetic beads, and was reverse transcribed into first-strand cDNA using a random hexamer primer. Then, DNA polymerase I and RNase H were used for second-strand cDNA synthesis. After end-blunting and adenylation of the 3′ ends, the DNA fragments were ligated with a sequencing adaptor and then size fractionation and

selection (1–2, 2–3, and 3–6 kb) were performed using the BluePippin Size Selection System (Sage Science, Beverly, MA, United States). SMRTbell libraries were constructed with the PacBio DNA Template Prep Kit 2.0, and SMRT sequencing was then performed on the PacBio Sequel platform with three cells. Multiple size-fractionated cDNAs (1–2, 2–3, and 3–6 kb) were prepared to construct the faba bean Iso-Seq libraries.

PacBio Long-Read Processing

Long reads from the PacBio platform were processed into error-corrected reads of insert (ROIs) using the ToFU pipeline¹ with default parameters. ROIs were classified into non-full-length and full-length reads (including full-length non-chimeric and chimeric reads) based on the presence and location of the 3' primer, 5' primer, and polyA. We used ICE (iterative clustering for error correction²) to obtain consensus isoforms. Full-length consensus sequences from ICE were first polished using Quiver (Chin et al., 2013) and further corrected by Proovread (Hackl et al., 2014) software with the help of Illumina short reads. Then, the full-length transcripts with a post-correction accuracy above 99% were generated for further analysis. CD-HIT (Li and Godzik, 2006) was used to remove redundant high-quality full-length transcripts.

SSR Detection

Simple sequence repeats (SSRs) of the transcriptome were identified using MISA software with default parameters³. The parameters were set as follows: one base repeat 10 times or more; two base repeats six times or more; three base repeats five times or more; four base repeats five times or more; five base repeats five times or more; and six base repeats five times or more. If the base repeats occur five or more times, they were regarded as microsatellite sequences. Moreover, if the distance between two microsatellites is less than 100 bp, the two microsatellites were treated as composite microsatellites.

Coding Sequence Detection

TransDecoder⁴ was applied to identify candidate coding regions within the transcript sequences using the following criteria: (1) a minimum length of open reading frame (ORF) is found in a transcript sequence; (2) the log-likelihood score similar to what is computed by the GeneID⁵ software is > 0 ; (3) the above coding score is greatest when the ORF is scored in the first reading frame as compared to scores in the other five reading frames; (4) if a candidate ORF is found fully encapsulated by the coordinates of another candidate ORF, the longer one is reported. However, a single transcript can report multiple ORFs (allowing for operons, chimeras, etc.); and (5) optional, the putative peptide has a match to a Pfam (Protein family) domain above the noise cutoff score.

¹http://github.com/PacificBiosciences/cDNA_primer/

²<https://www.pacb.com/productsand-services/analytical-software>

³<http://pgrc.ipk-gatersleben.de/misa/>

⁴<https://github.com/TransDecoder/TransDecoder/releases>

⁵<http://genome.crg.es/software/geneid/>

Long Non-coding RNA Analysis

Four computational approaches—CPC (coding potential calculator) (Kong et al., 2007), CNCI (Coding-Non-Coding Index) (Sun et al., 2013), CPAT (Coding Potential Assessment Tool) (Wang et al., 2013), and the Pfam database (Finn et al., 2014)—were combined to sort the non-protein-coding RNA candidates from the putative protein-coding RNAs in the unknown transcripts. Putative protein-coding RNAs were filtered out using a minimum length and exon number threshold. Transcripts with lengths greater than 200 nt and have more than two exons were selected as long non-coding RNA (lncRNA) candidates and further screened using CPC/CNCI/CPAT/Pfam, which have the power to distinguish the protein-coding genes from the non-coding genes.

Quantification of the Gene Expression Levels and Differential Expression Analysis

The clean reads of each RNA sequencing (RNA-Seq) library were aligned to the reference transcriptome to obtain unique mapped reads by using the Bowtie2⁶ software (Langmead and Salzberg, 2012), a comparison tool for sequencing reads aligned with long reference sequences and with output in the form of counts. The gene expression levels were estimated by as FPKM (fragments per kilobase of transcript per million fragments mapped) using RSEM (Li and Dewey, 2011) software. Mismatches of no more than two bases were allowed in the alignment. Clean data were first mapped back onto the assembled transcriptome, and then the read counts for each gene were obtained from the mapping results.

Differential expression analysis of two groups of samples was performed using the DESeq (Anders and Huber, 2012) R package. DESeq provides statistical routines for determining differential expression in digital gene expression data using a model based on a negative binomial distribution. The resulting *p*-values were adjusted using the Benjamini and Hochberg approach for controlling the false discovery rate (FDR). The critical standard for significant differentially expressed genes (DEGs) was set at $FDR < 0.01$ and fold change ≥ 2 .

Functional Annotation and Enrichment Analysis

Gene function was annotated based on the following databases: Nr (NCBI non-redundant protein sequences⁷), Pfam⁸, KOG/COG/eggNOG (Clusters of Orthologous Groups of Proteins)⁹, Swiss-Prot (a manually annotated and reviewed protein sequence database)^{10,11}, KEGG (Kyoto Encyclopedia of Genes and Genomes)¹², and GO (Gene Ontology)¹³.

⁶<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>

⁷<http://www.ncbi.nlm.nih.gov/RefSeq/>

⁸<http://pfam.sanger.ac.uk/>

⁹<http://www.ncbi.nlm.nih.gov/COG/>

¹⁰<http://www.expasy.ch/sprot/>

¹¹<http://www.ebi.ac.uk/swissprot/Boeckmann>

¹²<http://www.genome.jp/kegg/>

¹³<http://www.geneontology.org>

GO enrichment analysis of the DEGs was implemented by the GSeq (Young et al., 2010) R package based on the Wallenius non-central hyper-geometric distribution, which can adjust for gene length bias in DEGs. We used KOBAS (Mao et al., 2015) software to test the statistical enrichment of the differential expression genes in the KEGG pathways.

Arabidopsis Transformation and Growth Conditions

Faba bean total RNA was isolated using the Trizol reagent (Invitrogen, Foster City, CA, United States). The first-strand cDNA was then synthesized using M-MLV reverse transcriptase (Promega, Madison, WI, United States). PCR-amplified DNA fragments were linked to T Vector for sequences. Then, the full-length coding sequence of *VfSOC1* was cloned into pEarley Gate103 driven by the 35S promoter. The plasmid was transformed into the *Agrobacterium tumefaciens* EHA105 strain. Overexpressed transgenic plants were generated by the floral dipping method (Qi et al., 2017) in the *Col* background through *A. tumefaciens*-mediated transformation and subsequent selection by hygromycin on 1/2 Murashige and Skoog medium.

Wild-type *Arabidopsis thaliana* (*Col-0*) was used as the background, and transgenic lines were grown on soil under long days (16-h light/8-h dark) at $24 \pm 2^\circ\text{C}$.

Expression Analysis

For the semi-quantitative RT-PCR analysis of *VfSOC1* expression in transgenic *Arabidopsis* lines, *VfSOC1* was detected by the primers *VfSOC1*-F-1 (GACCGATACCGCAGTCATAGCCG)

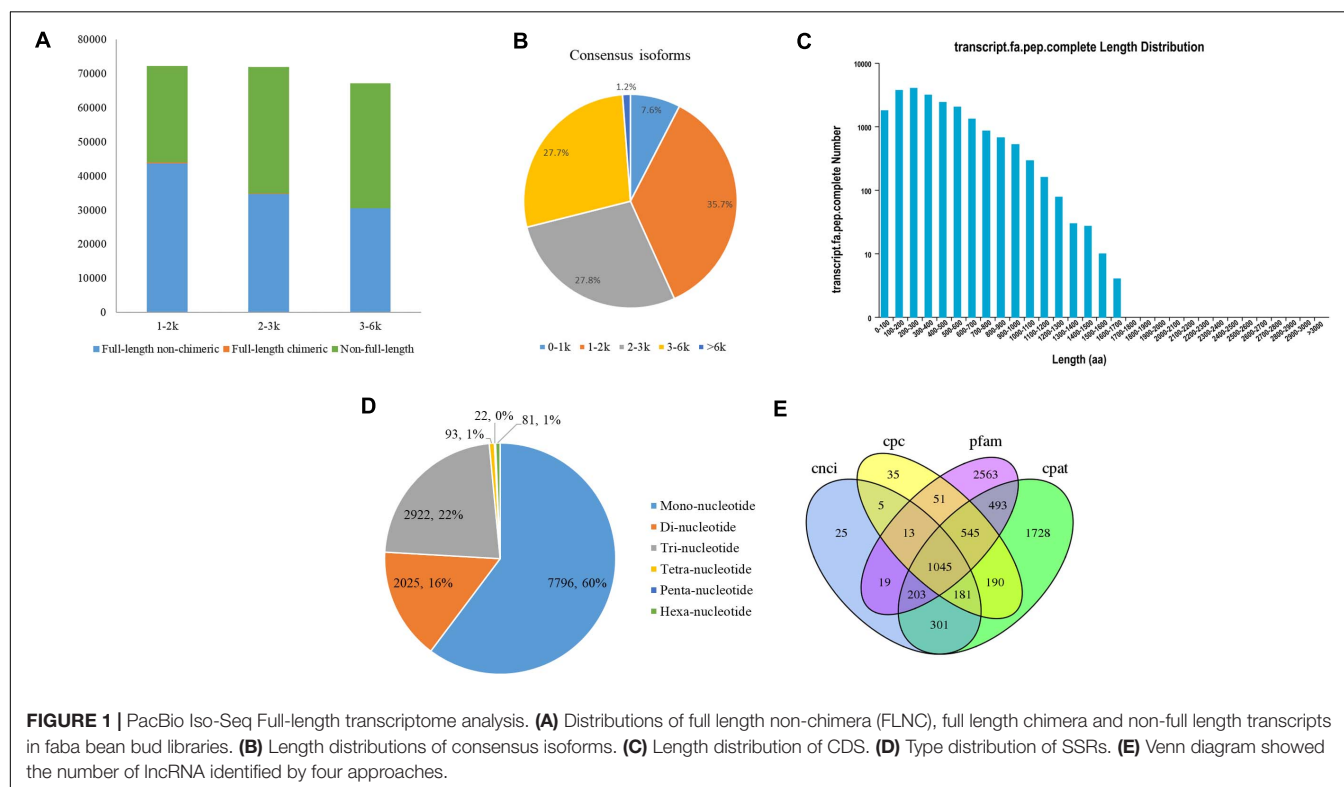
and *VfSOC1*-R-1 (GCATGACATTT TCAGCAACTAGGGCT). *Actin2* (*At3g18780*) was used as the reference gene (*Actin2*-F: TCCTTTGTTGCTGTTGACTACG, *Actin2*-R: ATTTTCT GTGAACGATTCTT GG).

RESULTS

Long-Length Transcriptome of Faba Bean

A total of 8.62 Gb of clean data were obtained from all three cells with 450,876 raw polymerase reads. After filtering the adaptor and low-quality sequences, a total of 3,671,393 subreads were obtained (Supplementary Table 1). A total of 233,493 high-quality ROIs were further generated from circular consensus sequencing (CCS) after filtering with full passes and accuracy. The numbers of ROIs from the faba bean libraries were 82,216 for 1–2 kb, 80,070 for 2–3 kb, and 71,207 for 3–6 kb. In total, 109,128 full-length non-chimeric (FLNC) reads were produced from the ROIs of the faba bean libraries, with average lengths of 1,346, 2,363, and 3,283 bp in the corresponding faba bean libraries (Figure 1A and Supplementary Table 2).

After classification and correction by the ICE and Quiver programs, 40,046 high-quality (accuracy > 0.99) and 10,727 low-quality polished consensus isoforms were generated from the ROIs. Consensus isoforms were further corrected using Proovread (Hackl et al., 2014) software with the help of Illumina short reads. Ultimately, a total of 29,203 unigene transcripts were obtained after CD-HIT classification. The length distribution of



the polished consensus isoforms is shown in **Figure 1B**, with almost 91.2% of the sequences ranging from 1 to 6 kb.

Faba Bean Transcriptome of NGS Short Reads

We used the same treated samples for SMRT, and the Illumina RNA-Seq libraries constructed from faba bean buds with different treatments were sequenced to correct the polished isoforms and to quantify the full-length transcripts obtained from PacBio Iso-Seq. After the trimming process, 86.99 Gb clean data were obtained from all samples. Over 214 million short reads were successfully mapped back to the full-length transcripts of PacBio Iso-Seq with an average mapping ratio of 73.3% (**Supplementary Table 3**), which suggested that the full-length transcripts derived by the PacBio Iso-Seq method represented most of the genetic information of faba bean.

Functional Annotations

Databases such as Nr, Swiss-Prot, GO, COG, KOG, Pfam, and KEGG were used to perform functional annotations on the 29,203 unigenes. A total of 27,903 unigenes (95.5%) were successfully matched to known sequences in at least one database (**Supplementary Table 4**). There were 95.4% unigenes matched in the Nr database, 76.9% in Swiss-Prot, and 41.9% in COG.

A total of 12,763 (43.7%) unigenes were annotated with 125 pathways in the KEGG pathway analysis. GO terms were assigned to 18,040 unigenes (61.8%) with 19 biological processes, 17 cellular components, and 14 molecular functions. In the biological processes, the top three GO terms were “metabolic process,” “cellular process,” and “single-organism process.” In the class of cellular component, “cell part” was dominant, and then “cell” and “organelle.” High percentages of the unigenes were enriched in “catalytic activity,” “binding,” and “transporter activity” in the class of molecular functions (**Supplementary Table 5**).

CDS, SSR, and lncRNA Prediction

The candidate coding sequence (CDS) in the PacBio Iso-Seq transcripts was analyzed by retaining a minimum length of ORFs using the TransDecoder software. As shown in **Figure 1C**, a total of 21,098 complete CDS were obtained from faba bean transcripts with lengths ranging from 50 to 1,669 bp and an average length of 390 bp.

SSRs are short (1–6 bp) tandem repeat DNA sequences. They are also known as microsatellites. In this study, we identified 12,939 SSRs in 8,939 unigenes (30.6% of the total unigenes), including six types of SSRs: mononucleotide (7,796, 60.3% of all SSRs), dinucleotide (2,025, 15.7%), trinucleotide (2,922, 22.6%), tetranucleotide (93, 0.7%), pentanucleotide (22, 0.2%), and hexanucleotide (81, 0.6%) (**Figure 1D**). Among them, 1,373 SSRs present in compound formation.

Four computational approaches—CPC, CNCI, CPAT, and Pfam—were combined to sort the non-protein-coding RNA candidates from the putative protein-coding RNAs in the faba bean transcripts and then identify the lncRNAs. A total of 1,045 lncRNA transcripts were predicted by these four methods

(**Figure 1E**). The lengths ranged from 541 to 14,958 bp, with an average length of 1,860 bp.

Identification of Differentially Expressed Genes

The expression levels of the transcripts from the PacBio Iso-Seq libraries of four groups of faba bean buds with different treatments were quantified using the FPKM method. Differences in expression were examined to identify the transcripts that may participate in the vernalization of faba bean. A total of 4,044 DEGs (fold change > 2, FDR < 0.01) in faba bean were identified through pairwise comparisons (**Figure 2A**), of which 1,553, 851, 1,443, and 197 DEGs were found between buds with cold treatment and dark culture (sample 1) and 1 cm control (sample 0), cold treated then transferred to incubators (sample 2) and 1 cm control (sample 0), cold treated and dark cultured (sample 1) and those cultured in incubators (sample 3), and cold treated then transferred to incubators (sample 2) and those cultured in incubators (sample 3), respectively. Six hundred eighty-nine genes were upregulated and 864 genes were downregulated under sample 1 compared with sample 0. Among these genes, the expression levels of *VfADO3* ($\log_2FC = 3.32$, FDR = 6.56×10^{-8}), *VfCRY1* ($\log_2FC = 1.87$, FDR = 3.17×10^{-4}), *VfSOC1* ($\log_2FC = 3.97$, FDR = 0.003), and *VfGI* ($\log_2FC = 5.05$, FDR = 5.94×10^{-6}) were significantly higher in sample 1 than those in sample 0 (**Supplementary Table 7**). A comparison of sample 2 with sample 0 revealed 569 genes that were upregulated (**Supplementary Table 9**), including *VfCDF3* ($\log_2FC = 2.69$, FDR = 9.79×10^{-4}), and 282 genes that were downregulated, including *VfADO3* ($\log_2FC = -3.54$, FDR = 1.83×10^{-4}) and *VfCOL9* ($\log_2FC = -3.63$, FDR = 1.19×10^{-7}) (**Supplementary Table 7**). Six hundred thirty genes were upregulated, including *VfCDF3* ($\log_2FC = 2.31$, FDR = 0.009), and 813 genes were downregulated, including *VfADO3* ($\log_2FC = -6.77$, FDR = 7.41×10^{-18}), *VfSOC1* ($\log_2FC = -5.55$, FDR = 1.15×10^{-4}), *VfPHYA* ($\log_2FC = -7.30$, FDR = 2.14×10^{-5}), and *VfGI* ($\log_2FC = -3.54$, FDR = 1.83×10^{-4}), under sample 1 compared with sample 3 (**Supplementary Table 6**). Moreover, 139 genes were upregulated and 58 genes were downregulated under sample 2 compared with sample 3 (**Supplementary Table 8**); however, only *VfPHYA*, the flowering-related gene, was significantly downregulated ($\log_2FC = -4.99$, FDR = 6.24×10^{-5}). Lastly, we chose *VfSOC1* for further functional analysis.

Furthermore, a general overview of the expression patterns was visualized in a heat map (**Figure 2B**), which provided an overall understanding of the changes in gene expression.

Functional Classification of the DEGs by GO and KEGG Pathway Analyses

To identify the flowering-related genes in faba bean, the DEGs were examined with GO and KEGG pathway analyses. The three faba bean bud samples with cold treatment and dark culture were compared to the three matched samples of faba bean bud cultured in incubators. Of the DEGs, 1,443 were screened out, with $\log_2$ fold changes ranging from -9.9 to 12.1 (**Supplementary Table 6**). The DEGs in comparison

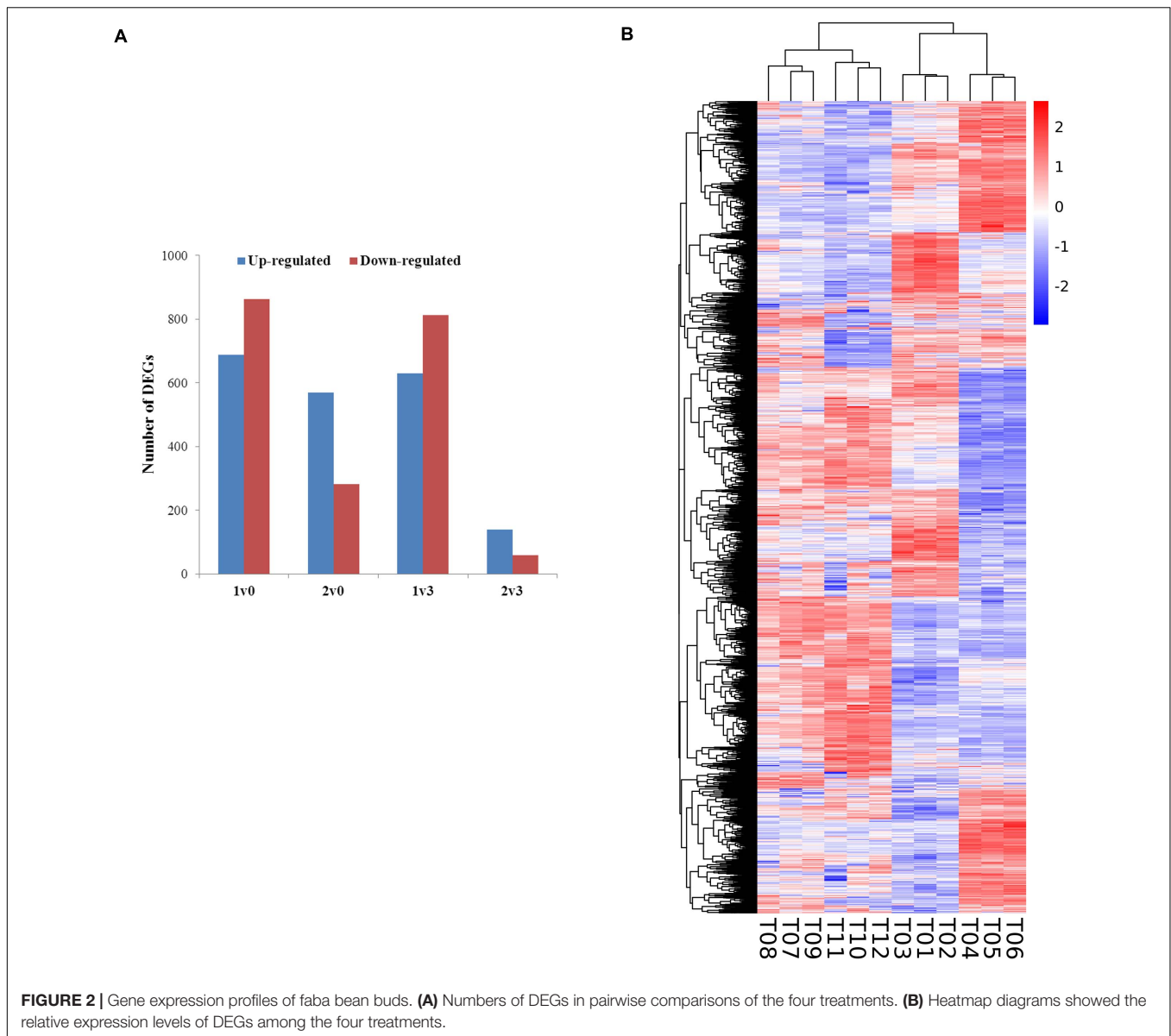


FIGURE 2 | Gene expression profiles of faba bean buds. **(A)** Numbers of DEGs in pairwise comparisons of the four treatments. **(B)** Heatmap diagrams showed the relative expression levels of DEGs among the four treatments.

were enriched in three main GO categories. In the biological process category, the GO terms were significantly enriched in reproduction and rhythmic process. In the molecular function category, binding transcription factor activity, electron carrier activity, receptor activity, and nutrient reservoir activity were significantly enriched. These proteins may play critical roles as transcription factors or at the transcriptional level in low-temperature stress and flowering signaling pathways (Figure 3). As for the KEGG pathway analysis, “photosynthesis,” “circadian rhythm—plant,” “monoterpenoid biosynthesis,” “glyoxylate and dicarboxylate metabolism,” “carotenoid biosynthesis,” and “nitrogen metabolism” were significantly enriched. Changes of the flowering phenotype of faba bean may be associated with circadian rhythm (Figure 3). In addition, 58 unigenes were specifically highly expressed in faba bean buds with cold treatment and dark culture, while no expression was observed in

any of the matched incubator-cultured samples. In contrast, 22 unigenes were highly expressed in the three incubator-cultured samples exclusively (Figure 3A and Supplementary Table 6).

A total of 1,553 significant DEGs (Figure 3B and Supplementary Table 7) were identified in faba bean between cold-treated and dark-cultured faba bean buds and 1 cm control, with log2 fold changes ranging from -9.0 to 10.3 . The GO terms were clustered. In the biological process category, the significantly enriched GO terms were similar to those enriched in cold-treated dark-cultured buds compared to buds cultured in incubators. As for molecular function, DEGs were specifically involved in “antioxidant activity” (Figure 3B). Enrichment analysis based on the KEGG database revealed that the most enriched terms were “phenylalanine metabolism,” “phenylpropanoid biosynthesis,” and “cutin, suberine, and wax biosynthesis.” In addition, 22 and 18 unigenes were exclusively

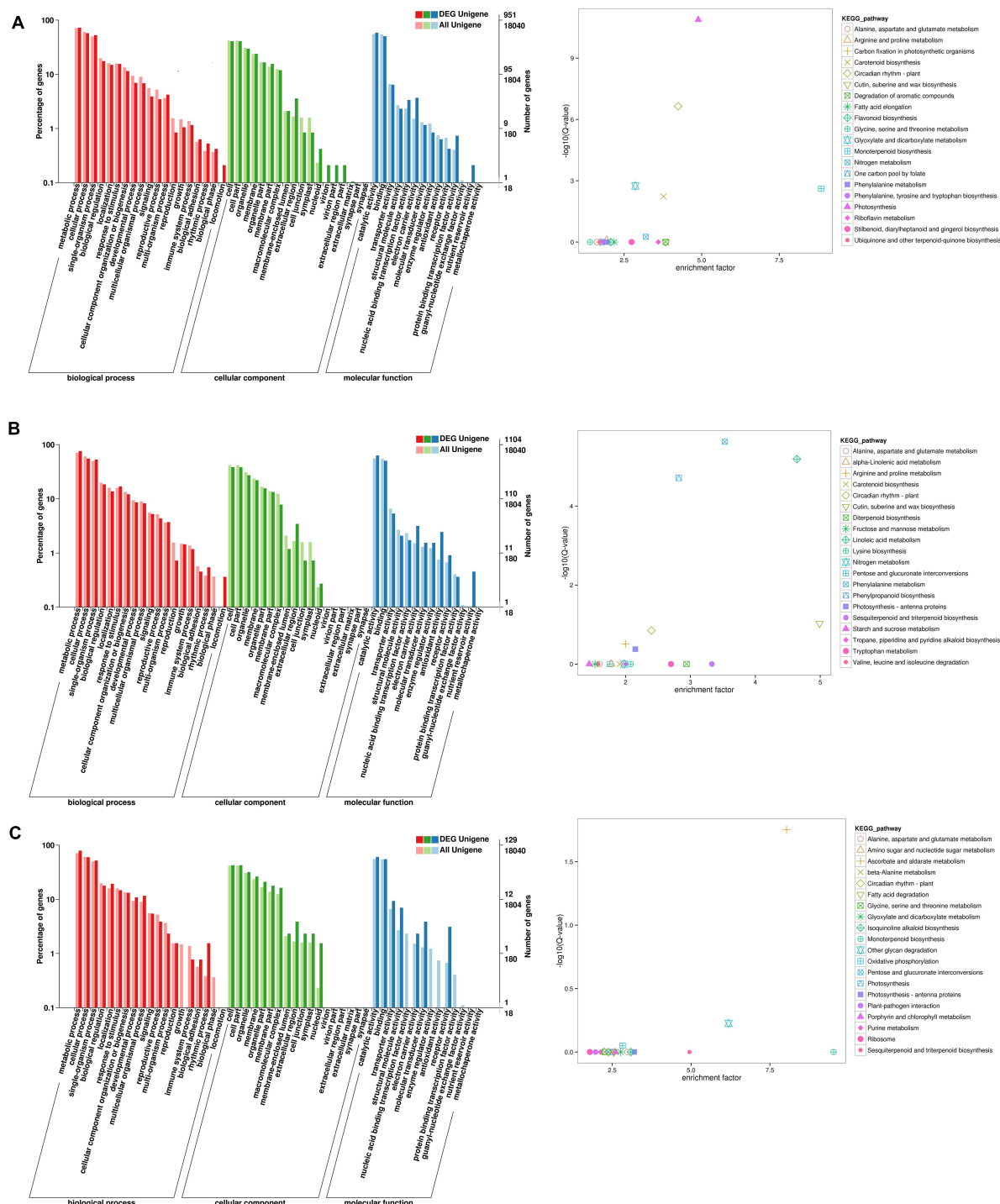


FIGURE 3 | GO and KEGG enrichment analyses of DEGs (move to Appendix). **(A)** GO and KEGG enrichment analyses of DEGs between cold treated dark cultured and incubators cultured faba bean buds. **(B)** GO and KEGG enrichment analyses of DEGs between cold treated dark cultured faba bean buds and control. **(C)** GO and KEGG enrichment analyses of DEGs between cold treated dark cultured then transferred to incubators and incubators cultured faba bean buds.

highly expressed in the cold-treated dark-cultured and control faba bean buds (**Supplementary Table 7**), respectively.

Three faba bean bud samples from the cold treatment and dark culture then transferred to incubators group were compared

to the three matched samples cultured in incubators in order to screen the key genes regulating the growth of faba bean that maintained expression. A total of 197 DEGs (**Figure 3C** and **Supplementary Table 8**) were identified with log2 fold

changes ranging from -11.7 to 10.1 . The most significantly enriched GO terms in biological process were “rhythmic process” and “growth.” In the molecular function category, the DEGs were significantly enriched in “structural molecule activity,” “molecular transducer activity,” “enzyme regulator activity,” and “receptor activity.” In addition, 21 unigenes were highly expressed in buds that were cold treated and dark cultured then transferred to incubators specifically, but were not expressed in the other treatments; seven unigenes were exclusively expressed in faba bean buds cultured in incubators.

Identification and Expression Analysis of Flower-Related DEGs in Faba Bean

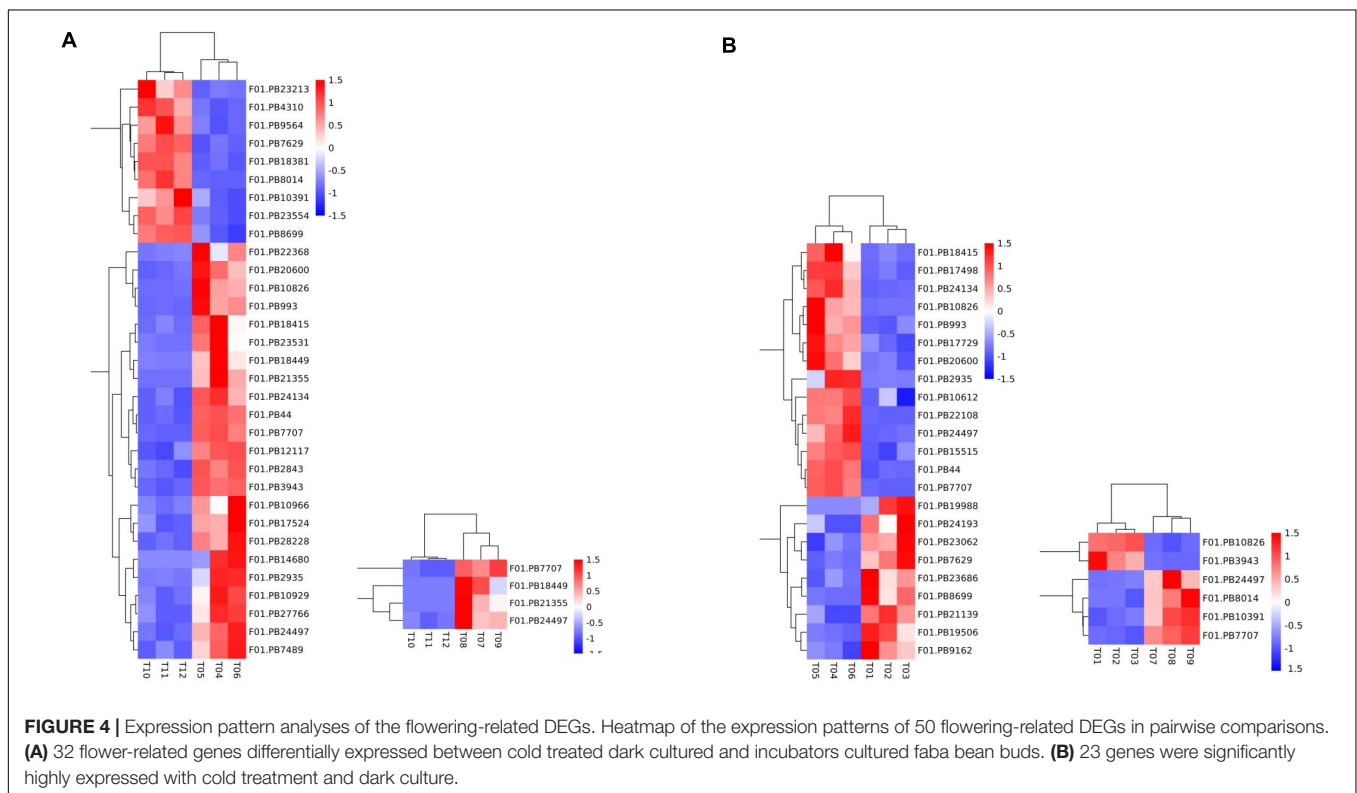
A total of 50 DEGs were functionally annotated as putative flowering-related genes (Supplementary Table 10). We analyzed the expression patterns of these differentially expressed transcripts (Figure 4) and identified 32 flower-related genes differentially expressed between faba bean buds that were cold treated dark cultured and buds cultured in incubators. Among them (Figure 4A), 23 genes were significantly highly expressed with cold treatment and dark culture (Figure 4B). *SOC1* (suppressor of overexpression of CO) is a key gene involved in the flowering pathway, while *ADO3* (Adagio protein 3) forms a complex with “GIGANTEA” (GI) to regulate “CONSTANS” (CO) expression and promotes CO expression during the light period of long daylight by decreasing the stability of *CDF1* and *CDF2* and by interacting directly with the CO protein and stabilizing it (Imaizumi et al., 2005). These genes may affect flower development when the faba bean buds were treated with cold.

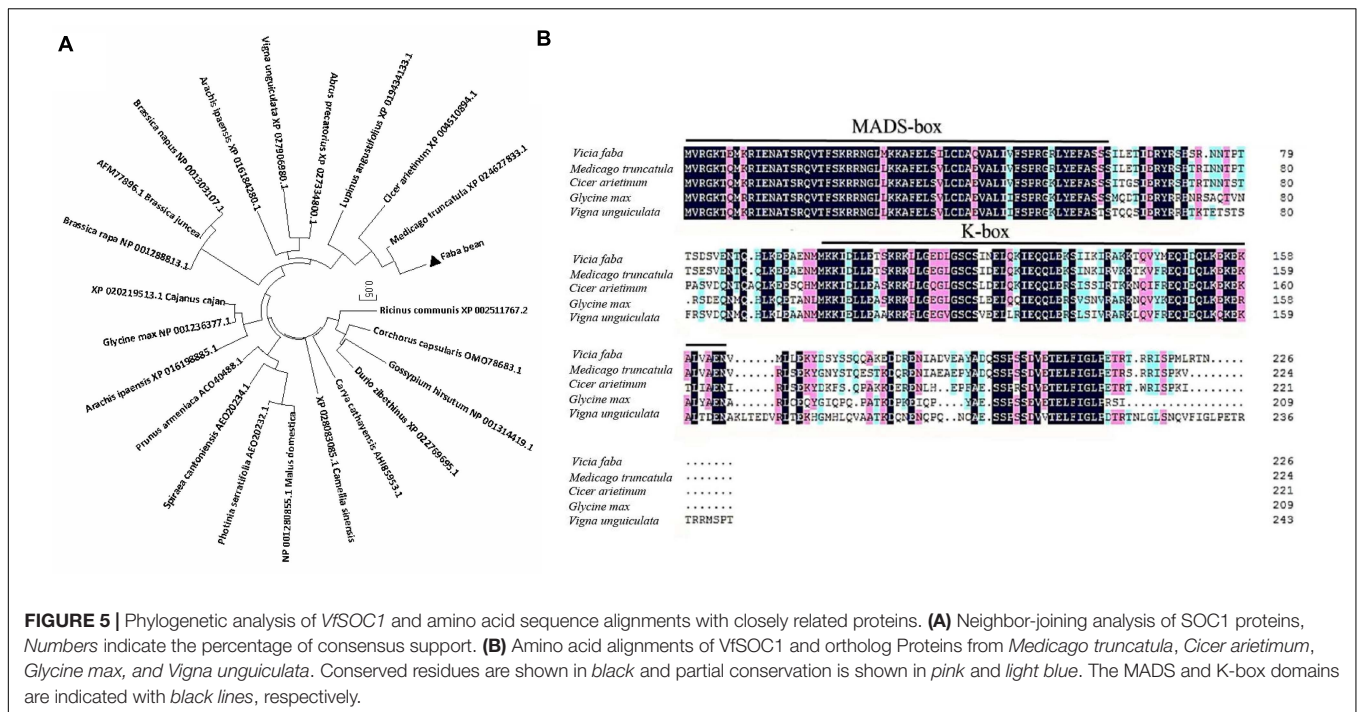
In cold-treated and dark-cultured buds vs. the 1 cm control, 23 flower-related genes were identified, 14 of which were significantly highly expressed in cold-treated faba bean. These three genes significantly highly expressed in cold-treated vs. buds cultured in incubators—*SOC1*, *ADO3*, and *GI*—were also highly expressed (Figure 4B). Interestingly, *COL5* (zinc finger protein CONSTANS-LIKE 5) and *CRY1* (cryptochromes) were also upregulated, and the stability of CO is positively regulated by the blue light receptor cryptochrome CRY1, suggesting that these genes may be involved in the process of flowering.

Only four flowering-related DEGs were identified in the samples treated with cold and dark cultured then transferred to incubators compared with the incubator-cultured samples, and all of them were upregulated in the former, including *PHYA* (*Phytochromes A*), which could positively regulate the stability of the CO protein (Figure 4A). *PHYA* has the same function as *CRY1*. Together, these two genes may be involved in the flowering pathway.

Validation of Flowering-Related Genes

The ORF of *VfSOC1* had 678 bases that encoded 226 amino acids. The full-length *VfSOC1* amino acid sequence was used as a query to search for predicted orthologs in GenBank. Phylogenetic analysis showed that *VfSOC1* is most similar to *MtSOC1c* in *Medicago*, with 86.5% identity (Figure 5A). *VfSOC1* contained a highly conservative N-terminal MADS domain and a *Keratin-like* domain (Figure 5B). In *Medicago*, the *SOC1* family genes had distinct and specific expression patterns (REF) (Jaudal et al., 2018). Thus, we determined the expression of





VfSOC1 in different tissues by RT-PCR; the results showed that *VfSOC1* was constitutively expressed in both vegetative and reproductive organs with high abundance in the leaves and flowers (**Figure 6A**). Since flowering is promoted by vernalization in faba bean, we checked the expression level of *VfSOC1* in the leaf under continuous low-temperature treatment. A pronounced elevation of the *VfSOC1* transcript was observed at day 7, and the expression level showed a persistent increase (**Figure 6B**). This indicated that *VfSOC1* expression is sensitive to vernalization in faba bean. To investigate the function of *VfSOC1*, a series of *Arabidopsis* transgenic plant lines expressing the CDS of *VfSOC1* driven by the CaMV 35S promoter were generated in the wild-type (*Col-0*) background. Three individual homozygous lines were obtained at T₂ generation, with an increasing expression level of *VfSOC1* (**Figures 6C,D**). Compared with the wild type, all three lines (*OE4-3*, *OE8-1*, and *OE13-1*) displayed an earlier flowering under the long-day condition (**Figure 6D**).

In summary, these results corroborate that *VfSOC1* can promote flowering and may play an important role in response to the vernalization of faba bean.

DISCUSSION

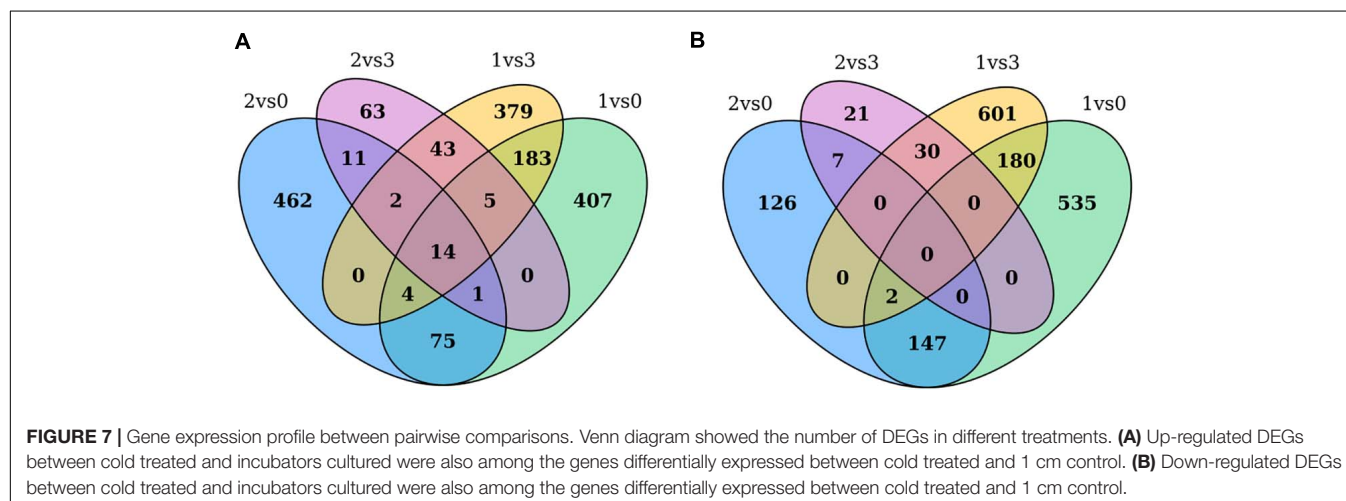
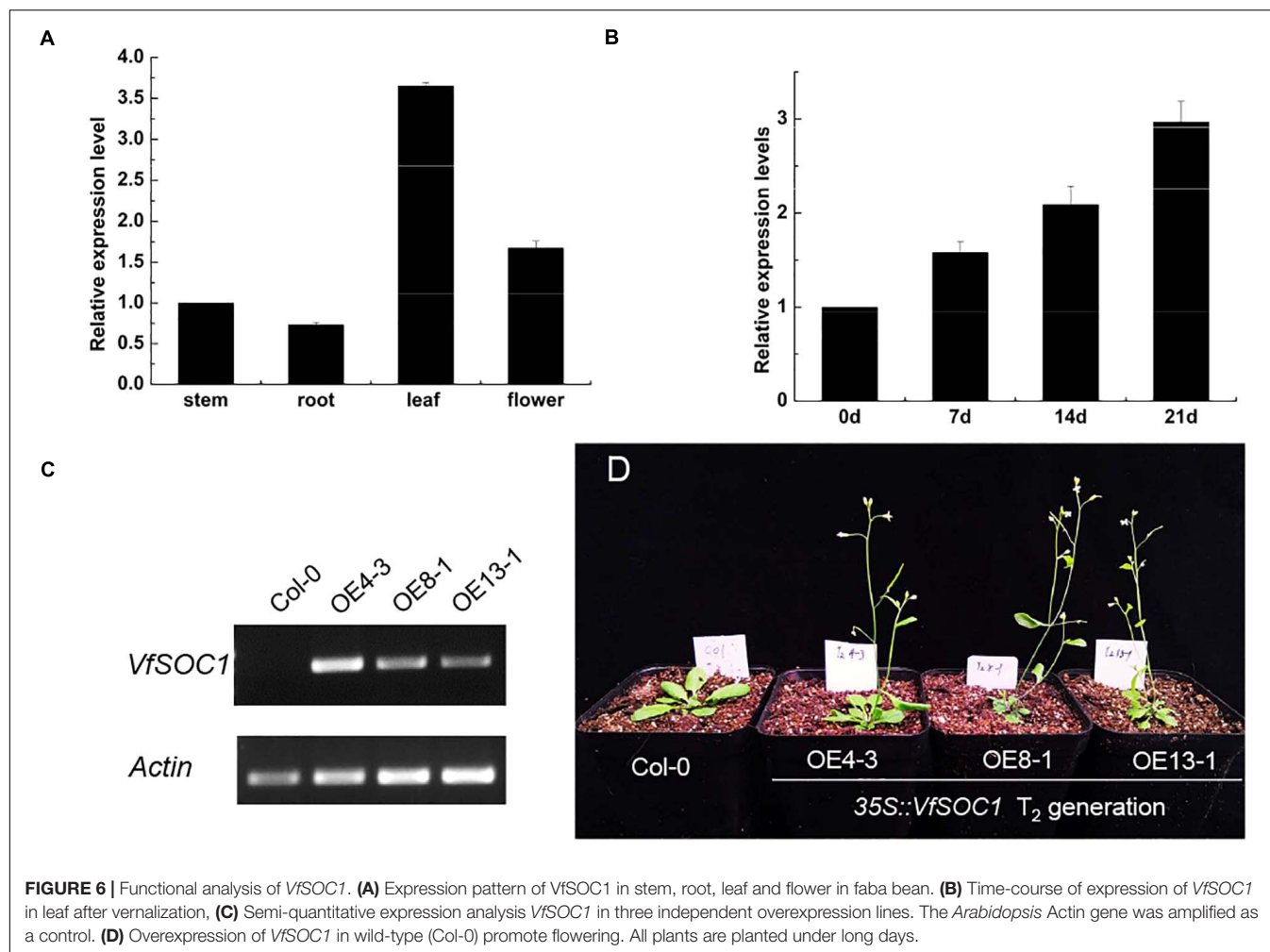
Compared with the next-generation sequencing (NGS) platform, the third-generation sequencing platform, SMRT sequencing, developed by PacBio RS is more appropriate for long reads with an average read length of greater than 10 kb, and the real length can reach 60 kb¹⁴. After the correction procedure

(self-corrected *via* CCS reads and corrected with the help of NGS reads), the error rate of SMRT sequencing is expected to be 1%. Furthermore, the assembly of the short NGS reads often encounters the complication of lacking reference genome sequences. This problem seems more severe for faba bean, which has a “giant genome” with an estimated genome size of greater than 13 Gb ($2n = 12$). Also, repetitive sequences in faba bean, such as retrotransposons and microsatellites, bring in further challenges for the accuracy of the short-read assembly. PacBio Iso-Seq can obtain a collection of high-quality full-length transcripts as reference sequences without assembly, thus overcoming these difficulties (Li and Godzik, 2006; Chin et al., 2013; Hackl et al., 2014).

In this study, NGS and the SMRT sequencing technology were combined to generate full-length transcriptomes of faba bean, and differentially expressed full-length transcripts of faba bean were identified and characterized. A total of 29,203 full-length faba bean transcripts were identified with an average length of 2,401.89 bp. Among these unigenes, 27,903 were functionally annotated with a functional annotation ratio of 95.5%. The high functional annotation ratio suggested that Iso-Seq could provide higher accuracy and is more effective in recovering full-length transcripts.

Vernalization plays an important role in the regulation of flowering time in higher plants, which is an evolutionarily adaptive mechanism preventing plant growth transition from the vegetative to the reproductive phase before winter and allowing flowering in favorable conditions in spring (Zhao et al., 2006). All winter crops generally have a vernalization requirement. In the past years, the vernalization process has been thoroughly studied in *Arabidopsis* (Michaels and Amasino, 1999; Gendall et al., 2001; Levy et al., 2002;

¹⁴www.pacb.com/smrt-science/smrt-sequencing/read-lengths/



Lee and Lee, 2010; Fadina and Khavkin, 2014; Eom et al., 2018) and in winter cereals such as wheat (Yan et al., 2003, 2006; Dubcovsky et al., 2006). However, the molecular regulatory mechanisms of vernalization in faba bean are still largely unknown. Recent advances in sequencing technologies

enable the generation of large volumes of sequencing data efficiently and cost effectively. Transcriptome analysis is a rapid approach to obtaining expressed gene sequences for faba bean, which has a large and poorly characterized genome. In this study, the transcriptome of faba bean under vernalization

or chilling treatment was sequenced using the combined approach of short-read NGS and long-read SMRT technology, and the genes were functionally annotated. It provided a basis for future researches on the vernalization responses and breeding of faba bean and established a reference full-length transcriptome for future studies on faba bean.

As SSR markers are important in variety identification, in this study, we identified 12,939 SSRs, although these SSRs have not been verified in natural populations. The SSRs were identified in vernalization-related traits, which would be useful in winter hardiness, DAF and yield traits in faba bean. In addition, the SSR markers identified in this study are useful for evaluating genetic relationships and the population structure of winter ecotype faba bean accessions of different origins and will facilitate utilizing various genetic resources.

Through pairwise comparisons, a total of 29,203 DEGs were identified in this study. To reveal the underlying molecular mechanism of buds of faba bean to flower under cold treatment, GO and KEGG pathway analyses of DEGs were conducted. DEGs were enriched in pathways of “binding transcription factor activity,” “electron carrier activity,” “receptor activity,” and “nutrient reservoir activity.” Among the 29,203 DEGs identified, 206 of the 630 upregulated DEGs between those cold treated and those cultured in incubators were also among the genes differentially expressed between those cold treated and the 1 cm control (**Figure 7A**). In addition, 182 of the 813 downregulated DEGs between those cold treated and incubator cultured were also among the genes differentially expressed between those cold treated and the 1-cm control (**Figure 7B**). Among these genes, the expression levels of *VfADO3*, *VfCRY1*, *VfSOC1*, and *VfGI* were significantly higher in sample 1 than those in sample 0 (**Supplementary Table 7**). Moreover, *VfSOC1* ($\log_2FC = -5.55$, $FDR = 1.15e-04$) was downregulated under sample 1 compared with sample 3 (**Supplementary Table 6**). In *Arabidopsis*, *SOC1* could be regulated by two antagonistic flowering regulators, *CONSTANS* (*CO*) and *FLC*, and acts as a floral activator and repressor (Lee and Lee, 2010). In order to study the function of *VfSOC1*, we implemented the transgenic work.

In the transgenic *Arabidopsis* lines (*OE4-3*, *OE8-1*, and *OE13-1*), all three lines displayed earlier flowering compared with *Col-0* (**Figures 6D,E**), which indicated that *VfSOC1* could promote flowering. Through this study, candidate genes related to vernalization were identified (Sallam et al., 2016; Catt et al., 2017).

Moreover, the molecular mechanism of *VfSOC1* is worth studying in the future. Further physiological and molecular experiments should be conducted to provide more precise insights into the mechanisms of vernalization in faba bean.

REFERENCES

- Alharbi, N. H., and Adhikari, K. N. (2020). Factors of yield determination in faba bean (*Vicia faba*). *Crop Pasture Sci.* 71, 305–321.
- Anders, S., and Huber, W. (2012). *Differential Expression of RNA-Seq Data at the Gene Level—the DESeq Package*. Heidelberg: European Molecular Biology Laboratory (EMBL).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI BioProject, accession no: PRJNA704197.

AUTHOR CONTRIBUTIONS

XY conceived the project. WS and XC supervised the work. XY, BY, and JZ performed most of the experiments, with assistance from QW, CX, JC, YL, and XZ. XC, XY, and QW wrote the manuscript, with contributions from all authors. All authors analyzed the data.

FUNDING

This work was supported by the China Agriculture Research System (CARS-08-G15) and the earmarked fund for Jiangsu Agricultural Industry Technology System [JATS, (2020)377].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.656137/full#supplementary-material>

Supplementary Table 1 | PacBio Iso-Seq data statistics.

Supplementary Table 2 | Summary of reads of insert (ROIs).

Supplementary Table 3 | Mapping statistics of Illumina RNA-seq data.

Supplementary Table 4 | Statistics of annotated transcripts in different databases.

Supplementary Table 5 | GO enrichment statistics.

Supplementary Table 6 | DEGs expressed in cold treated and dark cultured (sample 1) compared with incubators cultured faba bean buds (sample 3).

Supplementary Table 7 | DEGs expressed in cold treated and dark cultured faba bean buds (sample 1) compared with control (sample 0).

Supplementary Table 8 | DEGs expressed in cold treated and dark cultured then transferred to incubators (sample 2) compared with incubators cultured faba bean buds (sample 3).

Supplementary Table 9 | DEGs expressed in cold treated then transferred to incubators (sample 2) compared with control (sample 0).

Supplementary Table 10 | Flowering-related DEGs identified in this study.

- Cao, Y. Y., Bian, X. C., Chen, M. X., Xia, L. R., Zhang, J., Zhu, F. Y., et al. (2017). iTRAQ-based quantitative proteomic analysis in vernalization-treated faba bean (*Vicia faba* L.). *PLoS One* 12:e01874. doi: 10.1371/journal.pone.0187436
- Catt, S. C., Braich, S., Kaur, S., and Paull, J. G. (2017). QTL detection for flowering time in faba bean and the responses to ambient temperature and photoperiod. *Euphytica* 213, 1–13.

- Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474
- Dubcovsky, J., Loukoianov, A., Fu, D., Valarik, M., Sanchez, A., and Yan, L. (2006). Effect of photoperiod on the regulation of wheat vernalization genes VRN1 and VRNPlant. *Mol. Biol.* 60, 469–480.
- Eom, H., Park, S. J., Kim, M. K., Kim, H., Kang, H., and Lee, I. (2018). TAF15b, involved in the autonomous pathway for flowering, represses transcription of FLOWERING LOCUS C. *Plant J.* 93:79. doi: 10.1111/tpj.13758
- Fadina, O. A., and Khavkin, E. E. (2014). The vernalization gene FRIGIDA in cultivated *Brassica* species. *Russian J. Plant Physiol.* 61, 309–317. doi: 10.1134/s1021443714030030
- Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Fuchs, J., Strehl, S., Brandes, A., Schweizer, D., and Schubert, I. (1998). Molecular-cytogenetic characterization of the *Vicia faba* genome-heterochromatin differentiation, replication patterns and sequence localization. *Chrom. Res.* 6, 219–230. doi: 10.1023/a:1009215802737
- Gendall, A. R., Levy, Y. Y., and Wilson, A. (2001). The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. *Cell* 107:525. doi: 10.1016/s0092-8674(01)00573
- Hackl, T., Hedrich, R., Schultz, J., and Forster, F. (2014). Proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* 30, 3004–3011. doi: 10.1093/bioinformatics/btu392
- Heather, R., and Fawzy, G. (2010). A genomic approach to nutritional, pharmacological and genetic issues of faba bean (*Vicia faba*): prospects for genetic modifications. *GM Crops Food* 1, 99–106. doi: 10.4161/gmcr.1.2.11891
- Hepworth, J., Antoniou-Kourounioti, R. L., Bloomer, R. H., Selga, C., Berggren, K., Cox, D., et al. (2018). Absence of warmth permits epigenetic memory of winter in *Arabidopsis*. *Nat. Commun.* 9:639. doi: 10.1038/s41467-018-03065-7
- Imaizumi, T., Schultz, T. F., Harmon, F. G., Ho, L. A., and Kay, S. A. (2005). FKFI F-box protein mediates cyclic degradation of a repressor of CONSTANS in *Arabidopsis*. *Science* 309, 293–297.
- Jaudal, M., Zhang, L., Che, C., Li, G., Tang, Y., Wen, J., et al. (2018). A SOC1-like gene MtSOC1a promotes flowering and primary stem elongation in *Medicago*. *J. Exp. Bot.* 69, 4867–4880. doi: 10.1093/jxb/ery284
- Johnston, J. S., Bennett, M. D., Rayburn, A. L., Galbraith, D. W., and Price, H. J. (1999). Reference standards for determination of DNA content of plant nuclei. *Am. J. Bot.* 86, 609–613.
- Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., et al. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35, W345–W349. doi: 10.1093/nar/gkm391
- Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lazaro, A., Obeng-Hinneh, E., and Albani, M. C. (2018). Extended vernalization regulates inflorescence fate in *Arabidopsis* by stably silencing PERPETUAL FLOWERING. *Plant Physiol.* 176, 2819–2833. doi: 10.1104/pp.17.01754
- Lee, J., and Lee, I. (2010). Regulation and function of SOC1, a flowering pathway integrator. *J. Exp. Bot.* 61, 2247–2254. doi: 10.1093/jxb/erq098
- Levy, Y. Y., Mesnage, S., and Mylne, J. S. (2002). Gendall AR, dean C. multiple roles of *Arabidopsis* VRN1 in vernalization and flowering time control. *Science* 297:243. doi: 10.1126/science.1072147
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:3. doi: 10.1186/1471-2105-12-323
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Mahfooz, S., Limin, A. E., Ahakpaz, F., and Fowler, D. B. (2006). Phenological development and expression of freezing resistance in spring and winter wheat under field conditions in north-west Iran. *Field Crops Res.* 97, 182–187. doi: 10.1016/j.fcr.2005.09.012
- Mao, X., Cai, T., Olyarchuk, J. G., and Wei, L. (2015). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.1093/bioinformatics/bti430
- Michaels, S. D., and Amasino, R. M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11, 949–949. doi: 10.1105/tpc.11.5.949
- O'Sullivan, D. M., and Angra, D. (2016). Advances in faba bean genetics and genomics. *Front. Genet.* 7:1. doi: 10.3389/fgene.2016.001eCollection
- Qi, H., Xia, F. N., Xie, L. J., Yu, L. J., Chen, Q. F., Zhuang, X. H., et al. (2017). TRAF family proteins regulate autophagy dynamics by modulating AUTOPHAGY PROTEIN6 stability in *Arabidopsis*. *Plant Cell* 29, 890–899. doi: 10.1105/tpc.17.00056
- Rispail, N., Kal, P., Kiss, G. B., Ellis, T. H. N., Gallardo, K., Thompson, R. D., et al. (2010). Model legume contribute to faba bean breeding. *Field Crops Res.* 115, 253–269. doi: 10.1016/j.fcr.2009.03.014
- Sallam, A., Dhanapal, A. P., and Liu, S. (2016). Association mapping of winter hardiness and yield traits in faba bean (*Vicia faba* L.). *Crop Pasture Sci.* 67, 55–68.
- Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., et al. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 41:e166. doi: 10.1093/nar/gkt646
- Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J. P., and Li, W. (2013). CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res.* 41:e74. doi: 10.1093/nar/gkt006
- Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., et al. (2006). The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proc. Natl. Acad. Sci. U.S.A.* 103, 19581–19586.
- Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., and Dubcovsky, J. (2003). Positional cloning of the wheat vernalization gene VRN. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6263–6268. doi: 10.1073/pnas.09373991
- Ye, Y., Lang, L., Xia, M., and Tu, J. (2003). *Faba Bean in China*. Beijing: China Agriculture Press, 1.
- Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Geno. Biol.* 11:R14. doi: 10.1186/gb-2010-11-2-r14
- Zhao, Z., Zeng, Q., and Zhao, S. (2006). The molecular mechanism of vernalization in plants. *Chinese Bull. Bot.* 23:60. doi: 10.1360/aps040178

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

Copyright © 2021 Yuan, Wang, Yan, Zhang, Xue, Chen, Lin, Zhang, Shen and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Identification of Gray Leaf Spot Disease Candidate Gene in Narrow-Leafed Lupin (*Lupinus angustifolius* L.)

Gaofeng Zhou^{1,2}, Huaan Yang¹, Daniel Renshaw¹, Meilin Zou², Geoff Thomas¹ and Chengdao Li^{1,2*}

¹ Department of Primary Industries and Regional Development Western Australia, South Perth, WA, Australia, ² Western Crop Genetics Alliance, Murdoch University, Murdoch, WA, Australia

OPEN ACCESS

Edited by:

Longjiang Fan,
Zhejiang University, China

Reviewed by:

Ye Chu,
The University of Georgia,
United States
Paolo Franchini,
University of Konstanz, Germany

*Correspondence:

Chengdao Li
C.li@murdoch.edu.au

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 15 April 2021

Accepted: 06 July 2021

Published: 05 August 2021

Citation:

Zhou G, Yang H, Renshaw D,
Zou M, Thomas G and Li C (2021)
Identification of Gray Leaf Spot
Disease Candidate Gene in
Narrow-Leafed Lupin (*Lupinus
angustifolius* L.).
Front. Genet. 12:695791.
doi: 10.3389/fgene.2021.695791

Selection for resistance against gray leaf spot (GLS) is a major objective in the lupin breeding programs. A segregation ratio of 1:1 (resistant:susceptible) in F₈ recombinant inbred lines (RIL₈) derived from a cross between a breeding line 83A:476 (resistant to GLS) and a wild accession P27255 (susceptible to GLS) indicated that GLS was controlled by a single major gene. To develop molecular markers linked to GLS, in the beginning, only 11 resistant lines and six susceptible lines from the 83A:476 and P27255 population were genotyped with MFLP markers, and three MFLP markers were identified to be co-segregated with GLS. This method was very efficient, but the markers were located outside of the gene, and could not be used in other germplasms. Then QTL analysis and fine mapping were conducted to identify the gene. Finally, the gene was narrowed down to a 241-kb region containing two disease resistance genes. To further identify the candidate gene, DNA variants between accessions Tanjil (resistant to GLS) and Unicrop (susceptible to GLS) were analyzed. The results indicated that only one SNP was detected in the 241 kb region. This SNP was located in the TMV resistance protein N-like gene region and also identified between 83A:476 and P27255. Genotyping the Tanjil/Unicrop RIL₈ population showed that this SNP co-segregated with GLS resistance. The phylogenetic tree analysis of this gene among 18 lupin accessions indicates that Australian resistant breeding line/varieties were clustered into one group and carry two resistant alleles, while susceptible accessions were clustered into different groups.

Keywords: SNP, lupin, grey leaf spot, QTL, fine mapping, sequencing

INTRODUCTION

Molecular markers linked to genes of agronomic traits can be applied in plant breeding to hasten genetic improvement. Marker-assisted selection (MAS) requires that the markers should be cost effective, reliable, and be applicable in high-throughput systems for many samples. In addition, the marker genotypes must be consistent with the plant phenotypes across a diverse range of germplasm to enable wide application in breeding programs. Unfortunately, most molecular markers are developed from the outside of the gene region; in these cases, new recombination can happen between the gene and the marker in plant breeding programs. As a result, cultivars

exhibiting the ideal marker genotypes may not necessarily carry the targeted genes, which produce “false positive” (Sharp et al., 2001; Boersma et al., 2009). Due to the occurrence of “false positives” or “false negatives” in MAS practices, breeders must validate markers to determine which markers suit which crosses in breeding populations (Eagles et al., 2001; Sharp et al., 2001). This marker validation step can slow down the pace of MAS implementation and increase the cost in breeding programs. A key limiting factor of MAS in plant molecular breeding is to develop diagnostic markers that are applicable for a wide range of germplasm (Brown-Guedira, 2005). The best solution to the plight of a false result in MAS is to develop “diagnostic markers,” which have their genotypes consistent with their phenotypes. Diagnostic markers can be applied for MAS in breeding programs without the need of a marker validation step (Ogbonnaya et al., 2001). For functional genes where gene sequences are available, diagnostic PCR markers can be designed based on polymorphic sites within gene sequences (Zhou et al., 2013). For agronomic traits without the knowledge of gene sequences, one of the methodologies for developing widely applicable molecular markers is to design several markers linked to the target genes, followed by a validation step to test the candidate markers and select the best one for MAS (Yang et al., 2008). Using this concept, diagnostic markers have been developed and applied for anthracnose disease resistance (You et al., 2005) and phomopsis stem blight disease resistance (Yang et al., 2013a) selection in Australian national lupin breeding programs. However, the development of several markers associated with the genes of interest by gel-based DNA fingerprinting methods is very tedious and time consuming. Recently, the rapid advancement and the reduction in the cost of the next-generation sequencing technology (NGS) have made it feasible for many molecular laboratories to embrace the genome sequencing wave. Whole-genome sequencing is the ultimate approach to generate genome-wide, maximum number of molecular markers in plant species. For instance, over 55 million SNPs (single-nucleotide polymorphisms) were detected by genome sequencing and re-sequencing in maize (Chia et al., 2012; Jiao et al., 2014). Therefore, the whole-genome sequencing approach offers tremendous potential for developing diagnostic markers and/or identifying genes in plant breeding.

Narrow-leaved lupin (*Lupinus angustifolius* L.) was fully domesticated as a grain legume crop during the 1960s in Australia. Pedigree records showed that 31 varieties were released from 1967 to 2016. Lupin is a grain legume crop and provides a source of protein for animals and humans. Gray leaf spot (GLS) is a serious disease in narrow-leaved lupin (*Lupinus angustifolius* L.), caused by the fungal pathogens *Stemphylium botryosum* spp. (Vaghefi et al., 2020). It causes leaf lesions and defoliation (Figure 1) as well as infecting stems and pods. Dark-brown circular lesions can often appear on leaves and then progress to ash-brown necrosis, and in severe cases, plants lose all their leaves, which can lead to significant yield loss (Thomas et al., 2011). Genetic studies on three separate breeding populations in North America revealed that the resistance to GLS was regulated by a single gene (Forbes et al., 1961). The first three commercial cultivars (Uniwhite, Uniharvest, and

Unicrop) released in Australia before 1974 were susceptible to GLS. A collaborative breeding effort between USDA and the Department of Agriculture in Western Australia resulted in the introduction of the gray leaf spot-resistant gene into the Australian germplasm and the release of the first resistant variety in 1974. This resistant (R) gene had been integrated into all the subsequent commercial cultivars over the following 30 years (Thomas et al., 2011). The R gene *gl* is so effective against the disease that it has been effectively absent from Australian lupin crops for over 30 years; during that time, resistance screening for GLS was not carried out by the lupin breeding program. However, GLS re-emerged in Western Australia in 2006. Subsequent disease screening tests in 2008 and 2009 found that approximately 40% of the breeding lines in the Australian lupin breeding program were susceptible to GLS (Thomas et al., 2011). Since then, screening and selection for GLS has become an essential component of the program. The gel-based DNA fingerprinting method MFLP (microsatellite fragment length polymorphism) (Yang et al., 2001, 2002) has been applied in marker development for anthracnose and phomopsis resistance in lupin.

In this study, first, we used an efficient method and identified molecular markers linked to GLS by genotyping only 17 RIL lines. Second, we fine mapped the QTL region and identified the candidate gene for GLS using RIL₈ populations, and then an SNP diagnostic marker was developed within the gene region, and genotyping RIL population confirmed that 83A:476 and Tanjil shared the same QTL/locus. Finally, GLS alleles were investigated in 18 Australian varieties and European wild ecotypes, and a phylogenetic tree was constructed.

MATERIALS AND METHODS

Plant Materials

Two populations of F₈ recombinant inbred lines (RIL) were used in this study. They were developed using a single seed descent from their F₂ generations. The population (109 RILs) derived from a cross between the Australian domesticated breeding line 83A:476 (resistant to GLS) and Morocco wild-type P27255 (susceptible to GLS) was used for genetic map construction and QTL mapping for GLS confirmation. Another 93 RIL₈ population derived from the cross between two Australian cultivars Tanjil (resistant to GLS) and Unicrop (susceptible to GLS) was used to validate the QTL and the molecular markers in this study. Unicrop was one of the three susceptible varieties released prior to the introduction of GLS resistance, while Tanjil was released in 1998 as part of the lineage with the GLS resistance. A total of 18 historical and current commercial cultivars and breeding or wild ecotypes (Tanjil, Yorrel, Coromup, Mandelup, Merrit, 83A:476, Moonah, Unicrop, Kalya, Tallerack, P26167, P27255, Quillinock, 75A:258, P26603, Bo7212, P26668, and P27221) were used to analyze GLS alleles and the gene evolution. These germplasms are from Australian varieties and European wild ecotypes.

Phenotyping Gray Leaf Spot Resistance

Two isolates (*S. beticola*: WAC12986 and *S. vesicarium*: WAC13136) from the Western Australian Culture Collection



FIGURE 1 | Gray spot leaf disease and phenotyping in glasshouse. Left: disease in adult plant; Right: disease assessment in glasshouse at the seedling stage.

(Thomas et al., 2011; Vaghefi et al., 2020) were used for phenotyping in this study. Isolates were grown on 20% V8 agar medium at 22°C with a 12:12 h light/dark cycle for 14 days (Everts and Armentrout, 2001). Then spores were harvested and suspended in sterile distilled water with 0.1% Tween-20. The final concentration of spores was adjusted to $1.5 \times 10^5 \text{ ml}^{-1}$. Three replicate pots, each containing seven 14-day-old seedlings of each line were spray inoculated with the spore suspension. The pots were placed randomly. Inoculated plants were placed in an intermittent misting growth chamber with 90% of shade of natural daylight. Two days later, the plants were transferred to the greenhouse at 20°C. Disease scoring was conducted 14 days after inoculation. Disease resistance was scored on the first two fully expanded leaves on a 0–5 scale (0: no symptoms; 1: one to two lesions < 1 mm on each leaflet; 2: less than five small lesions on each leaflet; 3: 1- to 5-mm lesions, often coalescing with associated chlorosis; 4: some leaflets completely necrotic or fallen; 5: all leaflets completely necrotic or complete defoliation). Plants scoring 0 and 1 were considered resistant, while those scoring greater than 3 were considered susceptible (Thomas et al., 2011; Ahmad et al., 2016). Genotype effects on disease severity were analyzed by ANOVA with differences compared by LSD. The disease assessments of the two RIL populations were carried out in 2008, 2009, and 2013.

Identification of Markers Linked to Gray Leaf Spot Resistance by Microsatellite Fragment Length Polymorphism

Identification of MFLP markers linked to GLS resistance followed the same method as described by previous studies (Yang et al., 2002; Lin et al., 2009). MFLP method combines the concept of microsatellite-anchor primer and AFLP technique. Genomic DNA was digested by restriction enzyme *MseI*. One AFLP *MseI* adaptor was ligated onto the restriction fragments. PCR was performed using one *MseI* primer in combination with one microsatellite-anchor primer (Yang et al., 2001). Seventeen plants were used in the MFLP analysis. Eleven of these plants, including parental line 83A:476 and 10 RILs, were resistant to GLS. The other six plants, including parental line P27255 and five RILs, were susceptible to GLS. These plants were genotyped with MFLP

markers. The genotypes of the markers close to GLS should be associated with the disease resistance.

Sequencing Microsatellite Fragment Length Polymorphism Amplicons

PCR bands from the MFLP markers linked to GLS were excised from MFLP gels and boiled for 1 min in TE buffer. The products were re-amplified by PCR, and electrophoresed in 1% agarose gel. DNA fragments were isolated with agarose gel and purified with a gel extraction kit (Qiagen), and then sequenced using the BigDye Terminator system (Applied Biosystems). Blastn searches were performed on the sequences in the lupin genome database (Wang et al., 2020). The chromosome and physical positions of MFLP markers were anchored.

Sequencing

The lupin parental lines (Unicrop, 83A:476, and P27255) were shotgun sequenced with 10 times coverage using the second-generation sequencing method, and the sequences were obtained

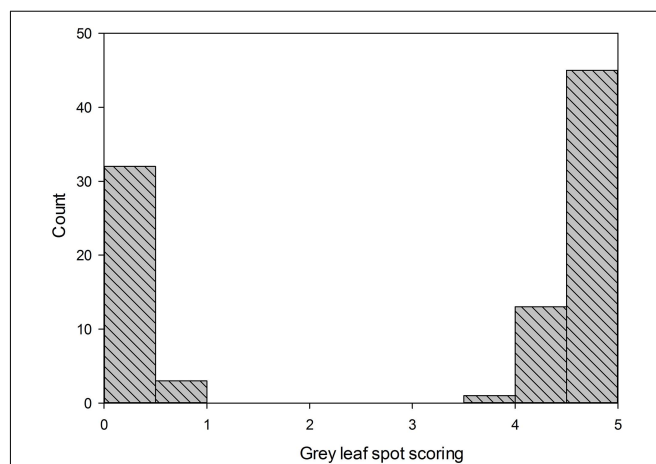


FIGURE 2 | Distribution of gray leaf spot scoring in F8 recombinant inbred lines (RILs) population. Line 83A:476 score was 0 (resistant), and wild-type P27255 was 4.65 (susceptible); $n = 111$.

TABLE 1 | Microsatellite fragment length polymorphism (MFLP) markers linked to gray leaf spot (GLS) resistance.

Lines	Source	Marker-D145 (LG-19 3.097 Mb)	Marker-D280 (LG-19 3.033 Mb)	Marker-D300 (LG-19 3.033 Mb)	GLS score
83A:476	Parent 1	+	–	–	0.0
Line13	RILs	+	–	–	0.0
Line14	RILs	+	–	–	0.36
Line19	RILs	+	–	–	0.0
Line20	RILs	+	–	–	0.0
Line35	RILs	+	–	–	0.0
Line49	RILs	+	–	–	0.0
Line74	RILs	+	–	–	0.0
Line81	RILs	+	–	–	0.0
Line95	RILs	+	–	–	0.0
Line105	RILs	+	–	–	0.0
P27255	Parent 2	–	+	+	4.6
Line28	RILs	–	+	+	4.9
Line29	RILs	–	+	+	4.9
Line70	RILs	–	+	+	4.6
Line87	RILs	–	+	+	4.8
Line88	RILs	–	+	+	4.7

from the China National GeneBank Database¹ and were available under the project accession CNP0001034. The RIL₈ population

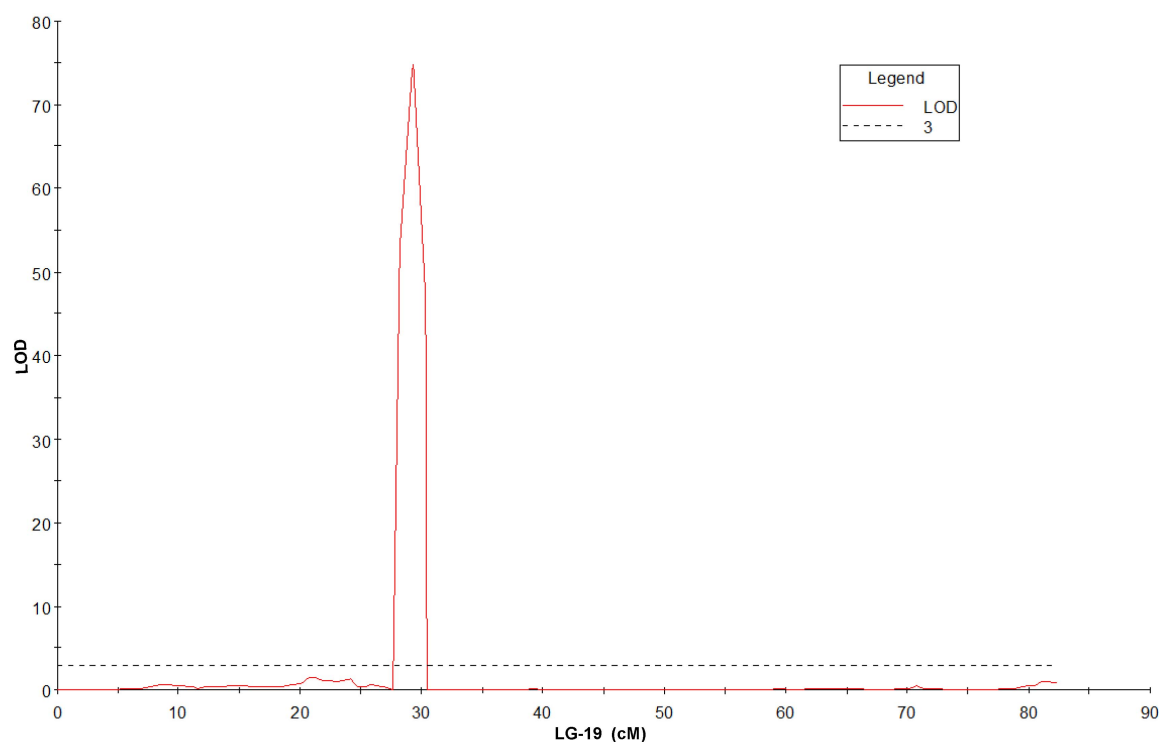
¹<https://db.cngb.org>

from the cross of 83A:476 and P27255 were genotyped using the RAD method (Yang et al., 2013b; Zhou et al., 2018). The protocol was described by Chutimanitsakun et al. (2011) with minor modifications that *EcoRI* was used. Sequencing libraries (100 bp) were constructed by using a unique multiplex identifier (MID) barcode (Baird et al., 2008). The RAD product sequencing was carried out on Solexa HiSeq2000. Sequencing raw data were divided into different lupin lines according to the MID barcodes (Baird et al., 2008), and then the MID barcodes were removed from the raw sequencing data. The 92-bp RAD reads with the same DNA sequences were clustered into one read tag. Tags containing over 100 reads were considered to be from repetitive regions, so they were removed to avoid the detection of SNP/indel variants (Catchen et al., 2011). Genome assembly of the two parental lines was conducted before alignment.

QTL Mapping and Fine Mapping

MapQTL5.0 was used to detect QTL for gray leaf spot disease. The genetic map of the population between 83A:476 and P27255 was obtained from our previous studies (Zhou et al., 2018). Formatted files including the genetic map, genotypic data, and phenotypes were imported into MapQTL5.0. The threshold of LOD value for QTL mapping was 3.0. First, interval mapping was performed, and then the marker with the highest LOD value was selected for MQM mapping analysis.

High-density markers across the QTL region were genotyped in the RIL population. The recombinant lines were selected to narrow down the QTL for gray leaf spot resistance. The

**FIGURE 3** | QTL for gray leaf spot disease resistance on LG-19 in narrow-leaved lupin (LOD = 3).

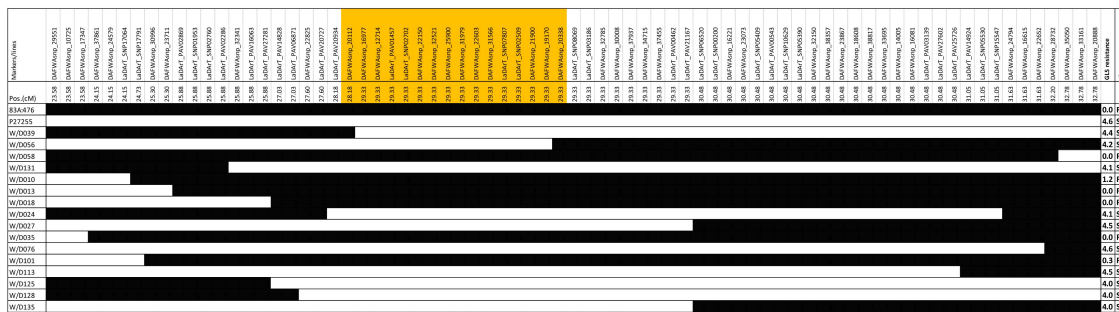


FIGURE 4 | Recombinant lines for fine-mapping of gray leaf spot resistance.

left border and right border of the gene were determined by comparing the genotypes and phenotypes.

Variant Detection

The software CLC Genomic Workbench 21.0.3 was used to map the sequence data to the reference narrow-leaved lupin “Tanjil” genome. The 17 accessions excluding the reference cv. Tanjil were shotgun sequenced with 10 times coverage. All the sequencing read data were obtained from the China National GeneBank Database (see text footnote 1) and were available under the project accession CNP0001034. The parameters were as follows: match score 1, mismatch cost 2, linear gap cost, length fraction 0.5, similarity fraction 0.8, and non-specific match handling with map randomly. Tracks were created, and all the genome sequences used in this study were combined. The SNP and indel variants were detected and generated. For variant detection, fixed ploidy variant parameters were as follows: ploidy 2 and required variant probability 90%, ignore broken pairs, minimum coverage 10, minimum count 2, and minimum frequency 20%.

Phylogenetic Tree Construction

Gray leaf spot (GLS) alleles and gene evolution were investigated among 18 narrow-leaved lupin accessions. SNPs were identified by mapping their reads to the Tanjil reference genome. SNP variants were obtained from the variants detection. The SNP genotype data were imported into Geneious 8.0 for phylogenetic tree construction. Geneious Tree Builder function was used, and the parameters were as follows: genetic distance model of Jukes–Cantor with the build method of UPGMA.

RESULTS

Single Gene Controls Gray Leaf Spot Resistance

For the gray leaf spot disease phenotypes, all the plants of parental line 83A:476 were resistant to GLS (scoring 0), while all the plants of P22755 were susceptible (scoring 4 or 5, mean 4.6). The parental phenotypes were significantly different ($p < 0.05$). Among the 109 tested F_8 RILs, 51 RILs were susceptible to GLS, and the other 58 RILs were resistant (Figure 2). The segregation of susceptible:resistant in the F_8 population fit the expected 1:1

ratio ($\chi^2 = 0.33$, $\chi^2_{0.05,1} = 3.84$, ($\chi^2 < \chi^2_{0.05,1}$), indicating that a single gene controls GLS resistance in 83A:476.

Identification of Microsatellite Fragment Length Polymorphism Markers Linked to Gray Leaf Spot Resistance

In the beginning, we wanted to develop molecular markers linked to GLS and use immediately in breeding programs, so an efficient method was adopted to identify the locus. A total of 1,657 polymorphic MFLP markers were identified between parental

TABLE 2 | List of annotated genes in the fine-mapping region.

Lupin gene ID	Annotated ID	Description
XM_019569076.1	LOC109333574	Probable sugar phosphate/phosphate translocator At5g04160
xm_019569260.1	LOC109333709	Uncharacterized
XM_019569303.1	LOC109333739	Uncharacterized
XM_019569319.1	LOC109333752	Zinc finger protein WIP2-like
XM_019569320.1	LOC109333754	Syntaxin-121-like
xm_019569325.1	LOC109333759	Zeatin O-glucosyltransferase-like
XM_019569329.1	LOC109333763	Remorin
XM_019569426.1	LOC109333847	Mulatexin
xm_019569492.1	LOC109333910	Uncharacterized mitochondrial protein AtMg00810-like
XM_019569493.1	LOC109333911	F-box/kelch-repeat protein At3g23880-like
XM_019570034.1	LOC109334326	TMV resistance protein N-like
XM_019570035.1	LOC109334327	Putative disease resistance protein RGA3
xm_019570323.1	LOC109334506	Uncharacterized
XM_019570326.1	LOC109334508	Uncharacterized
XM_019570390.1	LOC109334557	Adenylosuccinate synthetase 2, chloroplastic
xm_019570483.1	LOC109334615	1-acyl-sn-glycerol-3-phosphate acyltransferase 2
XM_019570535.1	LOC109334656	DNA-binding protein SMUBP-2
XM_019570536.1	LOC109334658	Dehydration-responsive element-binding protein 2F
XM_019570540.1	LOC109334660	Uncharacterized
xm_019570782.1	LOC109334820	Guanylate kinase 2
XM_019570938.1	LOC109334924	Calcium-dependent protein kinase 20

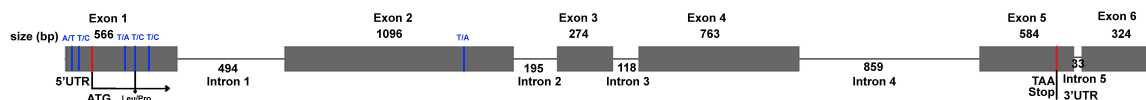


FIGURE 5 | Candidate gene LOC109334326 DNA structure. Introns and exons are represented with continuous lines and gray rectangular boxes. Start codon and stop codon are marked with red lines. 5'-UTR and 3'-UTR are shown in the figure. SNP variations are marked with blue lines and letters (83A:476 on the left, and P27255 on the right). One amino acid substitution is indicated in the figure.

lines 83A:476 and P27255. These markers were genotyped in 10 resistant RILs and 5 susceptible RILs as well. Among these markers, three markers (D145, D280, and D300) co-segregated with GLS resistance in these RILs (Table 1). All the resistant RILs exhibited the same genotype to 83A:476, and the susceptible RILs showed the same genotypes to P27255.

Sequencing of Linked Microsatellite Fragment Length Polymorphism Markers and Anchor to Physical Map

The PCR products from PAGE gel were amplified and sequenced. The amplicon sequences of the three markers are listed in Supplementary Table 1. The markers did not perform very well (data not shown) among the lupin germplasms maybe because these markers were located outside of the GLS gene, so we continued to identify the gene and develop diagnostic markers. Recently, the release of the lupin genome sequence enables us to anchor the MFLP markers to their physical positions and to mine the genes based on the lupin genome (Wang et al., 2020). The sequences excluding forward and reverse primers were used to conduct a blast in the lupin cv. Tanjil genome. All the three markers (D145, D280, and D300) were anchored to the lupin linkage group LG-19 with a physical position of 2.156–2.218 Mb region. D280 and D300 were anchored to the 2.156 Mb region, while D145 was mapped to the 2.218 Mb region.

QTL Mapping and Fine Mapping

QTL mapping and fine mapping were used to validate the candidate gene. The genetic map from 83A:476 and P27255 population was constructed in our previous studies (Zhou et al., 2018). The genetic map contains 27,892 markers (Supplementary Data Sheet 1). A simplified map containing one marker for each locus was used for QTL mapping. QTL analysis was conducted in the RIL population. A single major QTL for gray leaf spot resistance was identified on chromosome 19. The closest marker DAFWAsnp_16977 (29.33 cM) can explain 97.9% phenotypic variations and with the LOD value of 74.83 (Figure 3 and Supplementary Figure 1).

The recombinant lines around the QTL region (23.58 to 32.78 cM) were investigated in the RILs population. A total of 16 recombinant lines were obtained based on the genotypes (Figure 4). Figure 4 showed that the genotypes of QTL marker DAFWAsnp_16977 were consistent with their gray leaf spot resistance in this population. From these recombinant lines, the gene was narrowed down by the lines W/D039 and W/D056. The recombination point of W/D039 was between the markers DAFWAsnp_20112 (R genotype) and DAFWAsnp_16977 (S

genotype), while the GLS resistance of this line was susceptible, so the gene must be on the right side of the marker DAFWAsnp_20112. For the line W/D056, the recombination point was between the markers DAFWAsnp_19170 (S genotype) and DAFWAsnp_20338 (R genotype), and the GLS resistance was susceptible, so the gene should be on the left side of the marker DAFWAsnp_20338. In a word, the gene was narrowed down between the two markers DAFWAsnp_20112 and DAFWAsnp_20338. These markers were anchored to narrow-leaved lupin cv. Tanjil genome, and the physical positions of the markers were 2,157,779 and 2,399,114 bp, respectively. The QTL was finally fine-mapped to a 241-kb region. The three MFLP markers (D145, D280, and D300) overlapped with this fine mapping region.

Genes in the Region

A total of 21 genes were annotated in the QTL mapping region (Table 2). Two genes, LOC109334326 and LOC109334327, were associated with disease resistance. LOC109334326 is a TMV resistance protein N-like, and LOC109334327 is a putative disease resistance protein RGA3 (Wang et al., 2020).

Candidate Gene Identification

Genomic sequence analysis revealed that there were six SNPs in the LOC109334326 region between 83A:476 and P27255

TABLE 3 | Variations in the gene LOC109334327 region.

Mapping	Position (bp)	Length (bp)	83A:476 (R)	P27255 (S)	Unicrop (S)
LG19	2178851	1	C	T	C
LG19	2179252	2	CT	GG	CT
LG19	2179279	1	A	G	A
LG19	2179290	1	G	C	G
LG19	2179876	1	A	G	A
LG19	2179880	1	G	C	G
LG19	2180038	1	T	G	T
LG19	2180046	1	T	G	T
LG19	2180445	1	T	C	T
LG19	2180462	1	G	A	T
LG19	2180774	1	G	A	G
LG19	2180780	1	T	C	T
LG19	2181125	1	T	C	T
LG19	2181150	1	G	T	G
LG19	2181221	1	T	C	T
LG19	2181386	1	A	G	A
LG19	2181491	1	T	G	A

(Figure 5). Two SNPs were in the 5'-UTR regions, while the other four SNPs were in the coding region. Only one SNP (T/C) at 2,190,169 in the coding region induces amino acid substitution from leucine in 83A:476 (R) to proline in P27255 (S) (Figure 5).

To further investigate these two genes, we used two additional Australian cultivars Tanjil and Unicrop. Tanjil is resistant to gray leaf spot disease, while Unicrop is susceptible. Unicrop was re-sequenced and mapped to the Tanjil reference genome. The segregation of susceptible:resistant in the F₈ population fit the expected 1:1 ratio ($\chi^2 = 3.76$, $\chi^2_{0.05,1} = 3.84$), ($\chi^2 < \chi^2_{0.05,1}$). It indicated that a single gene controlled GLS resistance in the Tanjil and Unicrop population. Interestingly, the result indicated that there was only one SNP variation from T (Tanjil) to C (Unicrop) at 2,190,196 bp in the 241-kb fine-mapping region on chromosome 19, and this SNP variation from the gene LOC109334326 results in the same amino acid substitution from leucine in Tanjil (R) to proline in Unicrop (S). It indicates that this is the candidate gene for gray leaf spot resistance.

For another disease resistance RGA3 gene LOC109334327, 17 SNPs were detected between the parental lines 83A:476 and P27255 (Table 3), but no variations exist between Tanjil and Unicrop. Furthermore, no SNPs were identified between the

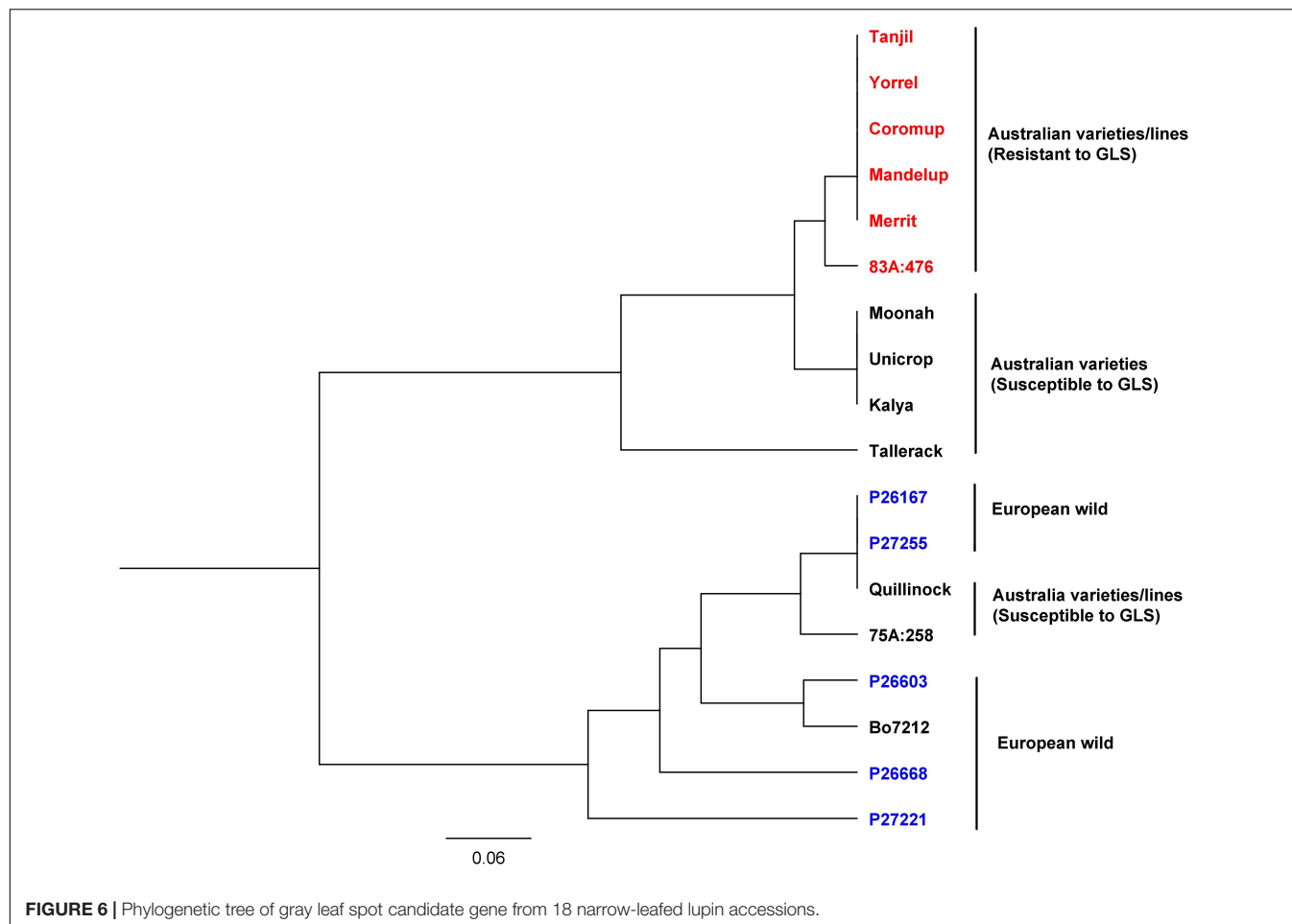
tolerant line 83A:476 and the susceptible cultivar Unicrop. It indicates that this gene is not the candidate gene.

Single-Nucleotide Polymorphism Marker Development for Gray Leaf Spot Resistance

To confirm the Tanjil carrying the same GLS QTL/locus, which was mapped from 83A:476 and P27255 population, a dominant SNP resistant marker was developed for the GLS candidate gene (forward: TCTGATGGCGTACATGTGTAA, reverse: TACTCTGCCCTCACTGACCT). This marker was tested in 93 RIL₈ derived from a cross between Tanjil and Unicrop. The genotypes co-segregated with their GLS resistance (Supplementary Table 2).

Phylogenetic Tree

The gene alleles were investigated among Australian and European lines. SNP variants of the gene (LOC109334326) were analyzed among 18 narrow-leaved lupin accessions consisting of Australian varieties, Australian breeding lines, and European wild lupin (Tanjil, Yorrel, Coromup, Mandelup, Merrit, 83A:476, Moonah, Unicrop, Kalya, Tallerack, P26167, P27255, Quillinock, 75A:258, P26603, Bo7212, P26668, P27221).



Quillinock, 75A:258, P26603, Bo7212, P26668, and P27221). The genomic sequences of these accessions were obtained from the China National GeneBank Database (see text footnote 1) and were available under the project accession CNP0001034. The sequence reads were mapped to the reference genome Tanjil, and SNP variants were obtained (see **Supplementary Data**). The phylogenetic tree (**Figure 6**) showed that they were classified into two major groups. All the European wild lupin accessions, together with the Australian variety/breeding lines Quillinock and 75A:258, were in one group, while all the other Australian varieties/breeding lines were clustered into another group. Furthermore, all the resistant (Tanjil, Yorrel, Coromup, Mandelup, Merrit, and 83A:476) and susceptible varieties (Moonah, Unicrop, Kalya, and Tallerack) from the major group were divided into different sub-groups. For the resistant gene, two alleles were identified. Australian resistant cultivars Tanjil, Yorrel, Coromup, Mandelup, and Merrit carried the same allele, while the breeding line 83A:476 carried a different GLS allele. It indicated that new GLS-resistant allele from lupin germplasm was incorporated into the breeding line 83A:476 during the breeding program.

DISCUSSION

In the present study, the rapid mapping method was used to develop markers for the GLS resistance gene in lupin. Based on the GLS segregation ratio of 1:1 in 109 RILs, it was confirmed that only one major resistance locus existed in this population. At the first step, only 15 RILs from resistance and susceptible groups were selected for gene mapping. This method significantly reduced cost and saved time. Eleven resistant lines and six susceptible lines were genotyped with MFLP markers. A total of 1,657 polymorphic markers were identified in the RILs. Finally, three co-segregated MFLP markers (D145, D280, and D300) were identified.

The sequences of the three MFLP amplicons were obtained and anchored to the lupin linkage group LG19 with the physical position of 2.156 and 2.218 Mb by running blast against the lupin cv. Tanjil reference genome (Wang et al., 2020). It indicated that the three MFLP were in the same chromosome and the same region in lupin.

To compare the result with the QTL mapping result. The bio-parental lines were re-sequenced using a RAD method. A major advantage of RAD sequencing is that it can generate tens of thousands of SNPs and InDels (insertions and deletions), which can be used for MAS in plant-breeding programs. Furthermore, all the variants detected from RAD sequencing have had the DNA sequenced already. Therefore, conversion of variants into PCR-based markers does not need PCR amplification, ligation, transformation, and sequencing as required from traditional markers such as AFLP (Shan et al., 1999), RAPD (Paran and Michelmore, 1993), and MFLP (Yang et al., 2002). One single QTL was detected, and the closest marker DAFWAsnp_16977 was mapped to 2.167 Mb on chromosome LG19. This result is consistent with the positions of the MFLP markers (2.156–2.218 Mb). It indicated that rapid

genotyping of a few lines can be an effective and efficient way to develop disease-linked markers if the trait is controlled by a single major gene, and phenotypes are very stable for quantitative traits.

A high-density genetic map and many population lines will facilitate us to narrow down the QTL region. In this study, the QTL was mapped to a 241-kb region. There were two genes (LOC109334326 and LOC109334327) associated with disease resistance. Even though there were 17 variations in LOC109334327 between 83A:476 and P27255 (**Table 3**), no variations were detected between Tanjil (R) and Unicrop (S) in the gene region including 5'-UTR and 3'-UTR. Furthermore, no variations were identified between the tolerant breeding line 83A:476 and the susceptible variety Unicrop. It indicated that this gene was not associated with gray leaf spot resistance. For the other gene LOC109334326, one SNP variation resulted in one amino acid substitution between 83A:476 (R) and P27255 (S), and this SNP was also the only variation between Tanjil (R) and Unicrop (S) in the 241-kb QTL region based on re-sequence mapping result. We assumed that Tanjil and 83A:476 shared the same gene or locus. After the candidate gene was identified, molecular markers should be used to validate that the gene marker co-segregated with the resistance in the Tanjil and Unicrop population.

A dominant SNP marker for GLS was developed based on the SNP variation in the candidate gene LOC109334326 and validated in the RIL population derived from the cross between Tanjil and Unicrop. The genotypes co-segregated with their phenotypes in the RIL population. It indicated that Tanjil and 83A:476 could have the same gene or locus. Currently, there are several methods for SNP marker development, such as McSNP (melting curve analysis of SNPs), KASP (kompetitive allele-specific PCR), and CAPS (cleaved amplified polymorphic sequences). This dominant marker can be converted to these kinds of markers depending on the laboratory instruments.

Phylogenetic tree results indicated that the resistance allele of GLS candidate gene was different from the alleles from the European wild ecotypes because the resistant allele was found from a United States wild ecotype of *Lupinus angustifolius*, which was later incorporated into the Australian varieties since the 1970s (Thomas et al., 2011). Two resistant alleles indicated that the breeding line 83A:476 incorporated a new allele from our lupin germplasm.

The results demonstrated that RAD sequencing is a rapid approach to identify molecular markers closely linked to the trait of interest for MAS in breeding programs. Whole-genome re-sequencing can facilitate us to identify the DNA variations and candidate genes. The transgenic complementation of the candidate gene and GLS resistance mechanism will be studied in the future.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

HY, DR, and GT performed the genotyping and phenotyping. GZ analyzed the data and wrote the manuscript. MZ analyzed the phylogenetic tree. HY and CL supervised this study. All authors have discussed and revised the manuscript.

ACKNOWLEDGMENTS

This work was conducted with support from the Western Australian State Agricultural Biotechnology Centre.

REFERENCES

- Ahmad, A., Thomas, G. J., Barker, S. J., and MacLeod, W. J. (2016). Genotype resistance, inoculum source and environment directly influence development of grey leaf spot (caused by *Stemphylium* spp.) and yield loss in narrow-leaved lupin (*Lupinus angustifolius*). *Crop Pasture Sci.* 67, 81–90. doi: 10.1071/cp15073
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376. doi: 10.1371/journal.pone.0003376
- Boersma, J., Nelson, M., Sivasithamparan, 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
- Brown-Guedira, G. (2005). Integrating MAS with conventional wheat breeding programs. *Phytopathology* 95:S139.
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: building and genotyping loci *de novo* from short-read sequences. *G3 Genes Genom. Genet.* 1, 171–182. doi: 10.1534/g3.111.000240
- Chia, J. M., Song, C., Bradbury, P. J., Costich, D., de Leon, N., Doebley, J., et al. (2012). Maize *HapMap2* identifies extant variation from a genome in flux. *Nat. Genet.* 44, 803–U238.
- Chutimanitsakun, Y., Nipper, R. W., Cuesta-Marcos, A., Cistue, L., Corey, A., Filichkina, T., et al. (2011). Construction and application for QTL analysis of a restriction site associated DNA (RAD) linkage map in barley. *BMC Genom.* 12:4. doi: 10.1186/1471-2164-12-4
- Eagles, H. A., Bariana, H. S., Ogonnaya, F. C., Rebetzke, G. J., Hollamby, G. J., Henry, R. J., et al. (2001). Implementation of markers in Australian wheat breeding. *Aust. J. Agric. Res.* 52, 1349–1356. doi: 10.1071/ar01067
- Everts, K. L., and Armentrout, D. K. (2001). Report of leaf spot of spinach caused by *Stemphylium botryosum* in Maryland and Delaware. *Plant Dis.* 85:1209. doi: 10.1094/pdis.2001.85.11.1209b
- Forbes, I., Wells, H. D., Edwards, D., and Ostazeski, S. (1961). Inheritance of resistance to gray leaf spot disease and of an antho-cyanin inhibitor in blue lupines. *Crop Sci.* 1, 1037–1038.
- Jiao, Y. P., Zhao, H. N., Ren, L. H., Song, W. B., Zeng, B., Guo, J. J., et al. (2014). Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* 46, 1039–1040. doi: 10.1038/ng0914-1039
- Lin, R. M., Renshaw, D., Luckett, D., Clements, J., Yan, G. J., Adhikari, K., et al. (2009). Development of a sequence-specific PCR marker linked to the gene “*pauper*” conferring low-alkaloids in white lupin (*Lupinus albus* L.) for

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.695791/full#supplementary-material>

Supplementary Figure 1 | QTL for gray leaf spot disease resistance in narrow-leaved lupin (Threshold LOD = 3).

Supplementary Table 1 | MFLP marker amplicon sequence.

Supplementary Table 2 | SNP genotyping and GLS phenotyping in RIL population derived from Tanjil and Unicrop. ? represents unclear phenotype.

Supplementary Table 3 | SNPs of LOC109334327 in narrow-leaved lupin germplasm.

Supplementary Data Sheet 1 | Genotypes and GLS phenotypes in mapping population.

- marker assisted selection. *Mol. Breed.* 23, 153–161. doi: 10.1007/s11032-008-9222-2
- Ogonnaya, F. C., Subrahmanyam, N. C., Moullet, O., de Majnik, J., Eagles, H. A., Brown, J. S., et al. (2001). Diagnostic DNA markers for cereal cyst nematode resistance in bread wheat. *Aust. J. Agric. Res.* 52, 1367–1374. doi: 10.1071/ar01031
- Paran, I., and Michelmore, R. W. (1993). Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theor. Appl. Genet.* 85, 985–993. doi: 10.1007/bf00215038
- Shan, X., Blake, T. K., and Talbert, L. E. (1999). Conversion of AFLP markers to sequence-specific PCR markers in barley and wheat. *Theor. Appl. Genet.* 98, 1072–1078. doi: 10.1007/s001220051169
- Sharp, P. J., Johnston, S., Brown, G., McIntosh, R. A., Pallotta, M., Carter, M., et al. (2001). Validation of molecular markers for wheat breeding. *Aust. J. Agric. Res.* 52, 1357–1366.
- Thomas, G. J., Buirchell, B. J., and Adcock, K. G. (2011). Re-emergence of grey leaf spot caused by *Stemphylium botryosum* and its implications for Western Australian lupins. *Austral. Plant Path.* 40, 537–539. doi: 10.1007/s13313-011-0078-2
- Vaghefi, N., Thompson, S. M., Kimber, R. B. E., Thomas, G. J., Kant, P., Barbetti, M. J., et al. (2020). Multi-locus phylogeny and pathogenicity of *Stemphylium* species associated with legumes in Australia. *Mycol. Prog.* 19, 381–396. doi: 10.1007/s11557-020-01566-8
- Wang, P., Zhou, G., Jian, J., Yang, H., Renshaw, D., Aubert, M. K., et al. (2020). Whole-genome assembly and resequencing reveal genomic imprint and key genes of rapid domestication in narrow-leaved lupin. *Plant J.* 105, 1192–1210. doi: 10.1111/tj.15100
- Yang, H., Renshaw, D., Thomas, G., Buirchell, B., and Sweetingham, M. (2008). A strategy to develop molecular markers applicable to a wide range of crosses for marker assisted selection in plant breeding: a case study on anthracnose disease resistance in lupin (*Lupinus angustifolius* L.). *Mol. Breed.* 21, 473–483. doi: 10.1007/s11032-007-9146-2
- Yang, H., Shankar, M., Buirchell, B. J., Sweetingham, M. W., Caminero, C., and Smith, P. M. C. (2002). Development of molecular markers using MFLP linked to a gene conferring resistance to *Diaporthe toxica* in narrow-leaved lupin (*Lupinus angustifolius* L.). *Theor. Appl. Genet.* 105, 265–270. doi: 10.1007/s00122-002-0925-1
- Yang, H., Sweetingham, M. W., Cowling, W. A., and Smith, P. M. C. (2001). DNA fingerprinting based on microsatellite-anchored fragment length polymorphisms, and isolation of sequence-specific PCR markers in lupin (*Lupinus angustifolius* L.). *Mol. Breed.* 7, 203–209.
- Yang, H. A., Tao, Y., Zheng, Z. Q., Shao, D., Li, Z. Z., Sweetingham, M. W., et al. (2013a). Rapid development of molecular markers by next-generation sequencing linked to a gene conferring phomopsis stem blight disease resistance for marker-assisted selection in lupin (*Lupinus angustifolius* L.) breeding. *Theor. Appl. Genet.* 126, 511–522. doi: 10.1007/s00122-012-1997-1

- Yang, H. A., Tao, Y., Zheng, Z. Q., Zhang, Q. S., Zhou, G. F., Sweetingham, M. W., et al. (2013b). Draft genome sequence, and a sequence-defined genetic linkage map of the legume crop species *Lupinus angustifolius* L. *PLoS One* 8:e64799. doi: 10.1371/journal.pone.0064799
- You, M., Boersma, J. G., Buirchell, B. J., Sweetingham, M. W., Siddique, K. H. M., Yang, H., et al. (2005). molecular marker applicable for marker-assisted selection for anthracnose disease resistance in lupin breeding. *Cell Mol. Biol. Lett.* 10, 123–134.
- Zhou, G. F., Jian, J. B., Wang, P. H., Li, C. D., Tao, Y., Li, X., et al. (2018). Construction of an ultra-high density consensus genetic map, and enhancement of the physical map from genome sequencing in *Lupinus angustifolius*. *Theor. Appl. Genet.* 131, 209–223. doi: 10.1007/s00122-017-2997-y
- Zhou, L., Chen, Z. J., Lang, X. Y., Du, B., Liu, K., Yang, G. C., et al. (2013). Development and validation of a PCR-based functional marker system for the brown planthopper resistance gene Bph14 in rice. *Breed. Sci.* 63, 347–352. doi: 10.1270/jsbbs.63.347

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

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

Copyright © 2021 Zhou, Yang, Renshaw, Zou, Thomas and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: frontiersin.org/about/contact



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership