



EVOLUTION AND FUNCTIONAL MECHANISMS OF PLANT DISEASE RESISTANCE

EDITED BY: Zhu-Qing Shao, Jia-Yu Xue, Frank L.W. Takken,
Takaki Maekawa and Madhav P. Nepal
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 978-2-88966-199-2

DOI 10.3389/978-2-88966-199-2

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

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

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

What are Frontiers Research Topics?

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

EVOLUTION AND FUNCTIONAL MECHANISMS OF PLANT DISEASE RESISTANCE

Topic Editors:

Zhu-Qing Shao, Nanjing University, China

Jia-Yu Xue, Jiangsu Province and Chinese Academy of Sciences, China

Frank L. W. Takken, University of Amsterdam, Netherlands

Takaki Maekawa, Max Planck Institute for Plant Breeding Research Cologne, Germany

Madhav P. Nepal, South Dakota State University Brookings, United States

Citation: Shao, Z.-Q., Xue, J.-Y., Takken, F. L. W., Maekawa, T., Nepal, M. P., eds. (2020). Evolution and Functional Mechanisms of Plant Disease Resistance. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-199-2

Table of Contents

- 05 Editorial: Evolution and Functional Mechanisms of Plant Disease Resistance**
Jia-Yu Xue, Frank L.W. Takken, Madhav P. Nepal, Takaki Maekawa and Zhu-Qing Shao
- 08 Genome- Wide Analysis of the Nucleotide Binding Site Leucine-Rich Repeat Genes of Four Orchids Revealed Extremely Low Numbers of Disease Resistance Genes**
Jia-Yu Xue, Tao Zhao, Yang Liu, Yang Liu, Yong-Xia Zhang, Guo-Qiang Zhang, Hongfeng Chen, Guang-Can Zhou, Shou-Zhou Zhang and Zhu-Qing Shao
- 20 Comparative Transcriptome Analysis Reveals Key Pathways and Hub Genes in Rapeseed During the Early Stage of Plasmodiophora brassicae Infection**
Lixia Li, Ying Long, Hao Li and Xiaoming Wu
- 35 Characterization of the Powdery Mildew Resistance Gene in the Elite Wheat Cultivar Jimai 23 and its Application in Marker-Assisted Selection**
Mengshu Jia, Hongxing Xu, Cheng Liu, Ruixi Mao, Haosheng Li, Jianjun Liu, Wenxiao Du, Wenrui Wang, Xu Zhang, Ran Han, Xiaolu Wang, Liru Wu, Xiao Liang, Jiancheng Song, Huagang He and Pengtao Ma
- 45 Expressing a Target Mimic of miR156fhl-3p Enhances Rice Blast Disease Resistance Without Yield Penalty by Improving SPL14 Expression**
Ling-Li Zhang, Yan Li, Ya-Ping Zheng, He Wang, Xuemei Yang, Jin-Feng Chen, Shi-Xin Zhou, Liang-Fang Wang, Xu-Pu Li, Xiao-Chun Ma, Ji-Qun Zhao, Mei Pu, Hui Feng, Jing Fan, Ji-Wei Zhang, Yan-Yan Huang and Wen-Ming Wang
- 59 Unlike Many Disease Resistances, Rx1-Mediated Immunity to Potato Virus X is Not Compromised at Elevated Temperatures**
Manon M. S. Richard, Marijn Knip, Thomas Aalders, Machiel S. Beijaert and Frank L. W. Takken
- 67 Genome-Wide Identification and Evolutionary Analysis of NBS-LRR Genes From Dioscorea rotundata**
Yan-Mei Zhang, Min Chen, Ling Sun, Yue Wang, Jianmei Yin, Jia Liu, Xiao-Qin Sun and Yue-Yu Hang
- 78 Genetic Diversity and Evolutionary Analyses Reveal The Powdery Mildew Resistance Gene Pm21 Undergoing Diversifying Selection**
Huagang He, Jian Ji, Hongjie Li, Juan Tong, Yongqiang Feng, Xiaolu Wang, Ran Han, Tongde Bie, Cheng Liu and Shanying Zhu
- 90 Wheat Encodes Small, Secreted Proteins That Contribute to Resistance to Septoria Tritici Blotch**
Binbin Zhou, Harriet R. Benbow, Ciarán J. Brennan, Chanemougasoundharam Arunachalam, Sujit J. Karki, Ewen Mullins, Angela Feechan, James I. Burke and Fiona M. Doohan

- 108** *Impact of DNA Demethylases on the DNA Methylation and Transcription of Arabidopsis NLR Genes*
Weiwen Kong, Xue Xia, Qianqian Wang, Li-Wei Liu, Shengwei Zhang, Li Ding, Aixin Liu and Honggui La
- 123** *Bulked Segregant RNA-Seq Reveals Distinct Expression Profiling in Chinese Wheat Cultivar Jimai 23 Responding to Powdery Mildew*
Tong Zhu, Liru Wu, Huagang He, Jiancheng Song, Mengshu Jia, Liancheng Liu, Xiaolu Wang, Ran Han, Liping Niu, Wenxiao Du, Xu Zhang, Wenrui Wang, Xiao Liang, Haosheng Li, Jianjun Liu, Hongxing Xu, Cheng Liu and Pengtao Ma
- 136** *Patterns of Sequence and Expression Diversification Associate Members of the PADRE Gene Family With Response to Fungal Pathogens*
Marie Didelon, Mehdi Khafif, Laurence Godiard, Adelin Barbacci and Sylvain Raffaele
- 152** *Enzymatic Functions for Toll/Interleukin-1 Receptor Domain Proteins in the Plant Immune System*
Adam M. Bayless and Marc T. Nishimura
- 168** *Distinct Evolutionary Patterns of NBS-Encoding Genes in Three Soapberry Family (Sapindaceae) Species*
Guang-Can Zhou, Wen Li, Yan-Mei Zhang, Yang Liu, Ming Zhang, Guo-Qing Meng, Min Li and Yi-Lei Wang
- 181** *Wheat Disease Resistance Genes and Their Diversification Through Integrated Domain Fusions*
Ethan J. Andersen, Madhav P. Nepal, Jordan M. Purintun, Dillon Nelson, Glykeria Mermigka and Panagiotis F. Sarris



Editorial: Evolution and Functional Mechanisms of Plant Disease Resistance

Jia-Yu Xue^{1,2}, Frank L.W. Takken³, Madhav P. Nepal⁴, Takaki Maekawa^{5†} and Zhu-Qing Shao^{6*}

¹ Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China, ² College of Horticulture, Nanjing Agricultural University, Nanjing, China, ³ Molecular Plant Pathology, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, Netherlands, ⁴ Department of Biology and Microbiology, South Dakota State University, Brookings, SD, United States, ⁵ Department of Plant Microbe Interactions, Max Planck Institute for Plant Breeding Research, Cologne, Germany, ⁶ School of Life Sciences, Nanjing University, Nanjing, China

Keywords: plant disease resistance, evolution, molecular mechanism, regulation mechanism, diversification

OPEN ACCESS

Edited by:

José M. Álvarez-Castro,
University of Santiago de
Compostela, Spain

Reviewed by:

Lóránt Király,
Hungarian Academy of
Sciences, Hungary

*Correspondence:

Zhu-Qing Shao
zhuqingshao@126.com

† Present address:

Takaki Maekawa,
Institute for Plant Sciences, University
of Cologne, Cologne, Germany

Specialty section:

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

Received: 10 August 2020

Accepted: 31 August 2020

Published: 06 October 2020

Citation:

Xue J-Y, Takken FLW, Nepal MP,
Maekawa T and Shao Z-Q (2020)
Editorial: Evolution and Functional
Mechanisms of Plant Disease
Resistance. *Front. Genet.* 11:593240.
doi: 10.3389/fgene.2020.593240

Editorial on the Research Topic

Evolution and Functional Mechanisms of Plant Disease Resistance

INTRODUCTION

Plants in their natural environment are inevitably subject to diseases caused by pathogenic microbes and pests. To survive and combat these threats, plants have evolved a sophisticated immune system. The plant immune system comprises of two layers of defense (pathogen-associated molecular patterns -triggered immunity [PTI] and effector-triggered immunity [ETI]) that can be triggered by hundreds of immune receptors upon detection of pathogen-derived signatures. ETI is induced by resistance (R) proteins, whose activation typically results in death of infected cells (DeYoung and Innes, 2006). Many R proteins belong to the NBS-LRR/NLR (nucleotide-binding domain and leucine-rich repeats) class of intracellular receptors (Jacob et al., 2013). As R genes render great value to crop protection, exploration and identification of functional R genes has greatly expanded over the past 30 years, with >300 functional R genes isolated since 1992 (Kourelis and van der Hoorn, 2018). Besides significant progress in R gene cloning, functional studies revealed nine distinct modes of pathogen perception by R genes (Kourelis and van der Hoorn, 2018). R genes and pathogens are involved in a dynamic “arms race” (Yang et al., 2013), which can be costly for the plant (Tian et al., 2003). As R proteins can trigger cell death, their expression is tightly controlled at both the transcriptional and post-transcriptional level by e.g., methylation, miRNA, and phasiRNAs (Richard et al., 2018a). Despite the large progress made so far, there are many unanswered questions related to, for instance, evolution and regulation of R genes, their mode of action and how these proteins initiate downstream immune signaling.

This topic presents recent advances in plant disease resistance studies and the 14 publications include one review and 13 research articles, contributed by 122 authors. Based on a comprehensive review of the literature, Bayless and Nishimura summarized recent advances on TIR-domain containing R proteins, covering their evolution, structure, function, role in various cell death- and immune-pathways, and their downstream signaling partners. The 13 original articles can be broadly attributed to three topic areas.

RESISTANCE TO POWDERY MILDEW IN WHEAT

Wheat is a globally important food crop. Three articles in this topic focus on resistance to powdery mildew in wheat. One study started with the confirmation of a single dominant gene (*PmJM23*) conferring mildew resistance. Subsequent bulked segregant RNA-Seq (BSR-Seq) analyses mapped the gene to the *Pm2* region on chromosome 5DS, where it could be narrowed down to six disease-related candidate genes (Zhu et al.). The other study revealed broad-spectrum resistance conferred by *PmJM23*, making it a useful source for resistance breeding. The linked markers were therefore used to screen breeding lines with high resistance to powdery mildew (Jia et al.). He et al. isolated 73 alleles of *Pm21*, another powdery mildew resistance gene from different wheat accessions, evaluated their genetic diversity, and discovered that the solvent-exposed LRR residues of proteins encoded by *Pm21* alleles had undergone diversifying selection. These studies will have implications in functional *R*-gene mining and utilization in wheat.

EXPRESSION AND REGULATORY MECHANISMS OF PLANT RESISTANCE SIGNALING

When infected by pathogens, plant gene expression is extensively altered. Distinct transcriptome profiles emerge before- and after infection of resistant and susceptible plant lines. Li et al. compared the transcriptomes of resistant (*R*) and susceptible (*S*) rapeseed (*Brassica napus*) after inoculation with *Plasmodiophora brassicae* and observed differences between the *R*- and *S*-lines during early infection stages, revealing a quick plant response to the pathogen. A detailed transcriptome analysis on genes encoding a conserved protein family harboring the Domain of Unknown Function 4228 (DUF4228) revealed that this gene family is highly responsive to infection by the fungal pathogen *Sclerotinia sclerotiorum* in *Arabidopsis thaliana*, castor bean and tomato. This finding suggests involvement of this uncharacterized protein family in disease resistance signaling to this devastating fungal pathogen (Didelon et al.). In wheat, small proteins were found to be specifically secreted into the apoplast upon infection with the fungal pathogen *Zymoseptoria tritici*. These proteins were experimentally verified to mediate recognition of pathogens and/or induce defense responses (Zhou B. et al.).

Expression of *R* genes is regulated by diverse factors (Richard et al., 2018a), such as small RNAs (Shivaprasad et al., 2012) and siRNAs affecting methylation of *NBS-LRR* genes (Richard et al., 2018b). In this topic, Zhang L.L. et al. discovered that a target mimic of miR156fhl-3p regulates miR156-5p and miR156-5p activity, which subsequently affect expression of *SPL14* and *WRKY45* enhancing rice blast disease resistance. Kong et al. showed that also DNA methylation is involved in regulating expression of *NBS-LRR* genes. Single, double and triple mutants of three *Arabidopsis* DNA demethylases (ROS1, DML2, and

DML3) were found to affect CG methylation of specific *NBS-LRR* promoters and to alter their transcription, providing a new layer in the complex regulatory network controlling *NBS-LRR* activity (Kong et al.). Besides transcriptional control, the activity of *NBS-LRR* genes is regulated post-transcriptionally at the protein level. For a large extent the functionality of receptor *NBS-LRR* proteins is dictated by their interactions with other *NBS-LRR* (helper) proteins (Bonardi et al., 2011; Wu et al., 2018). In Solanaceae, a small family of the NLR Required for Cell Death (*NRC*) genes has been identified to be required for the function of *NBS-LRR*- and non-*NBS-LRR* immune sensors (Adachi et al., 2019). Notably, resistance conferred by many of these immune receptors is temperature sensitive and compromised above 28°C, implying that the helper *NRCs* could be the plant Achilles' heel at elevated temperatures. In this issue, it is shown that Rx1, an *NBS-LRR* gene from potato conferring resistance to Potato Virus X, which signals via *NRC2*, *NRC3*, or *NRC4* (Wu et al., 2017), remains functional at high temperatures (Richard et al.). This finding implies that at least one helper *NRC* and its downstream signaling components are temperature tolerant, suggesting that thermosensitivity of the immune system is likely attributable to the receptors themselves.

EVOLUTION OF *R* GENES

The past two decades witnessed a tremendous growth of studies focused on the abundance, origin, and evolution of the *NBS-LRR* *R* genes (Shao et al., 2019), facilitated by the advancement of information technology. Particularly, genome sequencing and bioinformatics tools integrated in molecular biology are increasingly providing novel insights into the evolution of the plant immune system. Four articles in the topic utilized rigorous bioinformatics and cutting-edge wet-lab experiments to identify and characterize the *NBS-LRR* repertoire of 10 different plant species, including four species of orchids (Xue et al.), three species of soapberries (Zhou G.C. et al.), wheat (Andersen et al.) and African yam (Zhang Y.M. et al.). Among the ten species, hexaploid wheat has the highest number of *NBS-LRR* genes (>800) attributable to recent allopolyploidy events (Andersen et al.), while orchids have the lowest diversity attributable to gene-loss and less frequent gene duplication events (Xue et al.). In all species, most of the *NBS-LRR* genes were found to occur in gene clusters and a small portion of genes were singletons, whereas distinct expansion/contraction patterns were observed in the different species. Andersen et al. discovered that *R* protein gained integrated domains by alternative splicing, which allows creation of genetic diversity by *R* gene mRNAs that may encode for baits, decoys, and functional signaling components.

This Research Topic on Evolution and Functional Mechanisms of Plant Disease Resistance offers recent advances and insights in the field of plant disease resistance. We hope that with the rapid development of molecular technologies, analytical approaches, and methods, such as high-throughput sequencing, integrative multi-omics, and CRISPR gene editing, the reported advances help researchers to study and better understand Plant-Pathogen interactions.

AUTHOR CONTRIBUTIONS

J-YX compiled the contributions from all authors. All authors revised and approved the final version of the manuscript.

FUNDING

This work was supported by the National Natural Science Founding of China (32070243) to Z-QS. FLWT received funding from the NWO-Earth and Life Sciences-funded VICI project no. 865.14.003. MPN's contribution was supported by South Dakota Agriculture Experiment Station USDA-Hatch Project

and Wokini Fund (Grant #SD00H659-18). The Maekawa lab was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, SFB-1403 -414786233), a grant from the University of Cologne Centre of Excellence in Plant Sciences.

ACKNOWLEDGMENTS

We greatly appreciate the contributions from all authors and reviewers as well as the support of the editorial office of Frontiers in Genetics and Frontiers in Ecology and Evolutionary Biology.

REFERENCES

- Adachi, H., Derevnina, L., and Kamoun, S. (2019). NLR singletons, pairs, and networks: evolution, assembly, and regulation of the intracellular immunoreceptor circuitry of plants. *Curr. Opin. Plant Biol.* 50, 121–131. doi: 10.1016/j.pbi.2019.04.007
- Bonardi, V., Tang, S. J., Stallmann, A., Roberts, M., Cherkis, K., and Dangl, J. L. (2011). Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16463–16468. doi: 10.1073/pnas.1113726108
- DeYoung, B. J., and Innes, R. W. (2006). Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat. Immunol.* 7, 1243–1249. doi: 10.1038/ni1410
- Jacob, F., Vernaldi, S., and Maekawa, T. (2013). Evolution and conservation of plant NLR functions. *Front. Immunol.* 4:297. doi: 10.3389/fimmu.2013.00297
- 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. *Plant Cell* 30, 285–299. doi: 10.1105/tpc.17.00579
- Richard, M. M. S., Gratias, A., Meyers, B. C., and Geffroy, V. (2018a). Molecular mechanisms that limit the costs of NLR-mediated resistance in plants. *Mol. Plant Pathol.* 19, 2516–2523. doi: 10.1111/mpp.12723
- Richard, M. M. S., Gratias, A., Thareau, V., Kim, K. D., Balzergue, S., Joets, J., et al. (2018b). Genomic and epigenomic immunity in common bean: the unusual features of NB-LRR gene family. *DNA Res.* 25, 161–172. doi: 10.1093/dnares/dsx046
- Shao, Z. Q., Xue, J. Y., Wang, Q., Wang, B., and Chen, J. Q. (2019). Revisiting the origin of plant NBS-LRR genes. *Trends Plant Sci.* 24, 9–12. doi: 10.1016/j.tplants.2018.10.015
- Shivaprasad, P. V., Chen, H. M., Patel, K., Bond, D. M., Santos, B. A., and Baulcombe, D. C. (2012). A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell* 24, 859–874. doi: 10.1105/tpc.111.095380
- Tian, D., Traw, M. B., Chen, J. Q., Kreitman, M., and Bergelson, J. (2003). Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* 423, 74–77. doi: 10.1038/nature01588
- Wu, C. H., Abd-El-Halim, A., Bozkurt, T. O., Belhaj, K., Terauchi, R., Vossen, J. H., et al. (2017). NLR network mediates immunity to diverse plant pathogens. *Proc. Natl. Acad. Sci. U.S.A.* 114, 8113–8118. doi: 10.1073/pnas.1702041114
- Wu, C. H., Derevnina, L., and Kamoun, S. (2018). Receptor networks underpin plant immunity. *Science* 360, 1300–1301. doi: 10.1126/science.aat2623
- Yang, S., Li, J., Zhang, X., Zhang, Q., Huang, J., Chen, J. Q., et al. (2013). Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18572–18577. doi: 10.1073/pnas.1318211110

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 Xue, Takken, Nepal, Maekawa and Shao. 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-Wide Analysis of the Nucleotide Binding Site Leucine-Rich Repeat Genes of Four Orchids Revealed Extremely Low Numbers of Disease Resistance Genes

Jia-Yu Xue^{1,2,3}, Tao Zhao³, Yang Liu¹, Yang Liu⁴, Yong-Xia Zhang⁵, Guo-Qiang Zhang⁶, Hongfeng Chen⁷, Guang-Can Zhou^{8*}, Shou-Zhou Zhang^{1*} and Zhu-Qing Shao^{4*}

OPEN ACCESS

Edited by:

Horacio Naveira,
University of A Coruña, Spain

Reviewed by:

Serena Aceto,
University of Naples Federico II, Italy
Zhonghua Zhang,
Chinese Academy of Agricultural
Sciences, China

*Correspondence:

Guang-Can Zhou
zgcan2009@163.com
Shou-Zhou Zhang
shouzhouz@szbg.ac.cn
Zhu-Qing Shao
zhuqingshao@126.com;
zhuqingshao@nju.edu.cn

Specialty section:

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

Received: 02 July 2019

Accepted: 22 November 2019

Published: 08 January 2020

Citation:

Xue J-Y, Zhao T, Liu Y, Liu Y, Zhang Y-X,
Zhang G-Q, Chen H, Zhou G-C,
Zhang S-Z and Shao Z-Q (2020)
Genome-Wide Analysis of the
Nucleotide Binding Site Leucine-Rich
Repeat Genes of Four Orchids
Revealed Extremely Low Numbers of
Disease Resistance Genes.
Front. Genet. 10:1286.
doi: 10.3389/fgene.2019.01286

¹ Shenzhen Key Laboratory of Southern Subtropical Plant Diversity, Fairy Lake Botanical Garden, Shenzhen and Chinese Academy of Sciences, Shenzhen, China, ² Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China, ³ VIB-UGent Center for Plant Systems Biology and Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium, ⁴ State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing, China, ⁵ College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, China, ⁶ Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization at College of Landscape Architecture, Fujian Agriculture and Forestry University, Fuzhou, China, ⁷ South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China, ⁸ College of Agricultural and Biological Engineering (College of Tree Peony), Heze University, Heze, China

Orchids are one of the most diverse flowering plant families, yet possibly maintain the smallest number of the nucleotide-binding site-leucine-rich repeat (*NBS-LRR*) type plant resistance (*R*) genes among the angiosperms. In this study, a genome-wide search in four orchid taxa identified 186 *NBS-LRR* genes. Furthermore, 214 *NBS-LRR* genes were identified from seven orchid transcriptomes. A phylogenetic analysis recovered 30 ancestral lineages (29 *CNL* and one *RNL*), far fewer than other angiosperm families. From the genetics aspect, the relatively low number of ancestral *R* genes is unlikely to explain the low number of *R* genes in orchids alone, as historical gene loss and scarce gene duplication has continuously occurred, which also contributes to the low number of *R* genes. Due to recent sharp expansions, *Phalaenopsis equestris* and *Dendrobium catenatum* having 52 and 115 genes, respectively, and exhibited an “early shrinking to recent expanding” evolutionary pattern, while *Gastrodia elata* and *Apostasia shenzhenica* both exhibit a “consistently shrinking” evolutionary pattern and have retained only five and 14 *NBS-LRR* genes, respectively. *RNL* genes remain in extremely low numbers with only one or two copies per genome. Notably, all of the orchid *RNL* genes belong to the ADR1 lineage. A separate lineage, NRG1, was entirely absent and was likely lost in the common ancestor of all monocots. All of the *TNL* genes were absent as well, coincident with the *RNL* NRG1 lineage, which supports the previously proposed notion that a potential functional association between the *TNL* and *RNL* NRG1 genes.

Keywords: orchids, plant resistance genes, evolution, phylogeny, synteny

INTRODUCTION

Plants are exposed to the threat of pathogens on a day-to-day basis in their natural habitats. In order to survive, plants have developed systems to protect themselves from invading pathogens. Specifically, plants have evolved physical barriers like the surface composed of cuticle and wax, to block pathogens or the release chemical components like phenols, terpenes and compounds containing sulfur or nitrogen, to deter or dispose of invading enemies. Moreover, plants have an innate immune system for inducing rapid defense responses. This plant-specific immune system triggers a series of hypersensitive reactions after recognizing invading pathogens, resulting in apoptosis of infected cells, which halts the replication and spread of pathogen.

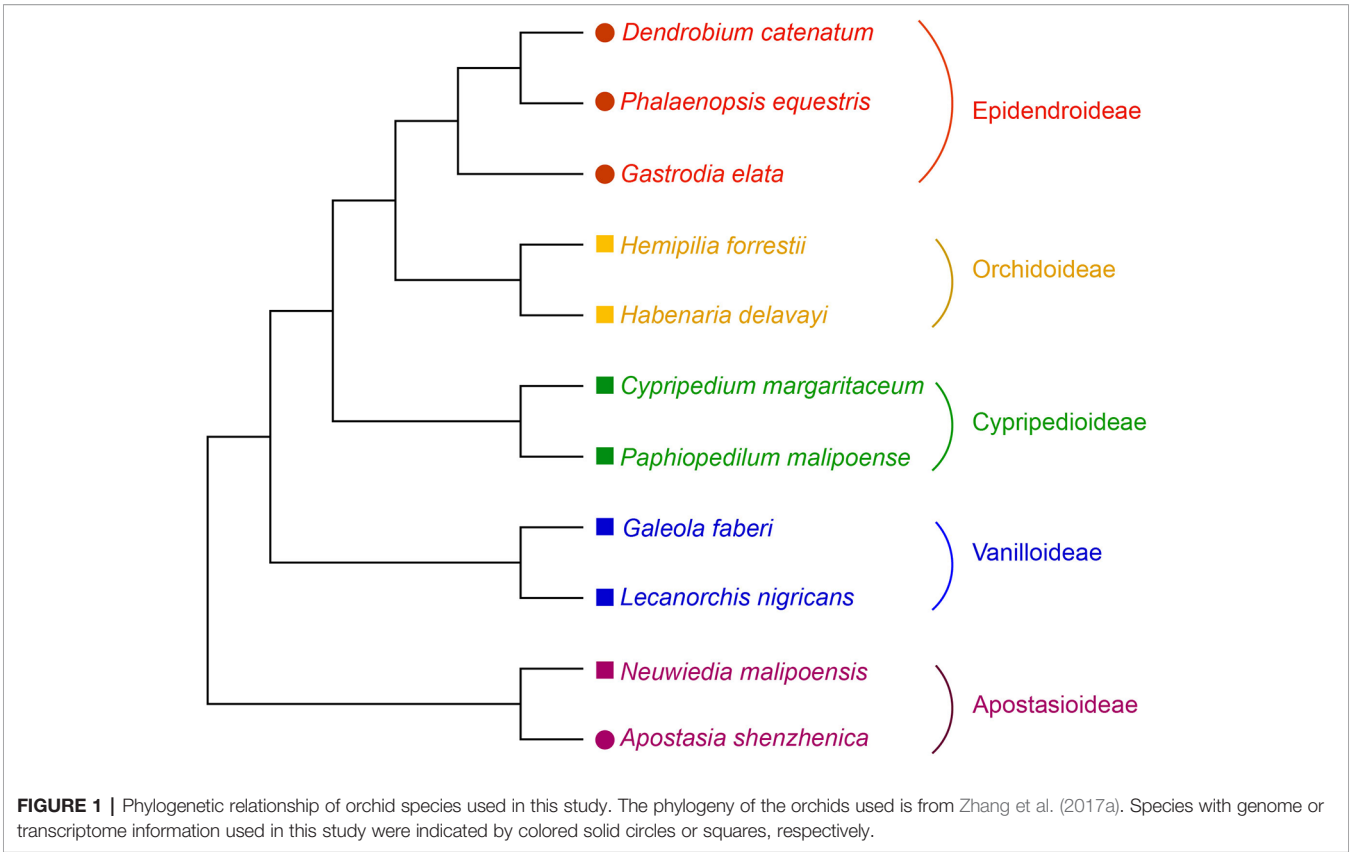
The core of this defending system involves a series of specific genes, namely, disease resistance (*R*) genes, which detect pathogens and trigger downstream resistance reactions. Five types of *R* genes have been discovered, including nucleotide-binding site and leucine-rich repeats (*NBS-LRR*), receptor-like protein (*RLP*), serine/threonine kinase (*STK*), receptor-like kinase (*RLK*) genes, and other genes that do not contain regular domains. Among all types of *R* genes, the *NBS-LRR* gene family is the largest and most important, containing over 60% of characterized *R* genes (Meyers et al., 2005; McHale et al., 2006; Friedman and Baker, 2007; Kourelis and van der Hoorn, 2018). This type of *R* genes originated early in the green plant lineage (Xue et al., 2012; Shao et al., 2019), and has expanded into a large gene family in angiosperms, usually consisting of hundreds of members in an individual genome. These members actively evolved with frequent recombinations occurring between paralogs, gene duplications and losses, and high substitution rates. Since the first genome-wide analysis was conducted on *NBS-LRR* genes in *Arabidopsis thaliana* (Meyers et al., 2003a), this gene family has been comprehensively studied across tens of plant genomes, most of which belong to the rosid lineage of the eudicots and Poaceae of the monocots (Yang et al., 2008; Porter et al., 2009; Li et al., 2010a; Lozano et al., 2012; Luo et al., 2012; Andolfo et al., 2013; Wu et al., 2014; Zhong et al., 2018).

NBS-LRR genes are divided into three classes, the *TIR-NBS-LRR* (*TNL*), *CC-NBS-LRR* (*CNL*) and *RPW8-NBS-LRR* (*RNL*), which are distinguished by the presence of a Toll/Interleukin-1 Receptor-like (*TIR*), coiled-coil (*CC*) or resistance to owdery mildew8 (*RPW8*) domain at the N-terminus of the translated proteins (Shao et al., 2016a; Shao et al., 2019). *RNL* genes were long considered to be part of the *CNL* genes due to some similarities between the sequences of the *CC* and *RPW8* domains (Meyers et al., 2003b), but *RNLs* have too few members to be easily detected. Recently, a functional characterization study found that *RNLs* do not function like regular *R* genes (Bonardi et al., 2011). A typical *R* gene, such as *TNL* or *CNL* genes, usually functions as a detector of certain pathogens and trigger resistance reactions, which is the beginning of the resistance pathway (McHale et al., 2006). The recognition of pathogens by the LRR domains of *TNL* and *CNL* proteins cause conformational changes in the *NBS* domain, which further promotes the multimerization of *TIR* or *CC*

domains that transfer defense signals, while *RNL* proteins appear to be more downstream and transduce signals from the *TNL* and *CNL* proteins through an undetermined pathway (Bonardi et al., 2011). Nevertheless, *RNLs* are clearly indispensable in the resistance pathway, otherwise resistance would be affected (Peart et al., 2005). Evolutionary studies also found strong evidence that supports *RNL* genes as a new class of *NBS-LRR* genes, equivalent to *TNL* and *CNL* genes (Wu et al., 2014; Shao et al., 2016b; Qian et al., 2017). Interestingly, although both *CNL* and *RNL* genes are always present in monocots and dicots, *TNL* genes are absent from monocots, which is likely due to an ancient gene loss event upon the split of this lineage (Meyers et al., 2003a; Tarr and Alexander, 2009; Andolfo et al., 2013; Shao et al., 2016b; Zhang et al., 2017b).

Diverse evolutionary patterns of *NBS-LRR* genes have been observed in different angiosperm lineages. For example, both Fabaceae and Rosaceae exhibit a consistently expanding pattern (Shao et al., 2014; Jia et al., 2015); whereas Brassicaceae exhibits a pattern of expansion followed by contraction (Zhang et al., 2016); Solanaceae demonstrates complicated patterns, potato shows a “consistent expansion” pattern, tomato exhibits a pattern of “first expansion and then contraction,” and pepper presents a “shrinking” pattern (Qian et al., 2017). Despite the absence of *TNLs*, the number of *NBS-LRR* genes analyzed in monocot genomes comparable to that of eudicots. For example, Asian rice *Oryza sativa* possesses 498 *NBS-LRR* genes, outnumbering most eudicots (Li et al., 2010a; Shao et al., 2016b). The discrepancy of retained gene number is drastic among species. Maize (*Zea mays*), a species from the same grass family as rice, possesses no more than 140 *NSB-LRR* genes, which shows a four-fold discrepancy between the two species and suggests an active evolutionary mode of *NBS-LRR* genes in Poaceae. The evolutionary history of *NBS-LRR* genes in Poaceae has been comprehensively studied: Li et al. used *NBS-LRR* genes from four sequenced genomes (Asian rice, maize, *Sorghum bicolor* and *Brachypodium distachyon*) to reconstruct the evolutionary history of *NBS-LRR* genes, and compared the gene tree with the systematic relationship of these four species, reconciling 496 ancestral lineages in the grass family. Varying numbers of gene gain and loss events resulted in the gene number discrepancy across these four species, indicating a shrinking pattern in this family (Li et al., 2010b).

To date, only *NBS-LRR* genes of the grass family have been well studied in monocots. Whether or not other monocot lineages exhibit different evolutionary patterns remains unanswered, as the sequenced genomes are not as prevalent in other monocot lineages as in Poaceae. Fortunately, in recent years the sequenced genomes in Orchidaceae (orchids) have rapidly increased and multiple genomes of this family, another monocot lineage, have been made readily available. In this study, a genome-wide analysis of *NBS-LRR* genes in the four sequenced orchid genomes and seven orchid transcriptomes was conducted (Figure 1). The goal of this study was to uncover the evolutionary features and modes of *NBS-LRR* genes in this family and further investigate the mechanisms that have shaped these evolutionary changes.



RESULTS

Identification and Domain Combination of NBS-LRR Genes From Four Orchid Genomes

A total of 186 NBS-LRR genes (Table 1 and Table S1) were identified from four orchid genomes following previously described procedures (Xue et al., 2012; Shao et al., 2016b; Shao et al., 2019), among which, the CNL genes (182) overwhelmingly outnumbered RNL genes (4). TNL genes were absent from all four genomes, in accordance with the hypothesis that an early and thorough loss of TNLs had occurred upon the divergence of the monocot lineage (Tarr and Alexander, 2009;

Shao et al., 2016b). RNL genes were found in three orchid genomes, except *G. elata*, but at extremely low numbers with one or two genes in each genome. Among the four orchids, the *D. catenatum* genome encoded the most NBS-LRR genes (115), followed by *P. equestris* (52) and *A. shenzhenica* (14), and *G. elata*, which encoded only five genes, the least among the four orchids and among all of the sequenced angiosperms. Since each orchid species had only one or two RNL genes, CNL genes must presumably be fully responsible for gene number variations among orchids.

Of the four orchids, intact NBS-LRR genes with all three domains (CC/RPW8-NBS-LRR) accounted for only 19.8% (37) of the total, whereas other genes either lacked a CC/RPW8 domain at the N-terminus, an LRR domain at the C-terminus, or lacked domains at both termini. *G. elata* had the highest proportion of intact genes (40.0%), while *P. equestris* had the lowest (15.4%). Several genomic changes, like recombination, fusion and pseudogenization, could result in real truncated genes, whereas other factors, such as sequencing, assembly errors and false annotations would elicit artificially “truncated” genes. Comparatively, the well-sequenced and annotated *A. thaliana* genome contains fewer (24.2%) truncated genes (Meyers et al., 2003b; Zhang et al., 2016).

Conserved Motifs of the NBS Domain in Orchids

The NBS domain contains several smaller motifs of 10 to 30 amino acids in length. including P-loop, kinase 2, kinase 3, RNBS-

TABLE 1 | The number of identified NBS-LRR genes in four orchid genomes.

Domain compositions	A. shenzhenica	G. elata	P. equestris	D. catenatum
CNL class	13 (93%)	5 (100%)	50 (96%)	114 (99%)
CNL (Intact)	2	2	6	23
CN	1	0	1	28
NL	8	0	11	13
N	2	3	32	51
RNL class	1 (7%)	0	2 (4%)	1 (1%)
RNL (Intact)	1	0	2	1
RN	0	0	0	0
NL	0	0	0	0
N	0	0	0	0
Total number	14	5	52	115

C, GLPL, and RNBS-D (DeYoung and Innes, 2006). Using MEME and WebLogo, these motifs in orchid CNL and RNL proteins were identified (**Figure 2**). Although RNL proteins are conserved along the whole NBS domain, six motifs exhibited different extents of variation in CNL proteins. Differences between the CNL and RNL proteins were observed in all six motifs, especially kinase 3 and RNBS-C, which exhibited the greatest discrepancy. These motifs can be used to distinguish orchid *NBS-LRR* genes without conducting phylogenetic analyses.

Phylogenetic Analysis of Orchid *NBS-LRR* Genes

To explore the evolutionary history of *NBS-LRR* genes in orchids, a phylogenetic analysis using the protein sequences of the NBS domain was conducted using three *Amborella* TNL proteins as outgroups. In order to obtain a more complete evolutionary pattern of *NBS-LRR* genes in orchid, 214 *NBS-LRR* genes from seven orchid transcriptomes were identified and involved for phylogenetic analysis (**Table S1**). The phylogeny revealed a deep divergence between the *RNL* and *CNL* genes, and the evolutionary rate of *RNL* genes was rather low, which was reflected by the short branches among species (**Figure 3**). Nevertheless, the branch separating *RNL* genes and *CNL* genes was long (**Figure 3**), supporting the hypothesis of ancient divergence between *RNL* and *CNL* genes.

Reconciling the phylogeny of orchid *NBS-LRR* genes by the species tree recovered 30 ancestral *NBS-LRR* lineages, including one *RNL* lineage and 29 *CNL* lineages. This represents the minimal number of ancestral *NBS-LRR* genes in the common ancestor of orchids, as the full *NBS-LRR* repertoire from the seven orchids could not be fully recovered from their transcriptomes. The reconciled *RNL* gene's phylogeny was consistent with that of the orchid species tree, suggesting that they are descendants of one ancestral gene from the common ancestor and experienced no shared gene duplication (**Figures S1 and S2**). *CNL* genes exhibited a more active evolutionary pattern with far more gene duplications and losses, as well as faster evolutionary rates, which was reflected by their longer branch lengths (**Figures S1 and S2**). In total, 29 *CNL* lineages were identified from the orchid ancestor (**Figure 3**). Species-specific expansions were observed in different branches of the phylogenetic tree, with *D. catenatum*-specific expansion in Lineage 8, 9, 18, and 30, and *P. equestris*-specific expansion in Lineage 30 (**Figures 3 and S2**). Moreover, independent gene losses occurred in the evolutionary history of the orchid family,

thus, none of the four species maintained all ancestral lineages. Both gene duplications and losses have contributed to the gene number variations among the different species.

The phylogenetic tree shows that 30 ancestral *NBS-LRR* lineages were not all retained by all four orchids, but differentially kept by different taxa (**Table 2**). *D. catenatum* maintained 17 lineages, *A. shenzhenica* had 10, *P. equestris* had seven, and *G. elata* had three (**Table 2**). Interestingly, *P. equestris* retained fewer ancestral lineages than *A. shenzhenica*, but developed more genes. For the 30 recovered ancestral lineages, 21 of them were inherited by at least one analyzed genomes. Lineage 29 is the only one lineage retained by all four orchids, 11 lineages are inherited in only one taxon, five lineages are shared by two taxa, and four lineages are reserved in three taxa.

Syntenic Analysis of *NBS-LRR* Genes in Orchid Genomes

The synteny analysis was performed both between and within the four orchid genomes. Results revealed that the *RNLs* reserved synteny among three species, except *G. elata*, which lost the *RNL* genes (**Figure 4A**; **Table S2**). These results were in accordance with the synteny analysis of the *RNL* genes in other angiosperms and supported the conservative evolutionary pattern of this *NBS-LRR* subclass (Shao et al., 2016b). Synteny of the *CNL* genes was also detected for some conservatively evolved *CNL* lineages. Lineage 29 *CNL* genes from the four orchid genomes were detected on syntenic blocks (**Figure 4B**; **Table S2**).

Within genome synteny was used to determine which lineages of the *NBS-LRR* genes were derived from whole genome duplications (WGDs) or segmental chromosomal duplications (**Figure 4**). Surprisingly, no segmentally duplicated *NBS-LRR* genes were identified in the four genomes, whereas 47, 22, and 7 tandemly duplicated genes were detected in *D. catenatum*, *P. equestris*, and *A. shenzhenica*, respectively. The remaining *NBS-LRR* genes from the four genomes may have been duplicated from other duplication types. Although no segmentally duplicated *NBS-LRR* genes were identified based on the within genome analysis, the role of this duplication mechanism in *NBS-LRR* gene evolution could not be ruled out. First, the syntenic relationship of *NBS-LRR* genes would be disrupted during long-term evolution. Second, the segmentally duplicated *NBS-LRR* genes may have been lost during evolution. Therefore, the contribution of segmental duplication may be underestimated in the within genome synteny analysis.

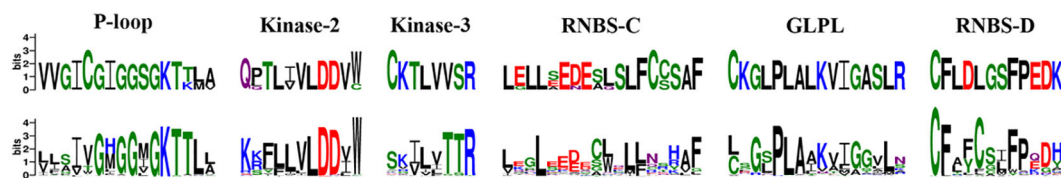


FIGURE 2 | Conserved motifs in the NBS domain of the four orchid species. The amino acids of the six conserved motifs are extracted. Larger letters indicate higher frequency.

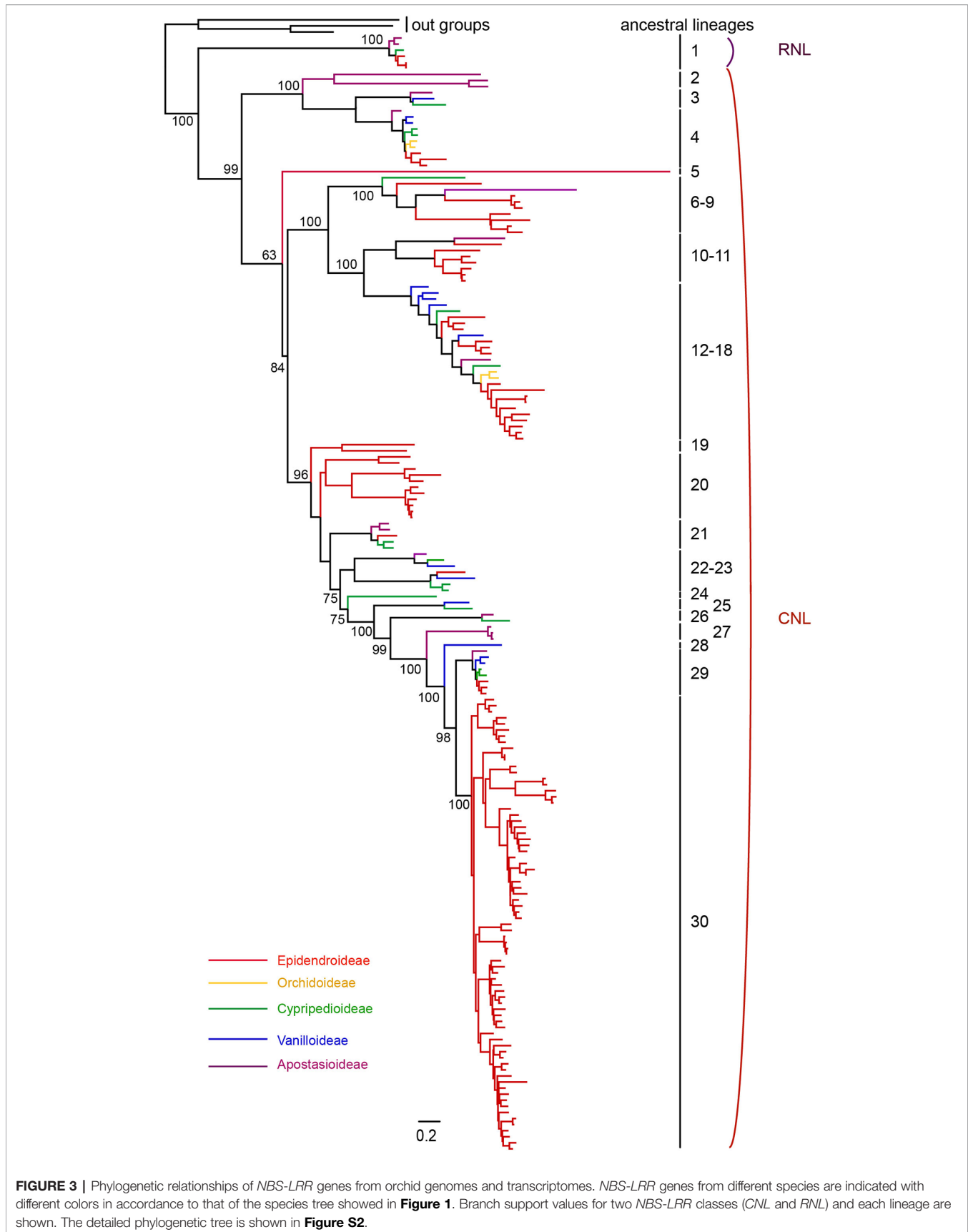


TABLE 2 | Reservation of 30 ancestral *NBS-LRR* lineages in four orchids

Lineage	<i>D. catenatum</i>	<i>P. equestris</i>	<i>G. elata</i>	<i>A. shenzhenica</i>
1	+	+	-	+
2	-	-	-	+
3	-	-	-	+
4	+	+	-	+
5	+	-	-	-
6	-	-	-	-
7	+	-	-	-
8	+	-	-	-
9	+	-	-	-
10	+	-	-	+
11	+	+	+	-
12	-	-	-	-
13	-	-	-	-
14	-	-	-	-
15	-	-	-	-
16	+	+	-	-
17	+	+	+	-
18	+	-	-	+
19	+	-	-	-
20	+	-	-	-
21	+	-	-	+
22	-	-	-	-
23	+	-	-	-
24	-	-	-	-
25	-	-	-	-
26	-	-	-	+
27	-	-	-	+
28	-	-	-	-
29	+	+	+	+
30	+	+	-	-

"+" indicates the presence of the lineage, "-" indicates the absence of the lineage.

Reconciliation of Gene Losses and Gains and the Evolutionary Patterns in Orchids

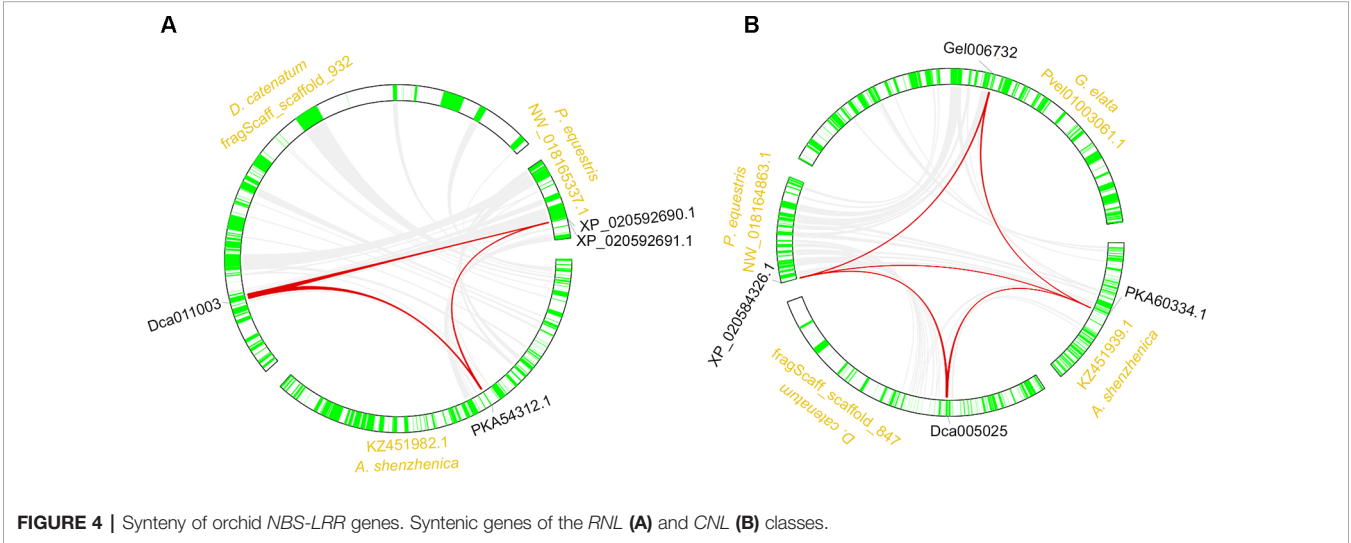
Based on the phylogenetic tree, it could be inferred that many independent gene gains and losses have occurred at different stages of orchid evolution (Figure 5). Starting from 30 ancestral genes, these four species have experienced considerably different evolutionary patterns: *A. shenzhenica*, the first split taxon, has undergone a process

of more gene losses (20) than duplications (4), resulting in 14 *NBS-LRR* genes in its genome today. This basal taxon overall exhibits a shrinking pattern of evolution (Figure 6). The one taxon with fewer genes than the common ancestor, *G. elata*, should have experienced more severe gene losses. Before its divergence, the *NBS-LRR* genes in the common ancestor of *G. elata*, *D. catenatum* and *P. equestris* was reduced to 17 and *G. elata* experienced additional gene loss after its split. Thus, *G. elata* has undergone a “consistent shrinking” pattern (Figure 6). *D. catenatum* and *P. equestris* both have more genes than the common ancestor of orchids. Along their evolutionary trajectories, these two taxa have gained more genes than they have lost and recent independent duplications have made major contribution to the gene number increase in these two species. Based on the phylogenetic tree, it is clear that that species-specific duplications have expanded the gene numbers of lineage 8, 9, 18 and 30 in *D. catenatum*, and Lineage 30 in *P. equestris*, outnumbering the other two taxa. Therefore, *D. catenatum* and *P. equestris* both exhibit an “early shrinking to recent expanding” pattern (Figure 6). Overall, the four orchids exhibit two different patterns of *NBS-LRR* evolution, and the discrepancy depends on whether a given taxon underwent recent expansions.

DISCUSSION

The NBS-LRR Gene Number of Orchids

NBS-LRR genes belong to a large gene family in angiosperms, which includes hundreds of members. Only a small number of angiosperm taxa contain less than 100 *NBS-LRR* genes in their genomes. For example, Shao et al. analyzed 22 angiosperm genomes, which all had more than 100 *NBS-LRR* genes, except one Brassicaceae species, *Thellungiella salsuginea*, which had 88 genes (Shao et al., 2016b). In this study, it was discovered that three orchids also belong to the minority of plants that encode less than 100 *NBS-LRR* genes, and only one taxon, *P. equestris*, encoded over 100 genes. The number of *T. salsuginea* *NBS-LRR* genes fall below 100 because it underwent more severe gene loss



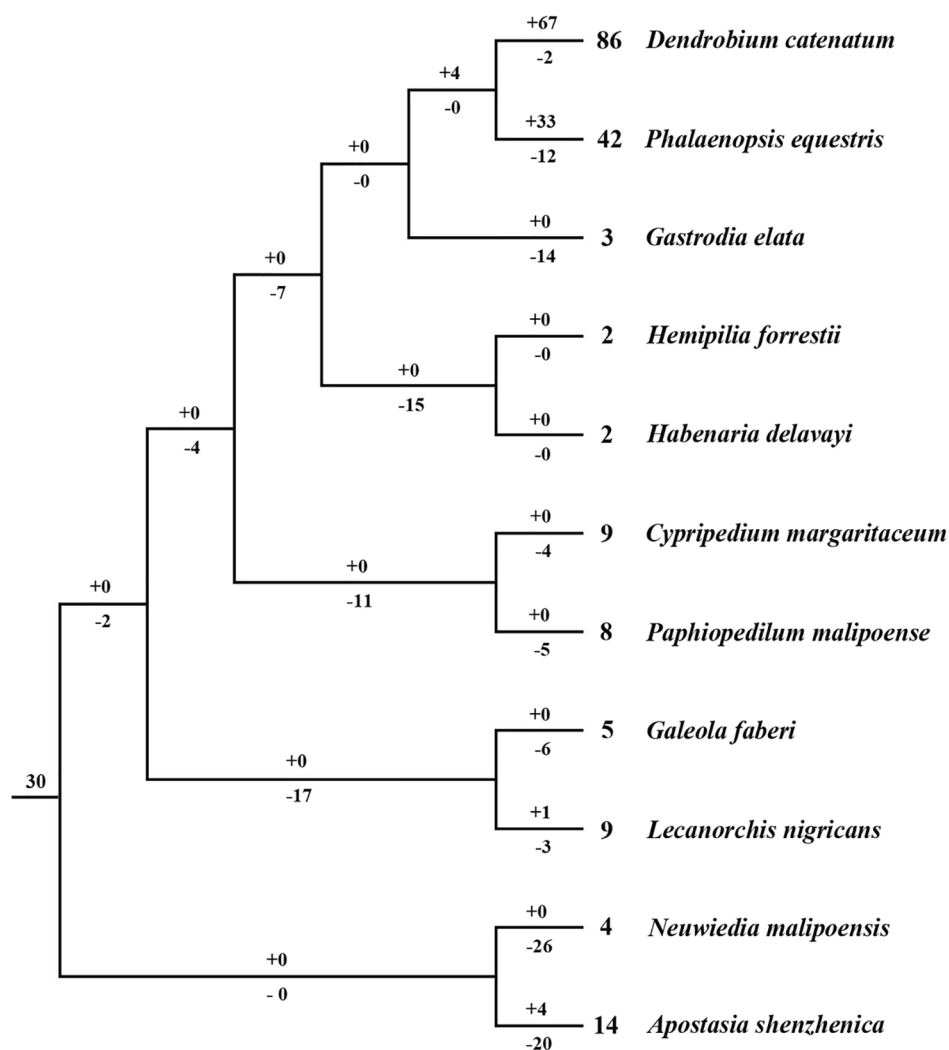


FIGURE 5 | Loss and gain events of NBS-LRR genes across orchid evolution. Gene losses and gains are indicated by numbers with '-' or '+' on each branch. Detailed information for gain and loss events of NBS-LRR genes is shown in **Figure S3**.

events than duplications, after all, its ancestor once had over 228 NBS-LRR genes (Zhang et al., 2016). The same situation was observed for three Cucurbitaceae species, *Cucumis sativus*, *C. melo*, and *Citrullus lanatus* (Lin et al., 2013). Orchids, however, are a different case. From a genetic and evolutionary perspective, as the reconciliation analysis suggests, the small number of orchid ancestral genes was mainly responsible for these results. In the orchid family, only 29 ancestral CNL genes and one RNL gene lineages were recovered as the family emerged, obviously far fewer than the 228 ancestral genes found in Brassicaceae (Zhang et al., 2016), 119 in Fabaceae (Shao et al., 2014), and 456 in Poaceae (Shao et al., 2016b). That's why although *D. catenatum* and *P. equestris* have gained more genes than lost, they have not yet reached a large number of genes, such as rice or soybeans (Bai et al., 2002; Shao et al., 2014).

The number of NBS-LRR genes varies drastically among different taxa, even among closely-related species or subspecies. For instance, potato and tomato, both belonging to Solanaceae

and have 447 and 255 NBS-LRR genes, respectively, showing a ratio of 1.75-fold difference in gene numbers (Qian et al., 2017). Intra-species variations of *Oryza*, *Glycine* and *Gossypium* reached a 5.4-fold discrepancy (Zhang et al., 2010). Therefore, the gene number variation observed in orchids. It is also noteworthy that the recent expansions are the main cause for this discrepancy was not surprising. Notably, recent expansions are the main cause for this discrepancy. In Fabaceae, Brassicaceae and Solanaceae, the majority of expansions are consequence of tandem duplications (Shao et al., 2014; Zhang et al., 2016; Qian et al., 2017). In this study, *D. catenatum* and *P. equestris* appear to have undergone recent abrupt expansions, but mechanically tandem and ectopic duplications, other than WGDs are responsible for such expansions, as no syntenic genes were detected in these two species. *A. shenzhenica* and *G. elata* have not experienced sharp duplications, which explains the low number of NBS-LRR genes in these genomes. *A. shenzhenica* represents the earliest split of orchids, and has a rather narrow

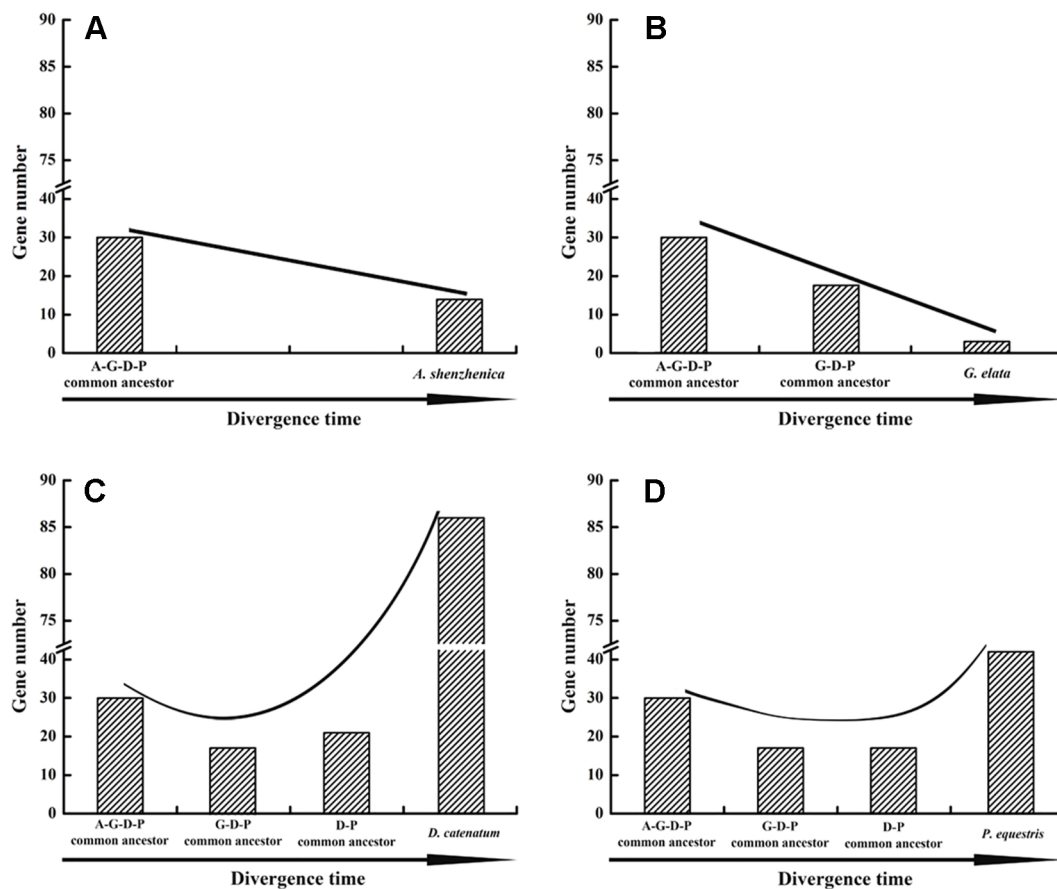


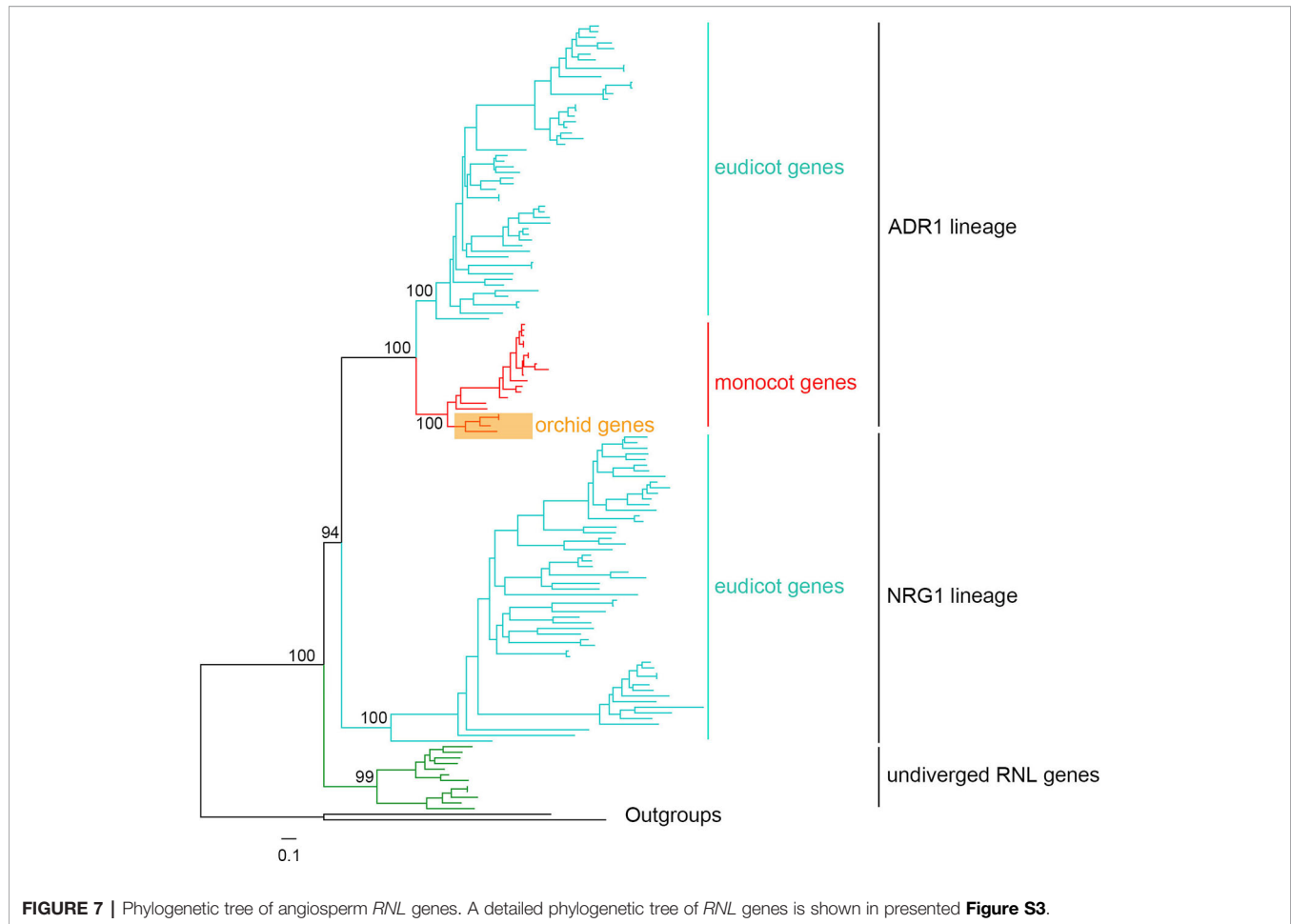
FIGURE 6 | Evolutionary patterns of NBS-LRR genes in four orchid species. Evolutionary patterns of NBS-LRR genes in four orchids: *A. shenzhenica* (A), *G. elata* (B), *P. equestris* (C), and *D. catenatum* (D). A-G-D-P indicates the common ancestor of all four orchids; GDP indicates the common ancestor of *G. elata*, *P. equestris* and *D. catenatum*, and D-P indicates the common ancestor of *P. equestris* and *D. catenatum*.

geographical distribution, as it is restricted to the Southeast Guangdong province, China (Zhang et al., 2017a). Its narrow distribution and stable habitat will likely lead fewer pathogen changes and stable pathogens diversity. Thus, *A. shenzhenica* has likely been battling a few of the same pathogens for a long period of time. Therefore, *A. shenzhenica* does not need to expand its *R* genes to face potential enemies. *G. elata*, despite its wide distribution, is an obligate mycoheterotrophic taxon, depending on a particular fungus *Armillaria mellea* to survive (Yuan et al., 2018), which probably does not allow *G. elata* to maintain many *R* genes. Coincidentally, two other obligate mycoheterotrophic taxa, *Cuscuta australis* (Sun et al., 2018) and *Epipogium roseum* (unpublished), all seem to show a global gene loss pattern, and reduction of *R* gene number is only a part of the consequences.

The Evolution of *RNLs* in Orchids and Other Angiosperms

According to a previous WGD, angiosperm *RNL* genes have diverged into two lineages, ADR1 and NRG1, based on *Aarbidopsis* and tobacco (Collier et al., 2011; Shao et al., 2016b). In this study, the comprehensive analysis of seed plant *RNL* genes

revealed an undiverged clade of gymnosperm genes at the basal position, followed by two diverged clades, ADR1 and NRG1, in angiosperms (Figures 7 and S3). Orchid *RNL* genes exclusively belong to the ADR1 lineage and have the shortest branch lengths among all of the angiosperms. Thus it is speculated that orchid *RNL* genes have been evolutionarily conserved since they have fewer diverse upstream signals to transduce. Orchids may be one of the plant lineages with the lowest number of *R* genes. The NRG1 lineage may have been lost as the origin of monocots, accompanied with the loss of an intron of the ADR1 lineage. The coincident loss of *TNL* genes and *RNL* NRG1 genes has been speculated to be due to their functional interdependence, as the resistance signals initiated by *TNL* genes are exclusively transduced by the NRG1 lineage (Collier et al., 2011). Several recent studies have suggested that nearly all test *TNL* genes are dependent on the *NRG1* gene for inducing hypersensitive reactions, although potential exceptions could exist (Qi et al., 2018; Castel et al., 2019; Wu et al., 2019). As a downstream gene with a conservative function, orchid *RNL* genes seem unnecessary to expand. Low copies are sufficient for maintaining a functional system. This may explain why *RNL* genes have remained in low numbers across the evolution of angiosperms.



MATERIALS AND METHODS

Identification and Classification of *NBS-LRR* Genes

The whole genomes of four orchid taxa, *A. shenzhenica*, *C. elata*, *P. equestris* and *D. catenatum*, were used in this study. Genomic sequences and annotation files of *A. shenzhenica*, *P. equestris* and *D. catenatum* were downloaded from the NCBI database (<https://www.ncbi.nlm.nih.gov/>) (accession nos. PRJNA310678, PRJNA389183, and PRJNA262478, respectively). The genomic sequences and annotation files of *C. elata* were obtained from the *G. elata* Genome Warehouse Database (<http://bigd.big.ac.cn/gwh/Assembly/129/show>). The identification of *NBS-LRR* genes involved a two-step process. First, BLAST and hidden Markov model (HMM) searches using the NB-ARC domain (Pfam accession No.: PF00931) as a query, were performed simultaneously to identify candidate genes in each genome. For the BLAST search, the threshold expectation value was set to 1.0. For the HMM search (<http://hmmer.org>), default parameter settings were used. Second, all of the obtained candidate genes using BLAST or HMM searches were merged together, and the redundant hits were removed. The remaining candidate genes were submitted for an online Pfam analysis (<http://pfam.sanger.ac.uk/>) to further confirm the presence of the NBS domain with

an E-value of 10^{-4} . When two or more transcripts were annotated for a gene from alternative splicing, the longest form with an NBS domain was selected. All of the identified *NBS-LRR* genes were analyzed using the NCBI's conserved domain database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) in order to determine the domains they possess.

Sequence Alignment and Conserved Motif Identification

The amino acid sequences of the NBS domain were extracted from the identified NBS-encoding genes and used for multiple alignments using ClustalW (Tamura et al., 2011) and Muscle (Edgar, 2004) integrated in MEGA 7.0 (Kumar et al., 2016) with default parameter settings. NBS domain sequences that were too short (i.e., shorter than two-thirds of a regular NBS domain) or too divergent (i.e. genes whose NBS domains could not be well aligned with others, and the aligned lengths are shorter than two-thirds of a regular NBS domain) were removed to prevent interference with the alignment and subsequent phylogenetic analysis. Resulting amino acid sequence alignments were manually edited in MEGA 7.0 (Kumar et al., 2016) for further improvement. Conserved protein motifs were analyzed by the online programs MEME (Multiple Expectation Maximization for Motif Elicitation) and WebLogo (Crooks et al., 2004; Bailey et al., 2006) with default parameter settings.

Phylogenetic Analysis and Reconciliation of Gene Loss/Duplication Events

In order to explore the relationships of *NBS-LRR* genes in the four orchids, a phylogenetic tree was reconstructed based on the aligned amino acid sequences of the conserved NBS domains. To avoid interference from “noisy characters,” too short or extremely divergent sequences were excluded from the phylogenetic analysis. Phylogenetic analyses were conducted using IQ-TREE and the maximum likelihood method (Nguyen et al., 2015). The best-fit model was estimated by ModelFinder (Kalyaanamoorthy et al., 2017). Branch support values were assessed with UFBoot2 tests (Minh et al., 2013). The scale bar indicated the genetic distance. *TNL* genes from the basal angiosperm, *Amborella trichopoda*, were used as outgroups. Additionally, gene loss/duplication events during the speciation of the four orchid taxa were recovered by reconciling the *NBS-LRR* gene phylogenetic tree with the real species tree using Notung software (Stolzer et al., 2012). The phylogenetic analysis of *RNL* genes used the full length amino acid sequences of *RNL* proteins of 45 seed plants downloaded from Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>).

Syntenic Analyses Within and Across the Four Orchid Genomes

A synteny network approach was employed in this study (Zhao et al., 2017; Zhao and Schranz, 2019). Briefly, pair-wise all-against-all blast of protein sequences from the four genomes (*Apostasia*, *Gastrodia*, *Phalaenopsis* and *Dendrobium*) was performed. The obtained results and gff annotation files were then subjected to MCScanX for intra- and interspecies microsynteny detection and gene duplication type determination (Wang et al., 2012). Microsynteny relationship was displayed by TBtools (<https://github.com/CJ-Chen/TBtools>).

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

J-YX, Z-QS, S-ZZ and G-CZ conceived and designed the project. JYX, G-CZ, TZ and Z-QS obtained and analyzed the data.

REFERENCES

- Andolfo, G., Sanseverino, W., Rombauts, S., Van de Peer, Y., Bradeen, J. M., Carputo, D., et al. (2013). Overview of tomato (*Solanum lycopersicum*) candidate pathogen recognition genes reveals important *Solanum* R locus dynamics. *New Phytol.* 197, 223–237. doi: 10.1111/j.1469-8137.2012.04380.x
- Bai, J., Pennill, L. A., Ning, J., Lee, S. W., Ramalingam, J., Webb, C. A., et al. (2002). Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Res.* 12, 1871–1884. doi: 10.1101/gr.454902
- YL (3rd Author), YL (4th Author), Y-XZ, G-QZ, HC and S-ZZ participated in the data analysis and discussion. J-YX drafted the initial manuscript. Z-QS and YL (3rd Author) complemented the writing. All authors contributed to discussion of the results, reviewed the manuscript and approved the final article.
- ## FUNDING
- This work was supported by grants from the Shenzhen Key Laboratory of Southern Subtropical Plant Diversity (SLPD-2018-3 to J-YX), the Jiangsu Key Laboratory for the Research and Utilization of Plant Resources (Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, KSPKLB201835 to J-YX), and the Strategic Priority Research Program of Chinese Academy of Sciences (XDA13020603 to HC).
- ## ACKNOWLEDGMENTS
- We also thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.
- ## SUPPLEMENTARY MATERIAL
- The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01286/full#supplementary-material>
- FIGURE S1** | A detailed ML phylogenetic tree with all sequence names and branch support values. The tree was reconstructed based on the NBS domain sequences of the *NBS-LRR* genes from the four genomes and seven transcriptomes.
- FIGURE S2** | Reconciled *NBS-LRR* gene tree with real species phylogeny and various loss and duplication events restored. “n3014” indicates a loss event that occurred in the common ancestor of Epidendroideae, Orchidoideae, Cyripedioideae, and Vanilloideae.
- FIGURE S3** | A detailed ML phylogenetic tree based on the full length sequences of *RNL* proteins from 45 seed plants.
- TABLE S1** | A list of *NBS-LRR* genes identified from seven orchid transcriptomes.
- TABLE S2** | Detailed information of the two synteny blocks.
- Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, W369–W373. doi: 10.1093/nar/gkl198
- Bonardi, V., Tang, S. J., Stallmann, A., Roberts, M., Cherkis, K., and Dangel, J. L. (2011). Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proc. Natl. Acad. Sci. United States America* 108, 16463–16468. doi: 10.1073/pnas.1113726108
- Castel, B., Ngou, P. M., Cevik, V., Redkar, A., Kim, D. S., Yang, Y., et al. (2019). Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. *New Phytol.* 222, 966–980. doi: 10.1111/nph.15659

- Collier, S. M., Hamel, L. P., and Moffett, P. (2011). Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Mol. Plant Microbe Interact.* 24, 918–931. doi: 10.1094/MPMI-03-11-0050
- Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- DeYoung, B. J., and Innes, R. W. (2006). Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat. Immunol.* 7, 1243–1249. doi: 10.1038/ni1410
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Friedman, A. R., and Baker, B. J. (2007). The evolution of resistance genes in multi-protein plant resistance systems. *Curr. Opin. Genet. Dev.* 17, 493–499. doi: 10.1016/j.gde.2007.08.014
- Jia, Y., Yuan, Y., Zhang, Y., Yang, S., and Zhang, X. (2015). Extreme expansion of NBS-encoding genes in Rosaceae. *BMC Genet.* 16, 48. doi: 10.1186/s12863-015-0208-x
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/Nmeth.4285
- 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. *Plant Cell* 30, 285–299. doi: 10.1105/tpc.17.00579
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Li, J., Ding, J., Zhang, W., Zhang, Y., Tang, P., Chen, J. Q., et al. (2010a). Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. *Mol. Genet. Genomics* 283, 427–438. doi: 10.1007/s00438-010-0527-6
- Li, J., Ding, J., Zhang, W., Zhang, Y., Tang, P., Chen, J. Q., et al. (2010b). Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. *Mol. Genet. Genomics* 283, 427–438. doi: 10.1007/s00438-010-0527-6
- Lin, X., Zhang, Y., Kuang, H. H., and Chen, J. J. (2013). Frequent loss of lineages and deficient duplications accounted for low copy number of disease resistance genes in Cucurbitaceae. *BMC Genomics* 14, 335. doi: 10.1186/1471-2164-14-335
- Lozano, R., Ponce, O., Ramirez, M., Mostajo, N., and Orjeda, G. (2012). Genome-wide identification and mapping of NBS-encoding resistance genes in *Solanum tuberosum* group phureja. *PLoS One* 7, e34775. doi: 10.1371/journal.pone.0034775
- Luo, S., Zhang, Y., Hu, Q., Chen, J., Li, K., Lu, C., et al. (2012). Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes in the grass family. *Plant Physiol.* 159, 197–210. doi: 10.1104/pp.111.192062
- McHale, L., Tan, X., Koehl, P., and Micheltore, R. W. (2006). Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 7, 212. doi: 10.1186/gb-2006-7-4-212
- Meyers, B. C., Kozik, A., Griego, A., Kuang, H., and Micheltore, R. W. (2003a). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15, 809–834. doi: 10.1105/tpc.009308
- Meyers, B. C., Kozik, A., Griego, A., Kuang, H. H., and Micheltore, R. W. (2003b). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15, 809–834. doi: 10.1105/tpc.009308
- Meyers, B. C., Kaushik, S., and Nandety, R. S. (2005). Evolving disease resistance genes. *Curr. Opin. Plant Biol.* 8, 129–134. doi: 10.1016/j.pbi.2005.01.002.S1369-5266(05)00006-3
- Minh, B. Q., Nguyen, M. A. T., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Peart, J. R., Mestre, P., Lu, R., Malcuit, I., and Baulcombe, D. C. (2005). NRG1, a CC-NB-LRR protein, together with N, a TIR-NB-LRR protein, mediates resistance against tobacco mosaic virus. *Curr. Biol.* 15, 968–973. doi: 10.1016/j.cub.2005.04.053
- Porter, B. W., Paidi, M., Ming, R., Alam, M., Nishijima, W. T., and Zhu, Y. J. (2009). Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. *Mol. Genet. Genomics* 281, 609–626. doi: 10.1007/s00438-009-0434-x
- Qi, T., Seong, K., Thomazella, D. P. T., Kim, J. R., Pham, J., Seo, E., et al. (2018). NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant immunity in *Nicotiana benthamiana*. *Proc. Natl. Acad. Sci. U.S.A.* 115, E10979–E10987. doi: 10.1073/pnas.1814856115
- Qian, L. H., Zhou, G. C., Sun, X. Q., Lei, Z., Zhang, Y. M., Xue, J. Y., et al. (2017). Distinct patterns of gene gain and loss: diverse evolutionary modes of NBS-encoding genes in three solanaceae crop species. *Genes Genomes Genet.* 7, 1577–1585. doi: 10.1534/g3.117.040485
- Shao, Z.-Q., Zhang, Y.-M., Hang, Y.-Y., Xue, J.-Y., Zhou, G.-C., Wu, P., et al. (2014). Long-term evolution of nucleotide-binding site-leucine-rich repeat genes: understanding gained from and beyond the legume family. *Plant Physiol.* 166, 217–234. doi: 10.1104/pp.114.243626
- Shao, Z. Q., Wang, B., and Chen, J. Q. (2016a). Tracking ancestral lineages and recent expansions of NBS-LRR genes in angiosperms. *Plant Signal Behav.* 11, e1197470. doi: 10.1080/15592324.2016.1197470
- Shao, Z. Q., Xue, J. Y., Wu, P., Zhang, Y. M., Wu, Y., Hang, Y. Y., et al. (2016b). Large-scale analyses of angiosperm nucleotide-binding site-Leucine-rich repeat Genes Reveal Three Anciently Diverged Classes with Distinct Evolutionary Patterns. *Plant Physiol.* 170, 2095–2109. doi: 10.1104/pp.15.01487
- Shao, Z. Q., Xue, J. Y., Wang, Q., Wang, B., and Chen, J. Q. (2019). Revisiting the origin of plant NBS-LRR Genes. *Trends Plant Sci.* 24, 9–12. doi: 10.1016/j.tplants.2018.10.015
- Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernot, B., and Durand, D. (2012). Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 28, i409–i415. doi: 10.1093/bioinformatics/bts386
- Sun, G., Xu, Y., Liu, H., Sun, T., Zhang, J., Hettenhausen, C., et al. (2018). Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. *Nat. Commun.* 9, 2683. doi: 10.1038/s41467-018-04721-8
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739.
- Tarr, D. E., and Alexander, H. M. (2009). TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. *BMC Res. Notes* 2, 197. doi: 10.1186/1756-0500-2-197
- Wang, Y. P., Tang, H. B., DeBarry, J. D., Tan, X., Li, J. P., Wang, X. Y., et al. (2012). MCS-X: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293
- Wu, P., Shao, Z. Q., Wu, X. Z., Wang, Q., Wang, B., Chen, J. Q., et al. (2014). Loss/retention and evolution of NBS-encoding genes upon whole genome triplication of *Brassica rapa*. *Gene* 540, 54–61. doi: 10.1016/j.gene.2014.01
- Wu, Z., Li, M., Dong, O. X., Xia, S., Liang, W., Bao, Y., et al. (2019). Differential regulation of TNF-mediated immune signaling by redundant helper CNLs. *New Phytol.* 222, 938–953. doi: 10.1111/nph.15665
- Xue, J. Y., Wang, Y., Wu, P., Wang, Q., Yang, L. T., Pan, X. H., et al. (2012). A primary survey on bryophyte species reveals two novel classes of nucleotide-binding site (NBS) genes. *PLoS One* 7, e36700. doi: 10.1371/journal.pone
- Yang, S., Zhang, X., Yue, J. X., Tian, D., and Chen, J. Q. (2008). Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genomics* 280, 187–198. doi: 10.1007/s00438-008-0355-0
- Yuan, Y., Jin, X. H., Liu, J., Zhao, X., Zhou, J. H., Wang, X., et al. (2018). The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nat. Commun.* 9, 1615. doi: 10.1038/s41467-018-03423-5
- Zhang, M. P., Wu, Y. H., Lee, M. K., Liu, Y. H., Rong, Y., Santos, T. S., et al. (2010). Numbers of genes in the NBS and RLK families vary by more than four-fold within a plant species and are regulated by multiple factors. *Nucleic Acids Res.* 38, 6513–6525. doi: 10.1093/nar/gkq524
- Zhang, Y. M., Shao, Z. Q., Wang, Q., Hang, Y. Y., Xue, J. Y., Wang, B., et al. (2016). Uncovering the dynamic evolution of nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes in Brassicaceae. *J. Integr. Plant Biol.* 58, 165–177. doi: 10.1111/jipb.12365
- Zhang, G. Q., Liu, K. W., Li, Z., Lohaus, R., Hsiao, Y. Y., Niu, S. C., et al. (2017a). The *Apostasia* genome and the evolution of orchids. *Nature* 549, 379–383. doi: 10.1038/nature23897
- Zhang, Y. M., Xue, J. Y., Liu, L. W., Sun, X. Q., Zhou, G. C., Chen, M., et al. (2017b). Divergence and conservative evolution of XTNX genes in land plants. *Front. Plant Sci.* 8, 1844. doi: 10.3389/fpls.2017.01844

- Zhao, T., and Schranz, M. E. (2019). Network-based microsynteny analysis identifies major differences and genomic outliers in mammalian and angiosperm genomes. *Proc. Natl. Acad. Sci. U. S. A.* 116, 2165–2174. doi: 10.1073/pnas.1801757116
- Zhao, T., Holmer, R., de Bruijn, S., Angenent, G. C., van den Burg, H. A., and Schranz, M. E. (2017). Phylogenomic synteny network analysis of MADS-Box transcription factor genes reveals lineage-specific transpositions, ancient tandem duplications, and deep positional conservation. *Plant Cell* 29, 1278–1292. doi: 10.1105/tpc.17.00312
- Zhong, Y., Zhang, X. H., and Cheng, Z. M. (2018). Lineage-specific duplications of NBS-LRR genes occurring before the divergence of six *Fragaria* species. *BMC Genomics* 19, 128. doi: 10.1186/S12864-018-4521-4

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 Xue, Zhao, Liu, Liu, Zhang, Zhang, Chen, Zhou, Zhang and Shao. 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.



Comparative Transcriptome Analysis Reveals Key Pathways and Hub Genes in Rapeseed During the Early Stage of *Plasmodiophora brassicae* Infection

Lixia Li, Ying Long, Hao Li and Xiaoming Wu*

Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crop Research Institute, Chinese Academy of Agricultural Sciences, Hubei, China

OPEN ACCESS

Edited by:

Jia-Yu Xue,
Jiangsu Province and Chinese
Academy of Sciences, China

Reviewed by:

Wenxing Pang,
Shenyang Agricultural University,
China
Arvind H. Hirani,
Kemin Industries, Inc, United States
Chunyu Zhang,
Huazhong Agricultural University,
China

*Correspondence:

Xiaoming Wu
wuxm@oilcrops.cn

Specialty section:

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

Received: 27 August 2019

Accepted: 19 November 2019

Published: 17 January 2020

Citation:

Li L, Long Y, Li H and Wu X (2020)
Comparative Transcriptome Analysis
Reveals Key Pathways and Hub
Genes in Rapeseed During the Early
Stage of *Plasmodiophora*
brassicae Infection.
Front. Genet. 10:1275.
doi: 10.3389/fgene.2020.01275

Rapeseed (*Brassica napus* L., AACC, 2n = 38) is one of the most important oil crops around the world. With intensified rapeseed cultivation, the incidence and severity of clubroot infected by *Plasmodiophora brassicae* Wor. (*P. brassicae*) has increased very fast, which seriously impedes the development of rapeseed industry. Therefore, it is very important and timely to investigate the mechanisms and genes regulating clubroot resistance (CR) in rapeseed. In this study, comparative transcriptome analysis was carried out on two rapeseed accessions of R- (resistant) and S- (susceptible) line. Three thousand one hundred seventy-one and 714 differentially expressed genes (DEGs) were detected in the R- and S-line compared with the control groups, respectively. The results indicated that the CR difference between the R- and S-line had already shown during the early stage of *P. brassicae* infection and the change of gene expression pattern of R-line exhibited a more intense defensive response than that of S-line. Moreover, Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of 2,163 relative-DEGs, identified between the R- and S-line, revealed that genes participated in plant hormone signal transduction, fatty acid metabolism, and glucosinolate biosynthesis were involved in regulation of CR. Further, 12 hub genes were identified from all relative-DEGs with the help of weighted gene co-expression network analysis. Haplotype analysis indicated that the natural variations in the coding regions of some hub genes also made contributed to CR. This study not only provides valuable information for CR molecular mechanisms, but also has applied implications for CR breeding in rapeseed.

Keywords: *Brassica napus*, *plasmodiophora brassicae*, transcriptome, hub genes, glucosinolate, plant hormone

INTRODUCTION

Plasmodiophora brassicae Wor. (*P. brassicae*), an obligate and biotrophic pathogen of Rhizaria (Schwelm et al., 2015), could infect over 3,700 species in Brassicaceae (Hwang et al., 2012), and lead clubroot which has caused significant economic losses every year (Dixon, 2009). The *P. brassicae* has been discovered in more than 60 countries or regions (Dixon, 2009), the life cycle of which consists of dormant stage of resting spores, germination stage of resting spores, and secondary zoospore

reinfection. Once the conditions are suitable, primary zoospores are released from the resting spores to infect the root hairs when feel the stimulation of relevant signaling molecules secreted by host plants (Aist and Williams, 1971; Kageyama and Asano, 2009; Rolfe et al., 2016). The primary plasma mass is formed in the root hairs, and then divided to form secondary sporangium, from which the secondary zoospores are released. Secondary zoospores directly infect cortical cells, where secondary plasma mass form. Finally, the secondary plasma mass is divided to form mature resting spores, which are scattered in the soil and become the initial infection source in the coming year (McDonald et al., 2014). The resting spores of *P. brassicae* can survive for at least 7 years in the soil (Karling, 1968). Once contaminated, the field will no longer suitable for Brassicaceae crops (Howard et al., 2010). As early as 1930, pathogenic specialization has been found in *P. brassicae* (Honig, 1931). There are great differences in the biological and molecular characteristics of different pathogenic strains, which impede the research progress of pathogenesis. Up to now, only a few genes considering as the pathogenic factors in *P. brassicae* have been identified (Ando et al., 2006; Bulman et al., 2006; Feng et al., 2010).

Plants could defense pathogens with the help of physical (such as cell wall, cuticle, waxy layer, and xylogen) or chemical barriers (such as phenols, saponins, and mustard oil). Once the above defense is breached, the plant activates its defense immune system immediately, which consists of pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) and effector-triggered immunity (ETI, Jones and Dangl, 2006). PTI is the basal immune response, which could be inhibited by the effectors secreted by pathogens. Then, effector could be recognized by R protein in plants, consequently triggering a more dramatic immune response ETI (Bent and Mackey, 2007). Most R proteins were reported containing conserved motifs such as toll-interleukin receptor (TIR), nucleotide-binding (NB), leucine-rich repeat (LRR), coiled-coil (CC), or leucine zipper (Liu et al., 2007). Despite many clubroot resistance (CR) sites were identified in Brassicaceae, only three of which, *CRa* (Ueno et al., 2012), *Crr1a* (Hatakeyama et al., 2013), and *CRb* (Hatakeyama et al., 2017), have been cloned and found containing TIR-NB-LRR or NB-LRR.

In recent years, many studies were focused on the molecular mechanisms of CR with the help of “-omics” approach, especially in *Arabidopsis thaliana* and *Brassica* species. In *Arabidopsis*, the expression level of genes related to growth, sugar-phosphate metabolism, defense, and plant hormone had undergone large-scale changes after inoculation (Siemens et al., 2006). Some studies also elucidated the role of genes associated with metabolism, hormonal signaling pathways, and stress response (Jubault et al., 2008; Ludwig-Muller et al., 2009; Agarwal et al., 2011; Schuller et al., 2014). In *Brassica*, some studies indicated that the genes involved in the signaling metabolism of jasmonate and ethylene, defensive deposition of callose, and the biosynthesis of indole-containing compounds were all significantly up-regulated in clubroot-resistant plants compared with susceptible cultivars (Chu et al., 2014). It was confirmed that genes associated with PAMPs, calcium ion influx, hormone signaling, pathogenesis related, and cell-wall modification

played important roles in the interactions between *Brassica rapa* and *P. brassicae* (Chen et al., 2015). More recently, proteomic analysis found that two proteins related to salicylic acid (SA) mediated systemic acquired resistance and two proteins related to jasmonic acid (JA)/ethylene (ET) mediated induced systemic resistance in Chinese cabbage (Ji et al., 2018). Compared with the susceptible accessions, the genes involved in cell wall, SA signal transduction, phytoalexin synthesis, chitinase synthesis, Ca^{2+} signaling, and reactive oxygen species were significantly activated in resistant cabbage (Devos et al., 2006; Zhang et al., 2016). Remarkably, a series of studies were conducted to explore the relationship of glucosinolate (GLS) and CR in different species, which could provide more valuable information on CR mechanisms. GLS had been proved associated with clubroot disease symptoms both in *Arabidopsis* and *Brassica* species (LudwigMuller et al., 1997; Ludwig-Muller et al., 1999a; Ludwig-Muller et al., 1999b; Ludwig-Muller et al., 2009).

Rapeseed (*Brassica napus* L., AACC, $2n = 38$) is one of the most important *Brassica* crops around the world, which provides not only edible oil for human, but also protein-rich feed for animals. With intensified rapeseed cultivation, the incidence and severity of clubroot has also increased, which impedes the development of the rapeseed industry seriously. Up to now, large number of studies are focused on screening of resistant materials or mapping of CR genes/quantitative trait loci (Manzanares-Dauleux et al., 2000; Tewari et al., 2005; Werner et al., 2008; Zhang et al., 2015), but the resistance mechanisms of rapeseed against clubroot is still not clear. The results of protein level changes on infected *B. napus* showed that there were differences in proteins related to lignin synthesis, cytokinins metabolism, glycolysis, intracellular calcium ion balance, and reactive oxygen species detoxification (Cao et al., 2008). MicroRNA (miRNA) analysis on infected *B. napus* root showed that differential expressed genes (DEGs) of miRNA targets were predicted involved in transcriptional factors activity, hormone, and plant defense response (Verma et al., 2014). Furthermore, it was pointed out that genes related to the IAA (indole-3-acetic acid), SA, and JA pathways were involved in the reaction of *B. napus* to *P. brassicae* (Xu et al., 2016; Prerostova et al., 2018). The latest research showed that phenylpropanoid pathway was instrumental in resistance to clubroot disease progression in resistant line (Irani et al., 2019).

In this study, we investigated the early defense response of different resistant level rapeseed accessions to *P. brassicae*. Comparative dynamic analysis of the number of DEGs in R- (resistant) and S- (susceptible) line suggested that the differences between the R- and S-line had already shown in the early stage and the R-line was more sensitive to the invasion of *P. brassicae*. Functional enrichment analysis of DEGs revealed the important pathways responsible for CR in rapeseed. Based on that, the hub genes in each important pathway were screened out combining with the weighted gene co-expression network analysis (WGCNA). Haplotype analysis indicated that the natural variations in the coding regions of some hub genes made contributed to CR. This study not only provided valuable information for CR molecular mechanisms, but also had applied implications for CR breeding in rapeseed.

MATERIALS AND METHODS

Plant Materials, Resistance Identification, and Sampling

Two rapeseed accessions of 28,669 (resistant, R-line; *B. napus*, $2n = 4x = 38$) and YJ-8 (susceptible, S-line; *B. napus*, $2n = 4x = 38$) with contrasting performance on resistance to *P. brassicae*, which have experienced several resistance identifications in different environments, were used in the present study. These two lines were semi-winter type rapeseed, which were collected from the National Mid-term Gene Bank for Oil Crops of China. The R-line was double-high oil quality with GLS (94.5 $\mu\text{mol/g}$) and erucic (24.5%), while the S-line was double-low oil quality with GLS (28.5 $\mu\text{mol/g}$) and erucic (0%). The pathogen used in this study was collected from the infected field (IF) of Dangyang, China, where the pathogen was reported as pathotype 4 based on Williams classification (Ren et al., 2012). The seeds of R- and S-line were germinated on wet filter paper for 7 days, then transferred into the plastic pots filled with 10 L Hoagland nutrient solution adding the $\text{Ca}(\text{NO}_3)_2$ with 0.945 gram (g) per L additionally for 1 month under a 16 hpi (h) photoperiod at 25°C. Then, the seedlings were transferred into the fermentative soil as the proportion of 10^6 resting spores per g dry soil with the same condition of the culture room. The methods of making *P. brassicae* suspension and fermentative soil were also as the previous study (Li et al., 2016). We choose the 12, 24, 60, and 96 h post-inoculation of *P. brassicae* as the sampling time points for RNA-seq based on the results of the study (Dobson and Gabrielson, 1983). Three biological replicates of each treatment with one mock-control were performed. The roots of 10 plants were sampled at each replication for RNA sequencing. To verify successful infection, 20 plants of each accession, which were transferred into the fermentative soil with *P. brassicae*, were remained for resistance identification until the 42 days post-inoculation. The evaluation of severity of disease was as reported before (Kuginuki et al., 1999).

RNA Extraction and Construction of cDNA Sequencing Library

Total RNAs of 32 samples were extracted using the TRIzol reagent (Life Technologies, Carlsbad, California). The RNA quality (degradation and DNA contamination) was monitored on 1% agarose gels electrophoresis. The RNA purity and concentration were checked using the NanoPhotometer[®] spectrophotometer (IMPLEN, CA, USA), and Qubit[®] RNA Assay Kit in Qubit[®] 2.0 Fluorometer (Life Technologies, CA, USA), respectively. The integrity of RNA was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). A total amount of 3 μg RNA per sample was used for library preparation using NEBNext[®] Ultra[™] RNA Library Prep Kit for Illumina[®] (NEB, USA). In order to select complementary DNA (cDNA) fragments of preferentially 150–200 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 μL USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C before PCR. Then

PCR was performed and the products were purified (AMPure XP system). The library quality was assessed using the Agilent Bioanalyzer 2100 system.

RNA Sequencing and Data Preprocessing

All the 32 libraries were sequenced on the Illumina HiSeq platform and 150 bp paired-end raw reads were generated. Before assemblies, various quality-controlling measures for raw data were conducted. High-quality clean data was obtained by removing reads containing adapter, reads containing ploy-N, and low-quality reads from raw data. The high-quality paired-end clean reads were aligned to the *B. napus* reference genome (Darmor-bzh, Chalhoub et al., 2014) using TopHatv2.0.12 (Trapnell et al., 2009). Only uniquely mapped reads were considered for further analyses. The reads number mapped to each gene was counted using HTSeq v0.6.1. Then, the fragment per kilobases of transcript per million reads (FPKM) of each gene was calculated based on the length of the gene and reads count (Trapnell et al., 2010). The software of Microsoft Excel 2010 was used to calculate the person correlation of biological replicates.

Identification of Differentially Expressed Genes and Real Time-PCR Verification

Differential expression analysis of two samples was performed using the DESeq R package (1.18.0). Genes with an adjusted p -value (padj) < 0.05 and $|\log_2(\text{fold change})| > 0$ found by DESeq were considered as DEGs. To distinguish the DEGs more clearly, the concept of relative differentially expressed genes (RDEGs) was introduced in this study. A gene was defined as a RDEG, when it was not only identified as DEG between R- and S-line, but also identified as DEG in R-line (compared with the R-mock) or in S-line (compared with the S-mock) at the corresponding time points. RT-PCR was carried out to confirm the RNA-seq results. cDNAs were synthesized from the same RNAs as for RNA-seq. The results analysis performed on LightCycler 480 SYBR Green I Mastermix, and a LightCycler 480II real-time PCR system (Roche, Switzerland). The transcript abundance calculated from three biological and three technical replicates with *Bna.ACTIN7* (*BnaA03g55890D*) as internal control. The fold change was estimated using the $2^{-\Delta\Delta\text{CT}}$ (Livak and Schmittgen, 2001). The gene-specific primers sequences were list (Table S1).

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Enrichment Analysis of Differentially Expressed Genes

Gene Ontology (GO) enrichment analysis was implemented by the GOrseq R package, in which gene length bias was corrected. GO terms with $\text{padj} \leq 0.05$ were considered as significantly enriched. Kyoto Encyclopedia of Genes and Genomes (KEGG) was carried out online (<http://www.genome.jp/kegg/>). KOBAS software was used to test the statistical enrichment of differential expression genes in KEGG pathways (Mao et al., 2005). The pathways with $\text{padj} \leq 0.01$ were considered as significantly enriched. The heat maps performed on the software of Genesis.

Co-Expression Network Analysis and Prediction of Hub-Genes

The co-expression network analysis was conducted using WGCNA version 1.61 package in R software (Langfelder and Horvath, 2008). Module identification was implemented after merging of modules whose expression profiles were similar with a merge CutHeight of 0.25. The interaction network of hub-genes in module was visualized using Cytoscape 3.5.1.

Haplotype Analysis

The primers were designed for PCR of three hub genes genomic sequences in a population including 130 accessions (Table S2). Single-nucleotide polymorphism (SNP) information of hub genes in the population was obtained by blasting the sequences of the PCR products. The software of Haploview was used to analyze the haplotype in the population. The phenotype data of 130 rapeseed accessions used for the haplotype analysis were obtained by artificial inoculation at seedling stage in greenhouse from Li et al. (2016). The significance of phenotype difference among different haplotypes was detected by *t*-test using Microsoft EXCEL 2010 and visualized by the violin plot (<http://shiny.chemgrid.org/boxplotr/>).

RESULTS

Phenotype Characterization of Two Rapeseed Accessions With Contrasting Resistance to Clubroot in Different Environments

Two rapeseed accessions, R- and S-line (the most constant accessions), were screened out from a natural population containing 472 accessions (Li et al., 2014), which were performed the CR evaluation in three environments (data were not shown on). The morphological differences between R- and S-line in greenhouse by artificial inoculation were showed in Figure 1A and that in IF by natural infection were showed in Figure 1B. The difference of disease index (DI) between R- and S-line was significant in any of environments (Figure 1C). The DI of R-line was 28.5 at the seedling stage in greenhouse by artificial inoculation, while the DI of S-line was 87.91; the DI of R-line was 6.64 at the seedling stage in IF, which was significant lower than that of S-line (46.35). At the flowering stage in IF, the DI of R-line was 10.54, while the S-line was 42.7. It was obvious that artificial inoculation in greenhouse can increase the DI greatly compared with natural infection in field. The result suggested that the resistance difference between these two accessions was highly stable, which were reliable for the further study.

RNA-Sequencing Analysis and Global Comparison of Transcriptomes on Infected Roots by *Plasmodiophora brassicae* Revealed the Difference of the Early Stages Between R- and S-Line

To explore the molecular basis difference of early defense response induced by *P. brassicae* on different resistant level

rapeseed accessions, RNA-seq analysis was conducted to generate transcriptome profiles. RNA was extracted from the roots of the R- and S-line at 12, 24, 60, and 96 h after infected by *P. brassicae* with three biological replicates per treatment, respectively. A total of 32 libraries were constructed and analyzed. In total, approximately 1.77 billion raw reads were generated from the 32 samples, and 1.74 billion high-quality clean reads with an average of 54.32 million clean reads (a total of 8.15G) for each sample were obtained after removing low-quality reads. The GC content of the sequence data from the 32 libraries were all around 46.6%, and the Q30 values were all above 90%, indicating that the quality and accuracy of sequencing data was sufficient for further analyses. On average, 88.9% clean reads were mapped on the *B. napus* reference genome, and about 95.7% of which were matched uniquely (Table S3). Pearson correlation coefficients of three biological replicates in each treatment for both the R- and S-lines were high ($R^2 > 0.90$ in most cases, Figure S1), which indicated that the RNA-seq data was of high quality and consistency. The number of transcript of each sample was showed after removing the genes with a FPKM value (average of three biological replicates) < 1 . In total, 55,240 and 54,538 transcripts were identified in mock-inoculated and inoculated samples by *P. brassicae* of R-line, respectively. Similarly, 55,821 and 55,560 transcripts were detected in S-line, respectively (Table 1). Overall, the numbers of expressed genes in any samples accounted for 47.5–50.5% of the 101,040 *B. napus* annotated genes. There was no significant difference in the number of transcript among the different sampling point both in R- and S-line. Similarly, the clean reads also mapped to the *P. brassicae* genome for analysis of the genome enrichment in rapeseed roots. However, it was failed to acquired enough reads to perform the further analysis maybe because that the pathogen invasion time is too short (the longest time-point was 96 h) to get *P. brassicae* genome information in rapeseed roots.

Identification of Differentially Expressed Genes and Validation of RNA-Sequencing by RT-PCR

Compared with the corresponding mock group, 3,171 and 714 DEGs were detected in R- and S-line with $\text{padj} < 0.05$, and $|\log_2(\text{fold change})| > 0$, respectively. A total of 159 genes were detected in both lines, while 3,012 and 555 genes were R-line specific and S-line specific, respectively (Figure 1D). In R-line, 240 DEGs were present at more than one time points (Figure S2A). Similarly, 42 DEGs presented at more than one points were identified in S-line (Figure S2B). Subsequently, the expression patterns of DEGs identified at different sampling time points in both lines were investigated. At 12 h after inoculated, 493 genes were up-regulated and 668 genes were down-regulated in R-line, while only 23 genes were up-regulated and 24 genes were down-regulated in S-line, respectively. The similar situation was also present at 96 h, 974 genes were up-regulated, and 976 genes were down-regulated in R-line, while only 19 genes were up-regulated and 81 genes were down-regulated in S-line, respectively (Figure 1E). The above results indicated that the difference between R- and S-line had already shown in the early stage after being attacked by *P. brassicae*. Overall, the change of gene expression

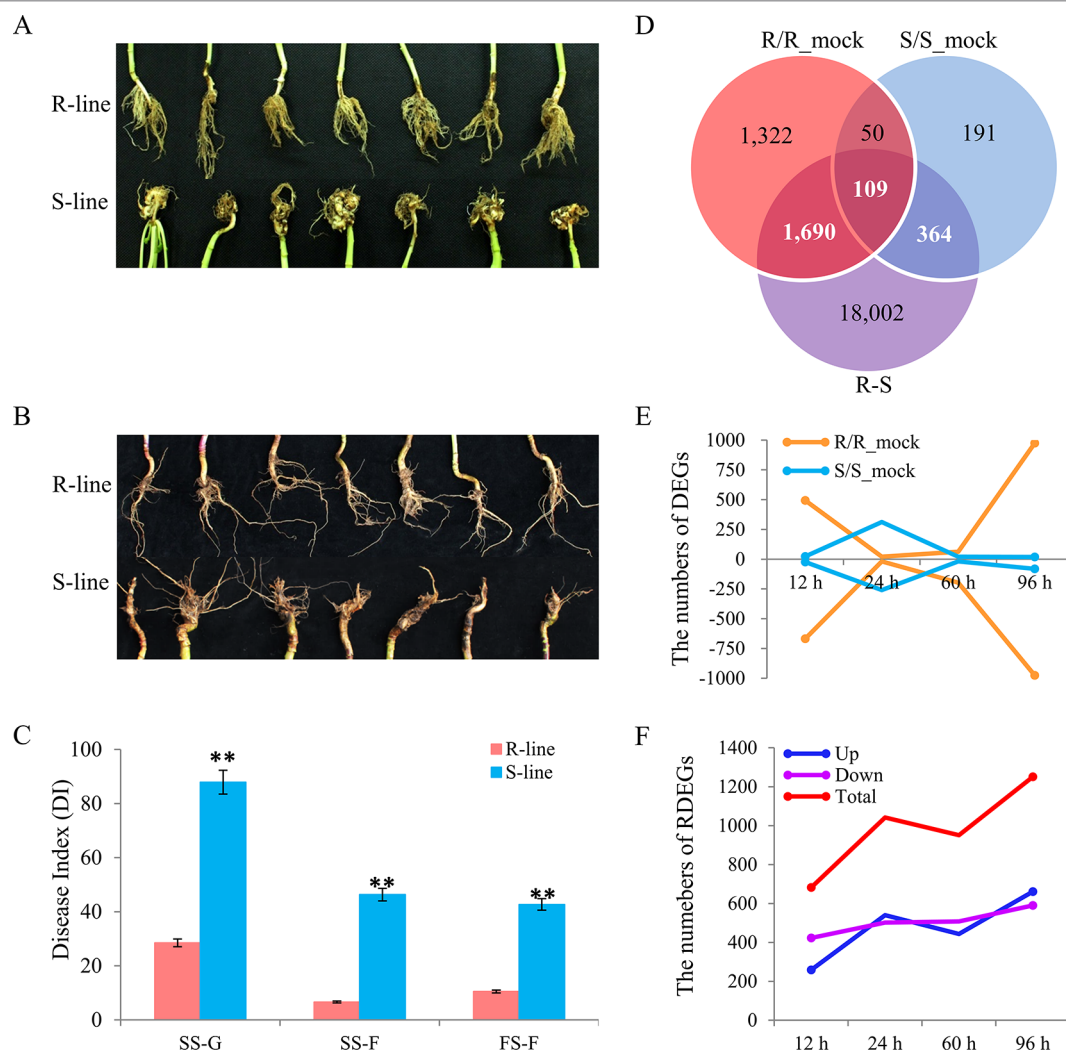


FIGURE 1 | Phenotype characterization in different environments, and identification of differentially expressed genes (DEGs) and relative differentially expressed genes (RDEGs) after infected by *Plasmodiophora brassicae*. Phenotype difference between the R- and S-line at seedling stage (A) and flowering stage (B). (C) Disease index of the two accessions in different environments. SS-G, SS-F, and FS-F indicated the seedling stage in greenhouse, field, and the flowering stage in field, respectively. (D) Venn diagram of DEGs in the R/R_mock, S/S_mock, and R/S. The sum of white and bold numeral represented the number of RDEGs. (E) Dynamic variation of up- and down-DEGs in each accession. The negative indicated the number of down-DEGs. The yellow and blue line indicated the DEGs of R- and S-line compared with the corresponding mock group, respectively. (F) Dynamic variation of RDEGs. The blue, purple, and red line indicated the up-, down-, and total RDEGs, respectively.

TABLE 1 | Transcripts statistics of different sampling point expressed in mock-inoculated and inoculated samples.

	12 h		24 h		60 h		96 h		Total	
	Mock inoculated	Inoculated	Mock inoculated	Inoculated	Mock inoculated	Inoculated	Mock inoculated	Inoculated	Mock inoculated	Inoculated
R-line	48,150	49,839	50,280	50,592	49,981	50,880	49,582	50,630	55,240	54,538
S-line	48,015	50,104	50,238	50,798	49,869	50,605	51,003	50,868	55,821	55,560

pattern of R-line exhibited a more intense defensive response than that of S-line, which manifested the R-line was more sensitive to the invasion of *P. brassicae*. To explore the key important genes responsible for the difference of CR between the two accessions, 4,567, 10,065, 7,453, and 9,477 DEGs were identified between R- and S-line at 12, 24, 60, and 96 h after inoculated, respectively.

Finally, a total of 2,163 RDEGs were screened out from all DEGs for further analyses (Figure 1D). In detail, compared with S-line, 259, 540, 443, and 661 up-regulated RDEGs were identified in R-line at 12, 24, 60, and 96 h after inoculation, respectively. Similarly, 423, 502, 508, and 590 down-regulated RDEGs were identified in that of S-line (Figure 1F).

To validate the quality of RNA-seq data and difference expressional level, 29 RDEGs were selected randomly for RT-PCR. The relative expression level measured by RT-PCR was converted to fold changes (R/S). All RT-PCR data was collected from three technical replicates for each sampling time point and the strong correlation between the RNA-seq and RT-PCR data were present ($R^2 = 0.852-0.986$, **Figure S3**), which indicated that the transcriptomic profiling data was of reliability.

Functional Enrichment Analyses of Differentially Expressed Genes and Relative Differentially Expressed Genes

To understand the biological mechanisms of CR deeply, GO enrichment analyses of the DEGs in R-line (compared with the R-mock) and S-line (compared with the S-mock) were conducted, respectively. In total, the 3,171 DEGs in R-line were assigned to 60 terms belonging to three categories: biological process (19 terms), cell components (5 terms), and molecular function (36 terms) significantly. Seven hundred fourteen DEGs in S-line were assigned to 26 terms belonging to two categories: biological process (14 terms) and molecular function (12 terms) (**Figure 2**). It was noticeable that 48 terms were enriched in R-line specifically, which were mainly related to immune system process, sulfate transport, extracellular region, cell wall, calcium ion binding, xyloglucosyl transferase activity, and others. While, only 14 terms were enriched in S-line specifically, which were involved in the amino sugar and chitin metabolism (**Figure 2**). In addition, GO analysis was conducted for the DEGs at each stage in both lines. The

results showed that 30 pathways were enriched at 12 h after inoculation, accounting for 50% of all pathways in R-line. However, the DEGs at 12 h after inoculation in S-line were not enriched into any pathway. On the contrary, 20 pathways were enriched at 24 h after inoculation, accounting for 76.9% of all pathways in S-line. Compared with the R-line, the S-line showed a delay in response to pathogen invasion, which was mainly reflected in the response to biological stimulation and oxidative stress (**Figure 2**).

Furthermore, KEGG analysis was performed on the RDEGs and 30 pathways were showed, only eight of which were enriched significantly with a cutoff value of $p_value < 0.05$ (**Figure 3A**). The results revealed that the pathways of GLS biosynthesis, pyridine alkaloid biosynthesis, fatty acid metabolism, plant hormone signal transduction, sulfur metabolism, tryptophan metabolism, and carotenoid biosynthesis, might involve in the regulation of CR in *B. napus*. In order to show the expression differences of the RDEGs enriched in above pathways, heat map was prepared. It is noticeable that the genes involved in GLS biosynthesis were up-regulation both in R- and S-line at 12 h after inoculation, which were also up-regulation in R-line compared with the S-line. At 96 h after inoculation, the genes involved in GLS biosynthesis were up-regulation in R-line, while down-regulation in S-line. In addition, the expression differences of genes enriched in fatty acid metabolism performed consistently at 96 h after inoculation, which were all up-regulation. In the pathway of plant hormone signal transduction, the genes encoding auxin-responsive protein family were down-regulation in R-line compared with that in S-line. On the contrary, the genes encoding

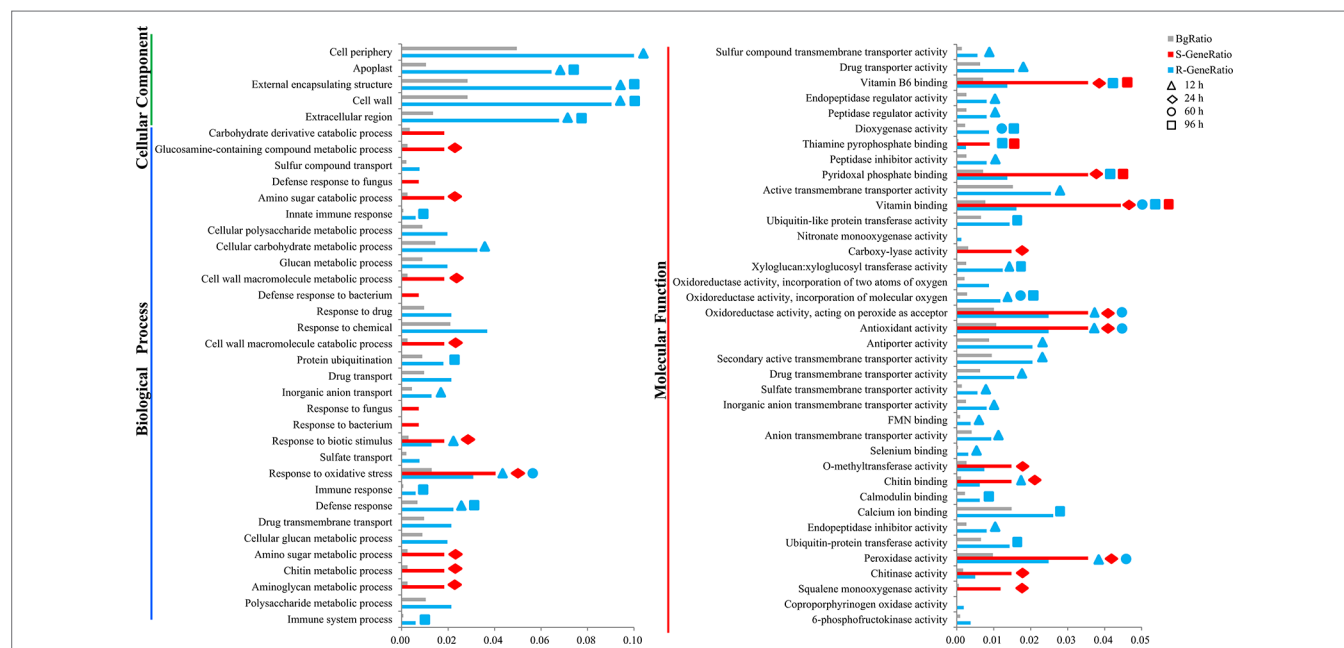


FIGURE 2 | Gene ontology (GO) enrichment analysis of the differentially expressed genes in the R- and S-line. The y-axis meant the GO terms enriched in the R- or S-line belonging to three categories, and the x-axis meant the percentage, which calculated by the formula that the gene number enriched in each term dividing that in corresponding categories. The gray, red, and blue pillars meant the percentage in the background, S-line, and R-line, respectively. The pillars marked with the shape of triangle, rhombus, circle, and square represented the terms enriched at 12, 24, 60, and 96 h after infection, respectively.

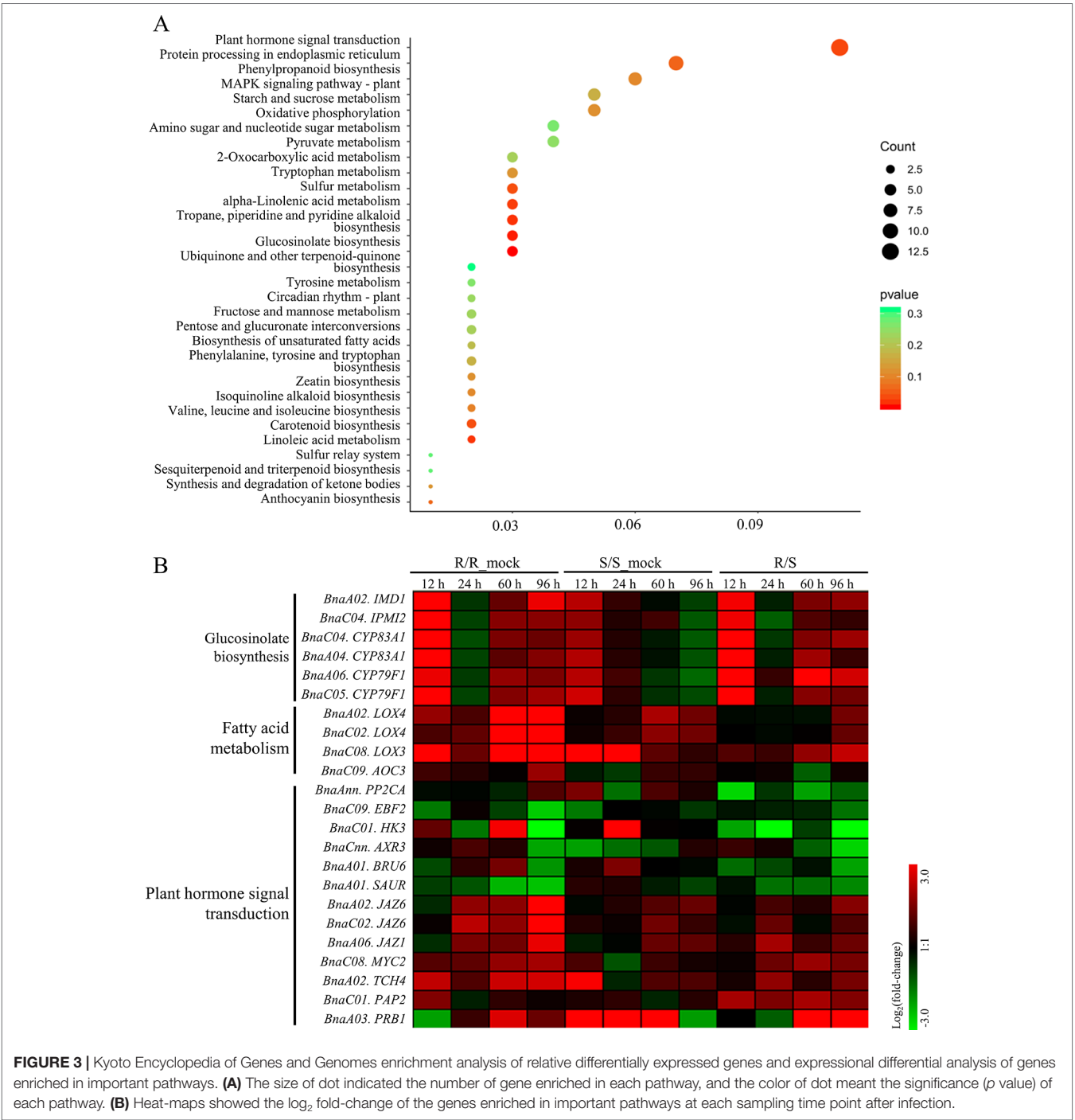


FIGURE 3 | Kyoto Encyclopedia of Genes and Genomes enrichment analysis of relative differentially expressed genes and expressional differential analysis of genes enriched in important pathways. **(A)** The size of dot indicated the number of gene enriched in each pathway, and the color of dot meant the significance (p value) of each pathway. **(B)** Heat-maps showed the \log_2 fold-change of the genes enriched in important pathways at each sampling time point after infection.

jasmonate-zim-domain protein were up-regulation in R-line compared with that in S-line (**Figure 3B**).

Construction of Gene Co-Expression Networks and Prediction of Hub Genes Related to Clubroot Resistance in *Brassica napus*

To obtain an insight for understanding the molecular mechanisms in depth and comprehensively, we carried the

WGCNA analysis to construct the gene co-expression network. All 2,163 RDEGs were assigned into 13 distinct modules labeled with different colors (**Figure 4A**), except 4 of which cannot be assigned into any module were put into gray module. The module was a cluster of highly interconnected genes with similar expression changes in a physiological process. The number of RDEGs that the modules harbored was varying from 30 (salmon) to 589 (turquoise). Then, we associated the modules with each of samples, which demonstrated that six modules (green, pink, salmon, magenta, blue, and yellow)

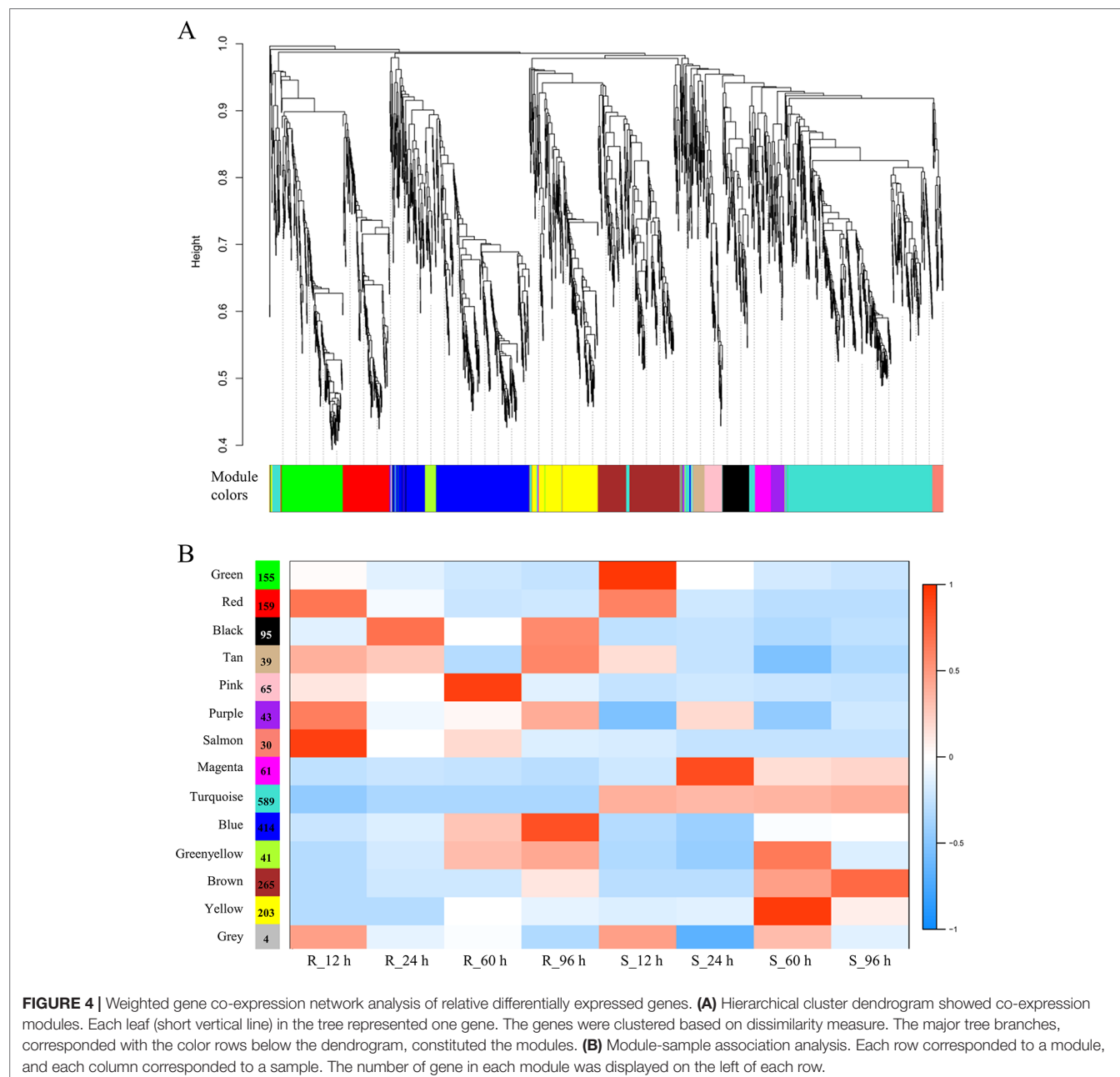


FIGURE 4 | Weighted gene co-expression network analysis of relative differentially expressed genes. **(A)** Hierarchical cluster dendrogram showed co-expression modules. Each leaf (short vertical line) in the tree represented one gene. The genes were clustered based on dissimilarity measure. The major tree branches, corresponded with the color rows below the dendrogram, constituted the modules. **(B)** Module-sample association analysis. Each row corresponded to a module, and each column corresponded to a sample. The number of gene in each module was displayed on the left of each row.

showed highly correlation with R-line or S-line (**Figure 4B**). Noticeably, the purple module performed opposite expression patterns between R and S-line at any sampling time point and the same simulation happened on turquoise module.

Furthermore, we performed KEGG enrichment analysis of the above eight modules. The genes of three modules were enriched in six pathways significantly, which were marked with the bold font in the **Table 2**. It was worth mentioning in particularly that the genes of blue (414 RDEGs) and purple (43 RDEGs) modules could be involved in fatty acid metabolism, plant hormone signal transduction, and GLS biosynthesis, which were accordance with the KEGG analysis of all RDEGs (**Table 2**). Subsequently, the interaction network of fatty acid metabolism, plant

hormone signal transduction, GLS biosynthesis genes in blue module, and purple module were constructed and visualization using Cytoscape3.6.1. The network showed the lipoxygenase (*LOX4*) in fatty acid metabolism, and jasmonate-zim-domain protein (*JAZ1* and *JAZ6*) in plant hormone signal transduction played pivotal role in the CR (**Figure 5A**). At the same time, isopropylmalate isomerase 2 (*IPMI2*), isopropylmalate dehydrogenase 1 (*IMD1*), methylthioalkylmalate synthase -in pathway of GLS biosynthesis also played an important role in resistance to *P. brassicae* (**Figure 5B**). In total, there were 12 genes were highlighted after WGCNA and interaction network analyses, which were considered to be the hub genes for CR in the R-line (**Figure 5**).

the DI of Hap2 (64.18) was significantly lower than that of the other three haplotypes among which there was no significantly difference (**Figure 6A**). The results indicated that Hap2 was the favorable haplotypes of *BnaA04.CYP83A1* for CR. Similarly, the

gene of *BnaA06.JAZ1*, involved in the pathway of plant hormone signal transduction, had four exons, three introns, 5' UTR, and 3'UTR. The entire *BnaA06.JAZ1* (1,547 bp) was also sequenced in the 130 accessions. One, three, three, and two SNPs were

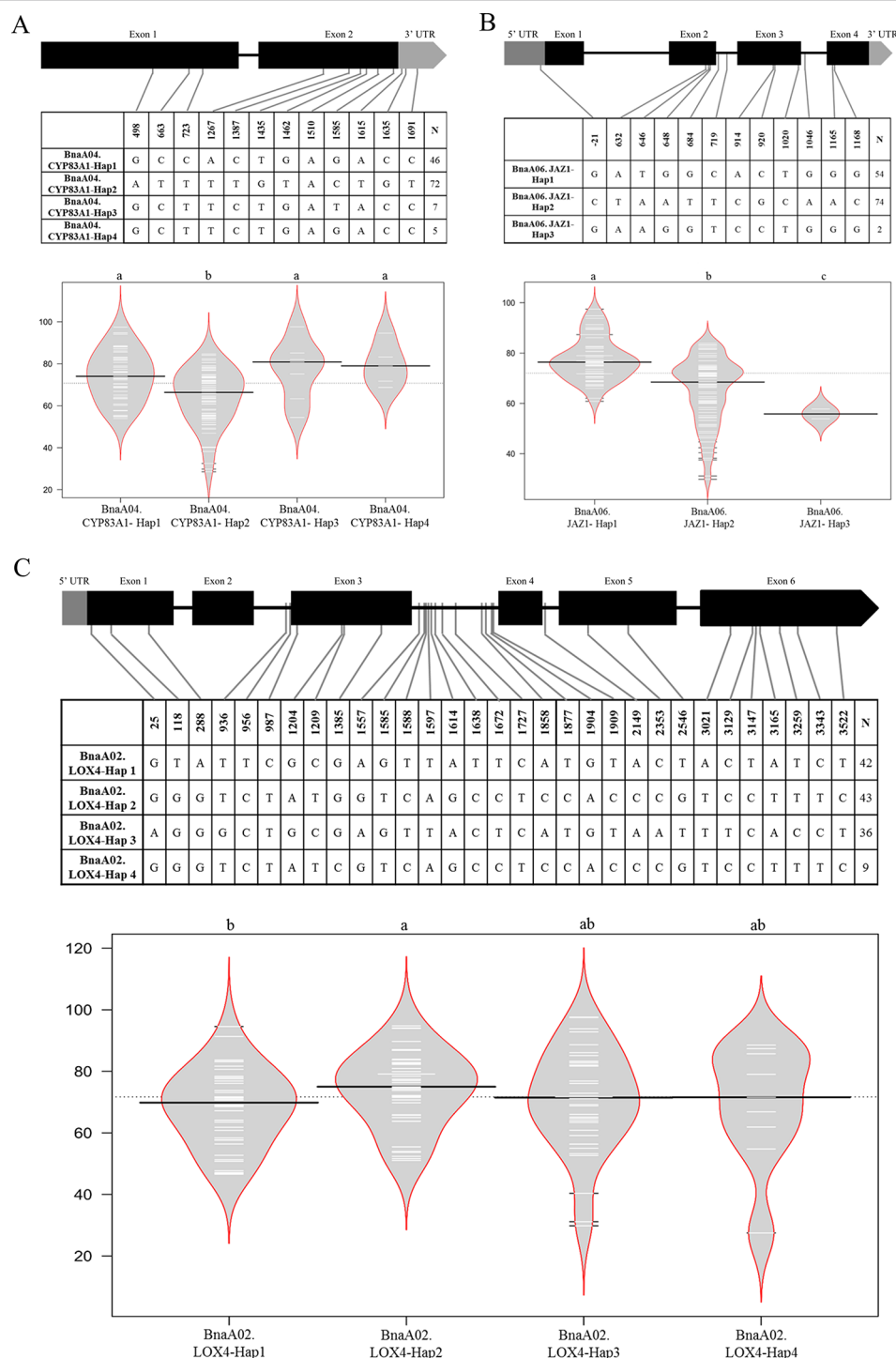


FIGURE 6 | Haplotype analyses of *BnaA04.CYP83A1* (**A**), *BnaA06.JAZ1* (**B**), and *BnaA02.LOX4* (**C**). The exon, and UTR in the gene structure were displayed with black, and gray box, respectively. The single nucleotide polymorphism (SNP) positions were connected to the haplotype table by lines, and the number of accessions carrying each haplotype was indicated in the columns of right. The difference significance analysis among haplotypes was displayed by the violin plot, in which the black and white horizontal lines represented the medians and individual data points, respectively.

detected in the 5'UTR, exon 2, exon 3, and exon 4, respectively. Haplotype analysis of *BnaA06.JAZ1* showed that three haplotypes were constructed. Hap1 and Hap2 were the prevalent haplotypes, represented by 54 and 74 accessions severally, while Hap3 was rare type represented only by 2 accessions. Phenotype difference of any two haplotypes was significantly and the Hap2 was the favorable haplotype of *BnaA06.JAZ1* for CR (**Figure 6B**). The gene of *BnaA02.LOX4* participated in fatty acid metabolism had six exons, five introns, and 5' UTR. PCR products were sequenced for *BnaA02.LOX4* (3,757 bp) in the 130 accessions. Three, four, two, and seven SNPs were in the exon 1, exon 3, exon 5, and exon 6, respectively. Haplotype analysis of *BnaA02.LOX4* showed that four haplotypes were constructed. Hap1, Hap2, and Hap3 were the prevalent haplotypes, represented by 42, 43, and 36 accessions severally, while Hap4 was rare type represented only by 9 accessions. Combined with the phenotype data of artificial inoculation identification, it was found that the DI of Hap1 (68.12) was significantly lower than that of Hap2 (73.92, **Figure 6C**), which indicated that Hap1 was the favorable haplotypes of *BnaA02.LOX4* for CR.

To better understand how the mutations were impact on the protein transcriptions or the functions (whether the mutations in the motifs or not), the online software (<https://www.genome.jp/tools/motif/> and <https://prosite.expasy.org/scanprosite/>) were used to predict motifs of above three genes. The results showed that, for *BnaA04.CYP83A1*, only the SNP (Pos_1510) caused missense mutation that from the isoleucine to methionine, which was not in the motif of this gene (CYTOCHROME_P450), though; for *BnaA06.JAZ1*, there were four missense mutations (Pos_646, 648, 914, and 920), also not in the motifs of the gene (TIFY and CCT_2); for *BnaA02.LOX4*, there were five missense mutations (Pos_25, 118, 1204, 1385, and 2546), also not in the motifs of the gene (LIPOXYGENASE_1 and LIPOXYGENASE_2).

DISCUSSION

Many studies focused on the middle or late phase/stage of clubroot course in *Arabidopsis*, *B. rapa* (Chinese cabbage), *Brassica oleracea*, and *B. napus* on the level of transcriptomics, proteomics, or others (Zhang et al., 2016; Hao et al., 2017; Irani et al., 2018; Ji et al., 2018; Prerostova et al., 2018; Su et al., 2018; Peng et al., 2019), while few studies were aimed at the early infection. In recent years, some researchers put forward that the early infection also played an important role. Both *Arabidopsis* and Chinese cabbage were carried out the study on the mechanisms of early infection (Chen et al., 2015; Zhao et al., 2017), which showed that part of pathways or proteins identified in the middle or late phase also could be detected in the early infection stage. In this study, we investigated the response of *B. napus* accessions with different resistance levels at the early stage of *P. brassicae* infection, which could provide more information about the mechanisms of early infection of *Brassica* crops to *P. brassicae*. It speculated that resistant genotype could sense pathogen invasion earlier because that more DEGs were detected in R-line and the 12 h/96 h after infection might be the key time points for the R-line. The dynamic change of RDEGs number

showed that the number at 12 h after inoculation was the least. We considered that most resistance reaction at this point in R-line might belong to the basal reaction, which could occur in S-line in spite that it would occur later. In addition, the number of RDEGs reached the maximum at 96 h after inoculation, and there was a growing trend over time. It was conjecture that the main reason for this change was that many specific responding of R-line revealed gradually in this process. Therefore, it illustrated that the sampling time point in this study was desirable.

In this study, the concept of RDEG was introduced. Compared with simple analysis of DEGs in R- and S-line, it was more convinced and targeted to discover the key genes that lead to the difference between R- and S- line. It could filter some genes belonging to basic resistant pathways. The expressional level of these genes had no difference between R- and S-line. Eight metabolic pathways were identified involved in the regulation of CR by KEGG enrichment analysis of RDEGs. Among that, two pathways, GLS biosynthesis and plant hormone signal transduction, have been repeatedly reported to be involved in the regulation of CR (Ludwig-Muller, 2009), which further proved the referential value of the information obtained in this study. It was worth mentioning that the pathway of tryptophan metabolism was considered synergy with above two pathways, because that it was a precursor of auxin and various secondary metabolites, such as camalexin and GLS. A new study indicated that suppression of tryptophan synthase could activate cotton immunity by triggering cell death *via* promoting SA synthesis (Miao et al., 2019). While, few study about another four pathways of fatty acid metabolism, pyridine alkaloid biosynthesis, carotenoid biosynthesis, and sulfur metabolism related to CR was reported. For all this, the four pathways were reported involved in other disease resistance reaction, which also provided some new ideas for study on CR mechanisms. Alkaloids, as an important natural phytoalexin, were accumulated in large amount when plants were stimulated by adverse environment. The relationship between alkaloids and disease resistance was mainly manifested as inhibition of spore germination and mycelial growth of pathogenic fungi, or inhibition and inactivation of enzymes or toxins produced by pathogens. Some plant hormones had inhibitory effects on the synthesis of alkaloids. For example, auxin negatively regulated the synthesis of alkaloids (Kutchan, 1995). Plant immune response was a very complex biological process, there were abundant genes participated in this process. Some key pathways, like GLS biosynthesis and plant hormone signal transduction, were considered as the key pathways in the plant immune response process. Our data also reflected the importance of GLS biosynthesis and plant hormone signal transduction on the CR resistance in rapeseed (**Table 2, Figures 4 and 5**). On the other hand, some pathways were indirect involved or influenced the plant immune, like cell wall, lipid metabolism, glycolysis/gluconeogenesis, which might affect the energy flux from pathway to pathway or as the physical barrier to impact the plant disease resistance response. Also, our RNA-Seq data identified the indirect involvement pathways on the CR resistance in rapeseed, like linoleic acid metabolism, sulfate transport, and gluconeogenesis (**Table 2**). It was indicated that these pathways also very important in the plant immune response, despite the effect was indirect. It

was also verified that transcriptome was a powerful method for understanding the complex biological questions.

Fatty acids were also involved in the regulation of plant responses to various biotic and abiotic stresses, primary metabolites of which played an important role in signal transduction of plant disease resistance. Oleic and linoleic acid can induce the activation of nicotinamide adenine dinucleoside phosphate oxidase mediated by protein kinase C, thus inducing the production of plant reactive oxygen species involving the plant disease resistance (Cury-Boaventura and Curi, 2005). JA was an important plant hormone derived from fatty acid metabolism. Therefore, it was not surprised that these two pathways were detected in the network of blue module, simultaneously. Fatty acid desaturation was an important part of plant defense reaction and two genes (*LOXs*) encoding lipoxygenase, which catalyzed the oxygenation of fatty acids, were screened out in this study. In the meanwhile, two genes (*JAZs*) encoding jasmonate-zim-domain protein, which involved in the JA signaling pathway were also screened out. JA and its derivatives played an important role in mediating plant resistance to various kinds of biological stress, as well as in the process of vegetative reproduction, cell cycle regulation. In addition, some studies suggested that *JAZ1* maybe connect the auxin and JA signaling pathways. JA was also believed to mediate the anabolism of alkaloids. The results presented a complex metabolic network formed by the interaction of multiple metabolic pathways, which could provide more information to explore the regulation of CR. Although Chen et al. (2015) also identified the *LOXs* and *JAZs* involved in the regulation of *B. rapa* against to *P. brassicae*. However, the results showed that *JAZs* were down-regulated in CR BJN3-2, and it was concluded that the SA signaling pathway, not the JA/ET played a crucial role in resistance of *B. rapa* against to *P. brassicae*. On the contrary, both our study and Zhang et al. (2016) showed that *JAZs* were up-regulated in R-line, inhibiting the JA signaling pathway, to improve the CR. It revealed the importance of JA signaling pathway to the CR. Numerous studies had shown that CR in *B. rapa* was considered as a quality trait, while CR in *B. oleracea* had been considered as a quantitative trait (Piao et al., 2009; Lee et al., 2016). It was a pity that the resistance mechanisms of *B. napus* to *P. brassicae* was still unknown. This study could further reveal the genetic mechanisms of CR in *B. napus*.

GLS, as a kind of secondary metabolite widely existing in cruciferous plants, were broadly existed in *Brassica* species. The degradation products of GLS had widely biological functions, which could not only regulate auxin metabolism, but also participate in plant defense reaction, preventing and controlling plant diseases. In general, the biosynthesis of GLS included three stages: extension of precursor amino side chain, formation of GLS core structure, and modification of side chain groups (Grubb and Abel, 2006). According to the different side chains, the GLS could be divided into aliphatic, aromatic, and indole. It had been proved that different group GLS and their corresponding isothiocyanates were only resistant to specific pathogen (Tierens et al., 2001). Up to now, most studies focused on indole-GLS, while information on aliphatic and aromatic-GLS was limited. Aliphatic-GLS was considered playing the role of defense by releasing toxic thiocyanate and isothiocyanates, while indole-GLS precursor was regarded as the synthesis of auxin, associated with the root of forming large size. It was supposed that indole-GLS may directly or indirectly promote

clubroot incidence degree (Ludwig-Muller, 2009). At present, the relevant major synthesis and regulation genes in biosynthesis of GLS have been verified, which were determined by *MAM*, *CYP79/CYP83*, *AOP*, and other synthetic gene families (Hull et al., 2000; Hansen et al., 2001; Mikkelsen et al., 2002; Hirai, 2008). The TFs of *MYB* gene family played the role of regulating the above genes (Yan and Chen, 2007; Gigolashvili et al., 2009). The gene, *CYP83A1*, which involved in the formulation of GLS core structure, was screened as a candidate gene by combining functional enrichment analysis with co-expression network analysis in this study. Meanwhile, the contribution of *CYP83A1* to CR was verified in a population by correlation analysis of CR phenotypes with the haplotypes of *CYP83A1* in rapeseed. Studies showed that *CYP83A1* had a high affinity for aliphatic acetaldoxime compared with its homologous gene *CYP83B1*, which had a high affinity for indole acetaldoxime (Bak and Feyereisen, 2001). The results of this study indicated that aliphatic-GLS also played an important role in the regulation of CR, which could extend the understanding on the contribution of different group of GLS to CR in *Brassica*.

In addition, WGCNA analysis also revealed genes interacting with above key genes, especially some important TFs and *R* genes, which were worthy of attention in the following study. Several TFs (*WRKY33*, *WRKY40*, *WRKY67*, *MYB15*, *MYB77*, and *MYC2*) were identified in the blue module network, which were reported to involve in defense to biotic stress (Lorenzo et al., 2004; Pandey et al., 2010; Birkenbihl et al., 2012; Chezem et al., 2017). In addition, two *R* genes (*TIR-NBS-LRR*) were identified from the network of blue related to the fatty acid metabolism and plant hormone signal transduction. The expressional data showed that both *R* genes were induced express after 72 h infected by *P. brassicae*. And compared with the S-line, both *R* genes were up-regulated in the R-line. WGCNA analysis could provide valuable information for the establishment of regulatory network of CR in rapeseed.

Through the RNA-Seq analysis at the early stage of the infection the *P. brassicae*, a total of 12 hub genes related to club root resistance were obtained by WGCNA and interaction network analyses. The information of these genes is powerful for CR improvement breeding in *Brassica*. Haplotype and mutations details resulted by SNPs analyses were performed in this study (Figure 6). It could be concluded that the natural variations of target genes could affect the CR, although these variations were not located on the motif of the genes. Even so, the mutations might cause the protein substrate binding ability or enzyme activity had slightly changes, thus results in the phenotype differences. The results appeared to be particularly important for understanding the mechanisms on CR resistance in *B. napus*.

DATA AVAILABILITY STATEMENT

The raw transcriptome reads have been deposited into NCBI Short Read Archive (SRA) under accession number PRJNA564005.

AUTHOR CONTRIBUTIONS

XW conceived the study. LL and XW designed the experiments. LL organized the implementation and analyzed

the data of experiment. YL and HL participated in the RT-PCR and phenotype identification. LL wrote the paper. All the authors have read and approved the publication of the manuscript.

FUNDING

The National Key Program for Research and Development (2016YFD0100202) and The Germplasm Resources Protection Project in China (2019NWB040) supported this work.

REFERENCES

- Agarwal, A., Kaul, V., Faggian, R., Rookes, J. E., Ludwig-Muller, J., and Cahill, D. M. (2011). Analysis of global host gene expression during the primary phase of the *Arabidopsis thaliana* - *Plasmodiophora brassicae* interaction. *Funct. Plant Biol.* 38 (6), 462–478. doi: 10.1071/Fp11026
- Aist, J. R., and Williams, P. H. (1971). The cytology and kinetics of Cabbage root hair penetration by *Plasmodiophora brassicae*. *Can. J. Bot.* 49 (11), 2023–2034. doi: 10.1139/b71-284
- Ando, S., Yamada, T., Asano, T., Kamachi, S., Tsushima, S., Hagio, T., et al. (2006). Molecular cloning of *PbSTKL1* gene from *Plasmodiophora brassicae* expressed during clubroot development. *J. Phytopathol.* 154 (3), 185–189. doi: 10.1111/j.1439-0434.2006.01078.x
- Bak, S., and Feyereisen, R. (2001). The involvement of two P450 enzymes, CYP83B1 and CYP83A1, in auxin homeostasis and glucosinolate biosynthesis. *Plant Physiol.* 127, 108–118. doi: 10.1104/pp.127.1.108
- Bent, A. F., and Mackey, D. (2007). Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. *Annu. Rev. Phytopathol.* 45, 399–436. doi: 10.1146/annurev.phyto.45.062806.094427
- Birkenbihl, R. P., Diezel, C., and Somssich, I. E. (2012). Arabidopsis *WRKY33* is a key transcriptional regulator of hormonal and metabolic responses toward *Botrytis cinerea* infection. *Plant Physiol.* 159 (1), 266–285. doi: 10.1104/pp.111.192641
- Bulman, S., Siemens, J., Ridgway, H. J., Eady, C., and Conner, A. J. (2006). Identification of genes from the obligate intracellular plant pathogen, *Plasmodiophora brassicae*. *FEMS Microbiol. Lett.* 264 (2), 198–204. doi: 10.1111/j.1574-6968.2006.00466.x
- Cao, T., Srivastava, S., Rahman, M. H., Kav, N. N. V., Hotte, N., Deyholos, M. K., et al. (2008). Proteome-level changes in the roots of *Brassica napus* as a result of *Plasmodiophora brassicae* infection. *Plant Sci.* 174 (1), 97–115. doi: 10.1016/j.plantsci.2007.10.002
- Chalhoub, B., Denoeud, F., Liu, S. Y., Parkin, I. A. P., Tang, H. B., Wang, X. Y., et al. (2014). Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* 345 (6166), 950–953. doi: 10.1126/science.1253435
- Chen, J., Pang, W., Chen, B., Zhang, C., and Piao, Z. (2015). Transcriptome analysis of *Brassica rapa* near-isogenic lines carrying clubroot-resistant and -susceptible alleles in response to *Plasmodiophora brassicae* during early infection. *Front. Plant Sci.* 6, 1–14. doi: 10.3389/fpls.2015.01183
- Chezem, W. R., Memon, A., Li, F. S., Weng, J. K., and Clay, N. K. (2017). SG2-type R2R3-MYB transcription factor *MYB15* controls defense-induced lignification and basal immunity in Arabidopsis. *Plant Cell* 29 (8), 1907–1926. doi: 10.1105/tpc.16.00954
- Chu, M., Song, T., Falk, K. C., Zhang, X., Liu, X., Chang, A., et al. (2014). Fine mapping of *Rcr1* and analyses of its effect on transcriptome patterns during infection by *Plasmodiophora brassicae*. *BMC Genomics* 15, 1166. doi: 10.1186/1471-2164-15-1166
- Cury-Boaventury, M. F., and Curi, R. (2005). Regulation of reactive oxygen species (ROS) production by C18 fatty acids in Jurkat and Raji cells. *Clin. Sci.* 108, 245–253. doi: 10.1042/cs20040281
- Devos, S., Laukens, K., Deckers, P., Dominique, V. D. S., Beeckman, T., Inze, D., et al. (2006). A hormone and proteome approach to picturing the initial metabolic events during *Plasmodiophora brassicae* infection on Arabidopsis. *Mol. Plant Microbe Interact.* 19 (12), 1431–1443. doi: 10.1094/MPMI-19-1431
- Dixon, G. R. (2009). The occurrence and economic impact of *Plasmodiophora brassicae* and clubroot disease. *J. Plant Growth Regul.* 28 (3), 194–202. doi: 10.1007/s00344-009-9090-y
- Dobson, R. L., and Gabrielson, R. L. (1983). Role of primary and secondary zoospores of *Plasmodiophora brassicae* in the development of clubroot in Chinese Cabbage. *Phytopathology* 73, 559–561. doi: 10.1094/Phyto-73-559
- Feng, J., Hwang, R., Hwang, S. F., Strelkov, S. E., Gossen, B. D., Zhou, Q. X., et al. (2010). Molecular characterization of a serine protease *Pro1* from *Plasmodiophora brassicae* that stimulates resting spore germination. *Mol. Plant Pathol.* 11 (4), 503–512. doi: 10.1111/j.1364-3703.2010.00623.x
- Gigolashvili, T., Berger, B., and Flügge, U. I. (2009). Specific and coordinated control of indolic and aliphatic glucosinolate biosynthesis by R2R3-MYB transcription factors in *Arabidopsis thaliana*. *Phytochem. Rev.* 8 (1), 3–13. doi: 10.1007/s11101-008-9112-6
- Grubb, C. D., and Abel, S. (2006). Glucosinolate metabolism and its control. *Trends Plant Sci.* 11 (2), 89–100. doi: 10.1016/j.tplants.2005.12.006
- Hansen, C. H., Wittstock, U., Olsen, C. E., Hick, A. J., Pickett, J. A., and Halkier, B. A. (2001). Cytochrome p450 CYP79F1 from Arabidopsis catalyzes the conversion of dihomomethionine and trihomomethionine to the corresponding aldoximes in the biosynthesis of aliphatic glucosinolates. *J. Biol. Chem.* 276 (14), 11078. doi: 10.1074/jbc.M010123200
- Hao, J., Wei, X. C., Yang, Y., Yuan, Y. X., Wei, F., Zhao, Y. Y., et al. (2017). Root RNA-seq analysis reveals a distinct transcriptome landscape between clubroot-susceptible and clubroot-resistant Chinese cabbage lines after *Plasmodiophora brassicae* infection. *Plant Soil* 421, 93–105. doi: 10.1007/s11104-017-3432-5
- Hatakeyama, K., Suwabe, K., Tomita, R. N., Kato, T., Nunome, T., Fukuoka, H., et al. (2013). Identification and characterization of *Crr1a*, a gene for resistance to clubroot disease (*Plasmodiophora brassicae* Woronin) in *Brassica rapa* L. *PLoS One* 8 (1), e54745. doi: 10.1371/journal.pone.0054745
- Hatakeyama, K., Niwa, T., Kato, T., Ohara, T., Kakizaki, T., and Matsumoto, S. (2017). The tandem repeated organization of NB-LRR genes in the clubroot-resistant *CRB* locus in *Brassica rapa* L. *Mol. Genet. Genomics* 292 (2), 397–405. doi: 10.1007/s00438-016-1281-1
- Hirai, M. Y. (2008). A robust omics-based approach for the identification of glucosinolate biosynthetic genes. *Phytochem. Rev.* 8 (1), 15–23. doi: 10.1007/s11101-008-9114-4
- Honig, F. (1931). Der Kohlkropferreger. (*Plasmodiophora brassicae* Wor.) Eine Monographie. *Die Gartenbauwissenschaft* 5, 116–225.
- Howard, R., Strelkov, S., and Harding, M. (2010). Clubroot of cruciferous crops – a new perspective on an old disease. *Can. J. Plant Pathol.* 32 (1), 43–57. doi: 10.1080/07060661003621761
- Hull, A. K., Vij, R., and Celenza, J. L. (2000). Arabidopsis cytochrome p450s that catalyze the first step of tryptophan-dependent indole-3-acetic acid biosynthesis. *Curr. Opin. Plant Biol.* 3 (3), 170–171. doi: 10.1016/s1369-5266(00)80021-7
- Hwang, S. F., Strelkov, S. E., Feng, J., Gossen, B. D., and Howard, R. J. (2012). *Plasmodiophora brassicae*: a review of an emerging pathogen of the Canadian canola (*Brassica napus*) crop. *Mol. Plant Pathol.* 13 (2), 105–113. doi: 10.1111/j.1364-3703.2011.00729.x

ACKNOWLEDGMENTS

The authors are very grateful to Dr. Guangqin Cai (Huazhong Agricultural University) for providing helpful on sampling preparation, data analysis, and article revision.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01275/full#supplementary-material>

- Irani, S., Trost, B., Waldner, M., Nayidu, N., Tu, J., Kusalik, A. J., et al. (2018). Transcriptome analysis of response to *Plasmodiophora brassicae* infection in the Arabidopsis shoot and root. *BMC Genomics* 19 (1), 23. doi: 10.1186/s12864-017-4426-7
- Irani, S., Todd, C. D., Wei, Y., and Bonham-Smith, P. C. (2019). Changes in phenylpropanoid pathway gene expression in roots and leaves of susceptible and resistant *Brassica napus* lines in response to *Plasmodiophora brassicae* inoculation. *Physiol. Mol. Plant Pathol.* 106, 196–203. doi: 10.1016/j.pmp.2019.02.007
- Ji, R., Wang, Y., Wang, X., Liu, Y., Shen, X., and Feng, H. (2018). Proteomic analysis of the interaction between *Plasmodiophora brassicae* and Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*) at the initial infection stage. *Scientia Hort.* 233, 386–393. doi: 10.1016/j.scienta.2018.02.006
- Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444 (7117), 323–329. doi: 10.1038/nature05286
- Jubault, M., Lariagon, C., Simon, M., Delourme, R., and Manzaneres-Dauleux, M. J. (2008). Identification of quantitative trait loci controlling partial clubroot resistance in new mapping populations of *Arabidopsis thaliana*. *Theor. Appl. Genet.* 117 (2), 191–202. doi: 10.1007/s00122-008-0765-8
- Kageyama, K., and Asano, T. (2009). Life Cycle of *Plasmodiophora brassicae*. *J. Plant Growth Regul.* 28 (3), 203. doi: 10.1007/s00344-009-9101-z
- Karling, J. S. (1968). The plasmodiophorales: including a complete host index, bibliography, and a description of diseases caused by species of this order. *Mol. Biol. Evol.* 22 (3), 582–588. doi: 10.1093/molbev/msi053
- Kuginuki, Y., Yoshikawa, H., and Hirai, M. (1999). Variation in virulence of *Plasmodiophora brassicae* in Japan tested with clubroot-resistant cultivars of Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Eur. J. Plant Pathol.* 105 (4), 327–332. doi: 10.1023/A:1008705413127
- Kutchan, T. M. (1995). Alkaloid biosynthesis: the basis for metabolic engineering of medicinal plants. *Plant Cell* 7 (7), 1059–1070. doi: 10.2307/3870057
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9 (1), 559. doi: 10.1186/1471-2105-9-559
- Lee, J., Izzah, N. K., Choi, B. S., Joh, H. J., Lee, S. C., Perumal, S., et al. (2016). Genotyping-by-sequencing map permits identification of clubroot resistance QTLs and revision of the reference genome assembly in cabbage (*Brassica oleracea* L.). *DNA Res.* 23, 29–41. doi: 10.1093/dnares/dsv034
- Li, F., Chen, B., Xu, K., Wu, J., Song, W., Bancroft, I., et al. (2014). Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). *DNA Res.* 21 (4), 355–367. doi: 10.1093/dnares/dsu002
- Li, L., Luo, Y., Chen, B., Xu, K., Zhang, F., Li, H., et al. (2016). A genome-wide association study reveals new loci for resistance to clubroot disease in *Brassica napus*. *Front. Plant Sci.* 7, 1483. doi: 10.3389/fpls.2016.01483
- Liu, J., Liu, X., Dai, L., Wang, G., Laboratory, R. G., University, H. A., et al. (2007). Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants. *J. Genet. Genomics* 34 (9), 765–776. doi: 10.1016/s1673-8527(07)60087-3
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C (T)) Method. *Methods* 25 (4), 402–408. doi: 10.1006/meth.20011262
- Lorenzo, O., Chico, J. M., Sanchez-Serrano, J. J., and Solano, R. (2004). JASMONATE-INSENSITIVE1 encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in Arabidopsis. *Plant Cell* 16 (7), 1938–1950. doi: 10.1105/tpc.022319
- Ludwig-Muller, J., Schubert, B., Pieper, K., Ihmig, S., and Hilgenberg, W. (1997). Glucosinolate content in susceptible and resistant Chinese cabbage varieties during development of clubroot disease. *Phytochemistry* 44 (3), 407–414. doi: 10.1016/s0031-9422(96)00498-0
- Ludwig-Muller, J., Bennett, R. N., Kiddle, G., Ihmig, S., Ruppel, M., and Hilgenberg, W. (1999a). The host range of *Plasmodiophora brassicae* and its relationship to endogenous glucosinolate content. *New Phytol.* 141, 443–458. doi: 10.1046/j.1469-8137.1999.00368.x
- Ludwig-Muller, J., Pieper, K., Ruppel, M., Cohen, J. D., Epstein, E., Kiddle, G., et al. (1999b). Indole glucosinolate and auxin biosynthesis in *Arabidopsis thaliana* (L.) Heynh. glucosinolate mutants and the development of clubroot disease. *Planta* 208 (3), 409–419. doi: 10.1007/s004250050576
- Ludwig-Muller, J., Prinsen, E., Rolfe, S. A., and Scholes, J. D. (2009). Metabolism and plant hormone action during clubroot disease. *J. Plant Growth Regul.* 28 (3), 229–244. doi: 10.1007/s00344-009-9089-4
- Ludwig-Muller, J. (2009). Glucosinolates and the clubroot disease: defense compounds or auxin precursors? *Phytochem. Rev.* 8 (1), 135–148. doi: 10.1007/s11101-008-9096-2
- Manzaneres-Dauleux, M. J., Divaret, I., Baron, F., and Thomas, G. (2000). Evaluation of French *Brassica oleracea* landraces for resistance to *Plasmodiophora brassicae*. *Euphytica* 113, 211–218. doi: 10.1023/A:1003997421340
- Mao, X., Cai, T., Olyarchuk, J. G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21 (19), 3787–3793. doi: 10.1093/bioinformatics/bti430
- McDonald, M. R., Sharma, K., Gossen, B. D., Deora, A., Feng, J., and Hwang, S. F. (2014). The role of primary and secondary infection in host response to *Plasmodiophora brassicae*. *Phytopathology* 104 (10), 1078–1087. doi: 10.1094/PHYTO-07-13-0189-R
- Miao, Y., Xu, L., He, X., Zhang, L., Shaban, M., Zhang, X., et al. (2019). Suppression of tryptophan synthase activates cotton immunity by triggering cell death via promoting SA synthesis. *Plant J.* 98 (2), 329–345. doi: 10.1111/tpj.14222
- Mikkelsen, M. D., Petersen, B. L., Olsen, C. E., and Halkier, B. A. (2002). Biosynthesis and metabolic engineering of glucosinolates. *Amino Acids* 22 (3), 279–295. doi: 10.1007/s007260200014
- Pandey, S. P., Roccaro, M., Schon, M., Logemann, E., and Somssich, I. E. (2010). Transcriptional reprogramming regulated by *WRKY18* and *WRKY40* facilitates powdery mildew infection of Arabidopsis. *Plant J.* 64 (6), 912–923. doi: 10.1111/j.1365-3113X.2010.04387.x
- Peng, Y., Gossen, B. D., Huang, Y., Al-Daoud, F., and McDonald, M. R. (2019). Development of *Plasmodiophora brassicae* in the root cortex of cabbage over time. *Eur. J. Plant Pathol.* 154, 727–737. doi: 10.1007/s10658-019-01696-0
- Piao, Z., Ramchiary, N., and Lim, Y. P. (2009). Genetics of clubroot resistance in Brassica species. *J. Plant Growth Regul.* 28, 252–264. doi: 10.1007/s00344-009-9093-8
- Prerostova, S., Dobrev, P., Konradova, V., Knirsch, V., Gaudinova, A., Kramna, B., et al. (2018). Hormonal responses to *Plasmodiophora brassicae* infection in *Brassica napus* cultivars differing in their pathogen resistance. *Int. J. Mol. Sci.* 19 (12), 4024. doi: 10.3390/ijms19124024
- Ren, L., Jia, J., Li, M., Liu, F., Cheng, Y., Zhang, M., et al. (2012). Distribution of rapeseed clubroot disease in Hubei Province and evaluation of yield loss. *Agric. Sci. Technol.* 13, 775–777.
- Rolfe, S. A., Strelkov, S. E., Links, M. G., Clarke, W. E., Robinson, S. J., Djavaheri, M., et al. (2016). The compact genome of the plant pathogen *Plasmodiophora brassicae* is adapted to intracellular interactions with host *Brassica* spp. *BMC Genomics* 17, 272. doi: 10.1186/s12864-016-2597-2
- Schuller, A., Kehr, J., and Ludwig-Muller, J. (2014). Laser microdissection coupled to transcriptional profiling of Arabidopsis roots inoculated by *Plasmodiophora brassicae* indicates a role for brassinosteroids in clubroot formation. *Plant Cell Physiol.* 55 (2), 392–411. doi: 10.1093/pcp/pct174
- Schwelm, A., Fogelqvist, J., Knaust, A., Julke, S., Lilja, T., Bonilla-Rosso, G., et al. (2015). The *Plasmodiophora brassicae* genome reveals insights in its life cycle and ancestry of chitin synthases. *Sci. Rep.* 5, 11153. doi: 10.1038/srep11153
- Siemens, J., Keller, I., Sarx, J., Kunz, S., Schuller, A., Nagel, W., et al. (2006). Transcriptome analysis of Arabidopsis clubroots indicate a key role for cytokinins in disease development. *Mol. Plant Microbe Interact.* 19 (5), 480–494. doi: 10.1094/MPMI-19-0480
- Su, T., Yu, S., Wang, W., Li, P., Zhang, F., Yu, Y., et al. (2018). iTRAQ analysis of protein profile during the secondary stage of infection of *Plasmodiophora brassicae* in Chinese cabbage (*Brassica rapa* subsp. *pekinensis*). *J. Plant Pathol.* 100, 533–542. doi: 10.1007/s42161-018-0121-z
- Tewari, J. P., Strelkov, S. E., Orchard, D., Hartman, M., Lange, R. M., and Turkington, T. K. (2005). Identification of clubroot of crucifers on canola (*Brassica napus*) in Alberta. *Can. J. Plant Pathol.* 27, 143–144. doi: 10.1080/07060660509507206
- Tierens, K. F., Thomma, B. P., Brouwer, M., Schmidt, J., Kistner, K., Porzel, A., et al. (2001). Study of the role of antimicrobial glucosinolate-derived isothiocyanates in resistance of Arabidopsis to microbial pathogens. *Plant Physiol.* 125 (4), 1688–1699. doi: 10.2307/4279801

- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25 (9), 1105–1111. doi: 10.1093/bioinformatics/btp120
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28 (5), 511–515. doi: 10.1038/nbt1621
- Ueno, H., Matsumoto, E., Aruga, D., Kitagawa, S., Matsumura, H., and Hayashida, N. (2012). Molecular characterization of the *CRA* gene conferring clubroot resistance in *Brassica rapa*. *Plant Mol. Biol.* 80 (6), 621–629. doi: 10.1007/s11103-012-9971-5
- Verma, S. S., Rahman, M. H., Deyholos, M. K., Basu, U., and Kav, N. N. (2014). Differential expression of miRNAs in *Brassica napus* root following infection with *Plasmodiophora brassicae*. *PloS One* 9 (1), e86648. doi: 10.1371/journal.pone.0086648
- Werner, S., Diederichsen, E., Frauen, M., Schondelmaier, J., and Jung, C. (2008). Genetic mapping of clubroot resistance genes in oilseed rape. *Theor. Appl. Genet.* 116 (3), 363–372. doi: 10.1007/s00122-007-0674-2
- Xu, L., Ren, L., Chen, K. R., Liu, F., and Fang, X. P. (2016). Putative role of IAA during the early response of *Brassica napus* L. @ to *Plasmodiophora brassicae*. *Eur. J. Plant Pathol.* 145 (3), 601–613. doi: 10.1007/s10658-016-0877-y
- Yan, X., and Chen, S. (2007). Regulation of plant glucosinolate metabolism. *Planta* 226 (6), 1343–1352. doi: 10.2307/23389821
- Zhang, H., Feng, J., Hwang, S. F., Strelkov, S. E., and Sun, R. (2015). Mapping of clubroot (*Plasmodiophora brassicae*) resistance in *Brassica napus* canola. *Plant Pathol.* 65 (3), 435–440. doi: 101111
- Zhang, X., Liu, Y., Fang, Z., Li, Z., Yang, L., Zhuang, M., et al. (2016). Comparative transcriptome analysis between Broccoli (*Brassica oleracea* var. italica) and Wild Cabbage (*Brassica macrocarpa* Guss.) in response to *Plasmodiophora brassicae* during different infection stages. *Front. Plant Sci.* 7, 1929. doi: 10.3389/fpls.2016.01929
- Zhao, Y., Bi, K., Gao, Z., Chen, T., Liu, H., Xie, J., et al. (2017). Transcriptome analysis of *Arabidopsis thaliana* in response to *Plasmodiophora brassicae* during early infection. *Front. Microbiol.* 8, 673. doi: 10.3389/fmicb.2017.00673

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, Long, Li 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.



Characterization of the Powdery Mildew Resistance Gene in the Elite Wheat Cultivar Jimai 23 and Its Application in Marker-Assisted Selection

OPEN ACCESS

Edited by:

Zhu-Qing Shao,
Nanjing University, China

Reviewed by:

Christina Cowger,
Plant Science Research Unit
(USDA-ARS), United States
Hong Zhang,
Northwest A&F University, China

*Correspondence:

Huagang He
hghe@mail.ujc.edu.cn
Pengtao Ma
ptma@ytu.edu.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: 15 December 2019

Accepted: 28 February 2020

Published: 02 April 2020

Citation:

Jia M, Xu H, Liu C, Mao R, Li H,
Liu J, Du W, Wang W, Zhang X,
Han R, Wang X, Wu L, Liang X,
Song J, He H and Ma P (2020)
Characterization of the Powdery
Mildew Resistance Gene in the Elite
Wheat Cultivar Jimai 23 and Its
Application in Marker-Assisted
Selection. *Front. Genet.* 11:241.
doi: 10.3389/fgene.2020.00241

Mengshu Jia^{1†}, Hongxing Xu^{2†}, Cheng Liu^{3†}, Ruixi Mao⁴, Haosheng Li³, Jianjun Liu³,
Wenxiao Du¹, Wenrui Wang¹, Xu Zhang¹, Ran Han³, Xiaolu Wang³, Liru Wu¹, Xiao Liang¹,
Jiancheng Song¹, Huagang He^{5*} and Pengtao Ma^{1*}

¹ School of Life Sciences, Yantai University, Yantai, China, ² State Key Laboratory of Crop Stress Adaptation
and Improvement, School of Life Sciences, Henan University, Kaifeng, China, ³ Crop Research Institute, Shandong Academy
of Agricultural Sciences, Jinan, China, ⁴ Shandong Seed Administration Station, Jinan, China, ⁵ School of Food
and Biological Engineering, Jiangsu University, Zhenjiang, China

Powdery mildew infection of wheat (*Triticum aestivum* L.), caused by *Blumeria graminis* f. sp. *tritici* (Bgt), is a destructive disease that threatens yield and quality worldwide. The most effective and preferred means for the control of the disease is to identify broad-spectrum resistance genes for breeding, especially the genes derived from elite cultivars that exhibit desirable agronomic traits. Jimai 23 is a Chinese wheat cultivar with superior agronomic performance, high-quality characteristics, and effective resistance to powdery mildew at all growth stages. Genetic analysis indicated that powdery mildew resistance in Jimai 23 was mediated by a single dominant gene, tentatively designated *PmJM23*. Using bulked segregant RNA-Seq (BSR-Seq), a series of markers was developed and used to map *PmJM23*. *PmJM23* was then located at the *Pm2* locus on the short arm of chromosome 5D (5DS). Resistance spectrum analysis demonstrated that *PmJM23* provided a broad resistance spectrum different from that of the documented *Pm2* alleles, indicating that *PmJM23* is most likely a new allele of *Pm2*. In view of these combined agronomic, quality, and resistance findings, *PmJM23* is expected to be a valuable resistance gene in wheat breeding. To efficiently use *PmJM23* in breeding, the closely linked markers of *PmJM23* were evaluated and confirmed to be applicable for marker-assisted selection (MAS). Using these markers, a series of resistant breeding lines with high resistance and desirable agronomic performance was selected from the crosses involving *PmJM23*, resulting in improved powdery mildew resistance of these lines.

Keywords: wheat powdery mildew, *PmJM23*, BSR-Seq, marker-assisted selection, agronomic trait

INTRODUCTION

Common wheat (*Triticum aestivum* L.) is one of the three major grain crops worldwide, and its high and stable yield plays an important role in food security. However, various diseases, including powdery mildew, rusts, and fusarium head blight, can have devastating impacts on yield (Huang and Pang, 2017; Ingvarsdén et al., 2019; Li et al., 2019b). Powdery mildew caused by *Blumeria graminis* f. sp. *tritici* (Bgt) is one of the most damaging diseases, typically decreasing wheat yield by 10–15% and up to 50% in severe cases (Morgounov et al., 2012; Xu et al., 2015). In China alone, the area of winter wheat affected annually by powdery mildew has exceeded 6 m ha in recent decades, causing 300,000 tons of crop loss each year¹.

Although the pesticides are commonly used for powdery mildew control, resistance to drugs and environmental pollution have become increasingly prominent (de Waard et al., 1986; Felsenstein et al., 2010). Improved host resistance provides an attractive opportunity for the development of an effective and environmentally acceptable means to control this disease (Ma et al., 2015a, 2018, 2019). In wheat production, the powdery mildew resistance (*Pm*) genes exhibit mainly race-specific resistance, which was often short lived, as they were defeated by the fast-evolving virulent pathogen (Xiao et al., 2013; El-Shamy et al., 2016). The ratio of broad-spectrum resistance in Chinese wheat cultivars/breeding lines is not yet satisfactory in wheat production (Li et al., 2011). Therefore, there is an urgent need to mine and utilize more effective resistance sources to increase the genetic diversity of *Pm* genes.

To date, more than 70 *Pm* genes/alleles (*Pm1*–*Pm65*, *Pm8* is allelic to *Pm17*, *Pm18* = *Pm1c*, *Pm22* = *Pm1e*, *Pm23* = *Pm4c*, *Pm31* = *Pm21*) have been identified in 60 loci from common wheat and its relatives (Li et al., 2019a; McIntosh et al., 2019). However, not all the genes can be directly used in resistance breeding. Many *Pm* genes have adverse pleiotropism, linkage drag, or competition lag due to their genetic characteristics. For example, the gene *Pm16* has broad-spectrum resistance to wheat powdery mildew, but the linkage drag associated with *Pm16* leads to a 15% yield loss (Summers and Brown, 2013). The gene *Pm8* derived from the 1RS chromosome of rye made a significant contribution to the control of wheat powdery mildew in 1990s, but the linked secalin glycopeptide in 1RS resulted in a decline in flour quality (Friebe et al., 1989; Lee et al., 1995). Additionally, the *Pm* genes derived from landraces usually have poor agronomic performance and need multigeneration backcrossing before acceptance by breeders (Xu et al., 2015; Li et al., 2020). Clearly, the value to breeding of a specific *Pm* gene depends not only on its effectiveness for disease control but also on the agronomic performance of its donor (Zhao et al., 2013; Ma et al., 2018). The discovery of new genes or new allelic variations from elite wheat cultivars offers an attractive prospect for the rapid genetic improvement of resistance.

While conventional wheat breeding has been remarkably successful in many respects, it is usually subjective, inefficient, and unable to achieve stable improvement (Gupta et al., 2010).

Molecular breeding programs worldwide can successfully provide a valuable complement to conventional breeding (Kuchel et al., 2007; Gupta et al., 2010). With the development of molecular markers, marker-assisted selection (MAS) can facilitate the exclusion of adverse genes in fewer generations and accelerate breeding progress (Jiang et al., 2017). Techniques including high-throughput single-nucleotide polymorphism (SNP) arrays, specific-locus amplified fragment sequencing (SLAF-Seq), and bulked segregant RNA-Seq (BSR-Seq) especially have been widely used for genetic mapping, thereby accelerating the cloning and utilization of superior wheat genes (Liu et al., 2018; Wu et al., 2018; Shi et al., 2019; Tan et al., 2019). Furthermore, advances continue to be made in whole genome sequencing of wheat. Common wheat (AABBDD) and its diploid (AA and DD genomes) and tetraploid (AABB) ancestors all have relatively perfect genome sequences (Avni et al., 2017; Luo et al., 2017; The International Wheat Genome Sequencing Consortium, 2018; Ling et al., 2018), which will greatly facilitate the effectiveness of SNP chip, BSR, and SLAF data in cloning of resistance genes and their utilization in breeding.

Jimai 23 is an elite wheat cultivar developed by the Shandong Academy of Agricultural Sciences (China) derived from Jimai 22, which is the most widely grown wheat cultivar in China during the last decade. Previous reports using a single Bgt isolate demonstrated that the powdery mildew resistance in Jimai 22 is controlled by a single dominant gene, *Pm52* (Yin et al., 2009; Qu et al., 2019). In recent years, Jimai 23 also showed highly effective resistance to powdery mildew, indicating that it is an attractive source for controlling wheat powdery mildew. To clarify the relationship between the *Pm* genes in Jimai 23 and 22 and to better use the powdery mildew resistance in Jimai 23, we report, in this work, the identification and dissection of the *Pm* gene(s) in Jimai 23, development of molecular markers for, and breeding with, the *Pm* gene(s) in Jimai 23.

MATERIALS AND METHODS

Plant Materials

The wheat cultivar Jimai 23 was bred from the cross of Jimai 22 and Yumai 34 by the Crop Research Institute, Shandong Academy of Agricultural Sciences, and used as the donor of resistant gene(s) against powdery mildew in this research. The wheat cultivar Tainong 18 was used as a susceptible parent and crossed with Jimai 23 to obtain an F₂ population and F_{2:3} families for genetic analysis of powdery mildew resistance in Jimai 23. Wheat cultivars Huixianhong and Mingxian 169, which were susceptible to all the Bgt isolates tested, were used as the susceptible controls for phenotypic assessment. Eight resistant donors with documented *Pm* genes (**Supplementary Table S1**) were used to compare their phenotypic responses to different Bgt isolates with those of Jimai 23 and to evaluate the breeding value of *Pm* gene(s) in Jimai 23 through resistance spectrum analysis. Jimai 22, one of the parents of Jimai 23, is a super-high yield and medium gluten wheat cultivar showing good resistance to wheat powdery mildew and stripe rust (Yin et al., 2009; Chen et al., 2016; Qu et al., 2019). In this study, Jimai 22 was also used

¹<http://cb.natesc.gov.cn>

as the control for evaluation of comprehensive traits of Jimai 23. Liangxing 99 with documented *Pm52* (Zhao et al., 2013) was used for comparing the relationship of its *Bgt* resistance gene with that in Jimai 22. Twenty-six susceptible wheat cultivars from different ecological regions of China were used to evaluate the availability of closely linked markers for MAS, five of which (Gaoyou 5766, SH4300, Tainong 2419, 125574, and Daimai 1503) had been crossed with Jimai 23 for MAS (**Supplementary Table S2**).

Evaluation of Comprehensive Traits

From 2012 to 2016, Jimai 23 and 22 were planted in the field at the Crop Research Institute, Shandong Academy of Agricultural Sciences (Jinan, China), for a comprehensive evaluation of their traits. Sowing and assessment were based on the methods of Xu et al. (2014). Seeds of Jimai 23 and 22 were sown in six rows (3.0 m in length and inter-row distance of 20 cm) with two rows of susceptible controls as guard rows on each side of the plot. Spike numbers per mu (SNM) (1 ha = 15 mu), kernel numbers per spike (KNS), thousand kernel weight (TKW), yield per mu (YM) (1 ha = 15 mu), bulk density (BD), gluten index (GI), sedimentation value (SV), and stable time of dough (STD) were analyzed to evaluate comprehensive traits of Jimai 23 using Jimai 22 as control. The method for assessing agronomic and yield traits, such as SNM, KNS, YM, and BD, are described in Teich (1984) and Xu et al. (2014), and grain quality traits, such as GI, SV, and STD, are described in Studnicki et al. (2018) and Kuchel et al. (2006). In each year, three replicates were sampled using the same procedure to confirm the phenotypic data. Analysis of variance (ANOVA) of each trait was performed using SPSS 16.0 software (SPSS Inc., Chicago, IL, United States), with a significance level of $p \leq 0.05$.

Phenotypic Evaluation of Reactions to Different *Bgt* Isolates

From 2017 to 2019, Jimai 23 and eight resistant donors with documented *Pm* genes (**Supplementary Table S1**) were planted in the greenhouse at Yantai University (Yantai, China) for disease assessment at the adult stage. They were planted in a plot with 30 rows (1.2 m in length and inter-row distance of 20 cm). Thirty seeds per row and four rows per cultivar/line were sown. Huixianhong and Mingxian 169 were used as a susceptible control and border plants and were sown in every 10th row and around the plot. In spring, the seedlings of Huixianhong and Mingxian 169 were inoculated with a mixture of the 20 *Bgt* isolates collected from major wheat production regions of China. At full heading and milk stages, infection types (ITs) were rated using a 0–9 scale, where 0–4 was resistant, and 5–9 susceptible (Ma et al., 2018).

At the seedling stage, Jimai 23 and eight resistant donors with documented *Pm* genes (**Supplementary Table S1**) were tested for their reaction patterns to 42 *Bgt* isolates with different avirulence/virulence patterns. They were collected from major wheat production regions of China and isolated into single spore for virulence evaluation by Prof. Hongxing Xu (**Supplementary Table S1**). Each isolate was put in an independent transparent glass tube with layer of gauze to prevent cross-contamination

among isolates. Five seeds of each genotype were sown in 128-cell rectangular trays in a growth chamber. The susceptible controls Huixianhong and Mingxian 169 were planted randomly in each tray. At the one leaf stage, the seedlings were inoculated with fresh conidiospores multiplied on Huixianhong seedlings, which were raised earlier and inoculated to provide a source of conidiophores for experimental inoculation. Then, the inoculated seedlings were incubated in a dark and independent chamber with high humidity at 18°C for 24 h. The trays were then placed in a climate incubator, set at a daily cycle of 14 h light at 22°C and 10 h of darkness at 18°C. ITs were surveyed when the spores were fully developed on the susceptible controls after about 10–14 days of inoculation using the 0–4 scale described by An et al. (2013), in which ITs 0, 0, 1, and 2 were regarded as resistant and ITs 3 and 4 as susceptible. Three repeats were tested using the same procedure.

To determine the inheritance of powdery mildew resistance in Jimai 23 and 22, all the *Bgt* isolates apart from those virulent to both Jimai 23 and 22 (**Supplementary Table S1**) were used to inoculate one-leaf seedlings of Jimai 23, Jimai 22, Tainong 18, and the F_1 , F_2 , and $F_{2:3}$ progenies of Jimai 23 \times Tainong 18 and Jimai 22 \times Tainong 18 for genetic analysis. For the disease assessment of parents and F_1 hybrids, 10 seeds were sown for inoculation with these isolates. For each $F_{2:3}$ family, 30 plants were tested against these isolates. Goodness of fit was analyzed using a Chi-squared (χ^2) test to assess deviations of the observed phenotypic data from theoretically expected segregation ratios using SPSS 16.0 software (SPSS Inc., Chicago, IL, United States) with a p -value level of 0.05.

Preliminary Confirmation of *Pm* Gene(s) in Jimai 23

After genetic analysis, $F_{2:3}$ families that were consistent with the ratio of monogenic segregation were selected for genotyping. Total genomic DNAs (gDNAs) of Jimai 23, Jimai 22, Tainong18, and the $F_{2:3}$ families selected above were isolated from leaves after phenotypic evaluation using the TE-boiling method (He et al., 2017). Resistant and susceptible DNA bulks were constructed. For each, equal amounts of gDNA from either 10 homozygous-resistant or 10 homozygous-susceptible $F_{2:3}$ families were pooled. Then, 50 simple sequence repeat (SSR) markers linked to documented *Pm* genes/alleles (Ma et al., 2016) were tested for polymorphisms between the parents and bulks. The polymorphic markers were then used to genotype the corresponding $F_{2:3}$ families.

Development of New Markers Using BSR-Seq

Total messenger RNA (mRNA) of Jimai 23, Tainong 18, and their $F_{2:3}$ families inoculated with *Bgt* isolate YT01 (avirulent to Jimai 23) were extracted using the mirVana miRNA Isolation Kit (Ambion, Thermo Fisher Scientific Inc., Waltham, MA, United States) following the manufacturer's protocol. The RNA integrity was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States), and the samples with RNA integrity number

(RIN) ≥ 7 were subjected to the subsequent construction of complementary DNA (cDNA) libraries. The cDNA libraries were constructed using TruSeq Stranded mRNA LTSample Prep Kit (Illumina, San Diego, CA, United States) according to the manufacturer's instructions. After quality control of the cDNA libraries using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States), the cDNA libraries were sequenced on the Illumina HiSeq sequencing platform (HiSeqTM 2500) by Biomarker Technologies Corporation (Beijing, China). After sequence assembly of the clean data with the reference genome of Chinese Spring (v1.0), SNP and small InDels in the targeted interval were obtained and used for marker development in BMK Cloud (developed by Biomarker Technologies Corporation).

Genotyping of the Mapping Population and Map Construction

The developed markers were tested for polymorphisms between the parents and bulks. The resulting markers were genotyped on the F_{2:3} population of Jimai 23 \times Tainong 18. Chi-squared (χ^2) test was then used to assess deviations of the observed phenotypic data of the F_{2:3} families from theoretically expected segregation ratios for goodness of fit. The linkage map of the powdery mildew resistance gene(s) was constructed based on Lincoln et al. (1993) and Kosambi (1943) using the MAPMAKER 3.0 and the Kosambi function.

Allelism Test

After the *Pm* gene in Jimai 23 was assigned to the *Pm2* interval, Jimai 23 was crossed with Wennong 14 and Liangxing 66, with documented *Pm2* alleles, to obtain F₂ populations. The isolate YT01, that was avirulent to Jimai 23, Wennong 14, and Liangxing 66, was used to inoculate the F₂ populations using Mingxian 169 and Huixianhong as susceptible controls. After the spores were fully developed on the susceptible controls, the number of resistant and susceptible plants of the F₂ populations were counted to evaluate the allelic relationships between the *Pm* gene(s) in Jimai 23 and documented *Pm* genes on the same interval based on the ratio of resistant and susceptible F₂ plants.

Screening of Markers Available for MAS

To evaluate the utility of the markers of *Pm* gene(s) in Jimai 23 for MAS, the closely linked SSR markers were used to test the polymorphisms between Jimai 23 and 26 susceptible wheat cultivars from China (Supplementary Table S2). The markers that stably amplified polymorphic bands between Jimai 23 and the susceptible cultivars were regarded to be effective for MAS in these genetic backgrounds.

Jimai 23 was then crossed with some of the susceptible cultivars evaluated above to construct breeding populations. In the earlier generations, the resistant plants were selected for further self-pollination using the markers available for MAS. Meanwhile, the plants with poor agronomic performance in the field were eliminated. When homozygous-resistant plants were confirmed, they were mainly selected by agronomic performance in field. In the F₃ or F₄ generation, homozygous-resistant plants with suitable agronomic performance were planted in head rows. In the F₅ generation, the best head rows were carried on to plot sowing. Finally, the stability of the breeding lines with superior agronomic and yield performance was confirmed once more by the markers and infection experiments.

RESULTS

Comprehensive Traits of Jimai 23 in the Field

Compared with Jimai 22, Jimai 23 has comprehensively excellent traits. The GY of Jimai 23 was routinely higher than that of Jimai 22 although not significantly. More profoundly, the GI, SV, and STD of Jimai 23, which relate to flour quality, were significantly superior to those of Jimai 22, indicating that Jimai 23 has better processing quality than Jimai 22 (Table 1). This suggested that Jimai 23 has improved quality traits while still kept high yield potential of Jimai 22, making Jimai 23 an attractive wheat cultivar for both yield and quality.

Evaluation of Powdery Mildew Resistance in Jimai 23

In the past three consecutive growing seasons (2017–2019), Jimai 23 showed high resistance to *Bgt* at the adult stage in the

TABLE 1 | Agronomic, yield, and quality traits of wheat cultivars Jimai 23 and its parent Jimai 22.

Years	Cultivars	SNM ($\times 10^4$)	KNS	TKW (g)	YM (kg)	BD (g/l)	GI	SV (ml)	STD (min)
2012–2013	Jimai 23	49.2 \pm 0.9 ^a	29.4 \pm 0.4 ^a	44.2 \pm 1.2 ^a	543.3 \pm 9.8 ^a	783.5 \pm 20.3 ^a	76.3 \pm 2.3 ^a	93.0 \pm 2.1 ^a	8.0 \pm 0.3 ^a
	Jimai 22	44.2 \pm 1.1 ^b	33.2 \pm 1.1 ^b	38.1 \pm 1.4 ^b	523.2 \pm 15.1 ^a	753.4 \pm 22.3 ^b	42.1 \pm 1.8 ^b	80.0 \pm 1.3 ^b	2.6 \pm 0.1 ^b
2013–2014	Jimai 23	45.5 \pm 1.5 ^a	33.1 \pm 1.5 ^a	49.0 \pm 1.2 ^a	618.3 \pm 10.8 ^a	811.0 \pm 15.3 ^a	64.0 \pm 1.6 ^a	37.2 \pm 0.5 ^a	4.6 \pm 0.2 ^a
	Jimai 22	45.5 \pm 0.7 ^a	35.6 \pm 0.6 ^b	44.6 \pm 0.5 ^b	597.2 \pm 18.3 ^b	800.3 \pm 21.3 ^a	30.0 \pm 0.9 ^b	26.7 \pm 0.6 ^b	3.1 \pm 0.1 ^b
2014–2015	Jimai 23	46.7 \pm 2.1 ^a	32.9 \pm 1.2 ^a	46.9 \pm 1.9 ^a	599.2 \pm 10.3 ^a	815.7 \pm 12.2 ^a	63.0 \pm 1.3 ^a	36.2 \pm 0.9 ^a	4.9 \pm 0.2 ^a
	Jimai 22	44.8 \pm 2.3 ^a	35.8 \pm 1.3 ^b	41.6 \pm 1.5 ^b	564.3 \pm 7.8 ^b	795.5 \pm 10.3 ^a	17.1 \pm 0.5 ^b	26.6 \pm 0.2 ^b	4.2 \pm 0.0 ^b
2015–2016	Jimai 23	43.4 \pm 0.9 ^a	33.4 \pm 0.5 ^a	49.0 \pm 1.3 ^a	617.7 \pm 17.7 ^a	798.8 \pm 25.3 ^a	62.0 \pm 1.6 ^a	37.0 \pm 1.1 ^a	8.5 \pm 0.3 ^a
	Jimai 22	41.3 \pm 1.3 ^a	36.0 \pm 0.2 ^b	45.7 \pm 2.1 ^b	597.5 \pm 14.5 ^b	788.4 \pm 21.6 ^a	40.3 \pm 1.2 ^b	32.0 \pm 1.4 ^b	3.1 \pm 0.2 ^b

^{a,b}Values in the same column followed by the same letter were not significantly different based on the test of least significant difference (LSD) at ($p < 0.05$). SNM, spike numbers per mu (1 ha = 15 mu); KNS, kernel numbers per spike; TKW, thousand kernel weight; YM, yield per mu (1 ha = 15 mu); BD, bulk density; GI, gluten index; SV, sedimentation value; STD, stable time of dough.

field, with an IT rating of 0–2, while Huixianhong was highly susceptible with an IT rating of 8–9. Compared with other resistance donors (**Supplementary Table S1**), Jimai 23 showed more satisfactory disease resistance. At the seedling stage, Jimai 23 was resistant to 39 of 42 *Bgt* isolates (92.9%) with diverse virulence profiles (**Supplementary Table S1**) and has a broader resistance spectrum than most of the documented resistant germplasm. This suggests that Jimai 23 is an elite resource for resistance breeding.

Inheritance of Powdery Mildew Resistance in Jimai 23 and Its Genealogical Relationship With Its Parent Jimai 22

When inoculated with isolate YT01 that was avirulent to both Jimai 23 and Liangxing 99 (*Pm52*), the F_1 plants of Jimai 23 \times Tainong 18 and Jimai 22 \times Tainong 18 were all resistant with an IT rating of 0, indicating the dominant nature of the resistance gene(s). The segregation ratio of Jimai 23 \times Tainong 18 population was consistent with the ratio for monogenic segregation of a dominant gene, while the segregation ratio of Jimai 22 \times Tainong 18 population was consistent with the independent separation of Mendelian law for two dominant genes (**Table 2**). While using isolates YT20 that was avirulent on Jimai 23 and virulent to Liangxing 99 (*Pm52*), the segregation ratios of Jimai 23 \times Tainong 18 and Jimai 22 \times Tainong 18 populations were both consistent with the ratio for monogenic segregation of a dominant gene (**Table 2**). Furthermore, using isolate YT03 that was avirulent to Liangxing 99 (*Pm52*) and virulent to Jimai 23, the segregation ratio of Jimai 22 \times Tainong 18 population was consistent with the ratio for monogenic segregation of a dominant gene, while the while that of Jimai 23 \times Tainong 18 population were all susceptible (**Table 2**). To confirm the segregation ratios of the populations when inoculated with other *Bgt* isolates that were avirulent to both Jimai 23 and Liangxing 99 (*Pm52*), three isolates YT05, YT23, and YT35, were selected to inoculate the F_2 and $F_{2:3}$ populations of Jimai 22 \times Tainong 18 and Jimai 23 \times Tainong 18. The results were consistent with those of YT01. These results suggested that the powdery mildew resistance in Jimai 23 may be controlled by a single dominant gene, whereas powdery mildew resistance in Jimai 22 is controlled by two dominant genes that fit independent separation of Mendelian law and that Jimai 23 most likely inherited one of the two *Pm* genes of Jimai 22.

Genetic Mapping of *PmJM23* and Its Genealogical Relationship With Its Parent Jimai 22

When inoculated with isolate YT20 that was virulent to Liangxing 99 (*Pm52*) and avirulent to Jimai 23, we detected a *Pm2*-linked marker polymorphism in both Jimai 23 \times Tainong 18 and Jimai 22 \times Tainong 18 populations using *Pm2*-linked marker *Cfd81*. When inoculated with isolate YT03 that was virulent to Jimai 23 and avirulent to Liangxing 99 (*Pm52*), we detected a *Pm52*-linked marker polymorphism in Jimai 22 \times Tainong 18 population using *Pm52*-linked

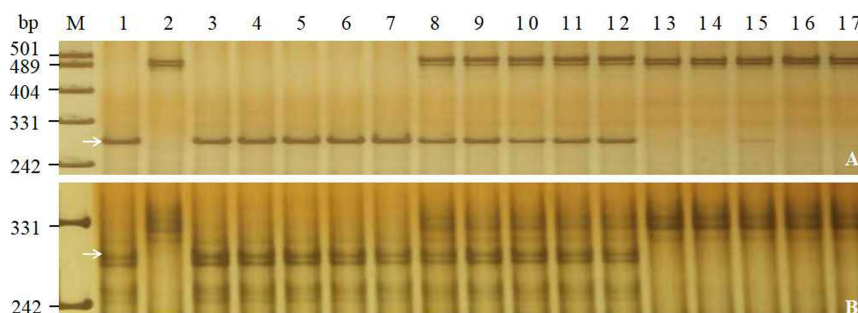
TABLE 2 | Segregation ratios of F_2 and $F_{2:3}$ generations of Jimai 23 \times Tainong 18 and Jimai 22 \times Tainong 18 following inoculation with different *Blumeria graminis* f. sp. *tritici* (*Bgt*) isolates at the seedling stage.

Crosses	YT01			YT20			YT03		
	Segregation ratio	χ^2	P	Segregation ratio	χ^2	P	Segregation ratio	χ^2	P
Jimai 22 \times Tainong 18 F_2	(RR + Rr):rr = 234:14	$\chi^2_{15:1} = 0.15$	0.69	(RR + Rr):rr = 218:69	$\chi^2_{3:1} = 0.14$	0.70	(RR + Rr):rr = 201:71	$\chi^2_{3:1} = 0.12$	0.73
Jimai 22 \times Tainong 18 $F_{2:3}$	RR:Rr:rr = 109:112:12	$\chi^2_{7:8:1} = 0.69$	0.71	RR:Rr:rr = 58:118:62	$\chi^2_{1:2:1} = 0.15$	0.93	RR:Rr:rr = 59:117:65	$\chi^2_{1:2:1} = 0.50$	0.78
Jimai 23 \times Tainong 18 F_2	(RR + Rr):rr = 175:56	$\chi^2_{3:1} = 0.07$	0.79	(RR + Rr):rr = 199:61	$\chi^2_{3:1} = 0.33$	0.57	(RR + Rr):rr = 0:242	–	–
Jimai 23 \times Tainong 18 $F_{2:3}$	RR:Rr:rr = 47:103:50	$\chi^2_{1:2:1} = 0.27$	0.87	RR:Rr:rr = 57:115:55	$\chi^2_{1:2:1} = 0.07$	0.96	RR:Rr:rr = 0:0:225	–	–

Values of χ^2 for statistical significance at $p < 0.05$ are 3.84 (1 df) and 5.99 (2 df); RR, homozygous resistant; Rr, segregating; rr, homozygous susceptible.

TABLE 3 | New polymorphic markers developed using BSR-Seq for the powdery mildew resistance gene *PmJM23*.

Markers	Forward primer (5'-3')	Reverse primer (5'-3')	SSR_START	SSR_END	Product size (bp)	Marker type
YTU201	GAATTCTGTTGCACCCCTGG	GGAGCTACCCGACAATCACT	Unknown	Unknown	272	Co-dominant
YTU3004	GACATGCCACCACACCATAC	TGAGCTACCTAACCCAACCG	41091652	41091661	290	Co-dominant
YTU3023	GGGACTGCAGGTTGAATAATGG	CATGTGCTTTCCAACAATGCA	41797078	41797152	242	Co-dominant
YTU3025	TGGCGAAGTTCCATATTTTGC	GAGTGGGTTAGTAGGCAGGG	41870024	41870034	330	Co-dominant
YTU3038	AGCACTTGTATTCGAGCCCT	GGCATGGGTTCTGACTAGGT	42356673	42356688	390	Co-dominant
YTU3042	AACACATGTAGGCCCCGAAG	GCGACCCCATTTATTACCAGG	42602783	42602792	235	Co-dominant
YTU3049	GAGATTGTGTTTCCATTGCGG	CAGGACACATCATCGCATCA	42914312	42914323	245	Co-dominant

**FIGURE 1** | Amplification patterns of the selected markers *Bwm20* (A) and *YTU3004* (B) in genotyping Jimai 23, Tainong 18, and random selected $F_{2:3}$ families of Jimai 23 \times Tainong 18. Lanes M, pUC18 *Msp*I; lanes 1–2: parents Jimai 23 and Tainong 18; lanes 3–7: homozygous-resistant $F_{2:3}$ families; lanes 8–12, heterozygous $F_{2:3}$ families; lanes 13–17, homozygous susceptible $F_{2:3}$ families. The white arrows indicate the 295 bp (A) and 290 bp (B) polymorphic bands in Jimai 23.

markers *Icssl326* and *Icssl795*, and when inoculated with isolate YT01 that was avirulent to both Jimai 23 and Liangxing 99 (*Pm52*), we detected *Pm2*-linked marker polymorphisms using five *Pm2*-linked markers (*Cfd81*, *Swgi069*, *Bwm20*, *Bwm21*, and *Bwm25*; Ma et al., 2018) and no *Pm52*-linked marker polymorphism using *Pm52*-linked markers *Icssl326* and *Icssl795* in Jimai 23 \times Tainong 18 population. By combining the genealogical relationship between Jimai 22 and 23 with the marker detection results, we confirmed that Jimai 23 has a *Pm* gene nearby or in the *Pm2* interval, which we tentatively designate as *PmJM23*. We suggest that Jimai 22 has two *Pm* genes, one of which is *PmJM23* and another is the reported *Pm52* (Yin et al., 2009; Qu et al., 2019).

To saturate the linkage map of the *PmJM23* interval, BSR-Seq was used to develop new markers linked to *PmJM23*. Using the SNPs and indels that distinguished or differed between resistant and susceptible parents and bulks in the targeted interval, 108 pairs of primers (Supplementary Table S3) were designed to screen polymorphic markers of *PmJM23* based on the SSR screening results in the interval (Supplementary Table S4). As a result, seven SSR markers (*YTU201*, *YTU3004*, *YTU3023*, *YTU3025*, *YTU3038*, *YTU3042*, and *YTU3049*) showed consistent polymorphisms between the parents and the resistant and susceptible bulks (Table 3). The markers were also linked with *PmJM23* after genotyping the $F_{2:3}$ population of Jimai 23 and Tainong 18 (Figure 1 and Table 3). A linkage map was then constructed using newly developed markers and also combining the documented *Pm2* markers

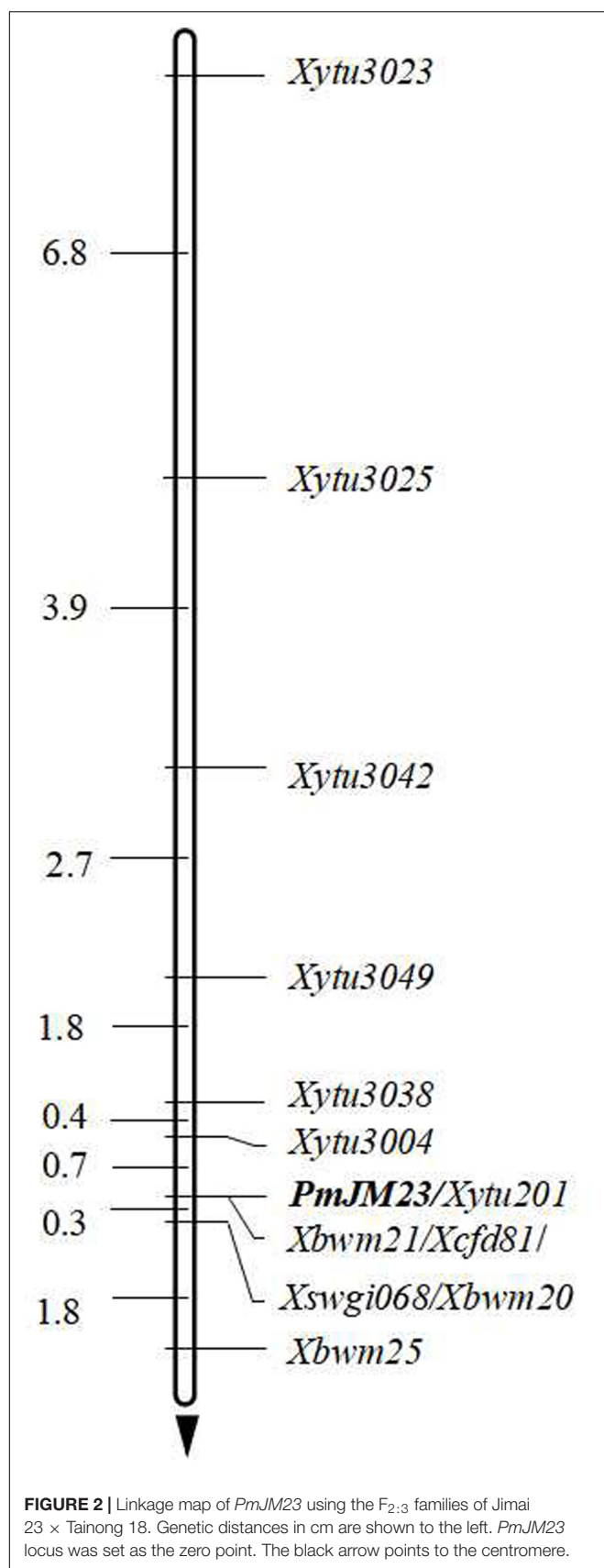
Cfd81, *Swgi068*, *Bwm20*, *Bwm21*, and *Bwm25* (Ma et al., 2018; Figure 2). *PmJM23* cosegregated with the markers *YTU201*, *Bwm21*, and *Cfd81* and was flanked by the markers *YTU3004* and *Swgi068/Bwm20* at genetic distances of 0.7 and 0.3 cM, respectively.

Allelic Test Between *PmJM23* and the Documented *Pm2* Alleles

Because *PmJM23* was assigned to the *Pm2* interval, the allelic relationship between *PmJM23* and documented *Pm2* alleles needed to be clarified using an allelic test. The phenotyping reactions of the 6,304 F_2 plants between Jimai 23 and Liangxing 66 (*PmLX66*, allelic to *Pm2*), and 5,869 F_2 plants between Jimai 23 and Wenong 14 (*PmW14*, allelic to *Pm2*) were surveyed. No susceptible plants were within all the tested F_2 populations, suggesting that the *PmJM23* locus is most likely allelic to the *Pm2* locus.

Comparisons of *PmJM23* and the Documented *Pm* Genes/Alleles at the *Pm2* Interval

When tested against 42 *Bgt* isolates, Jimai 23 showed a different response spectrum from other genotypes with documented *Pm* genes on chromosome arm 5DS, and especially from those of commercial cultivars (Supplementary Table S1). This information, combined with the allelic relationship between *PmJM23* and *Pm2*, demonstrated that *PmJM23* is most likely a new *Pm2* allele.



Evaluation of Closely Linked Markers for MAS and Their Application in Breeding

Combined with comprehensive agronomic traits of Jimai 23, *PmJM23* is a valuable gene for *Pm* breeding. To transfer *PmJM23* to susceptible cultivars using MAS, the closely linked markers of *PmJM23* (*YTU201*, *YTU3004*, *YTU3038*, *YTU3049*, *Cfd81*, *Swgi068*, *Bwm20*, *Bwm21*, and *Bwm25*) were first tested for their suitability for MAS through detecting 26 susceptible cultivars. The results indicated that all the tested markers can amplify polymorphic bands between Jimai 23 and these tested cultivars, suggesting that once *PmJM23* is transferred into these genetic backgrounds through hybridization, these markers can be used to detect *PmJM23* (Figure 3 and Supplementary Table S2). In other words, these markers could be used in MAS for *PmJM23* in those genetic backgrounds.

To check the effectiveness of the markers available for MAS, Jimai 23 was crossed with a series of susceptible cultivars/breeding lines, including Gaoyou 5766, SH4300, Tainong 2419, 125574, Daimai 1503, Jimai 20, Hanmai 13, Yannong 19, and Luyuan 502. The F_2 and F_3 plants with linked marker alleles were selected using the corresponding markers. Combined with selecting for agronomic performance in field, head rows were planted in F_4 generations. From the head rows, the hybridized combinations 19P084 (Jimai 23 × Daimai 1503), 19P085 (Jimai 23 × Gaoyou 5766), 19P086 (125574 × Jimai 23), 19P088 (Jimai 23 × SH4300), and 19P091 (Tainong 2419 × Jimai 23) with the best agronomic performance in the field were carried on to a field plot experiment. After comprehensive evaluation, two wheat breeding lines with superior agronomic performance in the field was selected. They were highly resistant to powdery mildew at both seedling and adult stages and have superior field performance (Figure 4). In the future, we expect these two lines to be evaluated in trials at both the provincial and national levels.

DISCUSSION

Jimai 23 is an elite wheat cultivar showing high resistance to powdery mildew during the whole growing season. The powdery mildew resistance in Jimai 23 is controlled by a single dominant gene, *PmJM23*. Using BSR-Seq, a series of new markers was developed and used to map *PmJM23* to a genetic interval, corresponding to the *Pm2* locus on chromosome arm 5DS. Resistance spectrum analysis demonstrated that *PmJM23* is a broad-spectrum *Pm* gene and hence is valuable for resistance breeding. A series of *Pm2* alleles have been identified in different wheat genotypes. Further genealogy tracing indicated that the *Pm2* alleles in landraces or breeding lines, such as *Pm2a* (Qiu et al., 2006), *Pm2b* (Ma et al., 2015a), and *Pm2c* (Xu et al., 2015), have unclear genealogy and that several wheat cultivars having the *Pm2* alleles, such as Liangxing 66 (genealogy: Ji91102/Jimai 19) with *PmLX66* (Sun Y.L. et al., 2015), Wennong 14 (genealogy: 84139/9215/876161) with *PmW14* (Sun Y.L. et al., 2015), Yingbo 700 (genealogy: Taigu sterility line/Jimai 19) with *PmYB* (Ma et al., 2015b), and Zhongmai 155 (genealogy: Jimai 19/Lumai 21) with *PmZ155* (Sun H.G. et al., 2015), have different genealogy

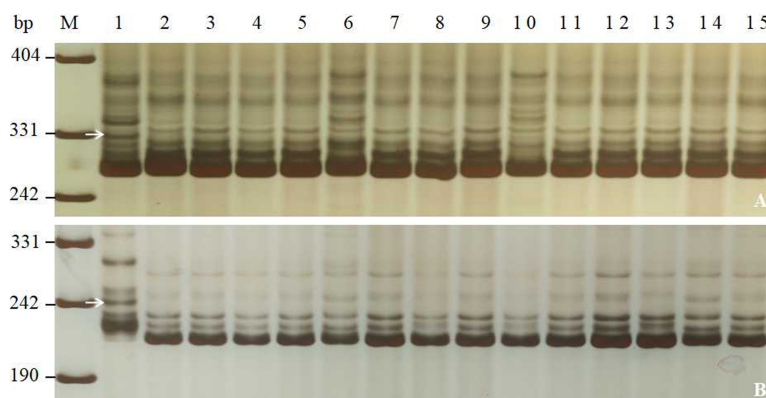


FIGURE 3 | Amplification patterns of *PmJM23*-linked markers *YTU3025* (A) and *YTU3049* (B) in Jimai 23, Tainong 18, and selected wheat cultivars/breeding lines. M, DNA marker pUC18 *Msp*I; lanes 1 and 2, Jimai 23 and Tainong 18; lanes 3–15, wheat cultivars with sequential order of Huixianhong, Hanmai 13, Huaimai 0226, Xinong 979, Luchen 185, Jimai 268, Tainong 1014, Jinan 17, Liangxing 619, Shimai 15, Jimai 21, Jimai 20, and Shixin 633. The white arrows indicate the 330 bp (A) and 245 bp (B) polymorphic bands in Jimai 23.

with Jimai 23 (Jimai 22/Yumai 13) and Jimai 22 (genealogy: 935024/935106). Even resistance genes from the same original donor may also have variations diversified under selective pressure. Thus, these genotypes exhibited significantly different resistance spectra (Ma et al., 2018). However, whether a resistance gene can be used efficiently and rapidly in resistance breeding depends not only on the breadth of the resistance spectrum but also the comprehensive agronomic and quality characters of the donor. Wheat genotypes that exhibit resistance but which have poor or defective agronomic traits will slow down the breeding cycle because of the number of backcrosses required, which is not acceptable to breeders (Summers and Brown, 2013). Although many *Pm* genes, including *Pm2* alleles, have been identified, relatively few genes have been successfully used in modern breeding because of an imperfect balance between resistance and the agronomic traits of the resistance donors (Bie et al., 2015; Shah et al., 2018). In this study, Jimai 23 exhibited broader resistant spectrum than four of the six currently grown cultivars, and with isolate resistance similar to one cultivar (Supplementary Table S1). We found excellent correspondence between resistance and agronomic traits in Jimai 23. Moreover, *PmJM23* has a broader resistance spectrum than most of the resistant genes currently in production. We propose that Jimai 23 is an ideal resistance resource for wheat breeding.

In this study, *PmJM23* was mapped to the *Pm2* locus. Although this locus has been cloned by mutant chromosome sequencing (Sánchez-Martín et al., 2016) and analysis of the annotated genes in the mapping interval using the reference genome of Chinese Spring (Chen et al., 2019), no transgenic evidence was provided to confirm that the cloned sequence is the unique functional element conferring resistance to powdery mildew. Additionally, other reports showed that all the homologous sequences in different *Pm2* allele donors have the exactly the same sequence, yet these *Pm2* alleles exhibit significantly different resistance spectra that cannot be explained by the background differences of the resistant germplasms (Jin et al., 2018; Ma et al., 2018). In this study, the marker order across the *PmJM23* interval showed inversion and recombination

phenomena (Figure 2 and Table 3). We speculate that the *Pm2* locus is most likely a complex locus that contains multiple elements conferring resistance to powdery mildew. To identify this resistance locus, map-based cloning using genomic information of the donor itself is imperative. In this study, we developed additional markers based on BSR-Seq that are located in the *Pm2* interval. These markers will accelerate map-based cloning of this locus and leading to the identification of the element(s) responding to powdery mildew.

Obviously, the complex *Pm2* locus has not been fully characterized, but this does not affect MAS of *PmJM23*. Using the markers newly developed in this work, the *PmJM23* locus as a breeding module can be efficiently transferred into susceptible cultivars. We showed that the newly developed markers are very suitable for MAS in different genetic backgrounds. Using these markers, we efficiently selected more than 20 breeding lines with greatly shortened breeding cycles and increased breeding efficiency. Five of these lines have superior agronomic traits and high resistance to powdery mildew. Thus, our study provides a successful model starting with evaluating the gene by resistance spectrum analysis, identifying the gene by genetic analysis, mapping the gene using closely linked markers and, finally, successfully transferring the gene by MAS.



FIGURE 4 | Field performance of Jimai 23 derived breeding line 19P084 (A) (Jimai 23 × Daimai 1503), 19P085 (Jimai 23 × Gaoyou 5766) (B), that contained *PmJM23*. The left bottom in each figure are the susceptible controls Daimai 1503 and Gaoyou 5766.

CONCLUSION

In this study, we identified and characterized a broad-spectrum *Pm* gene, *PmJM23*, in the elite cultivar Jimai 23 and successfully used it in MAS. This work is valuable as it provides the means of accelerating the utilization of *PmJM23* in breeding programs to control powdery mildew in wheat.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

PM and HH conceived the research. MJ, WD, WW, XZ, LW, and XL performed the experiments. HX, CL, and JS analyzed the data. RM, HL, JL, RH, and XW performed the MAS. PM wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was financially supported by the Shandong Agricultural Seed Improvement Project (2019LZGC016),

Key Research and Development Program of Yantai City (2019YT06000470), Taishan Scholars Project (tsqn201812123), National Natural Science Foundation of China (31971874), Jiangsu Agricultural Science and Technology Innovation Fund (CX(19)2042), and Priority Academic Program Development of Jiangsu Higher Education Institutions and Natural Science Foundation of China (31671771).

ACKNOWLEDGMENTS

We are grateful to Prof. John Clemens and Prof. Paula Jameson at University of Canterbury, New Zealand, for constructive comments and English editing of the manuscript. Prof. Jameson is also the team leader of the Shandong Provincial “Double-Hundred Foreign Talent Plan” and is currently employed at Yantai University as distinguished professor.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00241/full#supplementary-material>

REFERENCES

- An, D. G., Zheng, Q., Zhou, Y. L., Ma, P. T., Lv, Z. L., Li, L. H., et al. (2013). Molecular cytogenetic characterization of a new wheat-rye 4R chromosome translocation line resistant to powdery mildew. *Chromosome Res.* 21, 419–432. doi: 10.1007/s10577-013-9366-8
- Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S. O., Gundlach, H., et al. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* 357, 93–97. doi: 10.1126/science.aan0032
- Bie, T. D., Zhao, R. H., Zhu, S. Y., Chen, S. L., Cen, B., Zhang, B. Q., et al. (2015). Development and characterization of marker *MBH1* simultaneously tagging genes *Pm21* and *PmV* conferring resistance to powdery mildew in wheat. *Mol. Breeding* 35:189. doi: 10.1007/s11032-015-0385-3
- Chen, C., He, Z. H., Lu, J. L., Li, J., Ren, Y., Ma, C. X., et al. (2016). Molecular mapping of stripe rust resistance gene *YrJ22* in Chinese wheat cultivar Jimai 22. *Mol. Breeding* 36:118. doi: 10.1007/s11032-016-0540-5
- Chen, F., Jia, H. Y., Zhang, X. J., Qiao, L. Y., Li, X., Zheng, J., et al. (2019). Positional cloning of *PmCH1357* reveals the origin and allelic variation of the *Pm2* gene for powdery mildew resistance in wheat. *Crop J.* 6, 771–783. doi: 10.1016/j.cj.2019.08.004
- de Waard, M. A., Kipp, E. M. C., Horn, N. M., and van Nistelrooy, J. G. M. (1986). Variation in sensitivity to fungicides which inhibit ergosterol biosynthesis in wheat powdery mildew. *Neth. J. Plant Pathol.* 92, 21–32. doi: 10.1007/bf01976373
- El-Shamy, M. M., Emara, H. M., and Mohamed, M. E. (2016). Virulence analysis of wheat powdery mildew (*Blumeria graminis* f. sp. tritici) and effective genes in middle Delta. *Egypt. Plant Dis.* 100, 1927–1930. doi: 10.1094/PDIS-01-16-0130-RE
- Felsenstein, F., Semar, M., and Stammler, G. (2010). Sensitivity of wheat powdery mildew (*Blumeria graminis* f. sp. tritici) towards metrafenone. *Gesunde Pflanz.* 62, 29–33. doi: 10.1007/s10343-010-0214-x
- Friebe, B., Heun, M., and Bushuk, W. (1989). Cytological characterization, powdery mildew resistance and storage protein composition of tetraploid and hexaploid 1BL/1RS wheat-rye translocation lines. *Theor. Appl. Genet.* 78, 425–432. doi: 10.1007/BF00265307
- Gupta, P. K., Langridge, P., and Mir, R. R. (2010). Marker-assisted wheat breeding: present status and future possibilities. *Mol. Breeding* 261, 145–161. doi: 10.1007/s11032-009-9359-7
- He, H. G., Ji, Y. Y., Zhu, S. Y., Li, B., Zhao, R. H., Jiang, Z. Z., et al. (2017). Genetic, physical and comparative mapping of the powdery mildew resistance gene *Pm21* originating from *Dasypyrum villosum*. *Front. Plant Sci.* 8:1914. doi: 10.3389/fpls.2017.01914
- Huang, S., and Pang, F. (2017). “Biocontrol agents for controlling wheat rust,” in *Wheat rust diseases. Methods in Molecular Biology*, Vol. 1659, ed. S. Periyannan (New York, NY: Humana Press).
- Ingvarsdson, C. R., Massange-Sánchez, J. A., Borum, F., Uauy, C., and Gregersen, P. L. (2019). Development of *mlo*-based resistance in tetraploid wheat against wheat powdery mildew. *Theor. Appl. Genet.* 132, 3009–3022. doi: 10.1007/s00122-019-03402-4
- Jiang, Y., Schulthess, A. W., Rodemann, B., Ling, J., Plieske, J., Kollers, S., et al. (2017). Validating the prediction accuracies of marker-assisted and genomic selection of Fusarium head blight resistance in wheat using an independent sample. *Theor. Appl. Genet.* 130, 471–482. doi: 10.1007/s00122-016-2827-7
- Jin, Y. L., Xu, H. X., Ma, P. T., Fu, X. Y., Song, L. P., Xu, Y. F., et al. (2018). Characterization of a new *Pm2* allele associated with broad-spectrum powdery mildew resistance in wheat line Subtil. *Sci. Rep.* 8:475. doi: 10.1038/s41598-017-18827-4
- Kosambi, D. D. (1943). The estimation of map distance from recombination values. *Ann. Eugen.* 12, 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x
- Kuchel, H., Fox, R., Reinheimer, J., Mosione, L., Willey, N., Bariana, H., et al. (2007). The successful application of a marker-assisted wheat breeding strategy. *Mol. Breeding* 20, 295–308. doi: 10.1007/s11032-007-9092-z
- Kuchel, H., Langridge, P., Mosione, L., Williams, K., and Jefferies, S. P. (2006). The genetic control of milling yield, dough rheology and baking quality of wheat. *Theor. Appl. Genet.* 112:1487. doi: 10.1007/s00122-006-0252-z
- Lee, J. H., Graybosch, R. A., and Peterson, C. J. (1995). Quality and biochemical effects of a 1BL/1RS wheat-rye translocation in wheat. *Theor. Appl. Genet.* 1995, 105–112. doi: 10.1007/BF00221002
- Li, G. Q., Cowger, C., Wang, X. W., Carver, B. F., and Xu, X. Y. (2019a). Characterization of *Pm65*, a new powdery mildew resistance gene on

- chromosome 2AL of a facultative wheat cultivar. *Theor. Appl. Genet.* 132, 2625–2632. doi: 10.1007/s00122-019-03377-2
- Li, G. Q., Zhou, J. Y., Jia, H. Y., Gao, Z. X., Fan, M., Luo, Y. J., et al. (2019b). Mutation of a histidine-rich calcium-binding-protein gene in wheat confers resistance to Fusarium head blight. *Nat Genet* 51, 1106–1112. doi: 10.1038/s41588-019-0426-7
- Li, Y. H., Shi, X. H., Hu, J. H., Wu, P. P., Qiu, D., Qu, Y. F., et al. (2020). Identification of a recessive gene *PmQ* conferring resistance to powdery mildew in wheat landrace Qingxinmai using BSR-Seq analysis. *Plant Dis.* 104, 743–751. doi: 10.1094/PDIS-08-19-1745-RE
- Li, H. J., Wang, X. M., Song, F. J., Wu, C. P., Wu, X. F., Zhang, N., et al. (2011). Response to powdery mildew and detection of resistance genes in wheat cultivars from China. *Acta Agron. Sin.* 37, 943–954. doi: 10.1016/S1875-2780(11)60026-6
- Lincoln, S., Daly, M., and Lander, E. (1993). *Constructing Genetic Maps With Mapmaker/EXP3.0 Whitehead Institute Techn Rep*, 3rd Edn. Cambridge, MA: Whitehead Institute.
- Ling, H. Q., Ma, B., Shi, X., Liu, H., Dong, L. L., Sun, H., et al. (2018). Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature* 557, 424–428. doi: 10.1038/s41586-018-0108-0
- Liu, L. Q., Luo, Q. L., Li, H. W., Li, B., Li, Z. S., and Zheng, Q. (2018). Physical mapping of the blue-grained gene from *Thinopyrum ponticum* chromosome 4Ag and development of blue-grain-related molecular markers and a FISH probe based on SLAF-seq technology. *Theor. Appl. Genet.* 131, 2359–2370. doi: 10.1007/s00122-018-3158-7
- Luo, M., Gu, Y., Puiu, D., Wang, H., Twardziok, S. O., Deal, K. R., et al. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops auschii*. *Nature* 551, 498–502. doi: 10.1038/nature24486
- Ma, P. T., Han, G. H., Zheng, Q., Liu, S. Y., Han, F. P., Wang, J., et al. (2019). Development of novel wheat-rye chromosome 4R translocations and assignment of their powdery mildew resistance. *Plant Dis.* 104, 260–268. doi: 10.1094/PDIS-01-19-0160-RE
- Ma, P. T., Xu, H. X., Li, L. H., Zhang, H. X., Han, G. H., Xu, Y. F., et al. (2016). Characterization of a new *Pm2* allele conferring powdery mildew resistance in the wheat germplasm line FG-1. *Front. Plant Sci.* 7:546. doi: 10.3389/fpls.2016.00546
- Ma, P. T., Xu, H. X., Xu, Y. F., Li, L. H., Qie, Y. M., Luo, Q. L., et al. (2015a). Molecular mapping of a new powdery mildew resistance gene *Pm2b* in Chinese breeding line KM2939. *Theor. Appl. Genet.* 128, 613–622. doi: 10.1007/s00122-015-2457-5
- Ma, P. T., Xu, H. X., Xu, Y. F., Song, L. P., Liang, S. S., Sheng, Y., et al. (2018). Characterization of a powdery mildew resistance gene in wheat breeding line 10V-2 and its application in marker-assisted selection. *Plant Dis.* 102, 925–931. doi: 10.1094/PDIS-02-17-0199-RE
- Ma, P. T., Zhang, H. X., Xu, H. X., Xu, Y. F., Cao, Y. W., Zhang, X. T., et al. (2015b). The gene *PmYB* confers broad-spectrum powdery mildew resistance in the multi-allelic *Pm2* chromosome region of the Chinese wheat cultivar YingBo 700. *Mol. Breeding* 35:124. doi: 10.1007/s11032-015-0320-7
- McIntosh, R. A., Dubcovsky, J., Rogers, W. J., Xia, X. C., and Raupp, W. J. (2019). “Catalogue2 of gene symbols for wheat: 2019 supplement,” in *Annual Wheat Newsletter*, ed. W. J. Raupp (Manhattan, NY: The Wheat Genetic and Genomic Resources at Kansas State University), 98–113.
- Morgounov, A., Tufan, H. A., Sharma, R., Akin, B., Bagci, A., Braun, H. J., et al. (2012). Global incidence of wheat rusts and powdery mildew during 1969–2010 and durability of resistance of winter wheat variety Bezostaya 1. *Eur. J. Plant Pathol.* 132, 323–340. doi: 10.1007/s10658-011-9879-y
- Qiu, Y. C., Sun, X. L., Zhou, R. H., Kong, X. Y., Zhang, S. S., and Jia, J. Z. (2006). Identification of microsatellite markers linked to powdery mildew resistance gene *Pm2* in wheat. *Cereal Res. Commun.* 34, 1267–1273. doi: 10.1556/CRC.34.2006.4.268
- Qu, Y. F., Wu, P. P., Hu, J. H., Chen, Y. X., Shi, Z. L., Qiu, D., et al. (2019). Molecular detection of the powdery mildew resistance genes in winter wheats DH51302 and Shimai 26. *J. Integr. Agr.* 18, 2–11. doi: 10.1016/S2095-3119(19)62644-4
- Sánchez-Martín, J., Steuernagel, B., Ghosh, S., Herren, G., Hurni, S., Adamski, N., et al. (2016). Rapid gene isolation in barley and wheat by mutant chromosome sequencing. *Genome Biol.* 17:221. doi: 10.1186/s13059-016-1082-1
- Shah, L., Rehman, S., Ali, A., Yahya, M., Riaz, M. W., Si, H. Q., et al. (2018). Genes responsible for powdery mildew resistance and improvement in wheat using molecular marker-assisted selection. *J. Plant Dis. Protect.* 125, 145–158. doi: 10.1007/s41348-017-0132-6
- Shi, L. J., Jiang, C. C., He, Q., Habeku, A., Ordon, F., Luan, H. Y., et al. (2019). Bulk segregant RNA-sequencing (BSR-seq) identified a novel rare allele of *eIF4E* effective against multiple isolates of BaYMV/BaMMV. *Theor. Appl. Genet.* 132, 1777–1788. doi: 10.1007/s00122-019-03314-3
- Studnicki, M., Wijata, M., Sobczyński, G., Samborski, S., and Rozbicki, J. (2018). Assessing grain yield and quality traits stability of spring wheat cultivars at different crop management levels. *Cereal Res. Commun.* 46, 180–190. doi: 10.1556/0806.45.2017.066
- Summers, R. W., and Brown, J. K. M. (2013). Constraints on breeding for disease resistance in commercially competitive wheat cultivars. *Plant Pathol.* 62, 115–121. doi: 10.1111/ppa.12165
- Sun, H. G., Song, W., Sun, Y. L., Chen, X. M., Liu, J. J., Zou, J. W., et al. (2015). Resistance to powdery mildew in the wheat cultivar Zhongmai 155: effectiveness and molecular detection of the resistance gene. *Crop Sci.* 55, 1017–1025. doi: 10.2135/cropsci2014.05.0355
- Sun, Y. L., Zou, J. W., Sun, H. G., Song, W., Wang, X. M., and Li, H. J. (2015). *PmLX66* and *PmW14*: new alleles of *Pm2* for resistance to powdery mildew in the Chinese winter wheat cultivars Liangxing 66 and Wennong 14. *Plant Dis.* 99, 1118–1124. doi: 10.1094/PDIS-10-14-1079-RE
- Tan, C., Liu, Z. Y., Huang, S. N., and Feng, H. (2019). Mapping of the male sterile mutant gene *ftms* in *Brassica rapa* L. ssp. *pekinensis* via BSR-Seq combined with whole-genome resequencing. *Theor. Appl. Genet.* 132, 355–370. doi: 10.1007/s00122-018-3223-2
- Teich, A. H. (1984). Heritability of grain yield, plant height and test weight of a population of winter wheat adapted to Southwestern Ontario. *Theor. Appl. Genet.* 68, 21–23. doi: 10.1007/BF00252304
- The International Wheat Genome Sequencing Consortium (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191
- Wu, J. H., Zeng, Q. D., Wang, Q. L., Liu, S. J., Yu, S. Z., Mu, J. M., et al. (2018). SNP-based pool genotyping and haplotype analysis accelerate fine-mapping of the wheat genomic region containing stripe rust resistance gene *Yr26*. *Theor. Appl. Genet.* 131, 1481–1496. doi: 10.1007/s00122-018-3092-8
- Xiao, M. G., Song, F. J., Jiao, J. F., Wang, X. M., Xu, H. X., and Li, H. J. (2013). Identification of the gene *Pm47* on chromosome 7BS conferring resistance to powdery mildew in the Chinese wheat landrace Hongyanglazi. *Theor. Appl. Genet.* 126, 1397–1403. doi: 10.1007/s00122-013-2060-6
- Xu, H. X., Yi, Y. J., Ma, P. T., Qie, Y. M., Fu, X. Y., Xu, Y. F., et al. (2015). Molecular tagging of a new broad-spectrum powdery mildew resistance allele *Pm2c* in Chinese wheat landrace Niamai. *Theor. Appl. Genet.* 128, 2077–2084. doi: 10.1007/s00122-015-2568-z
- Xu, Y. F., Wang, R. F., Tong, Y. P., Zhao, H. T., Xie, Q. E., Liu, D. C., et al. (2014). Mapping QTLs for yield and nitrogen-related traits in wheat: influence of nitrogen and phosphorus fertilization on QTL expression. *Theor. Appl. Genet.* 127, 59–72. doi: 10.1007/s00122-013-2201-y
- Yin, G. H., Li, G. Y., He, Z. H., Liu, J. J., Wang, H., and Xia, X. C. (2009). Molecular mapping of powdery mildew resistance gene in wheat cultivar Jimai 22. *Acta Agron. Sin.* 35, 1425–1431. doi: 10.3724/SP.J.1006.2009.01425
- Zhao, Z., Sun, H., Song, W., Lu, M., Huang, J., Wu, L. F., et al. (2013). Genetic analysis and detection of the gene *MILX99* on chromosome 2BL conferring resistance to powdery mildew in the wheat cultivar Liangxing 99. *Theor. Appl. Genet.* 126, 3081–3089. doi: 10.1007/s00122-013-2194-6

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 © 2020 Jia, Xu, Liu, Mao, Li, Liu, Du, Wang, Zhang, Han, Wang, Wu, Liang, Song, He and Ma. 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.



Expressing a Target Mimic of miR156fhl-3p Enhances Rice Blast Disease Resistance Without Yield Penalty by Improving *SPL14* Expression

Ling-Li Zhang^{1†}, Yan Li^{1,2†}, Ya-Ping Zheng^{1†}, He Wang¹, Xuemei Yang¹, Jin-Feng Chen¹, Shi-Xin Zhou¹, Liang-Fang Wang¹, Xu-Pu Li¹, Xiao-Chun Ma¹, Ji-Qun Zhao^{1,2}, Mei Pu^{1,2}, Hui Feng^{1,2}, Jing Fan^{1,2}, Ji-Wei Zhang^{1,2}, Yan-Yan Huang^{1,2} and Wen-Ming Wang^{1,2*}

¹ Rice Research Institute and Key Lab for Major Crop Diseases, Sichuan Agricultural University at Wenjiang, Chengdu, China,

² State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University at Wenjiang, Chengdu, China

OPEN ACCESS

Edited by:

Zhu-Qing Shao,
Nanjing University, China

Reviewed by:

Fengming Song,
Zhejiang University, China
Guo-dong Lu,
Fujian Agriculture and Forestry
University, China

*Correspondence:

Wen-Ming Wang
j316wenmingwang@sicau.edu.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: 03 January 2020

Accepted: 19 March 2020

Published: 23 April 2020

Citation:

Zhang L-L, Li Y, Zheng Y-P,
Wang H, Yang X, Chen J-F, Zhou S-X,
Wang L-F, Li X-P, Ma X-C, Zhao J-Q,
Pu M, Feng H, Fan J, Zhang J-W,
Huang Y-Y and Wang W-M (2020)
Expressing a Target Mimic
of miR156fhl-3p Enhances Rice Blast
Disease Resistance Without Yield
Penalty by Improving *SPL14*
Expression. *Front. Genet.* 11:327.
doi: 10.3389/fgene.2020.00327

MicroRNAs (miRNAs) play essential roles in the regulation of plant growth and defense responses. More and more, miRNA-3ps are reported to act in plant development and immunity. miR156 is a conserved miRNA, and most previous studies focus on its roles in plant growth, development, and yield determinacy. Here, we show that expressing a target mimic of miR156fhl-3p led to enhanced rice blast disease resistance without a yield penalty. miR156fhl-3p was differentially responsive to *Magnaporthe oryzae* in susceptible and resistant accessions. Transgenic lines expressing a target mimic of miR156fhl-3p (MIM156-3p) exhibited enhanced rice blast disease resistance and increased expression of defense-related genes. MIM156-3p also enhanced the mRNA abundance of *SPL14* and *WRKY45* by down-regulating miR156-5p and pre-miR156. Moreover, MIM156-3p lines displayed a decreased number of second rachis branches per panicle but enlarged grains, leading to unchanged yield per plant. Consistently, overexpressing miR156h (OX156) led to enhanced susceptibility to *M. oryzae* and decreased the expression of *SPL14* and *WRKY45*. Our results indicate that miR156fhl-3p mounts a regulatory role on miR156-5p, which subsequently regulates the expression of *SPL14* and *WRKY45* to improve rice blast disease resistance.

Keywords: microRNA, osa-miR156, rice blast disease resistance, *Magnaporthe oryzae*, *SPL14*, *WRKY45*

INTRODUCTION

MicroRNAs (miRNAs) are a kind of 20- to 24-nucleotide (nt) non-coding RNA molecule that regulate the expression of genes with sequences complementary to the miRNAs (Bartel, 2004; Yu et al., 2017). One miRNA isoform can be transcribed from one or more *MIR* gene loci (Li et al., 2019). A *MIR* gene is first transcribed into a pri-miRNA, which is processed by Dicer-like proteins (DCLs) to form pre-miRNA (Rogers and Chen, 2013). The pre-miRNA is further cleaved by DCLs, generating a miRNA-5p/miRNA-3p (previously miRNA/miRNA*) duplex (Liu et al., 2005). Then, the duplex separates into a mature miRNA arm (-5p or -3p) that is bound by

Argonaute (AGO) proteins to form RNA-induced silencing complex (RISC) (Ji et al., 2011; Zhu et al., 2011; Kobayashi and Tomari, 2016). RISC binds to the sequences reversely complementary to the miRNA and triggers DNA methylation, mRNA cleavage, or translational inhibition, respectively (Iwakawa and Tomari, 2015; Yu et al., 2017). The selection of a mature miRNA (-5p or -3p arm) can be different across tissues, developmental stages, and abiotic or biotic stresses (Guo et al., 2014; Hu et al., 2014). Some *MIR* genes are only processed into miRNA-5p, and some miRNAs are processed into miRNA-5p and miRNA-3p. The miRNA-3p was once considered as a useless by-product of the miRNA biogenesis. Some recent studies showed that a miRNA-5p and a miRNA-3p could target different gene families, each of which is involved in diverse biological processes (Jones-Rhoades et al., 2006; Huang et al., 2019). For example, in *Arabidopsis*, miR393 can be induced by a well-studied pathogen-associated molecular pattern (PAMP) molecule, flg22, and positively contributes to resistance against *Pseudomonas syringae* pv. *tomato* DC3000 (*Pst* DC3000) via miR393-5p-mediated suppression of the expression of *Transport Inhibitor Response 1 (TIR1)*, *Auxin signaling F-Box proteins 2 (AFB2)*, and *AFB3* (Navarro et al., 2006). On the other hand, miR393-3p targets *MEMB12*, a gene encoding a Gogi-localized SNARE (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) protein involved in protein trafficking and regulating plant immune responses (Zhang et al., 2011). However, it is unclear whether the accumulation of a -3p correlates with a -5p.

Rice feeds half of the world population. Rice production is continually threatened by blast disease that is caused by *Magnaporthe oryzae* (syn. *Pyricularia oryzae*) (Zhang et al., 2018). In the past decade, more than 70 miRNAs have been identified to be responsive to *M. oryzae* or its elicitors, and 11 of them have been functionally characterized to play positive or negative roles in rice blast disease resistance (Li et al., 2019; Zhou et al., 2019). For example, miR160 positively regulates rice immunity against *M. oryzae* via up-regulating defense-related gene expression and hydrogen peroxide accumulation at the infection site (Li et al., 2014). miR167d facilitates *M. oryzae* infection by repressing the expression of *Auxin Responsive Factor 12 (ARF12)* (Zhao et al., 2019). miR169 negatively regulates rice immunity by repressing *Nuclear Factor Y-A (NF-YA)* genes (Wu et al., 2009; Li et al., 2017). miR319 targets the transcription factor *Teosinte Branched1/Cycloidea/Proliferating Cell Factor1 (OsTCP21)*, which manipulates JA synthesis via *OsLOX2* and *OsLOX5*, the key synthetic components of JA, to facilitate the blast fungus infection (Hao et al., 2012; Zhang et al., 2018). miR396 targets *Growth Regulating Factor (GRF)* genes to balance growth, yield traits, and immunity against *M. oryzae* (Chandran et al., 2019). Different miR156-3p isoforms are identified to be responsive to the infection of *M. oryzae* (Li et al., 2014), but their roles are unknown.

miR156 is a conserved and highly abundant miRNA in most plants and acts as a master regulator during the whole growth period (Axtell and Bowman, 2008). In *Arabidopsis*, miR156 is highly expressed in young seedlings and miR156-5p target *SQUAMOSA promoter-binding protein-like transcription factor*

(*SPL*) genes to regulate growth and development (Cardon et al., 1999; Wu and Poethig, 2006; Addoquaye et al., 2008; Shikata et al., 2009; Xu et al., 2016). These *Arabidopsis* *SPL* genes can be classified into three functionally distinct groups. The first group includes *SPL2*, *SPL9*, *SPL10*, *SPL11*, *SPL13*, and *SPL15*, which regulate juvenile-to-adult vegetative and vegetative-to-reproductive transition. The second group includes *SPL3*, *SPL4*, and *SPL5*, which could promote the floral meristem identity transition. The third group is *SPL6*, which may be important for physiological processes (Xie et al., 2006; Riese et al., 2007; Preston and Hileman, 2013). In rice, the *Osa-miR156* gene family has 12 members that generate three mature miR156-5p isoforms and four mature miR156-3p isoforms. miR156-5p targets 11 rice *SPL* genes that are involved in the regulation of important agronomic traits and blast disease resistance (Xie et al., 2006). For example, *SPL13* positively contributes to rice yield via improving grain size and panicle size (Si et al., 2016). *SPL16* positively regulates grain size via binding to GW7, which acts as a critical regulator in rice architecture and grain development (Wang et al., 2012; Wang S. et al., 2015). *SPL18* negatively regulates grain number through binding to the promoter of *DEP1* that acts as an important regulator of panicle architecture (Huang et al., 2009; Taguchi-Shiobara et al., 2011; Yuan et al., 2019). *SPL14* is also known as *Ideal Plant Architecture 1 (IPA1)*, which plays a critical role in rice architecture and resistance to *M. oryzae* via phosphorylation status and subsequent DNA-binding specificity (Jiao et al., 2010; Wang et al., 2018). Before *M. oryzae* invasion, *SPL14* protein binds to the *DEP1* promoter that contains the GTAC motif, to regulate tiller development, leading to plants with fewer unproductive tillers and more grains per panicle, supporting higher yield (Wang et al., 2018). Upon *M. oryzae* invasion, *SPL14* protein is phosphorylated and binds to the *WRKY45* promoter that contains TGGGCC motif, leading to enhanced resistance to the blast disease (Wang et al., 2018).

WRKY45 encodes a WRKY transcription factor that binds to W-box or W-like box-type cis-elements in response to pathogens and plays a vital role in plant-pathogen interaction (Li et al., 2004; Eulgem and Somssich, 2007). Overexpression of *WRKY45* markedly enhances rice blast resistance, whereas knock out of *WRKY45* compromises blast resistance (Shimono et al., 2007). The *WRKY45* transcript responds strongly to SA and BTH, but its role is neither downstream nor upstream of *NH1*, a rice homolog of *NPR1*. In contrast, a glutathione S-transferase and a cytochrome P450 are found to be downstream of *WRKY45*. Elevated *WRKY45* mRNA level could remarkably enhance resistance to *M. oryzae* (Shimono et al., 2007; Goto et al., 2015, 2016). However, constitutive *WRKY45* overexpression-conferred blast resistance remarkably penalizes yield (Goto et al., 2015). Expressing *WRKY45* under the control of pathogen-responsive promoters in combination with a translational enhancer derived from a 5'-untranslated region (UTR) of rice *alcohol dehydrogenase (ADH)* has strong disease resistance without penalty yield (Goto et al., 2016). Therefore, it is highly anticipated that miR156 plays an essential role in the regulation of rice blast disease resistance via the *SPL14-WRKY45* module. However, experimental evidence is lacking, and whether miR156-3p plays roles in blast resistance is unclear.

To investigate the function of miR156-3p in rice blast disease resistance, we constructed the transgenic lines expressing a target mimic of miR156fhl-3p (MIM156-3p) and found that MIM156-3p showed enhanced blast disease resistance. Surprisingly, the MIM156-3p plants exhibited significantly less pre-miR156 and miR156-5p accumulation than control, indicating that blocking miR156fhl-3p interfered with the production of miR156-5p, which in turn led to increase in the expression of *SPL14* and *WRKY45*. In contrast, elevated miR156h expression (OX156) enhanced susceptibility to *M. oryzae*, which could be due to the suppression of *SPL14* and decreased *WRKY45* expression. Moreover, MIM156-3p plants displayed an unchanged yield per plant. Together, our data indicate that miR156fhl-3p and miR156-5p may be mutually regulated via an unknown mechanism, and miR156 negatively regulates rice blast disease resistance via *SPL14* and *WRKY45*. Our results provide a potential tool to improve blast resistance without yield penalty and a new phenomenon in miRNA synthesis/accumulation.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Rice (*Oryza sativa*) materials used in this study include Lijiangxin Tuan Heigu (LTH) and International Rice Blast Line Pyricularia-Kanto51-m-Tsuyake (IRBLkm-Ts). LTH is a susceptible accession sensitive to over 1,300 regional isolates of *M. oryzae* worldwide (Lin et al., 2001). IRBLkm-Ts is a resistant accession, carrying a resistance locus *Pi-km* (Hiroshi et al., 2000). The rice cultivar Nipponbare was used as wild type (WT) and transgenic background. For germination, rice seeds were immersed in water for 2 d at 37°C in darkness. Then, seedlings were grown in an air-conditioned growth room at 26°C and 70% relative humidity with a 14 h light/10 h dark cycle. *Nicotiana benthamiana* plants were planted in an air-conditioned growth room at 24°C with a 16 h light/8 h dark cycle. *N. benthamiana* plants were used for *Agrobacterium*-infiltration experiments.

Plasmid Construction and Genetic Transformation

The artificial target mimic of miR156fhl-3p was acquired by annealing with primers miR156-mimic-F and miR156-mimic-R (Supplementary Figure S1 and Supplementary Table S1). The fragments were inserted into *IPS1* (*INDUCED BY PHOSPHATE STARVATION1*) to replace the miR399 target site as described previously (Franco-Zorrilla et al., 2007; Wu et al., 2013). Then, the DNA fragments were cloned into *Bam*HI/*Bgl*II site of pCambia1300-35S, resulting in overexpressing construct p35S:MIM156-3p. To generate transgenic plants overexpressing miR156h, we amplified the genomic sequences containing 487 bp upstream and 403 bp downstream of the *MIR156h* gene from Nipponbare genomic DNA with primers miR156h-F and miR156h-R (Supplementary Table S1). The amplified fragments were digested with *Kpn*I/*Sal*I. After purified by DNA Purification Kit (Invitrogen), the fragments were cloned into vector pCambia1300, resulting in the overexpressing

construct p35S:miR156h. The plasmids were introduced into the Nipponbare background via *Agrobacterium* (strain GV3101)-mediated transformation. Positive transformants were screened following a previous report (Li et al., 2014).

Disease Assay and Microscopy Analysis

Three *M. oryzae* strains, including Zhong10-8-14-GFP (GZ8), 97-27-2, and FJ08-09-1, were used, depending on the availability at the time for inoculation. GZ8 and 97-27-2 were used for punch/spray-inoculation assay, and GZ8 was used for rice sheath inoculation assay in transgenic rice plants in Nipponbare background; FJ08-09-1 was incompatible with resistance accession IRBLkm-Ts and used for spray-inoculation assay in susceptible accession LTH and IRBLkm-Ts. For sporulation, *M. oryzae* strains were cultured in complete medium and oatmeal and tomato media (OTA) at 28°C with 16 h/8 h light/dark. For disease assays via spray-inoculation, the spores of the indicated strains were collected and adjusted to 1×10^5 spores mL⁻¹ to spray onto three-leaf seedlings. The inoculated seedlings were incubated in darkness with 100% humidity for 24 h and then put in an air-conditioned growth room (Li et al., 2017). Leaves were collected at 12 and 24 h post-inoculation (hpi) for analysis on the expression of defense-related genes. For disease assays via punch-inoculation, the collected spores (3×10^5 spores mL⁻¹) were drop-inoculated on the punch-wound rice leaves and cultured in buffer containing 0.1% 6-Benzylaminopurine (6-BA). The leaves were incubated in darkness for 24 h, and then continually incubated in 100% humidity and 12 h/12 h light/dark conditions. The disease phenotypes via punch/spray-inoculation were recorded at 5 days post-inoculation (dpi) (Chandran et al., 2019; Zhao et al., 2019). Disease severity was measured by the relative fungal mass that was calculated using the DNA level of *MoPot2* against the rice *ubiquitin* gene via quantitative PCR (Li et al., 2017). For rice-sheath inoculation assays, 7 cm-long leaf sheaths were injected with GZ8 (1×10^5 spores mL⁻¹) spore suspension (Kankanala et al., 2007). The epidermal layer from the inoculated leaf sheaths was analyzed by Laser Scanning Confocal Microscopy (LSCM) (Nikon A1) at 24 and 36 hpi, respectively.

Gene Expression Analyses

Total RNAs were extracted from rice leaves using TRIzol reagent (Invitrogen), according to the manufacturer's instructions. The first-strand cDNA was synthesized by using Primescript RT reagent Kit with gDNA Eraser, following the manufacturer's instructions [TaKaRa Biotechnology (Dalian) Co., Ltd., Japan]. RT-qPCR was performed by using specific primers and SYBR Green mix (SYBR Green PCR Kit, Bimake) with BIO-RAD C1000TM Thermal Cycler. Rice *ubiquitin* (*UBQ*) gene was used as an internal reference for data normalization. The total RNA was reverse-transcribed by using a specific stem-loop RT primer (miR156 stem-loop RT primer) to examine the expression of mature miR156. The RT products were subsequently used as the template for RT-qPCR by using a miR156-specific forward primer and a universal reverse primer. The stem-loop primers were designed as described previously (Chen et al., 2005). The

snRNA U6 was used as an internal reference for the detection of miRNAs. RT-qPCR was carried out by using the indicated primers (Supplementary Table S1).

Transient Expression Assay in *N. benthamiana*

The *Agrobacterium* strain GV3101 carrying the individual construct in the binary vector pCambia1300 was incubated in LB media containing specific antibiotics (rifampin, 50 µg/mL; kanamycin, 50 µg/mL) at 28°C incubation. The bacteria were collected at 5,000 rpm for 5 min and resuspended in MMA buffer (10 mM MES, 10 mM MgCl₂, and 100 µM AS). The *Agrobacteria* containing the expression constructs were infiltrated into leaves of *N. benthamiana* for transient expression assay. Leaves were examined for image acquisition by using the Zeiss fluorescence microscope (Zeiss imager A2) between 36 and 72 hpi. Western blot analyses were performed following a previous protocol (Li et al., 2017). Total proteins were extracted from equal amounts of leaves with the protein extraction buffer with 1 × loading buffer (0.05 mg/mL Bromophenol blue; 0.065 M Tris-HCl, pH 6.8; 0.02 g/mL SDS; 0.05 mL/mL 2-mercaptoethanol; 0.1 mL/mL glycerol; 1 × protease inhibitor cocktail EDTA-free, Bimake, b14002). Total proteins were separated by 10% SDS-PAGE and transferred to PVDF membrane (Millipore) by using Trans-Blot Turbo (BIO-RAD). Protein blot was hybridized with the rabbit anti-GFP to determine YFP accumulation, and membranes were stained with Ponceau S as the loading control.

Agronomic Traits Measurement

The rice plants for agronomic traits assay were planted in a paddy field in Chengdu, China (36° N, 103° E), during the rice-growing season from April to September. Five representative plants for each line were sampled for measurement of agronomic traits, including plant height, number of tillers, panicle length, number of primary rachis branches (PBs) per panicle, number of secondary rachis branches (SBs) per panicle, number of seeds per panicle, seed length, seed width, 1,000-grain weight, and grain

yield per plant on a SC-G seed-counting and grain-weighing device (Wanshen Ltd., Hangzhou, China). The data were analyzed by a one-way ANOVA followed by *post hoc* Tukey HSD analysis with significant differences ($P < 0.05$) by using SPSS statistics software.

RESULTS

miR156fhl-3p Is Responsive to Chitin Treatment and *M. oryzae* Infection

Rice genome contains twelve *MIR156* genes generating three mature miR156-5p isoforms and four miR156-3p isoforms (Supplementary Figure S1A)¹. One major miR156-5p isoform was transcribed from *Osa-miR156a*, *Osa-miR156b*, *Osa-miR156c*, *Osa-miR156d*, *Osa-miR156e*, *Osa-miR156f*, *Osa-miR156g*, *Osa-miR156h*, *Osa-miR156i*, and *Osa-miR156j* (designated miR156-5p); whereas four miR156-3p isoforms were respectively transcribed from *Osa-miR156f*, *Osa-miR156h*, and *Osa-miR156l* (designated miR156fhl-3p), *Osa-miR156c* and *Osa-miR156g* (designated miR156cg-3p), *Osa-miR156b* (miR156b-3p), and *Osa-miR156j* (miR156j-3p) (Supplementary Figure S1A). In addition, *Osa-miR156f* and *Osa-miR156h* generated an identical pre-miRNA (designated pre-miR156fh). In our previous work, we detected the different expression patterns of miRNAs in resistance accession IRBLkm-Ts and susceptible accession LTH following the *M. oryzae* induction by high-throughput RNA-Seq assay. The abundance of miR156fhl-3p was remarkably higher than that of miR156b-3p, miR156cg-3p, and miR156j-3p, and all of their abundances were significantly lower in resistance accession IRBLkm-Ts than in susceptible accession LTH (Supplementary Figure S2; Li et al., 2014). Therefore, we focused on miR156fhl-3p in this study. To confirm miR156fhl-3p is responsive to the fungal pathogen, we conducted a time-course analysis of the accumulation pattern of miR156fhl-3p in LTH following chitin treatment, as well as in LTH and IRBLkm-Ts

¹www.mirbase.org

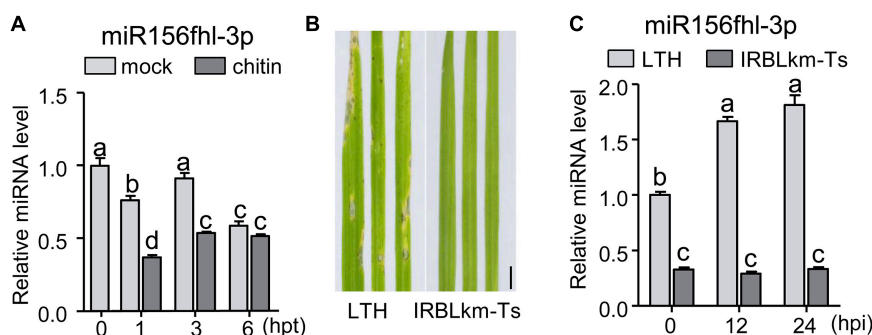


FIGURE 1 | miR156fhl-3p is differentially responsive to chitin and *M. oryzae*. **(A)** Expression pattern of miR156fhl-3p upon chitin treatment. **(B)** Representative leaf sections from LTH and IRBLkm-Ts at 5 dpi upon *M. oryzae* strain FJ08-09-1. Scale bars, 1 cm. **(C)** Expression pattern of miR156fhl-3p in LTH and IRBLkm-Ts following *M. oryzae* inoculation. RNAs were extracted from leaves at the indicated time points. miR156fhl-3p was analyzed using Stem-loop RT-qPCR. SnRNA U6 served as the internal reference. Error bars indicate standard deviation (SD) ($n = 3$). Different letters above the bars indicate significant differences ($P < 0.05$) determined by one-way ANOVA analysis followed by *post hoc* Tukey HSD analysis. These experiments were repeated two times with similar results.

following infection of an *M. oryzae* strain FJ08-09-1 (**Figure 1B**). Chitin is a component of the fungal cell wall and can induce immune responses (Kawano and Shimamoto, 2013; Li et al., 2014; Kawasaki et al., 2017). Upon application of chitin, the abundance of miR156fhl-3p was significantly decreased at 1 and 3 h post-treatment (hpt) in comparison with mock treatment (**Figure 1A**), indicating that miR156fhl-3p is responsive to chitin treatment. In contrast, upon *M. oryzae* inoculation, the amount of miR156fhl-3p was significantly increased in susceptible accession LTH, and unchanged in IRBLkm-Ts at 12 and 24 hpi (**Figure 1C**). However, the amount of miR156fhl-3p was significantly less in resistance accession IRBLkm-Ts

than that in susceptible accession LTH (**Figure 1C**). These results suggested that miR156fhl-3p is responsive to chitin and *M. oryzae*.

Blocking miR156fhl-3p via a Target Mimic Compromises Rice Susceptibility to *M. oryzae*

To detect the roles of miR156fhl-3p in rice resistance to blast disease, we constructed transgenic lines overexpressing a target mimic of miR156fhl-3p (MIM156-3p), which was complementary to miR156fhl-3p with 3-nt insertion between

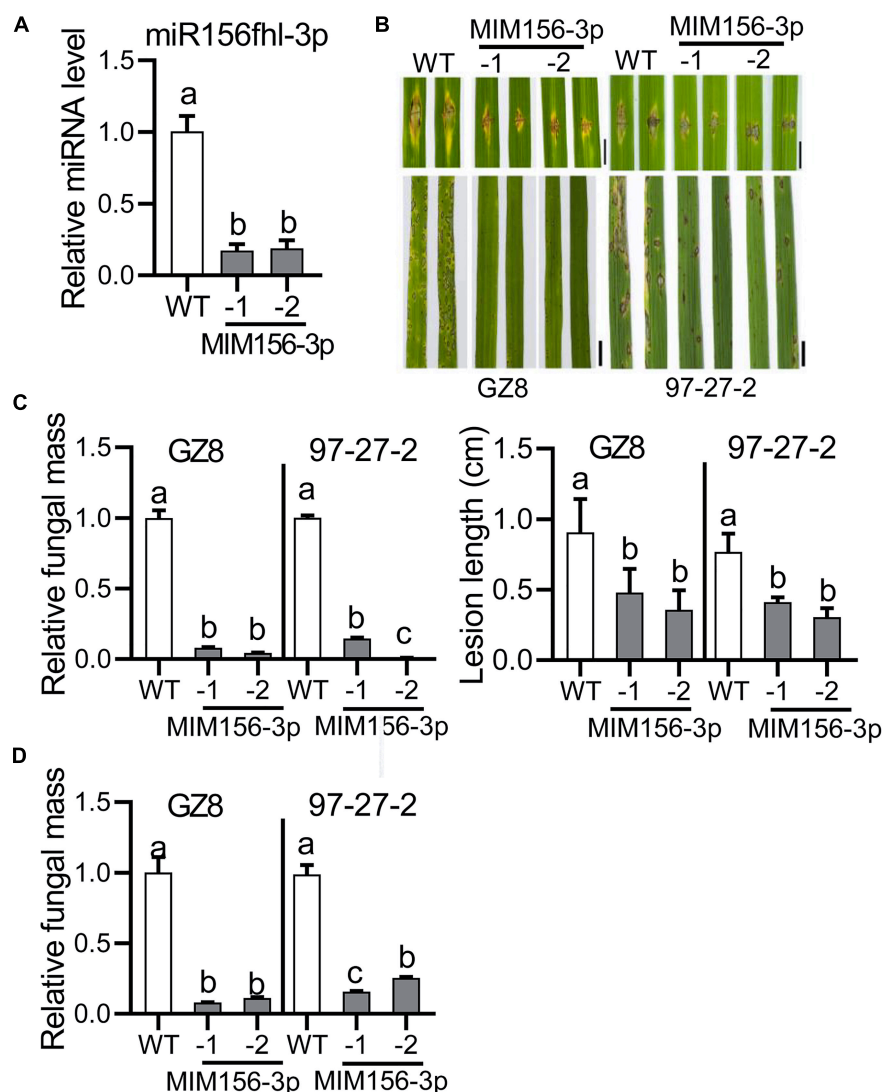


FIGURE 2 | Expression of a target mimic of miR156fhl-3p leads to compromised susceptibility to *M. oryzae*. **(A)** Relative amounts of miR156fhl-3p in transgenic lines expressing a target mimic of miR156fhl-3p (MIM156-3p) and WT plants. The amounts of mature miR156fhl-3p were analyzed using stem-loop RT-qPCR. **(B)** Disease phenotype of control and MIM156-3p lines. Three-leaf-stage seedlings were punch/spray-inoculated with *M. oryzae* strains GZ8 and 97-27-2. Leaf phenotypes were observed at 5 dpi. Scale bars: Punch-inoculation assays are 5 mm; spray-inoculation assays are 1 cm. **(C)** Relative fungal mass represented by measuring *MoPot2* gene against *OsUbiquitin* DNA level and average lesion length via punch-inoculation. **(D)** Relative fungal mass via spray-inoculation. Error bars indicate SD ($n = 3$). Different letters above the bars indicate significant differences ($P < 0.05$) determined by one-way ANOVA analysis followed by *post hoc* Tukey HSD analysis. These experiments were repeated two times with similar results.

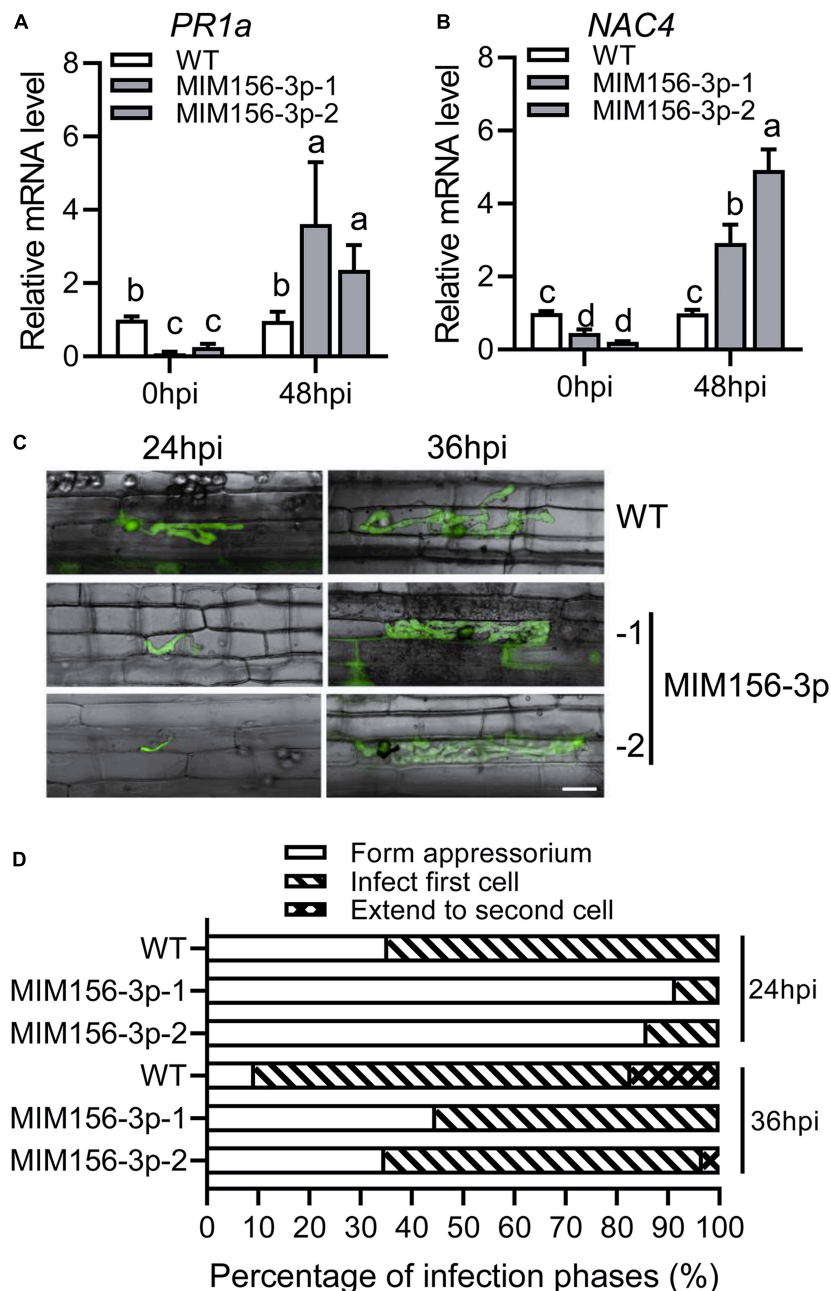


FIGURE 3 | Expression of a target mimic of miR156fhl-3p improves the induction of defense-related genes and delays the *M. oryzae* infection process. **(A,B)** Induction of defense-related genes *PR1a* **(A)** and *NAC4* **(B)** upon *M. oryzae* infection in WT and MIM156-3p lines. Error bars indicate SD ($n = 3$). Different letters above the bars indicate significant differences ($P < 0.05$), as determined by one-way ANOVA analysis followed by *post hoc* Tukey HSD analysis. **(C)** Representative LSCM images show the development of *M. oryzae* strain GZ8 at 24 and 36 hpi in sheath cells of WT and MIM156-3p lines. Scale bars, 20 μ m. **(D)** Statistical analysis of *M. oryzae* growth in **(C)**. More than 200 conidia in each line were analyzed. These experiments were repeated two times with similar results.

position 10 and 11 nts and acted as a sponge to hold miR156fhl-3p (Franco-Zorrilla et al., 2007; **Supplementary Figure S1B**). Out of seven lines, we selected two lines for further investigation, namely MIM156-3p-1 and MIM156-3p-2, which displayed significantly decreased miR156fhl-3p accumulation in comparison with control Nipponbare (WT) plants (**Figure 2A**). Both transgenic lines showed much smaller disease lesions and

less fungal mass than control via punch-inoculation of the *M. oryzae* strain GZ8 and 97-27-2 (**Figures 2B,C**). Consistently, both transgenic lines displayed compromised susceptibility with smaller disease lesions and less fungal growth following spray-inoculation with GZ8 and another strain 97-27-2 (**Figures 2B,D**). These data indicate that blocking miR156fhl-3p compromises rice susceptibility to *M. oryzae*.

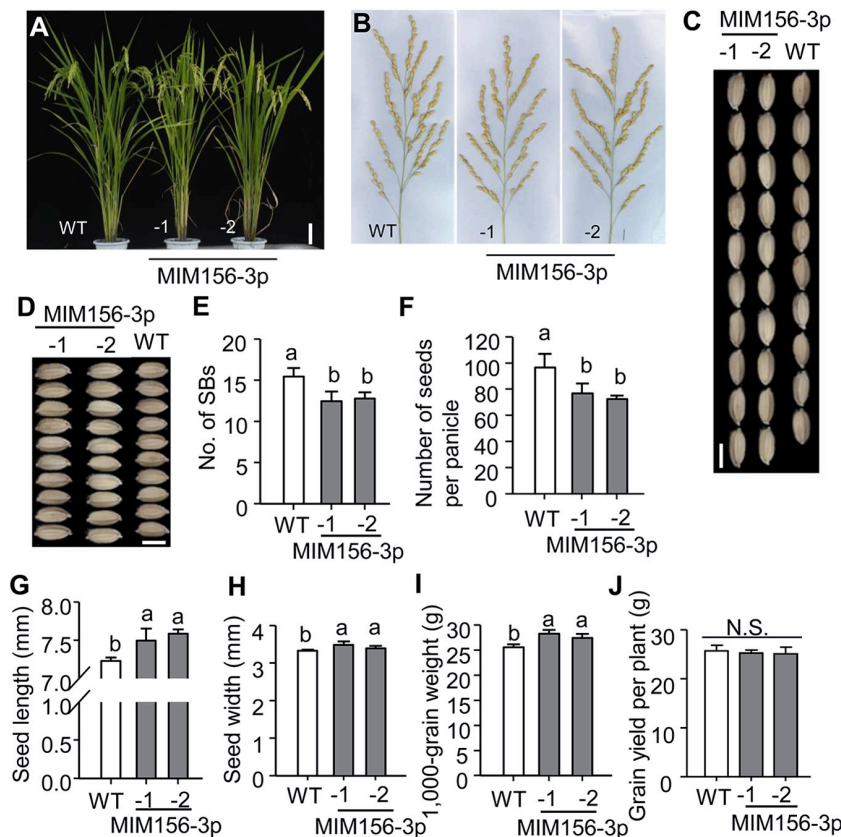


FIGURE 4 | Blocking miR156fhl-3p leads to changed rice agronomic traits without effect on yield. (A) Gross morphology of WT and MIM156-3p lines. Size bar, 10 cm. (B) Panicle morphology of WT and MIM156 lines. Size bar, 1 cm. (C,D) Grain length and width of WT and MIM156-3p. Size bar, 5 mm. (E–J) Statistical analysis of the number of secondary rachis branches per panicle (No. of SBs) (E), the number of seeds per panicle (F), seed length (G), seed width (H), 1,000-grain weight (I), and grain yield per plant (J) of WT and MIM156-3p lines. Error bars indicate standard deviation ($n = 5$). Different letters above the bars indicate significant differences ($P < 0.05$) determined by one-way ANOVA analysis followed by *post hoc* Tukey HSD analysis.

Expression of a Target Mimic of miR156fhl-3p Delays *M. oryzae* Infection and Enhances the Induced Expression of Defense-Related Genes

To understand why blocking miR156fhl-3p leads to compromised susceptibility to blast disease, we examined the expression of the two defense-related response marker genes, *PR1a* and *NAC4*, upon *M. oryzae* strain GZ8 infection. The expression of both genes was induced to significantly higher levels by GZ8 in MIM156-3p than that in control at 48 hpi (Figures 3A,B). Then, we compared the infection process in leaf sheath of WT control and MIM156-3p plants with a LSCM after the inoculation of the GFP-tagged strain GZ8. The growth of GZ8 was delayed in MIM156-3p at 24 and 36 hpi in comparison with that in control (Figure 3C). Quantified infection phase assay showed that more than 60% of spores formed the invasive hyphae in local cells at the infection site in WT, whereas only 10% spores formed invasive hyphae in the local cells in MIM156-3p (Figure 3D). At 36 hpi, more than 70% of hyphae infected into the first cell, and 17% of hyphae extend to the neighbor cells in control. In contrast, about 45% of spores formed the

appressorium, and 55% spores formed invasive hyphae and infected into the local cells in MIM156-3p (Figure 3D). These observations suggested that blocking miR156fhl-3p delayed the infection process of *M. oryzae* and promoted the induction of defense-related genes.

Expression of a Target Mimic of miR156fhl-3p Does Not Penalize Agronomic Traits

Activation of blast resistance often penalizes growth and agronomic traits that lead to yield formation (Heil and Baldwin, 2002; Walters and Boyle, 2005; Shimono et al., 2007; Walters and Heil, 2007). To detect whether blocking miR156fhl-3p via target mimicry affected rice growth, we assayed the agronomic traits of MIM156-3p plants. Intriguingly, MIM156-3p showed similar gross plant morphology as WT (Figures 4A,B). Statistical analysis indicated that several agronomic traits were slightly but not significantly decreased in MIM156-3p in comparison with those in WT (Supplementary Figure S3), including plant height, number of tillers per plant, panicle length, and number of primary rachis branches per panicle (No. of

PBs). The number of secondary rachis branches per panicle (No. of SBs) was significantly decreased in MIM156-3p plants (Figure 4E), leading to a reduced number of grains per panicle (Figure 4F). In contrast, the seed length (Figures 4C,G) and width (Figures 4D,H) in MIM156-3p were increased, resulting in an increased 1,000-grain weight (Figure 4I). As a result, the grain yield per plant was unchanged in MIM156-3p (Figure 4J) in comparison with that in WT. Altogether, our results indicated that blocking miR156fhl-3p could improve rice blast disease resistance without penalty of yield.

Expression of a Target Mimic of miR156fhl-3p Suppresses miR156-5p Accumulation and Increases *SPL14* Expression

To uncover how miR156fhl-3p regulated rice resistance against *M. oryzae*, we first predicted potential target genes of miR156fhl-3p by using psRNATarget², an online tool for prediction of miRNA targets. We chose two best-matched candidate genes, namely *LOC_Os02g17280* and *LOC_Os07g37580*, for examination (Supplementary Figure S4A). However, we failed to detect changes in their expressions in both MIM156-3p lines (Supplementary Figures S4B,C), indicating that the two candidates were possibly not regulated by miR156fhl-3p at mRNA level. We then tested whether the expression of MIM156-3p affected the accumulation of miR156-5p. Surprisingly, the abundance of mature miR156-5p was significantly decreased in MIM156-3p in comparison with that in WT (Figure 5A). Consistently, the abundance of pre-miR156fh was also significantly decreased in MIM156-3p in comparison with that in WT (Figure 5B). These data indicate that the blocking of miR156fhl-3p by MIM156fhl-3p could suppress the accumulation of miR156-5p. Previously, miR156-5p was reported to regulate the expression of *SPL14* (Jiao et al., 2010). The *SPL14* protein is phosphorylated during *M. oryzae* infection, and the phosphorylated *SPL14* binds to the promoter of *WRKY45* to promote its expression. In turn, *WRKY45* positively regulates rice immunity against *M. oryzae* (Shimono et al., 2007; Wang et al., 2018). We speculated that the expression of *SPL14* and *WRKY45* should be altered in MIM156-3p lines. As expected and reversely to the decreased accumulation of miR156, the mRNA amounts of *SPL14* and *WRKY45* were significantly higher in MIM156-3p than in control, and the expression of *WRKY45* was further significantly increased compared to that in control at 48 hpi of *M. oryzae* (Figures 5C,D). These results implied that the mRNA abundance of *SPL14* could be regulated by miR156fhl-3p via interference with miR156-5p, thus boosting the expression of *WRKY45* to enhance rice blast disease resistance.

To confirm the interference of miR156fhl-3p with miR156-5p, we established a YFP-based reporter system. In the system, we made a construct expressing a YFP reporter that was fused with the miR156 binding site of *SPL14* (*SPL14*_{TBS}) at the N-terminus (*SPL14*_{TBS}-YFP) and a construct expressing a mutant target

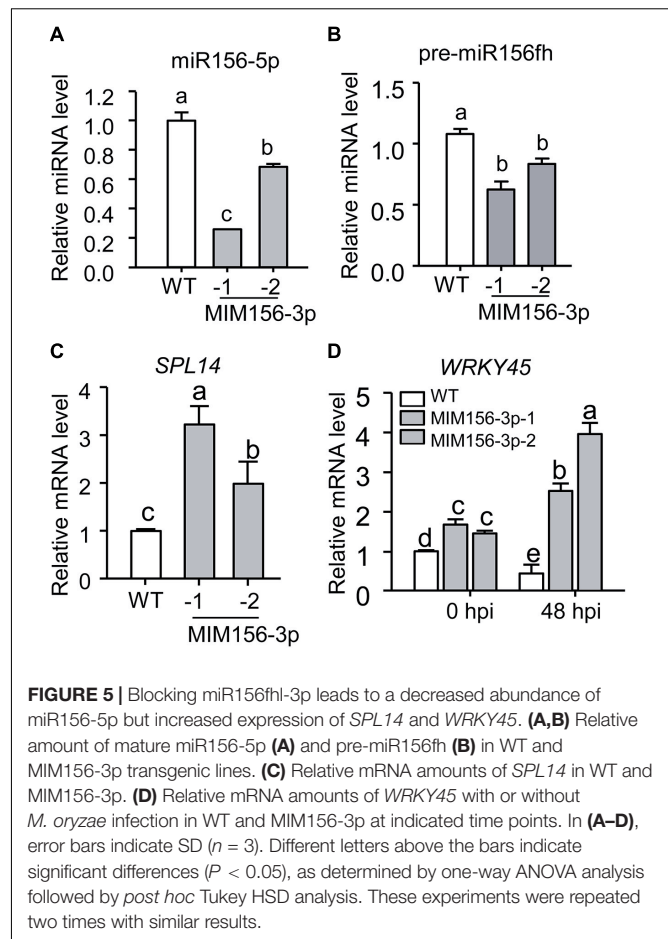


FIGURE 5 | Blocking miR156fhl-3p leads to a decreased abundance of miR156-5p but increased expression of *SPL14* and *WRKY45*. **(A,B)** Relative amount of mature miR156-5p **(A)** and pre-miR156fh **(B)** in WT and MIM156-3p transgenic lines. **(C)** Relative mRNA amounts of *SPL14* in WT and MIM156-3p. **(D)** Relative mRNA amounts of *WRKY45* with or without *M. oryzae* infection in WT and MIM156-3p at indicated time points. In **(A–D)**, error bars indicate SD ($n = 3$). Different letters above the bars indicate significant differences ($P < 0.05$), as determined by one-way ANOVA analysis followed by *post hoc* Tukey HSD analysis. These experiments were repeated two times with similar results.

site of *SPL14* (*SPL14*_{MBS}-YFP). *SPL14*_{TBS}-YFP was separately expressed or co-expressed with miR156 and MIM156-3p in *N. benthamiana* leaves. Then the YFP reporter was examined by YFP intensity and Western blotting. The YFP intensity and protein accumulation were highly accumulated when *SPL14*_{TBS}-YFP was expressed alone. In contrast, the YFP intensity and protein accumulation were significantly decreased when co-expressed with miR156 (Figures 6A,B), indicating suppression of miR156 on *SPL14*_{TBS}-YFP. Intriguingly, the YFP intensity and protein accumulation were recovered when MIM156-3p was added in the co-transient expression (Figures 6A,B). In contrast, the YFP intensity and protein accumulation of *SPL14*_{MBS}-YFP kept unchanged when co-expressed with miR156 or with both miR156 and MIM156-3p (Figures 6A,B). These data imply that MIM156-3p could release the suppression of *SPL14* by miR156 via interference with miR156-5p, leading to activation of *SPL14*.

miR156 Negatively Regulates Rice Blast Resistance

As MIM156-3p plants exhibited enhanced blast resistance via suppressing the accumulation of miR156-5p, we speculated that miR156-5p acts in blast resistance and should be responsive

²<http://plantgrn.noble.org/psRNATarget/>

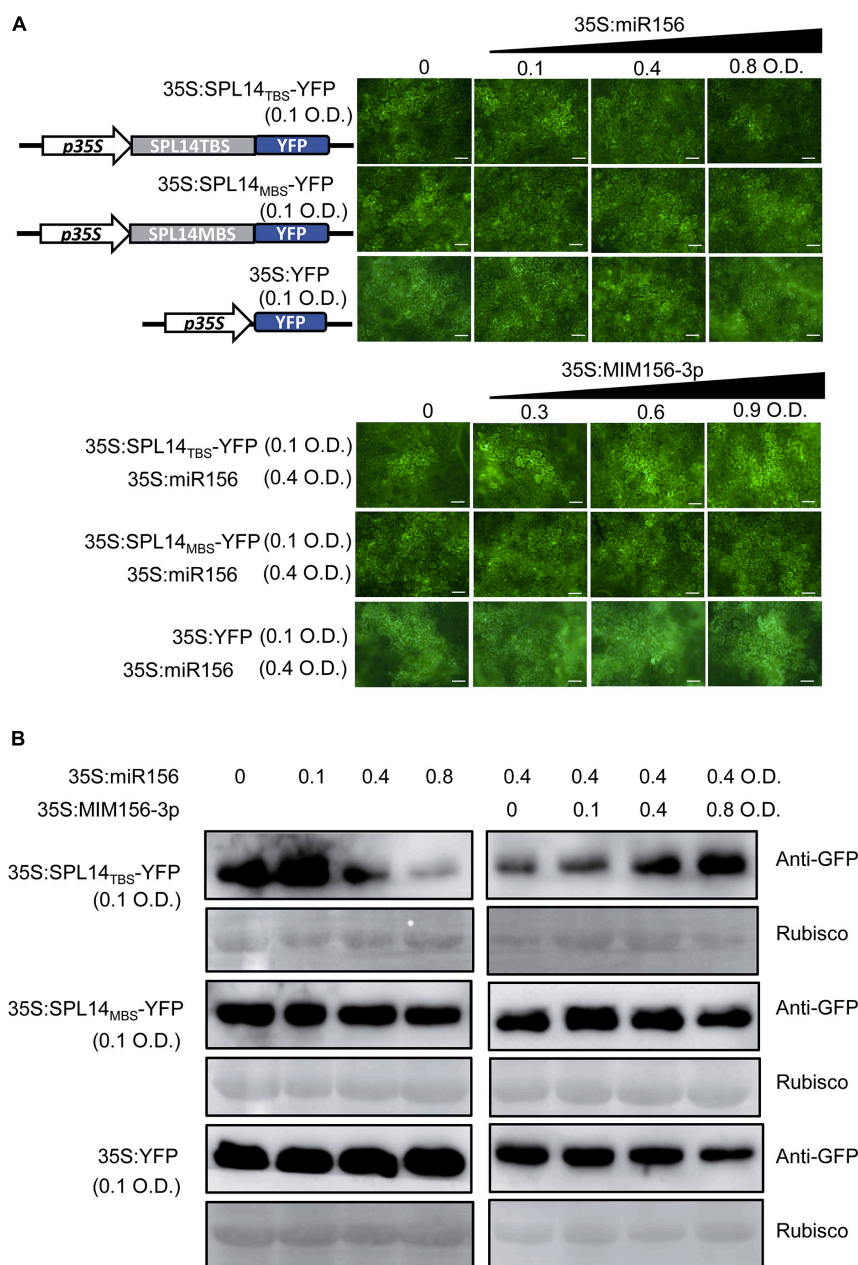


FIGURE 6 | Blocking miR156fhl-3p releases *SPL14* from suppression by miR156-5p. **(A)** Confocal images show the protein levels of SPL14_{TBS}-YFP, SPL14_{MBS}-YFP, and YFP co-expressed with miR156 or with miR156 and MIM156-3p. The indicated SPL14_{TBS}-YFP, SPL14_{MBS}-YFP, and YFP reporter constructs were transiently expressed alone or co-expressed with miR156 and MIM156-3p in *N. benthamiana* leaves by *Agrobacterium*-mediated transformation at the indicated optical density (O.D.) concentration. Size bars, 50 μ m. **(B)** Western blotting analysis shows the protein abundance of SPL14_{TBS}-YFP, SPL14_{MBS}-YFP, and YFP. Protein extracts from infiltrated leaves were collected for Western blot analysis using anti-GFP sera and anti-Rubisco sera, respectively. These experiments were repeated two times with similar results.

to *M. oryzae*. To confirm this hypothesis, we first examined the expression pattern of miR156-5p following chitin treatment in LTH and following *M. oryzae* invasion in LTH and IRBLkm-Ts. We compared the mock and chitin treatment in inducing miR156 expression. The abundance of miR156-5p was increased at 3 and 6 hpt in comparison with mock treatment (**Figure 7A**), indicating miR156-5p is responsive

to chitin. Moreover, miR156-5p was significantly increased in LTH upon *M. oryzae* infection, whereas it was decreased at 12 and 24 hpi in IRBLkm-Ts (**Figure 7B**), implying a negative role in pathogen-induced immunity. Then, we constructed transgenic lines overexpressing miR156h (OX156), which formed the same mature sequence as miR156abcdegij (miR156-5p) (**Supplementary Figure S1**). OX156 plants showed

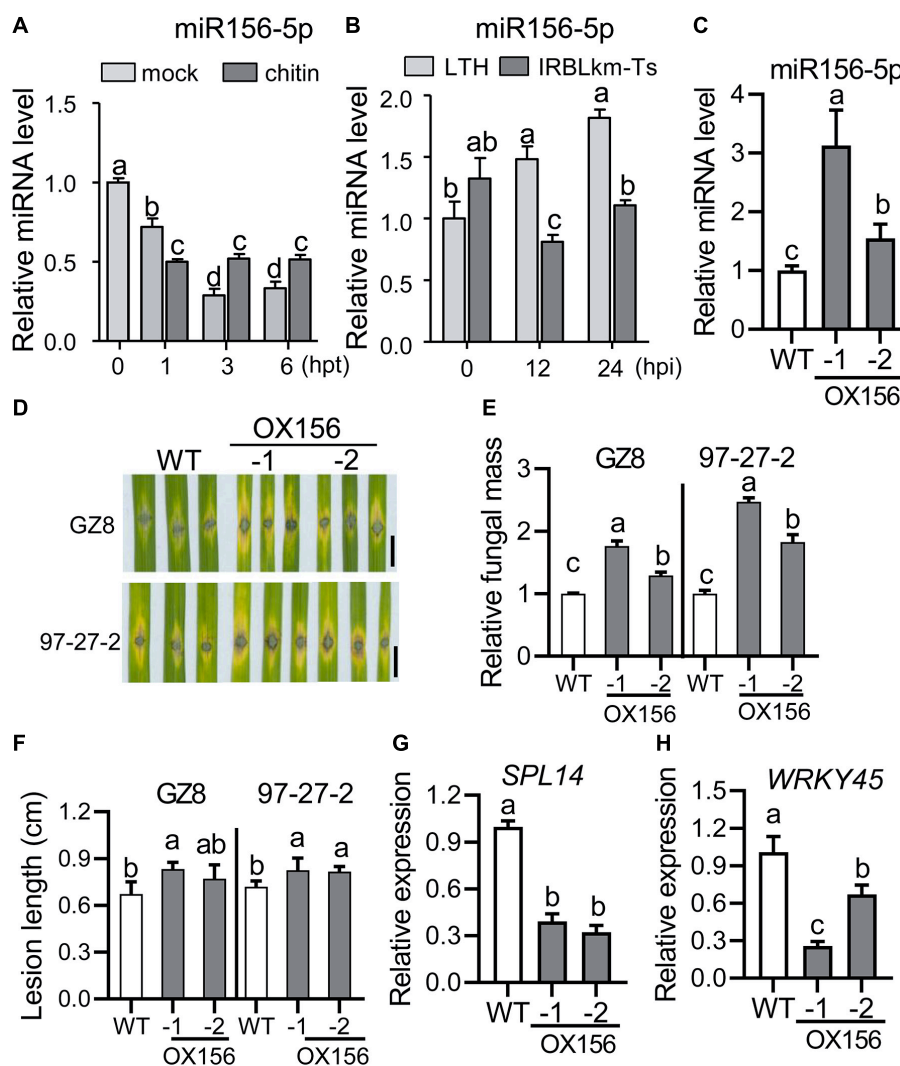


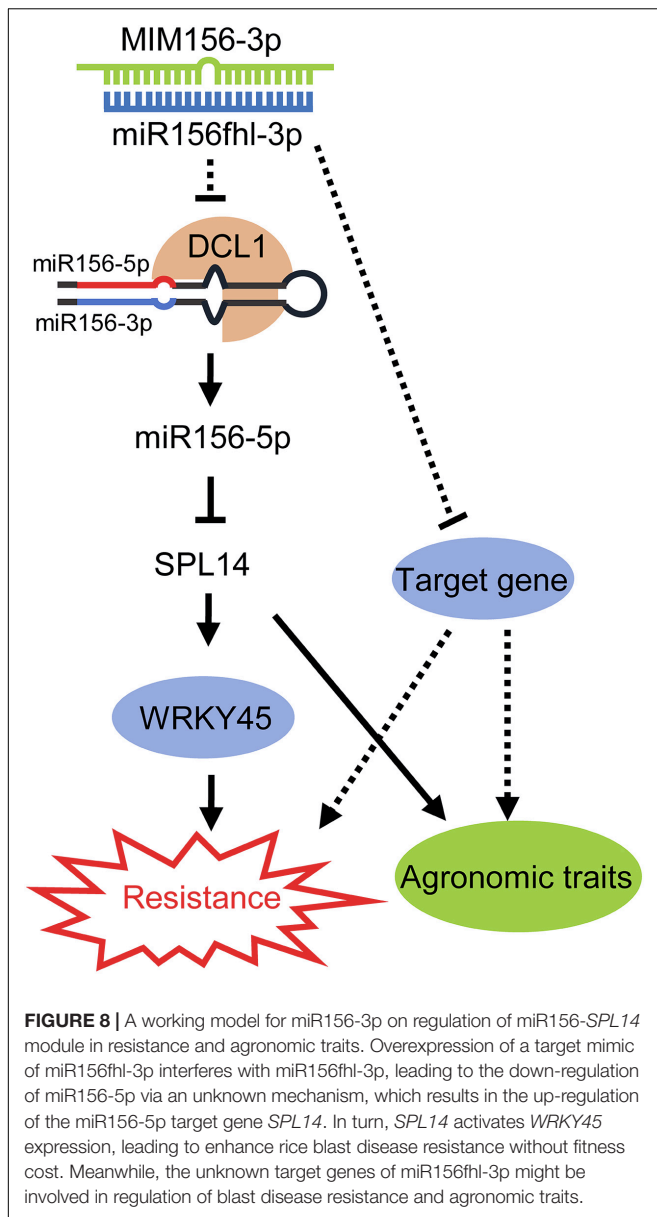
FIGURE 7 | Overexpression of miR156h results in enhanced susceptibility to *M. oryzae*. **(A)** Expression pattern of miR156-5p upon chitin treatment. **(B)** Expression pattern of miR156-5p in LTH and IRBLkm-Ts following *M. oryzae* inoculation. **(C)** Relative mRNA amount of miR156-5p in WT and OX156 lines. RNAs were extracted from leaves of three-leaf-stage plants from the indicated lines. The amounts of mature miR156 were analyzed using Stem-loop RT-qPCR. SnRNA U6 served as the internal reference. **(D)** Disease phenotypes of the indicated lines were observed at 5 dpi of *M. oryzae* strains GZ8 and 97-27-2. Scale bars, 5 mm. **(E)** Relative fungal mass represented by measuring *MoPot2* gene against *OsUbiquitin* DNA level. **(F)** Average lesion length. **(G,H)** RT-qPCR analysis of relative mRNA amounts of *SPL14* **(F)** and *WRKY45* **(G)** in WT and OX156 lines. For **(A–C,E–H)**, error bars indicate standard deviation ($n = 3$). Different letters above the bars indicate significant differences ($P < 0.05$) determined by one-way ANOVA analysis followed by *post hoc* Tukey HSD analysis.

remarkably decreased plant height (Supplementary Figure S6A) and other agronomic traits (Supplementary Figures S6B–J), as previously reported (Wang L. et al., 2015). We chose two transgenic lines, OX156-1 and OX156-2, for further study (Figure 7C and Supplementary Figure S6A). Both transgenic lines showed enhanced susceptibility to *M. oryzae* strain GZ8 and 97-27-2, with more significant disease lesions (Figures 7D,F) and fungal mass (Figure 7E) in comparison with control via punch-inoculation. The disease lesions in OX156-1 were even bigger than that in OX156-2 plants, which was consistent with the more accumulation of miR156-5p in OX156-1 than that in OX156-2 (Figures 7C–F). Furthermore, we found that the mRNA amounts of *SPL14*

and *WRKY45* were decreased in OX156 lines (Figures 7G,H), indicating that miR156-5p could facilitate the growth of *M. oryzae* in a dosage-dependent manner via suppressing the expression of *SPL14*, which in turn down-regulates *WRKY45* expression.

DISCUSSION

Previously, miRNA-3p was marked as miRNA* that was considered to be a side-product subjected to decay (Jones-Rhoades et al., 2006; Krol et al., 2010). Later on, miRNA-3p was found to target genes different from those of miRNA-5p



and thus performed different fine-tuning functions (Navarro et al., 2006; Zhang et al., 2011). Here, we found that expressing a target mimic of miR156fhl-3p affects the abundance of miR156-5p, which in turn regulates a target gene of miR156-5p to fine-tune the tradeoffs between blast disease resistance and yield traits.

Two techniques provide efficient tools to block the activity of a mature miRNA. One is target mimicry technology that is established based on the modulation of *IPS1* on the miR399 activity in *Arabidopsis*. The other is called short tandem target mimicry (STTM), which exploits two target mimics of a miRNA to increase the efficacy of a target mimic in blocking the function of a miRNA (Yan et al., 2012; Zhang et al., 2017). In our previous study, we found that the abundance of miR156fhl-3p was lower

in a resistance accession than in a susceptibility accession; still, its abundance was much more than those of miR156b-3p, miR156cg-3p, and miR156j-3p (Supplementary Figure S2; Li et al., 2014). Here, we exploited target mimicry to explore the function of miR156fhl-3p. Transgenic MIM156-3p plants showed compromised susceptibility to *M. oryzae* with less fungal mass and shorter lesion length than control (Figures 2B–D). Consistently, defense-related genes, such as *PR1a* and *NAC4*, were significantly up-regulated higher in MIM156-3p than in control (Figures 3A,B), and the fungal invasion process was obviously delayed in MIM156-3p plants (Figures 3C,D). Moreover, transgenic MIM156-3p plants also produced a grain yield comparable to that of control, which was built on the orchestration of the number of panicles, panicle size, and grain weight (Figure 4). Therefore, these data indicate that expressing a target mimic of miR156fhl-3p improves blast disease resistance without the yield penalty.

The expression of either a target mimic or an STTM of a miRNA can lead to a decrease in the abundance of the miRNA via an unknown mechanism (Zhang et al., 2017). Here, we found that the abundance of miR156fhl-3p was significantly down-regulated in MIM156-3p lines (Figure 2A). Surprisingly, the abundance of the mature miR156-5p was also significantly down-regulated (Figure 5A), which represents the transcripts from *Osa-miR156a*, *Osa-miR156b*, *Osa-miR156c*, *Osa-miR156d*, *Osa-miR156e*, *Osa-miR156f*, *Osa-miR156g*, *Osa-miR156h*, *Osa-miR156i*, and *Osa-miR156j* (Supplementary Figure S1). It is well-known that a *MIR* gene is transcribed into a pri-miRNA that is processed into the precursor of a mature miRNA, namely a pre-miRNA (Rogers and Chen, 2013). A pre-miRNA folds into a hairpin that is cleaved by a DCL protein and exported to the cytoplasm, where the DCL processes it into a miRNA-5p/miRNA-3p duplex (Liu et al., 2005). One or both of the strands of the miRNA-5p/miRNA-3p duplex are incorporated into AGO proteins to mediate the silencing of target genes via sequence complementarity (Zhu et al., 2011; Kobayashi and Tomari, 2016). We also found the down-regulation of pre-miR156fhl in MIM156-3p lines (Figure 5B), indicating that the target mimic of miR156fhl-3p may interfere with the expression of miR156 family genes at either the transcription or the procession stage. Previously, the expression of *SPL14* in a mutant, i.e., *ipa1* (*Ideal Plant Architecture1*) that carries a mutation at the miR156 target site, was not influenced by the infection of *M. oryzae*; instead, the phosphorylation of IPA1 was changed during *M. oryzae* infection, leading to improvement of both blast disease resistance and yield (Wang et al., 2018). We confirmed such interference of a -3p with -5p using an *SPL14*_{TBS}-YFP-based reporter system, through which we confirmed that the overexpression of a target mimic of miR156fhl-3p interfered with miR156-5p miRNA level, which did release the suppression of miR156-5p on the *SPL14*_{TBS}-YFP accumulation (Figure 6). Consistently, the expression of both *SPL14* and its downstream gene *WRKY45* was increased in MIM156-3p (Figures 5C,D). However, we currently could not tell whether the inference

of a -3p with a -5p occurs at the transcription or the post-transcription step. Alternatively, the phenotypes observed in MIM156-3p could be due to altered expression of the target genes of miR156fhl-3p. We searched for the target genes of miR156fhl-3p through website prediction. Among these candidate genes, we selected two genes for expression analysis (**Supplementary Figure S4A**) and found that their expressions were not significantly altered in MIM156-3p (**Supplementary Figure S4B**). It is unclear whether the target genes of miR156fhl-3p are involved in regulating rice resistance and agronomic traits. Therefore, future work should focus on searching miR156fhl-3p target genes and exploit Northern blotting to examine the abundance of each pri-miR156 and pre-miR156 to figure out which step is impacted by the target mimic of miR156fhl-3p. Also, future research should focus on addressing whether such interference of a -3p with a -5p is specific to *MIR156* genes or a common phenomenon among different miRNAs.

MIR156 family genes were classified into two groups based on their functions in regulating rice agronomic traits. Group I includes miR156d, miR156e, miR156f, miR156g, miR156h, miR156i, and miR156j, acting in controlling grain size (Miao et al., 2019); group II includes miR156a, miR156b, miR156c, miR156l, and miR156k, contributing to shoot architecture (**Supplementary Figure S5**; Miao et al., 2019). Intriguingly, miR156a, miR156b, miR156c, miR156d, miR156e, miR156f, miR156g, miR156h, miR156i, and miR156j are transcribed into the same -5p mature sequences (**Supplementary Figure S1**). miR156fh is in group I, and miR156l is in group II (**Supplementary Figure S5**). Transgenic OX156 lines overexpressing miR156h showed dwarfish phenotypes that were similar to previous reports (**Supplementary Figure S6**; Xie et al., 2006; Dai et al., 2018; Liu et al., 2019). In contrast to MIM156-3p, which exhibited enhanced blast resistance, OX156 showed enhanced susceptibility, which may be due to the down-regulation of *SPL14* and *WRKY45* (**Figure 7**). Because miR156h represents the major and the most abundant isoform of miR156, these data indicate that miR156 negatively regulates blast disease resistance via the *SPL14-WRKY45* transcription factor module.

Taken together, we propose a model to summarize our discovery (**Figure 8**). The expression of a miR156fhl-3p target mimic interferes with the accumulation of both miR156fhl-3p and miR156-5p by an unknown mechanism. Previously, miR156-5p was reported to regulate *SPL14* expression via mRNA cleavage or transcription suppress (Jiao et al., 2010). Overexpression of *SPL14* enhances rice resistance against *M. oryzae* (Wang et al., 2018). Here, miR156-5p was down-regulated in MIM156-3p plants, leading to the up-regulation of *SPL14*, which in turn promotes the expression of *WRKY45* to enhance resistance against *M. oryzae*. However, we could not exclude that the target genes of miR156fhl-3p might contribute to rice resistance to *M. oryzae* and be involved in the regulation of agronomic traits from our current data. In conclusion, our findings demonstrate another layer of regulation on *SPL14* expression by miR156 in fine-tuning the tradeoffs between blast disease resistance and yield.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

W-MW and YL conceived the study. L-LZ, Y-PZ, HW, S-XZ, L-FW, XY, J-FC, X-PL, and X-CM performed the experiment. J-QZ, MP, and HF carried out the field trial. JF, J-WZ, and Y-YH guided the experimental operation. YL, L-LZ, and W-MW contributed to manuscript writing and editing.

FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. 31430072 and 31672090 to W-MW).

ACKNOWLEDGMENTS

We thank Dr. Cai-Lin Lei (Institute of Crop Science, Chinese Academy of Agricultural Sciences) for providing the monogenic resistant lines IRBLkm-Ts.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00327/full#supplementary-material>

FIGURE S1 | Schematic diagram of miR156 isoforms and artificial mimicry target of miR156fhl-3p. **(A)** Sequence alignment of the miR156-5p and miR156-3p isoforms. **(B)** Alignment of MIM156h-3p with miR156fhl-3p.

FIGURE S2 | RNA-Seq revealed the differential amounts of miR156-3p isoforms following *M. oryzae* inoculation at indicated time points.

FIGURE S3 | miR156fhl-3p has no significant effect on several agronomic traits. Statistics data of plant height **(A)**, number of tillers **(B)**, panicle length **(C)** number of primary rachis branches per panicle (No. of PBs) **(D)**, of WT and MIM156-3p lines. Error bars indicate SD ($n = 5$). Different letters above the bars indicate significant differences ($P < 0.05$) determined by one-way ANOVA analysis followed by *post hoc* Tukey HSD analysis.

FIGURE S4 | The predicted target genes of miR156fhl-3p are not up-regulated significantly in MIM156. **(A)** Sequence alignments of miR156fhl-3p and predict target sites of indicated genes. **(B,C)** RT-PCR analysis of the predicted target genes expression in WT and MIM156-3p plants. Error bars indicate SD ($n = 3$). Different letters above the bars indicate significant differences ($P < 0.05$) determined by one-way ANOVA analysis followed by *post hoc* Tukey HSD analysis. These experiments were repeated two times with similar results.

FIGURE S5 | Phylogenetic tree of rice *MIR156* gene based on DNA sequences. 100 bp genomic sequences (20 bp upstream sequences corresponding to mature miR156 and 20 bp downstream sequences) were used to construct the phylogenetic tree.

FIGURE S6 | miR156 regulates rice agronomic traits. **(A)** Gross morphology of WT and transgenic plants over-expressing miR156h (OX156). Scale bars, 10 cm. **(B-I)** Statistical analysis of plant height **(B)**, number of tillers **(C)**, panicle length

(D), number of primary rachis branches per panicle (No. of PBs) (E), number of second rachis branches per panicle (No. of SBs) (F), number of seeds per panicle (G), seed width (H), seed length (I), 1,000-grain weight (J) from the indicated lines. Error bars indicate SD ($n = 5$). Different letters above the bars indicate

significant differences ($P < 0.05$) determined by one-way ANOVA analysis followed by *post hoc* Tukey HSD analysis.

TABLE S1 | The primers used in this study.

REFERENCES

- Addoquaye, C., Eshoo, T. W., Bartel, D. P., and Axtell, M. J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Curr. Biol.* 18, 758–762. doi: 10.1016/j.cub.2008.04.042
- Axtell, M. J., and Bowman, J. L. (2008). Evolution of plant microRNAs and their targets. *Trends Plant Sci.* 13, 343–349. doi: 10.1016/j.tplants.2008.03.009
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297.
- Cardon, G. H., Hohmann, S., Klein, J., Nettesheim, K., Saedler, H., and Huijser, P. (1999). Molecular characterisation of the *Arabidopsis* SBP-box genes. *Gene* 237, 91–104. doi: 10.1016/S0378-1119(99)00308-X
- Chandran, V., Wang, H., Gao, F., Cao, X.-L., Chen, Y.-P., Li, G.-B., et al. (2019). miR396-OSGRFs module balances growth and rice blast disease-resistance. *Front. Plant Sci.* 9:1999. doi: 10.3389/fpls.2018.01999
- Chen, C., Ridzon, D. A., Broomer, A. J., Zhou, Z., Lee, D. H., Nguyen, J. T., et al. (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.* 33, e179. doi: 10.1093/nar/gni178
- Dai, Z., Wang, J., Yang, X., Lu, H., Miao, X., and Shi, Z. (2018). Modulation of plant architecture by the miR156f-OsSPL7-OsGH3.8 pathway in rice. *J. Exp. Bot.* 69, 5117–5130. doi: 10.1093/jxb/ery273
- Eulgem, T., and Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signaling. *Curr. Opin. Plant Biol.* 10, 366–371. doi: 10.1016/j.pbi.2007.04.020
- Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., et al. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39:1033. doi: 10.1038/ng2079
- Goto, S., Sasakura-Shimoda, F., Suetsugu, M., Selvaraj, M. G., Hayashi, N., Yamazaki, M., et al. (2015). Development of disease-resistant rice by optimized expression of WRKY45. *Plant Biotechnol. J.* 13, 753–765. doi: 10.1111/pbi.12303
- Goto, S., Sasakura-Shimoda, F., Yamazaki, M., Hayashi, N., Suetsugu, M., Ochiai, H., et al. (2016). Development of disease-resistant rice by pathogen-responsive expression of WRKY45. *Plant Biotechnol. J.* 14, 1127–1138. doi: 10.1111/pbi.12481
- Guo, L., Zhang, H., Zhao, Y., Yang, S., and Chen, F. (2014). Selected isomiR expression profiles via arm switching? *Gene* 533, 149–155. doi: 10.1016/j.gene.2013.09.102
- Hao, J., Tu, L., Hu, H., Tan, J., Deng, F., Tang, W., et al. (2012). GbTCP, a cotton TCP transcription factor, confers fibre elongation and root hair development by a complex regulating system. *J. Exp. Bot.* 63, 6267–6281. doi: 10.1093/jxb/ers278
- Heil, M., and Baldwin, I. T. (2002). Fitness costs of induced resistance: emerging experimental support for a slippery concept. *Trends Plant Sci.* 7, 61–67. doi: 10.1016/S1360-1385(01)02186-0
- Hiroshi, T., Yanoria, M. J. T., Ebron, L. A., Nagao, H., Ikuo, A., Hiroshi, K., et al. (2000). Development of monogenic lines of rice for blast resistance. *Breed. Sci.* 50, 229–234. doi: 10.1270/jsbbs.50.229
- Hu, W., Wang, T., Yue, E., Zheng, S., and Xu, J.-H. (2014). Flexible microRNA arm selection in rice. *Biochem. Biophys. Res. Commun.* 447, 526–530. doi: 10.1016/j.bbrc.2014.04.036
- Huang, C.-Y., Wang, H., Hu, P., Hamby, R., and Jin, H. (2019). Small RNAs—big players in plant-microbe interactions. *Cell Host Microbe* 26, 173–182. doi: 10.1016/j.chom.2019.07.021
- Huang, X., Qian, Q., Liu, Z., Sun, H., He, S., Luo, D., et al. (2009). Natural variation at the DEP1 locus enhances grain yield in rice. *Nat. Genet.* 41, 494–497. doi: 10.1038/ng.352
- Iwakawa, H.-O., and Tomari, Y. (2015). The functions of microRNAs: mRNA decay and translational repression. *Trends Cell Biol.* 25, 651–665. doi: 10.1016/j.tcb.2015.07.011
- Ji, L., Liu, X., Yan, J., Wang, W., Yumul, R. E., Kim, Y. J., et al. (2011). ARGONAUTE10 and ARGONAUTE1 regulate the termination of floral stem cells through two microRNAs in *Arabidopsis*. *PLoS Genet.* 7:e1001358. doi: 10.1371/journal.pgen.1001358
- Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., et al. (2010). Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat. Genet.* 42, 541–544. doi: 10.1038/ng.591
- Jones-Rhoades, M. W., Bartel, D. P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* 57, 19–53. doi: 10.1146/annurev.arplant.57.032905.105218
- Kankanala, P., Czymmek, K., and Valent, B. (2007). Roles for rice membrane dynamics and plasmodesmata during biotrophic invasion by the blast fungus. *Plant Cell* 19, 706–724. doi: 10.1105/tpc.106.046300
- Kawano, Y., and Shimamoto, K. (2013). Early signaling network in rice PRR-mediated and R-mediated immunity. *Curr. Opin. Plant Biol.* 16, 496–504. doi: 10.1016/j.pbi.2013.07.004
- Kawasaki, T., Yamada, K., Yoshimura, S., and Yamaguchi, K. (2017). Chitin receptor-mediated activation of MAP kinases and ROS production in rice and *Arabidopsis*. *Plant Signal. Behav.* 12:e1361076. doi: 10.1080/15592324.2017.1361076
- Kobayashi, H., and Tomari, Y. (2016). RISC assembly: coordination between small RNAs and Argonaute proteins. *Bioch. Biophys. Acta Gene Regul. Mech.* 1859, 71–81. doi: 10.1016/j.bbagrm.2015.08.007
- Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* 11:597. doi: 10.1038/nrg2843
- Li, J., Brader, G., and Palva, E. T. (2004). The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. *Plant Cell* 16, 319–331. doi: 10.1105/tpc.016980
- Li, Y., Jeyakumar, J. M. J., Feng, Q., Zhao, Z.-X., Fan, J., Khaskheli, M. I., et al. (2019). The roles of rice microRNAs in rice-Magnaporthe oryzae interaction. *Phytopathol. Res.* 1:33. doi: 10.1186/s42483-019-0040-8
- Li, Y., Lu, Y.-G., Shi, Y., Wu, L., Xu, Y.-J., Huang, F., et al. (2014). Multiple rice MicroRNAs are involved in immunity against the blast fungus. *Plant Physiol.* 164, 1077–1092. doi: 10.1104/pp.113.230052
- Li, Y., Zhao, S.-L., Li, J.-L., Hu, X.-H., Wang, H., Cao, X.-L., et al. (2017). Osa-miR169 negatively regulates rice immunity against the blast fungus *Magnaporthe oryzae*. *Front. Plant Sci.* 8:2. doi: 10.3389/fpls.2017.00002
- Lin, Z. Z., Jiang, W. W., Wang, J. L., and Lei, C. L. (2001). Research and utilization of universally susceptible property of japonica rice variety lijiaxingxintuanheigu. *Sci. Agric. Sin.* 34, 116–117.
- Liu, B., Li, P., Li, X., Liu, C., Cao, S., Chu, C., et al. (2005). Loss of function of OsDCL1 affects microRNA accumulation and causes developmental defects in rice. *Plant Physiol.* 139, 296–305. doi: 10.1104/pp.105.063420
- Liu, M., Shi, Z., Zhang, X., Wang, M., Zhang, L., Zheng, K., et al. (2019). Inducible overexpression of Ideal Plant Architecture1 improves both yield and disease resistance in rice. *Nat. Plants* 5, 389–400. doi: 10.1038/s41477-019-0383-2
- Miao, C., Wang, Z., Zhang, L., Yao, J., Hua, K., Liu, X., et al. (2019). The grain yield modulator miR156 regulates seed dormancy through the gibberellin pathway in rice. *Nat. Commun.* 10, 1–12.
- Navarro, L., Dunoyer, P., Jay, F., Arnold, B., Dharmasiri, N., Estelle, M., et al. (2006). A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Sci.* 312, 436–439. doi: 10.1126/science.1126088
- Preston, J. C., and Hileman, L. (2013). Functional evolution in the plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) gene family. *Front. Plant Sci.* 4:80. doi: 10.3389/fpls.2013.00080
- Riese, M., Höhmann, S., Saedler, H., Münster, T., and Huijser, P. (2007). Comparative analysis of the SBP-box gene families in *P. patens* and seed plants. *Gene* 401, 28–37. doi: 10.1016/j.gene.2007.06.018
- Rogers, K., and Chen, X. (2013). Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell* 25, 2383–2399. doi: 10.1105/tpc.113.13159
- Shikata, M., Koyama, T., Mitsuda, N., and Ohmetakagi, M. (2009). *Arabidopsis* SBP-Box genes SPL10, SPL11 and SPL2 control morphological change in

- association with shoot maturation in the reproductive phase. *Plant Cell Physiol.* 50, 2133–2145. doi: 10.1093/pcp/pcp148
- Shimono, M., Sugano, S., Nakayama, A., Jiang, C., Ono, K., Toki, S., et al. (2007). Rice WRKY45 plays a crucial role in benzothiadiazole-inducible blast resistance. *Plant Cell* 19, 2064–2076. doi: 10.1105/tpc.106.046250
- Si, L., Chen, J., Huang, X., Gong, H., Luo, J., Hou, Q., et al. (2016). OsSPL13 controls grain size in cultivated rice. *Nat. Genet.* 48, 447–456. doi: 10.1038/ng.3518
- Taguchi-Shiobara, F., Kawagoe, Y., Kato, H., Onodera, H., Tagiri, A., Hara, N., et al. (2011). A loss-of-function mutation of rice DENSE PANICLE 1 causes semi-dwarfness and slightly increased number of spikelets. *Breed. Sci.* 61, 17–25. doi: 10.1270/jsbbs.61.17
- Walters, D., and Boyle, C. (2005). Induced resistance and allocation costs: what is the impact of pathogen challenge? *Physiol. Mol. Plant Pathol.* 66, 40–44. doi: 10.1016/j.pmpp.2005.04.002
- Walters, D., and Heil, M. (2007). Costs and trade-offs associated with induced resistance. *Physiol. Mol. Plant Pathol.* 71, 3–17. doi: 10.1016/j.pmpp.2007.09.008
- Wang, J., Zhou, L., Shi, H., Chern, M., Yu, H., Yi, H., et al. (2018). A single transcription factor promotes both yield and immunity in rice. *Science* 361, 1026–1028. doi: 10.1126/science.aat7675
- Wang, L., Sun, S., Jin, J., Fu, D., Yang, X., Weng, X., et al. (2015). Coordinated regulation of vegetative and reproductive branching in rice. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15504–15509. doi: 10.1073/pnas.1521949112
- Wang, S., Li, S., Liu, Q., Wu, K., Zhang, J., Wang, S., et al. (2015). The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat. re Genet.* 47, 949–954. doi: 10.1038/ng.3352
- Wang, S., Wu, K., Yuan, Q., Liu, X., Liu, Z., Lin, X., et al. (2012). Control of grain size, shape and quality by OsSPL16 in rice. *Nat. Genet.* 44, 950–954. doi: 10.1038/ng.2327
- Wu, G., and Poethig, R. S. (2006). Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. *Development* 133, 3539–3547. doi: 10.1242/dev.02521
- Wu, H.-J., Wang, Z.-M., Wang, M., and Wang, X.-J. (2013). Widespread long noncoding RNAs as endogenous Target Mimics for MicroRNAs in plants. *Plant Physiol.* 161, 1875–1884. doi: 10.1104/pp.113.215962
- Wu, L., Zhang, Q., Zhou, H., Ni, F., Wu, X., and Qi, Y. (2009). Rice microRNA effector complexes and targets. *Plant Cell* 21, 3421–3435. doi: 10.1105/tpc.109.070938
- Xie, K., Wu, C., and Xiong, L. (2006). Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiol.* 142, 280–293. doi: 10.1104/pp.106.084475
- Xu, M., Hu, T., Zhao, J., Park, M.-Y., Earley, K. W., Wu, G., et al. (2016). Developmental functions of miR156-regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) genes in *Arabidopsis thaliana*. *PLoS Genet.* 12:e1006263. doi: 10.1371/journal.pgen.1006263
- Yan, J., Gu, Y., Jia, X., Kang, W., Pan, S., Tang, X., et al. (2012). Effective small RNA destruction by the expression of a short tandem target mimic in *Arabidopsis*. *Plant Cell* 24, 415–427. doi: 10.1105/tpc.111.094144
- Yu, Y., Jia, T., and Chen, X. (2017). The 'how' and 'where' of plant microRNAs. *New Phytol.* 216, 1002–1017. doi: 10.1111/nph.14834
- Yuan, H., Qin, P., Hu, L., Zhan, S., Wang, S., Gao, P., et al. (2019). OsSPL18 controls grain weight and grain number in rice. *J. Genet. Genomics* 46, 41–51. doi: 10.1016/j.jgg.2019.01.003
- Zhang, H., Zhang, J., Yan, J., Gou, F., Mao, Y., Tang, G., et al. (2017). Short tandem target mimic rice lines uncover functions of miRNAs in regulating important agronomic traits. *Proc. Natl. Acad. Sci. U.S.A.* 114, 5277–5282. doi: 10.1073/pnas.1703752114
- Zhang, X., Bao, Y., Shan, D., Wang, Z., Song, X., Wang, Z., et al. (2018). Magnaporthe oryzae induces the expression of a MicroRNA to suppress the immune response in rice. *Plant Physiol.* 177, 352–368. doi: 10.1104/pp.17.01665
- Zhang, X., Zhao, H., Gao, S., Wang, W.-C., Katiyar-Agarwal, S., Huang, H.-D., et al. (2011). *Arabidopsis* Argonaute 2 regulates innate immunity via miRNA393*-mediated silencing of a Golgi-localized SNARE gene. *MEMB12. Mol. Cell* 42, 356–366. doi: 10.1016/j.molcel.2011.04.010
- Zhao, Z. X., Feng, Q., Cao, X. L., Zhu, Y., Wang, H., Chandran, V., et al. (2019). Osa-miR167d facilitates infection of Magnaporthe oryzae in rice. *J. Integrat. Plant Biol.* doi: 10.1111/jipb.12816
- Zhou, S.-X., Zhu, Y., Wang, L.-F., Zheng, Y.-P., Chen, J.-F., Li, T.-T., et al. (2019). Osa-miR1873 fine-tunes rice immunity against *Magnaporthe oryzae* and yield traits. *J. Integrat. Plant Biol.* doi: 10.1111/jipb.12900
- Zhu, H., Hu, F., Wang, R., Zhou, X., Sze, S.-H., Liou, L. W., et al. (2011). *Arabidopsis* Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. *Cell* 145, 242–256. doi: 10.1016/j.cell.2011.03.024

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, Li, Zheng, Wang, Yang, Chen, Zhou, Wang, Li, Ma, Zhao, Pu, Feng, Fan, Zhang, Huang 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.



Unlike Many Disease Resistances, Rx1-Mediated Immunity to Potato Virus X Is Not Compromised at Elevated Temperatures

Manon M. S. Richard, Marijn Knip, Thomas Aalders, Machiel S. Beijaert and Frank L. W. Takken*

Swammerdam Institute for Life Sciences (SILS), Molecular Plant Pathology, University of Amsterdam, Amsterdam, Netherlands

OPEN ACCESS

Edited by:

Horacio Naveira,
University of A Coruña, Spain

Reviewed by:

Dalia Gamil Aseel,
Arid Lands Cultivation Research
Institute (ALCRI), Egypt
Jacek Hennig,
Institute of Biochemistry
and Biophysics (PAN), Poland
Martin Cann,
Durham University, United Kingdom

*Correspondence:

Frank L. W. Takken
f.l.w.takken@uva.nl

Specialty section:

This article was submitted to
Evolution and Functional Mechanisms
of Plant Disease Resistance,
a section of the journal
Frontiers in Genetics

Received: 11 October 2019

Accepted: 02 April 2020

Published: 24 April 2020

Citation:

Richard MMS, Knip M, Aalders T,
Beijaert MS and Takken FLW (2020)
Unlike Many Disease Resistances,
Rx1-Mediated Immunity to Potato
Virus X Is Not Compromised
at Elevated Temperatures.
Front. Genet. 11:417.
doi: 10.3389/fgene.2020.00417

Specificity in the plant immune system is mediated by Resistance (R) proteins. Most R genes encode intracellular NLR-type immune receptors and these pathogen sensors require helper NLRs to activate immune signaling upon pathogen perception. Resistance conferred by many R genes is temperature sensitive and compromised above 28°C. Many Solanaceae R genes, including the potato NLR Rx1 conferring resistance to Potato Virus X (PVX), have been reported to be temperature labile. Rx1 activity, like many Solanaceae NLRs, depends on helper-NLRs called NRC's. In this study, we investigated Rx1 resistance at elevated temperatures in potato and in *Nicotiana benthamiana* plants stably expressing Rx1 upon rub-inoculation with GFP-expressing PVX particles. In parallel, we used susceptible plants as a control to assess infectiousness of PVX at a range of different temperatures. Surprisingly, we found that Rx1 confers virus resistance in *N. benthamiana* up to 32°C, a temperature at which the PVX::GFP lost infectiousness. Furthermore, at 34°C, an Rx1-mediated hypersensitive response could still be triggered in *N. benthamiana* upon PVX Coat-Protein overexpression. As the Rx1-immune signaling pathway is not temperature compromised, this implies that at least one *N. benthamiana* helper NRC and its downstream signaling components are temperature tolerant. This finding suggests that the temperature sensitivity for Solanaceous resistances is likely attributable to the sensor NLR and not to its downstream signaling components.

Keywords: sensor NLR, helper NLR, plant immunity, temperature, virus, thermotolerance, disease triangle

INTRODUCTION

Plants have developed a multi-layered immune system activated by receptor proteins that detect pathogen-generated molecules. Immune receptors can be classified into two main groups: (i) the extracellular receptors, mainly Receptor like-Kinases (RLK) or -Proteins (RLP), commonly associated with either recognition of pathogens' conserved features (microbial- or pathogen associated molecular patterns, MAMP or PAMP) or by pathogen inflicted damage (damage associated molecular patterns or DAMP) and (ii) Intracellular receptors. Members of this latter group often encode Nucleotide-binding domain and leucine-rich repeat (NLR) proteins that recognize specific pathogen encoded avirulence factors (Avr) (Dodds and Rathjen, 2010). NLRs can be divided into two major sub-groups according to their N-terminal domain, the TNLs, with a Toll/interleukin-1

receptor (TIR) and the CNLs, with a Coiled Coil (CC) domain (Monteiro and Nishimura, 2018). NLRs have been described as molecular switches that turn ON immune signaling after pathogen perception (Takken et al., 2006). NLR activation often triggers local cell death, the so-called hypersensitive response (HR) (Balint-Kurti, 2019).

In plant genomes, NLRs are typically encoded by a large gene family consisting of several hundreds of genes. NLRs can be categorized into two operative groups, the sensors (e.g., NLRs responsible for pathogen perception) and the transducers (or helpers). The latter group has recently been highlighted and is responsible for relaying or translating the upstream signal from the sensor NLR to the downstream signaling components (Wu et al., 2018). In Solanaceae, a phylogenetically related NLR family, consisting of NLR Required for Cell death (NRC) genes, have been described as helper NLRs (Adachi et al., 2019). Required by a large number of sensor NLRs that mediate resistance against diverse pathogens, they constitute a complex network of immune receptors (Wu et al., 2017). For example, the sensor NLRs Mi-1 from tomato, Rpi-blb2 and R1 from potato rely exclusively on NRC4 to trigger resistance responses, while the tomato NLR Prf and the potato NLR GPA2 can trigger HR via NRC2 or NRC3. Other NLRs such as Rx1 from potato or Bs2 from pepper can pair with either NRC2, NRC3, or NRC4 (Wu et al., 2017). Interestingly, the founder NRC, NRC1, has been initially identified to be required for resistance mediated by the non-NLR, Cf-4 that confers resistance toward the fungus *Cladosporium fulvum* in tomato (Gabriels et al., 2007). This finding suggests a potential role of these helpers to integrate immune signaling from both intra- and extracellular immune receptors (Leibman-Markus et al., 2018).

Although environmental conditions, such as temperature, have a crucial impact on the outcome of the diseases, this third component of the disease triangle (plant, pathogen, and environment) is often overlooked in plant-pathogen interaction studies. A temperature dependency of disease resistance has been reported in several cases involving different kind of pathogens, such as viruses, fungi, oomycetes, bacteria, or nematodes. For example, the tobacco NLR N is unable to confer resistance to the Tobacco mosaic virus (TMV) above 28°C (Whitham et al., 1996). Resistance to the nematode *Meloidogyne incognita* mediated by the NLR Mi-1 in tomato is compromised by exposure to 35°C for 3 h preceding inoculation (De Carvalho et al., 2015). Tsw-mediated resistance fails to trigger resistance to Tomato spotted wilt virus (TSWV) at 32°C and above in pepper plants (Moury et al., 1998). The NLR Bs2 from pepper, conferring resistance to the bacterium *Xanthomonas axonopodis* pv. *vesicatoria*, shows compromised resistance and HR at 32°C (Romero et al., 2002). Interestingly, non-NLR mediated resistance; such as resistance mediated by the transmembrane receptor like proteins Cf-4 and Cf-9 against *C. fulvum* is also impaired at elevated temperature (Cai et al., 2001; De Jong et al., 2002).

While temperature sensitivity of resistance seems widespread, it is not trivial to study this aspect in many plant pathogen interactions. One reason is that pathogen fitness and virulence can also be affected by (elevated) temperatures (Velasquez et al., 2018), complicating identification of temperature sensitive

components in an interaction. Therefore, as a proxy for immune activation at elevated temperatures, the capacity of R genes to trigger HR upon (over)expression of their corresponding *Avr* is often used. For example, in Wang et al. (2009) N- and Rx1-temperature sensitivity is assessed by monitoring loss of HR when co-expressed with the corresponding *Avrs*; *p50* from TMV and the *Coat Protein* (CP) from Potato mosaic virus (PVX) at 28 and 30°C, respectively (Wang et al., 2009). However, HR is not always correlated with functional resistance. For example, HR triggered by the recognition of *Pseudomonas syringae* pv. *tomato* DC 3000 (PtoDC3000) HopZ1a or *AvrRpt2* in *Arabidopsis thaliana* is suppressed at 30°C, while resistance to the bacteria is unaffected (Menna et al., 2015). Additionally, co-expression of R and/or *Avr* gene(s) in heterologous systems often relies on *Agrobacterium*-mediated transient transformation assays (ATTA). A drawback of this system is the temperature sensitivity of T-DNA transfer by *Agrobacterium tumefaciens*, which makes plant transformation above 27°C highly inefficient (Dillen et al., 1997).

Many Solanaceae resistances, mediated by NLR- or non-NLR sensors that depend on NRC helpers, are reported to be compromised at elevated temperature. However, the temperature sensitive component of their molecular signaling pathways (sensor, helper, or downstream signaling) remains unknown. Since this temperature sensitivity concerns resistance triggered by different types of receptors (NLR and non-NLR, such as the RLPs Cf-4 and Cf-9), it is tempting to speculate that shared downstream signaling components, such as the helpers NRCs, could be the Achilles' heel of the immune signaling at elevated temperatures. Rx1 is a well-studied NLR from potato and is a perfect model to challenge our hypothesis since it can signal via NRC2, NRC3, or NRC4 (Wu et al., 2017).

The NLR Rx1 triggers resistance to PVX upon recognition of its CP in potato and in *N. benthamiana* stably expressing Rx1 from its native promoter (Bendahmane et al., 1999). Rx1 confers a so called "extreme resistance" response that prevents viral replication without triggering cell death (Bendahmane et al., 1999). Overexpression of the avirulent CP (CP106) in an Rx1-expressing plant nonetheless can trigger HR in heterologous species such as *N. benthamiana*, whereas CP105, a CP variant of an Rx1 resistance breaking strain of PVX, does not (Goulden et al., 1993; Bendahmane et al., 1999). Rx1 has been reported to be temperature sensitive as it was unable to trigger HR at 30°C upon ATTA-mediated CP-expression in *N. benthamiana* (Wang et al., 2009). However, the capacity of Rx1 to mount resistance against PVX at elevated temperature is not known.

Here we investigate Rx1-mediated resistance to PVX in *N. benthamiana* plants stably expressing Rx1 from its native promoter and in potato. Infection with the GFP-expressing PVX particles (PVX::GFP) was done by rub inoculation and is independent of *Agrobacterium* transformation. We observed that, in *N. benthamiana*, Rx1 is preventing PVX::GFP replication up to temperatures above which PVX::GFP is no longer infectious. The temperature resilience of Rx1 in potato could not be assessed as the virus itself is not efficiently infecting potato at 30°C in our experimental set-up. Besides, we assessed the capacity of Rx1 to trigger HR at elevated temperatures using Rx1 *N. benthamiana* plants stably transformed with a dexamethasone

(DEX)-inducible *CP* construct. We observed that Rx1-mediated HR can still be observed at temperatures up to 34°C. Altogether, our results imply that Rx1 is a thermotolerant R protein – as are its downstream signaling components – providing new insights in the mechanisms underlying thermosensitivity in plant immunity.

MATERIALS AND METHODS

Plant Lines, *N. benthamiana* Transformation and Growing Conditions

Wild-type and transgenic *Rx1:4xHA* (Lu et al., 2003), expressing Rx1 under control of its native promoter, *N. benthamiana* plants were used for PVX bioassays. To monitor HR *Rx1:4xHA+DEX::CP106* 9-4 (referred to as *Rx1D106*, internal identifier #FP1807) and *Rx1:4xHA+DEX::CP105* 6-6 (referred to as *Rx1D105*, internal identifier #FP1810) stable transgenic lines were generated. For this, *N. benthamiana Rx1:4xHA* plants were transformed using the dexamethasone (DEX) inducible PVX-CP constructs, either *DEX::CP106* or *DEX::CP105* described in Knip et al. (2019), using *Agrobacterium*-mediated transformation as described in Sparkes et al. (2006). Briefly, *A. tumefaciens* infiltrated leaves were surface sterilized, cut into 2 cm² diamond shape pieces, and placed on shoot-induction medium supplemented with 14.8 µg/ml hygromycin for selection. Shoots from putative transformants were transferred to root-induction medium containing 14.8 µg/ml hygromycin. Ten and seven candidate transformants for *DEX::CP106* or *DEX::CP105* constructs, respectively, were selected for seed production. Segregation for hygromycin resistance of the obtained T1 progeny was assessed on selective medium and seven and five lines with a single insertion were identified for *DEX::CP106* or *DEX::CP105* constructs, respectively. Homozygosity of T = 2 plants was evaluated by Real-Time PCR using gDNA by estimation of t-DNA copy number (by amplification of the Hygromycin resistance gene, using the oligonucleotides FP7722-HP ThygroFW: GTTCGGGGATTCCCAATACGAGGTC and FP7723-HPThygroRV: ATCGAAATTGCCGTCAACCAAGTC) compared to a endogenous reference gene (NRG1, amplified using the oligonucleotides FP8254: GTGTCCGACCACTAAGCATGGAACCTA and FP8255: CTGCTGGTGCATCCTTTCTGGAAATC). The Real-Time PCRs were performed in QuantStudioTM3 (Thermo Fisher Scientific). The 10 µL PCR contained 0.2 µM of each primer, 100 ng of gDNA, 0.05 µl of DNA polymerase (DreamTaq, Thermo Fisher Scientific), 1x Evagreen dye (Solis Biodyne), 1x ROX passive reference, dNTPs, and water. The cycling program was set to initial denaturation 2 min at 94°C, 40 cycles; denaturation for 15 s at 95°C, annealing for 20 s at 58°C, elongation for 30 s at 72°C, followed a melting curve analysis of 15 s at 95°C, 1 min at 60°C, 15 s at 95°C. The copy number analysis was performed using the online Thermo Fisher Scientific application. Plants being phenotypically similar to the parental *Rx1:4xHA* plants, and showing expression of the *CP* upon dexamethasone treatment were selected, resulting in three and two independent

homozygous *Rx1D106* and *Rx1D105* lines, respectively. One out of the three *Rx1D106* and one out of the two *Rx1D105*, were used for this study as the different *Rx1D106* and *Rx1D105* lines showed identical responses when DEX treated (data not shown).

N. benthamiana plants were grown under long-day conditions in a climate chamber (22°C, 70% humidity, 16 h/8 h light/dark) for 3–4 weeks. Potato tubers from two diploid potato genotypes RH 89-039-16 (RH) and SH82-93-488 (SH) (Van der Voort et al., 1997) were planted in soil and plants were grown for 6 weeks under the conditions described above. One day before treatment (either PVX::GFP rub-inoculation or dexamethasone treatment, see below), plants were transferred and incubated in a MD1400 MODULAR CLIMATE CHAMBER (Snijders Labs) at the indicated temperatures under a constant humidity of 80% at a 12/12h light/dark regime.

PVX::GFP Rub-Inoculation and *in planta* Virus Detection in *N. benthamiana*

To produce infectious PVX::GFP particles, leaves of 4 weeks old WT *N. benthamiana* plants were agroinfiltrated with an *A. tumefaciens* GV3101 strain containing the *pJIC SA_Rep* helper plasmid and the PVX::GFP construct (internal identifier BglFP#4081) according to Ma et al. (2012). PVX::GFP was obtained by inserting the *LSS-msfGFP* ORF¹ after a duplicated CP promoter, into SgsI – NotI restriction sites of pGR106, a binary vector containing an infectious PVX clone (Jones et al., 1999). Two weeks after agroinfiltration, systemically infected leaves were either snap frozen with liquid nitrogen and stored at –80°C or directly used for rub-inoculation. PVX::GFP inoculum was made by grinding a fresh, or frozen, PVX::GFP infected leaf with a mortar in 4 mL 50 mM potassium phosphate buffer pH 7. The youngest fully expanded leaves of 3 weeks old *N. benthamiana* plants (WT or *Rx1*) were mechanically inoculated by rubbing the adaxial side with a piece of miracloth (Merck; pore size of 22–25 µm) soaked in the PVX::GFP inoculum using Carborundum as an abrasive. Five minutes post-inoculation the inoculated leaves were rinsed with tap water and excess water was removed using paper tissues. Ten days after rub-inoculation, plants were photographed using a Panasonic Lumix DMC-LX15 camera placed in a dark chamber (Extraneous Light Protector and RS 1 stand, Kaiser, Germany) illuminated with UV light (RB 5003 UV Lighting Unit code n°5591, Kaiser, Germany).

RNA Isolation and RT-PCR

For PVX::GFP RNA detection, systemic leaves from PVX::GFP rub-inoculated plants were sampled 10 days post-inoculation. To verify induction of PVX-CP transcription after dexamethasone treatment, leaves of *Rx1D105* and *Rx1D106* were sampled at 0, 2, and 4 h post-dexamethasone induction (hpdi). Total RNA was extracted using TRIzol LS reagent (Thermo Fisher Scientific, Waltham, MA, United States). The RNA was treated with DNase (Thermo Fischer Scientific) according to the supplier's protocol and RNA concentrations were determined by measuring the absorbance at 260 nm on a NanoDrop (Thermo Fisher Scientific). cDNA was synthesized from 1.5 µg of total RNA using

¹<https://www.addgene.org/112941/>

RevertAid H reverse transcriptase and Oligo-dT (Eurofins) in the presence of the RNase inhibitor Ribolock (Thermo Fisher Scientific) following the supplier's protocol and diluted 5 times in Milli-Q H₂O.

Semi quantitative Reverse-Transcriptase (RT) PCR (35 cycles, annealing temperature of 60°C) was performed on 1 µl of diluted cDNA using DreamTaq DNA Polymerase (Thermo Fisher Scientific) following the supplier's protocol, using CP specific primers FP8371-PVX-CP-F: CACTGCAGGCGCAACTCC and FP8372-PVX-CP-R: GTCGTTGGATTGYGCCCT or EF1α primers FP8391-NbEF1α-F: AGCTTTACCTCCCAAGTCATC and FP8392-NbEF1α-R: AGAACGCCTGTCAATCTTGG as a positive internal control.

Potato Inoculation and PVX Detection by ELISA and Western Blot

Six terminal leaflets from 6 weeks old potato plants SH and RH genotypes were rub-inoculated with PVX::GFP as described above. Plants were kept at either 20 or 30°C prior to analysis. One week after inoculation, a quarter of each inoculated leaflet was sampled and homogenized in 50 mM Sodium Phosphate buffer pH 7, using a TissueLyser (QIAGEN) and three 3 mm steel beads at 30 Hz for twice 30 s. The virus concentration was determined by DAS-ELISA using PVX antibodies (Prime Diagnostics, Wageningen, The Netherlands). Plates (NUNC-Immuno Plates Maxisorp F96) were coated with a 1:1000 dilution of the PVX polyclonal antibody to bind the antigen. A second polyclonal PVX antibody, conjugated with alkaline phosphatase, was used for detection by monitoring the conversion of the p-nitrophenyl phosphate substrate. Absorbance of each well was recorded at 405 nm with a reference filter of 655 nm using a BioTek Synergy H1 Hybrid multi-mode microplate reader (BioTek).

Proteins were isolated from one centimeter of petiole of each inoculated leaf 1 week post-inoculation using the method described in Knip et al. (2019). PVX detection by Western blot was performed on these samples as described in Knip et al. (2019) using PVX-specific antibody (diluted 1 : 3000) (ref 110411, Bioreba, Reinach, Switzerland).

Rx1-Mediated HR Induction

One day after transfer of 3 weeks old *Rx1D106* and *Rx1D105* plants into the growth incubator at 20, 30, 32, 33, or 34°C, leaves were treated with 20 µM dexamethasone, 0.01% Silwet R-77 in Milli-Q H₂O. The dexamethasone solution was applied on an ~1 cm-diameter circle on the left side of each leaf. HR was assessed and pictures were taken 24 h post-dexamethasone application.

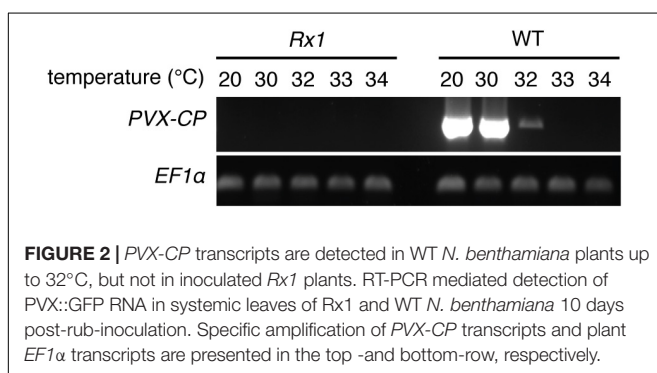
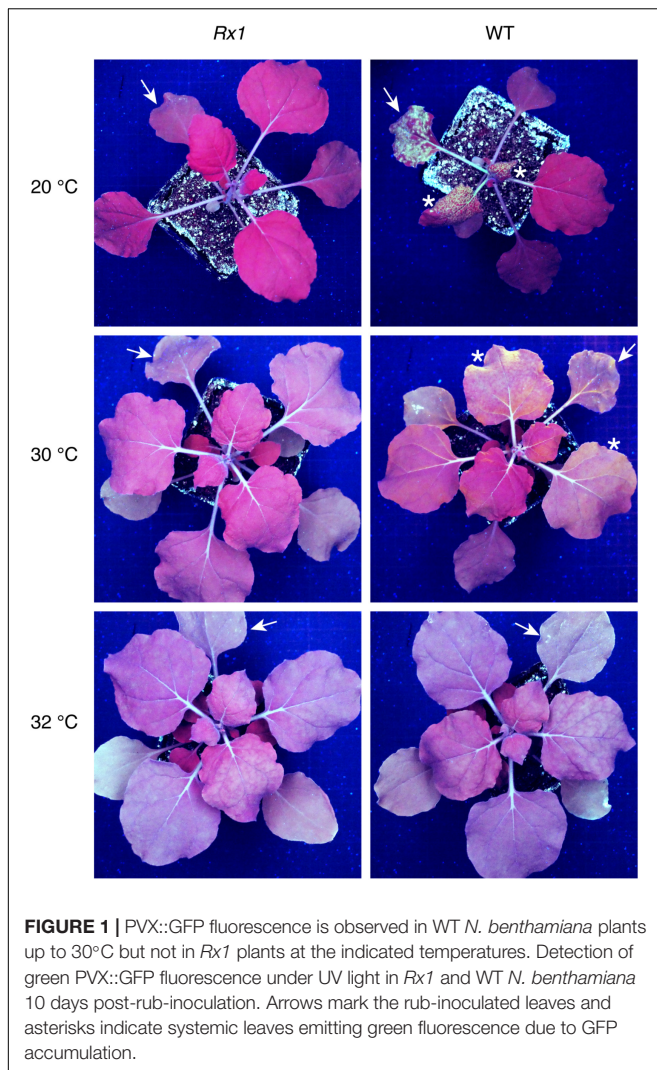
RESULTS

Virulence of PVX Is Compromised Above 32°C While Rx1-Mediated Resistance Is Unaffected at This Temperature

To determine a potential thermotolerance of Rx1-mediated resistance to PVX, *Rx1* transgenic *N. benthamiana* plants were

rub-inoculated with infectious PVX particles at 20, 30, 32, 33, or 34°C. To visualize infection and spread of the virus, a recombinant PVX::GFP strain was used that triggers production of green fluorescent protein in infected plant cells. As a positive control for infection, susceptible wildtype (WT) *N. benthamiana* were rub-inoculated with the virus and incubated at the indicated temperatures. Ten days after inoculation WT plants kept at 20°C displayed strong PVX symptoms (leaf deformations and stunting) that correlated with intense GFP fluorescence under UV light (**Figure 1**). Green fluorescence could be observed at infection foci on the inoculated leaf and its petioles, and around the vasculature of systemically infected leaves (**Figure 1**). As expected, no symptoms nor green fluorescence were observed in PVX::GFP-inoculated *Rx1* plants at 20°C due to the resistance conferred by Rx1 (**Figure 1**). Instead, *Rx1* plants appeared red under UV light due to chlorophyll autofluorescence. RT-PCR on systemic leaf material confirmed that viral transcripts were present in WT plants, but absent in *Rx1* plants at 20°C (**Figure 2**). At 30°C and above, visual PVX symptoms could no longer be discerned in WT plants. However, GFP fluorescence could still be observed in systemic leaves of WT plants at 30°C, attesting the usefulness of a GFP reporter virus (**Figure 1**). In comparison, no GFP fluorescence was observed in *Rx1* plants inoculated at 30°C (**Figure 1**). These differences were confirmed by semi quantitative RT-PCR revealing the presence of viral RNA in the WT plant, but not in the resistant *Rx1* line (**Figure 2**). At 32°C, and above (data not shown), in addition to an absence of PVX symptoms, no GFP fluorescence was observed in WT or *Rx1* plants (**Figure 1**). Plants inoculated at 33 and 34°C are phenotypically similar to those incubated at 32°C (data not shown). Only a very small amount of viral RNA could be detected in the systemic leaves of WT plants at 32°C, but no viral transcripts were observed at 33 and 34°C (**Figure 2**). These findings suggests that at temperatures above 32°C PVX::GFP is no longer able to infect and spread in *N. benthamiana*. In addition, as no viral RNAs were detected at 32°C or above in *Rx1* plants, this suggests that *Rx1* is able to confer resistance to PVX::GFP in *N. benthamiana* at least up to temperatures at which the virus is no longer infectious.

The thermotolerance of Rx1-mediated resistance to PVX::GFP was investigated in potato plants with or without Rx1 resistance, the genotypes SH (Rx1) and RH (no Rx1), respectively. As expected, no virus was detected in the resistant SH genotype inoculated leaves or petioles at either 20 or 30°C (**Supplementary Figures 1, 2**). Notably, while PVX::GFP was detectable in sensitive RH potato inoculated leaves at 20°C, no virus could be detected in their petioles suggesting that the virus did not move systemically yet (**Supplementary Figures 1, 2**). Furthermore, no virus could be detected in susceptible RH inoculated leaves nor petioles at 30°C (**Supplementary Figures 1, 2**). This suggests that the capacity of PVX::GFP to infect at elevated temperature is determined by the host, as in *N. benthamiana* plants present in the same compartment high viral titers could be detected at 30°C (**Supplementary Figures 1, 2**). The inability of PVX::GFP to infect potato plants at elevated temperature in our experimental set-up prevents further investigation of the thermotolerance of Rx1 resistance in potato.



Rx1 Triggers HR Upon Recognition of the Avirulent PVX Coat Protein Variant at 34°C

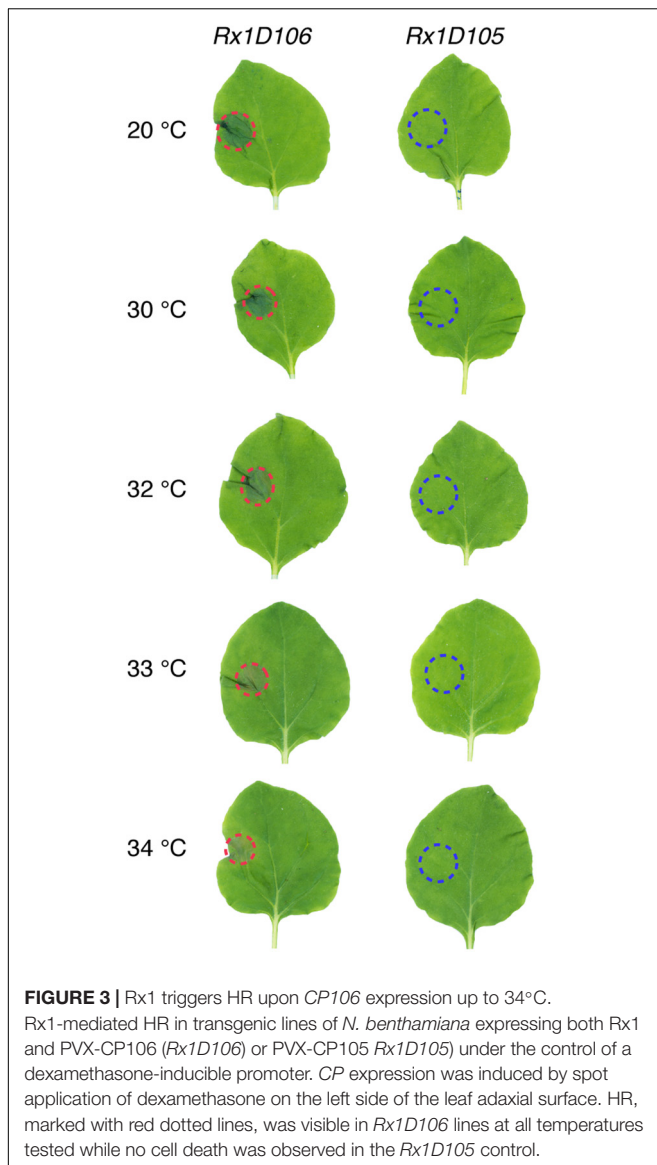
Since PVX infection is fully abolished above 32°C, an alternative approach was used to monitor Rx1 activity at higher temperatures. Rx1 activation by the Coat Protein 106

variant (CP106) of PVX triggers a Hypersensitive Response in *N. benthamiana* that is visible as a necrotic sector. To express CP106, or CP105 as a negative control, we used the CESSNA system to enable inducible expression upon dexamethasone application (Knip et al., 2019). As *Agrobacterium*-mediated transformation is compromised at 27°C and higher (Dillen et al., 1997) we generated stable transgenic plants expressing *Rx1* in combination with DEX::CP106 or DEX::CP105, referred to as *Rx1D106* and *Rx1D105*, respectively). In the absence of dexamethasone, the generated *Rx1D106* and *Rx1D105* transgenic plants were phenotypically identical to WT plants (data not shown). Upon dexamethasone treatment, expression of CP105 and CP106 was observed in *Rx1D105* and *Rx1D106*, respectively, within 2 h following application of dexamethasone (Supplementary Figure 3). As anticipated a clear HR was observed at the DEX treated sector in *Rx1D106* but not in *Rx1D105* leaves, as the latter does not trigger Rx1-mediated signaling (Figure 3). Using these hence validated transgenic lines, the capacity of Rx1 to trigger HR in response to CP106 at elevated temperature was examined. Dexamethasone was locally applied on the left side of one leaf of *Rx1D106* and *Rx1D105* plants placed at 20, 30, 32, 33, and 34°C. As shown in Figure 3, Rx1 triggered HR at all temperatures tested, but only upon dexamethasone mediated induction of CP106-, but not CP105-expression. These results show that the ability of Rx1 to trigger HR upon CP106 perception is not compromised at temperatures up to 34°C.

DISCUSSION

In this study, we show that Rx1-mediated resistance to PVX::GFP in *N. benthamiana* remains functional up to a temperature at which the virus was no longer infectious. We also show that the non-permissive temperatures for PVX::GFP infection differ between potato and *N. benthamiana* plants. Indeed, PVX::GFP could spread systemically up to 32°C in *N. benthamiana*, while at 30°C PVX::GFP multiplication in inoculated potato leaves was compromised (Supplementary Figure 1) and no systemically spreading virus could be detected in petioles of inoculated leaves (Supplementary Figure 2). In addition, Rx1-triggered HR was also observed at high temperatures (34°C) in *N. benthamiana*. This finding contrasts a previous study detailing that Rx1-mediated HR was abolished above 28°C (Wang et al., 2009). The main difference between both studies is the use of *A. tumefaciens* to express *R* and *Avr* constructs. In our study both genes are stably integrated in the plant genome and *Avr* expression is triggered by applying dexamethasone. In our system co-expressing *Rx1* with the non-recognized CP105 variant did not trigger cell death, thereby ruling out the possibility that dexamethasone itself triggered cell death at elevated temperatures. Considering the fact that *Agrobacterium*-mediated transformation efficiency strongly decreases at increased temperatures, and was shown to be fully compromised at 29°C (Dillen et al., 1997), we propose that the absence of HR observed in Wang et al. (2009) could be attributed to a lack or a poor expression of the construct(s) used.

The decrease of PVX titer in local, and especially in systemic leaves, with increasing temperature has been previously reported



in *Nicotiana* species (Ma et al., 2016). Furthermore, it is likely that PVX::GFP will have a slightly altered performance as compared to a natural strain. We observed that PVX::GFP infectiousness in potato was impaired at lower temperatures than in *N. benthamiana*. *N. benthamiana* is hypersusceptible to many RNA viruses due to a mutation in an RNA dependent RNA Polymerase that is important for antiviral defense based on RNA silencing (Yang et al., 2004). PVX replicates slower in potato than in *N. benthamiana*, which could explain the higher viral titers in *N. benthamiana* as compared to potato at 20°C (Supplementary Figure 1). At elevated temperature (e.g., 30°C) PVX::GFP was not detectable in potato 1 week after inoculation with the virus (Supplementary Figures 1, 2). We cannot exclude whether viral replication and spread is compromised in potato at the elevated temperature, but we could not prolong the incubation time as the plants started to collapse after 1 week. The mechanism underlying poor viral replication and spread

is unknown, but could be related to the observation that the RNA silencing machinery in plants is more active at elevated temperatures (Szittyá et al., 2003). Further study is required to resolve whether RNA silencing is responsible for impaired viral replication at elevated temperatures in potato.

NRCs are required for resistance mediated by both sensor CNLs and RLPs in Solanaceae (Gabriels et al., 2007; Wu et al., 2017). Several of these CNL- or RLP-mediated resistances are compromised at elevated temperatures (De Jong et al., 2002; Romero et al., 2002; De Carvalho et al., 2015), pointing to NRCs as potential suspects for thermosensitivity. However, our findings show that Rx1-mediated resistance against PVX::GFP is still efficient up to a temperature at which the virus is not infectious anymore. As Rx1 immunity is unaltered at elevated temperature, this means that Rx1 activation and its downstream signaling component are also functional. Rx1 has been shown to require NRC2, 3 or 4 downstream of its activation to trigger HR (Wu et al., 2017). Therefore, it is tempting to speculate that at least one of the Rx1-interacting NRCs functions at elevated temperature. Consequently, the immune sensors (sensor NLRs) are the most probable components that are affected in thermosensitive immune signaling. Notably, a similar observation has been made for two TNLs, SNC1 from *Arabidopsis* and N from tobacco (Zhu et al., 2010). Through genetic screens and targeted mutagenesis, it was shown that SNC1 and N are the thermosensitive components in immune signaling. Together these observations suggest that for both CNL-, TNL-, or RLP- triggered immunity, the sensor is the temperature sensitive element. How a sensor NLR is affected by elevated temperature and loses its activity is unknown. Of note, several studies show that an elevated ambient temperature reduces accumulation of sensor NLR proteins, such as SNC1 or RPS4 and N in the nucleus (Wang et al., 2009; Zhu et al., 2010; Mang et al., 2012; Hua, 2013). For some NLRs a nuclear location has been shown to be crucial for triggering immunity, providing a potential mechanism (Qi and Innes, 2013).

The premise of these observations, that the sensors are thermosensitive while downstream immune signaling components do function at elevated temperatures, is relevant as it provides leads on how to improve thermotolerance of disease resistances by identifying the immune receptor itself as main target for mutagenesis.

AUTHOR CONTRIBUTIONS

MR and FT designed the study and wrote the manuscript. MR, MK, TA, and MB performed experiments. MR analyzed the data and drafted all figures. All authors read and approved the final manuscript.

FUNDING

MR, MK, MB, and FT received funding from the NWO-Earth and Life Sciences-funded VICI project No. 865.14.003, and FT received funding from the European Union's Horizon

2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 676480 (Bestpass).

ACKNOWLEDGMENTS

We thank Marieke Mastop and Dr. Joachim Goedhart (University of Amsterdam) for providing lss-msf-GFP construct, Octavina Sukarta, and Dr. Aska Goverse (Wageningen University) for

sharing the ELISA protocol and Dr. Nico Tintor for providing the potato tubers.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00417/full#supplementary-material>

REFERENCES

- Adachi, H., Derevnina, L., and Kamoun, S. (2019). NLR singletons, pairs, and networks: evolution, assembly, and regulation of the intracellular immunoreceptor circuitry of plants. *Curr. Opin. Plant Biol.* 50, 121–131. doi: 10.1016/j.pbi.2019.04.007
- Balint-Kurti, P. (2019). The plant hypersensitive response: concepts, control and consequences. *Mol. Plant Pathol.* 20, 1163–1178.
- Bendahmane, A., Kanyuka, K., and Baulcombe, D. C. (1999). The Rx gene from potato controls separate virus resistance and cell death responses. *Plant Cell* 11, 781–791. doi: 10.1105/tpc.11.5.781
- Cai, X. Z., Takken, F. L. W., Joosten, M., and De Wit, P. (2001). Specific recognition of AVR4 and AVR9 results in distinct patterns of hypersensitive cell death in tomato, but similar patterns of defence-related gene expression. *Mol. Plant Pathol.* 2, 77–86. doi: 10.1046/j.1364-3703.2001.00053.x
- De Carvalho, L. M., Benda, N. D., Vaughan, M. M., Cabrera, A. R., Hung, K., Cox, T., et al. (2015). Mi-1-mediated nematode resistance in tomatoes is broken by short-term heat stress but recovers over time. *J. Nematol.* 47, 133–140.
- De Jong, C. F., Takken, F. L. W., Cai, X. H., De Wit, P., and Joosten, M. (2002). Attenuation of Cf-mediated defense responses at elevated temperatures correlates with a decrease in elicitor-binding sites. *Mol. Plant Microbe Interact.* 15, 1040–1049. doi: 10.1094/mpmi.2002.15.10.1040
- Dillen, W., De Clercq, J., Kapila, J., Zambre, M., Van Montagu, M., and Angenon, G. (1997). The effect of temperature on *Agrobacterium tumefaciens*-mediated gene transfer to plants. *Plant J.* 12, 1459–1463. doi: 10.1046/j.1365-313x.1997.12061459.x
- Dodds, P. N., and Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat. Rev. Genet.* 11, 539–548. doi: 10.1038/nrg2812
- Gabriels, S., Vossen, J. H., Ekengren, S. K., Van Ooijen, G., Abd-El-Halim, A. M., Van Den Berg, G. C. M., et al. (2007). An NB-LRR protein required for HR signalling mediated by both extra- and intracellular resistance proteins. *Plant J.* 50, 14–28. doi: 10.1111/j.1365-313x.2007.03027.x
- Goulden, M. G., Kohm, B. A., Cruz, S. S., Kavanagh, T. A., and Baulcombe, D. C. (1993). A feature of the coat protein of Potato Virus-X affects both induced virus-resistance in potato and viral fitness. *Virology* 197, 293–302. doi: 10.1006/viro.1993.1590
- Hua, J. (2013). Modulation of plant immunity by light, circadian rhythm, and temperature. *Curr. Opin. Plant Biol.* 16, 406–413. doi: 10.1016/j.pbi.2013.06.017
- Jones, L., Hamilton, A. J., Voinnet, O., Thomas, C. L., Maule, A. J., and Baulcombe, D. C. (1999). RNA-DNA interactions and DNA methylation in post-transcriptional gene silencing. *Plant Cell* 11, 2291–2301. doi: 10.1105/tpc.11.12.2291
- Knip, M., Richard, M. M. S., Oskam, L., Van Engelen, H. T. D., Alders, T., and Takken, F. L. W. (2019). Activation of immune receptor Rx1 triggers distinct immune responses culminating in cell death after 4 hours. *Mol. Plant Pathol.* 20, 575–588. doi: 10.1111/mpp.12776
- Leibman-Markus, M., Pizarro, L., Bar, M., Coaker, G., and Avni, A. (2018). NRC proteins – A critical node for pattern and effector mediated signaling. *Plant Signal. Behav.* 13:e1507404.
- Lu, R., Malcuit, I., Moffett, P., Ruiz, M. T., Peart, J., Wu, A. J., et al. (2003). High throughput virus-induced gene silencing implicates heat shock protein 90 in plant disease resistance. *EMBO J.* 22, 5690–5699. doi: 10.1093/emboj/cd9546
- Ma, L., Huang, X., Yu, R., Jing, X. L., Xu, J., Wu, C. A., et al. (2016). Elevated ambient temperature differentially affects virus resistance in two tobacco species. *Phytopathology* 106, 94–100. doi: 10.1094/phyto-11-14-0300-r
- Ma, L., Lukasik, E., Gawehns, F., and Takken, F. L. (2012). The use of agroinfiltration for transient expression of plant resistance and fungal effector proteins in *Nicotiana benthamiana* leaves. *Methods Mol. Biol. (Clifton, N.J.)* 835, 61–74. doi: 10.1007/978-1-61779-501-5_4
- Mang, H.-G., Qian, W., Zhu, Y., Qian, J., Kang, H.-G., Klessig, D. F., et al. (2012). Abscisic acid deficiency antagonizes high-temperature inhibition of disease resistance through enhancing nuclear accumulation of resistance proteins SNC1 and RPS4 in *Arabidopsis*. *Plant Cell* 24, 1271–1284. doi: 10.1105/tpc.112.096198
- Menna, A., Nguyen, D., Guttman, D. S., and Desveaux, D. (2015). Elevated temperature differentially influences effector-triggered immunity outputs in *Arabidopsis*. *Front. Plant Sci.* 6:995. doi: 10.3389/fpls.2015.00995
- Monteiro, F., and Nishimura, M. T. (2018). “Structural, functional, and genomic diversity of plant NLR proteins: an evolved resource for rational engineering of plant immunity,” in *Annual Review of Phytopathology*, Vol. 56, eds J. E. Leach, and S. E. Lindow, (Palo Alto, CA: Annual Reviews), 243–267. doi: 10.1146/annurev-phyto-080417-045817
- Moury, B., Selassie, K. G., Marchoux, G., Daubeze, A. M., and Palloix, A. (1998). High temperature effects on hypersensitive resistance to Tomato Spotted wilt Tospovirus (TSWV) in pepper (*Capsicum chinense* Jacq.). *Eur. J. Plant Pathol.* 104, 489–498.
- Qi, D., and Innes, R. W. (2013). Recent advances in plant NLR structure, function, localization, and signaling. *Front. Immunol.* 4:348. doi: 10.3389/fimmu.2013.00348
- Romero, A. M., Kousik, C. S., and Ritchie, D. F. (2002). Temperature sensitivity of the hypersensitive response of bell pepper to *Xanthomonas axonopodis* pv. vesicatoria. *Phytopathology* 92, 197–203.
- Sparkes, I. A., Runions, J., Kearns, A., and Hawes, 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
- Szittyá, G., Silhavy, D., Molnar, A., Havelda, Z., Lovas, A., Lakatos, L., et al. (2003). Low temperature inhibits RNA silencing-mediated defence by the control of siRNA generation. *EMBO J.* 22, 633–640. doi: 10.1093/emboj/cdg74
- Takken, F. L. W., Albrecht, M., and Tameling, W. I. L. (2006). Resistance proteins: molecular switches of plant defence. *Curr. Opin. Plant Biol.* 9, 383–390. doi: 10.1016/j.pbi.2006.05.009
- Van der Voort, J. N., Van Zandvoort, P., Van Eck, H. J., Folkertsma, R. T., Hutten, R. C. B., and Draaistra, J. (1997). Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. *Mol. Gen. Genet.* 255, 438–447. doi: 10.1007/s004380050516
- Velasquez, A. C., Castroverde, C. D. M., and He, S. Y. (2018). Plant-pathogen warfare under changing climate conditions. *Curr. Biol.* 28, R619–R634.
- Wang, Y., Bao, Z. L., Zhu, Y., and Hua, J. (2009). Analysis of temperature modulation of plant defense against biotrophic microbes. *Mol. Plant Microbe Interact.* 22, 498–506. doi: 10.1094/mpmi-22-5-0498
- Whitham, S., McCormick, S., and Baker, B. (1996). The N gene of tobacco confers resistance to tobacco mosaic virus in transgenic tomato. *Proc. Natl. Acad. Sci. U.S.A.* 93, 8776–8781. doi: 10.1073/pnas.93.16.8776
- Wu, C. H., Abd-El-Halim, A., Bozkurt, T. O., Belhaj, K., Terauchi, R., Vossen, J. H., et al. (2017). NLR network mediates immunity to diverse plant pathogens.

- Proc. Natl. Acad. Sci. U.S.A.* 114, 8113–8118. doi: 10.1073/pnas.1702041114
- Wu, C. H., Derevnina, L., and Kamoun, S. (2018). Receptor networks underpin plant immunity. *Science* 360, 1300–1301. doi: 10.1126/science.aat2623
- Yang, S. J., Carter, S. A., Cole, A. B., Cheng, N. H., and Nelson, R. S. (2004). A natural variant of a host RNA-dependent RNA polymerase is associated with increased susceptibility to viruses by *Nicotiana benthamiana*. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6297–6302. doi: 10.1073/pnas.0304346101
- Zhu, Y., Qian, W. Q., and Hua, J. (2010). Temperature modulates plant defense responses through NB-LRR proteins. *Plos Pathog.* 6:e1000844. doi: 10.1371/journal.ppat.1000844

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 Richard, Knip, Aalders, Beijaert and Takken. 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-Wide Identification and Evolutionary Analysis of *NBS-LRR* Genes From *Dioscorea rotundata*

Yan-Mei Zhang¹, Min Chen¹, Ling Sun¹, Yue Wang¹, Jianmei Yin², Jia Liu¹, Xiao-Qin Sun¹ and Yue-Yu Hang^{1*}

¹ Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China, ² Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China

OPEN ACCESS

Edited by:

Madhav P. Nepal,
South Dakota State University,
United States

Reviewed by:

Surendra Neupane,
University of Florida, United States
Yan Zhong,
Nanjing Agricultural University, China

*Correspondence:

Yue-Yu Hang
hangyueyu@cnbg.net

Specialty section:

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

Received: 13 February 2020

Accepted: 17 April 2020

Published: 07 May 2020

Citation:

Zhang Y-M, Chen M, Sun L,
Wang Y, Yin J, Liu J, Sun X-Q and
Hang Y-Y (2020) Genome-Wide
Identification and Evolutionary
Analysis of *NBS-LRR* Genes From
Dioscorea rotundata.
Front. Genet. 11:484.
doi: 10.3389/fgene.2020.00484

Dioscorea rotundata is an important food crop that is mainly cultivated in subtropical regions of the world. *D. rotundata* is frequently infected by various pathogens during its lifespan, which results in a substantial economic loss in terms of yield and quality. The disease resistance gene (*R* gene) profile of *D. rotundata* is largely unknown, which has greatly hampered molecular study of disease resistance in this species. *Nucleotide-binding site-leucine-rich repeat (NBS-LRR)* genes are the largest group of plant *R* genes, and they play important roles in plant defense responses to various pathogens. In this study, 167 *NBS-LRR* genes were identified from the *D. rotundata* genome. Subsequently, one gene was assigned to the *resistance to powdery mildew8 (RPW8)-NBS-LRR (RNL)* subclass and the other 166 genes to the *coiled coil (CC)-NBS-LRR (CNL)* subclass. None of the *Toll/interleukin-1 receptor (TIR)-NBS-LRR (TNL)* genes were detected in the genome. Among them, 124 genes are located in 25 multigene clusters and 43 genes are singletons. Tandem duplication serves as the major force for the cluster arrangement of *NBS-LRR* genes. Segmental duplication was detected for 18 *NBS-LRR* genes, although no whole-genome duplication has been documented for the species. Phylogenetic analysis revealed that *D. rotundata* *NBS-LRR* genes share 15 ancestral lineages with *Arabidopsis thaliana* genes. The *NBS-LRR* gene number increased by more than a factor of 10 during *D. rotundata* evolution. A conservatively evolved ancestral lineage was identified from *D. rotundata*, which is orthologs to the *Arabidopsis RPM1* gene. Transcriptome analysis for four different tissues of *D. rotundata* revealed a low expression of most *NBS-LRR* genes, with the tuber and leaf displaying a relatively high *NBS-LRR* gene expression than the stem and flower. Overall, this study provides a complete set of *NBS-LRR* genes for *D. rotundata*, which may serve as a fundamental resource for mining functional *NBS-LRR* genes against various pathogens.

Keywords: *Dioscorea rotundata*, *NBS-LRR* genes, pathogen defense, *R* gene, genome-wide analysis

INTRODUCTION

Yams (*Dioscorea* spp.) are important food crops in tropical and subtropical regions of the world, belonging to the *Dioscorea* genus in the family Dioscoreaceae of the order Dioscoreales (Salawu et al., 2014). Their starchy tubers have high nutritional content, containing carbohydrates, vitamin C, essential minerals, and dietary fiber (Muzac-Tucker et al., 1993). It was proposed that three

yam crops, *D. alata*, *D. trifida*, and *D. rotundata*, were domesticated independently and widely cultivated in Asia, America, and Africa (Scarcelli et al., 2019). In West Africa, yams serve as essential food crops, ranking second after cassava (Salawu et al., 2014). However, the productivity of yams is threatened by various pests and microbial pathogens, including nematodes, fungi, bacteria, and viruses (Mignouna et al., 2001; Coyne et al., 2006; Oyelana et al., 2011). The diseases caused by these pathogens not only severely reduce production but also affect the quality of the edible tissues (Salawu et al., 2014). This collectively contributes to the economic loss of the farmers. In the past, several studies have tried to collect germplasm resistant to various pathogens and to map the resistance loci (Amusa, 2001; Mignouna et al., 2001, 2002; Bhattacharjee et al., 2018). However, no functional disease resistance gene (*R* gene) has been cloned from yam crops so far.

Plant *R* genes are a group of genes that specifically function against invading pathogens. During the past 30 years, over 300 *R* genes have been cloned from many plant species (Kourelis and van der Hoorn, 2018). These *R* genes encode proteins (*R* proteins) with diverse domain structures (Kourelis and van der Hoorn, 2018). Among them, genes encoding nucleotide-binding site (NBS) and leucine-rich repeat (LRR) domains represent the largest class of known *R* genes; these genes are named *NBS-LRR* genes. The translated *NBS-LRR* proteins are intracellular receptors that recognize the presence of pathogens. *NBS-LRR* genes originated anciently during the evolution of green plants (Shao et al., 2019). Three subclasses diverged soon after the origin of this gene family. The characteristic N-terminal domains, including the Toll/interleukin-1 receptor-like (TIR), coiled coil (CC), and resistance to powdery mildew8 (RPW8) domains, were found in the three subclasses. Accordingly, they were named as *TIR-NBS-LRR* (*TNL*), *CC-NBS-LRR* (*CNL*), and *RPW8-NBS-LRR* (*RNL*) genes, respectively (Shao et al., 2016). *TNL* and *CNL* proteins usually function as sensors to detect pathogens. The binding of the LRR domain to pathogen effectors causes a conformational change of the *TNL* or *CNL* protein, which subsequently induces multimerization of the TIR or CC domain, resulting in immune system activation (Andersen et al., 2018; Kourelis and van der Hoorn, 2018). Alternatively, some *TNL* and *CNL* proteins may be activated by monitoring the state of specific host proteins. When the state of these host proteins is altered by pathogen effectors, such as by phosphorylation or degradation, the *CNL* and *TNL* proteins will perceive it and are activated in a similar way to when directly activated (Kourelis and van der Hoorn, 2018). The *RNL* subclass has two lineages, each named by a functional gene, namely *ADR1* and *NRG1* (Collier et al., 2011; Shao et al., 2016). Both *ADR1* and *NRG1* proteins function in immune signal transduction but not pathogen detection. Furthermore, *NRG1* proteins were found to act specifically in *TNL* signal transduction (Peart et al., 2005; Qi et al., 2018; Castel et al., 2019; Wu et al., 2019).

Generally, plant genomes harbor from dozens to more than a thousand *NBS-LRR* genes (Gu et al., 2015; Li et al., 2016; Shao et al., 2019). The maintenance of such a large number of *R* genes reflects a consequence of the long-term arms race

between the plant and pathogens. Genomic and evolutionary studies have provided insights into how functional *R* genes were generated and preserved during plant evolution. Collective studies revealed that a considerable number of *NBS-LRR* genes are clustered on the chromosome, which is a result of frequent tandem duplication events (Meyers et al., 2003; Shao et al., 2014). The clustering organization provides a unique opportunity for creating high-sequence diversity and generating functional *NBS-LRR* genes (Kuang et al., 2004; Wroblewski et al., 2007). Understanding genomic organization and evolutionary patterns has greatly promoted functional *R* gene identification and utilization in rice, soybean, and many other crops (Ashfield et al., 2012; Shao et al., 2014; Zhang et al., 2015). Therefore, elucidating the complete profile of *NBS-LRR* genes in a plant genome would be of great help for the mining and utilization of functional *R* genes.

D. rotundata (white Guinea yam) is the most popular yam species cultivated in the West and Central Africa. The publication of the *D. rotundata* genome provides a valuable resource for understanding the *R* gene profile in this important crop (Tamiru et al., 2017). Here, a systematic analysis was performed to understand the domain structure, chromosomal organization, evolutionary mechanism, duplication type, and expression pattern. The bioinformatic analysis of the *NBS-LRR* gene profile in this study provides a fundamental resource for further mining functional *R* genes in *D. rotundata* and understanding the evolutionary pattern of this gene family.

RESULTS

Identification of *NBS-LRR* Genes From the *D. rotundata* Genome

A total of 167 *NBS-LRR* genes (Supplementary Table S1) were identified from the *D. rotundata* genome using previously described criteria (see details in Materials and Methods section), accounting for approximately 0.6% of the 26,198 annotated genes. To assign these *NBS-LRR* genes to different subclasses, the protein sequences of all identified *NBS-LRR* genes were subjected to BLASTp analysis against the well-defined *Arabidopsis thaliana* *NBS-LRR* proteins (Zhang et al., 2016). The results showed that 166 of the 167 *D. rotundata* *NBS-LRR* genes belong to the *CNL* subclass, whereas only one belongs to the *RNL* subclass. None of the *TNL* genes were detected in the *D. rotundata* genome, which is consistent with reports of other monocot genomes that all lack *TNL* genes (Shao et al., 2016; Xue et al., 2020). One gene (*Dr02646.1*) encoding an atypical TIR domain and an atypical NBS domain was detected in the *D. rotundata* genome. However, this gene should be assigned to the *XTNX* gene family, not the *NBS-LRR* gene family according to the criteria described in our previous study (Zhang et al., 2017).

Based on the domain combinations of the translated proteins, the 167 *NBS-LRR* genes were classified into six groups as illustrated in Figure 1. Two of the six groups contain intact *RNL* (one gene) or *CNL* (64 genes) genes, respectively. Genes in these two groups each encode an NBS domain, an LRR domain, and an N-terminal RPW8 or

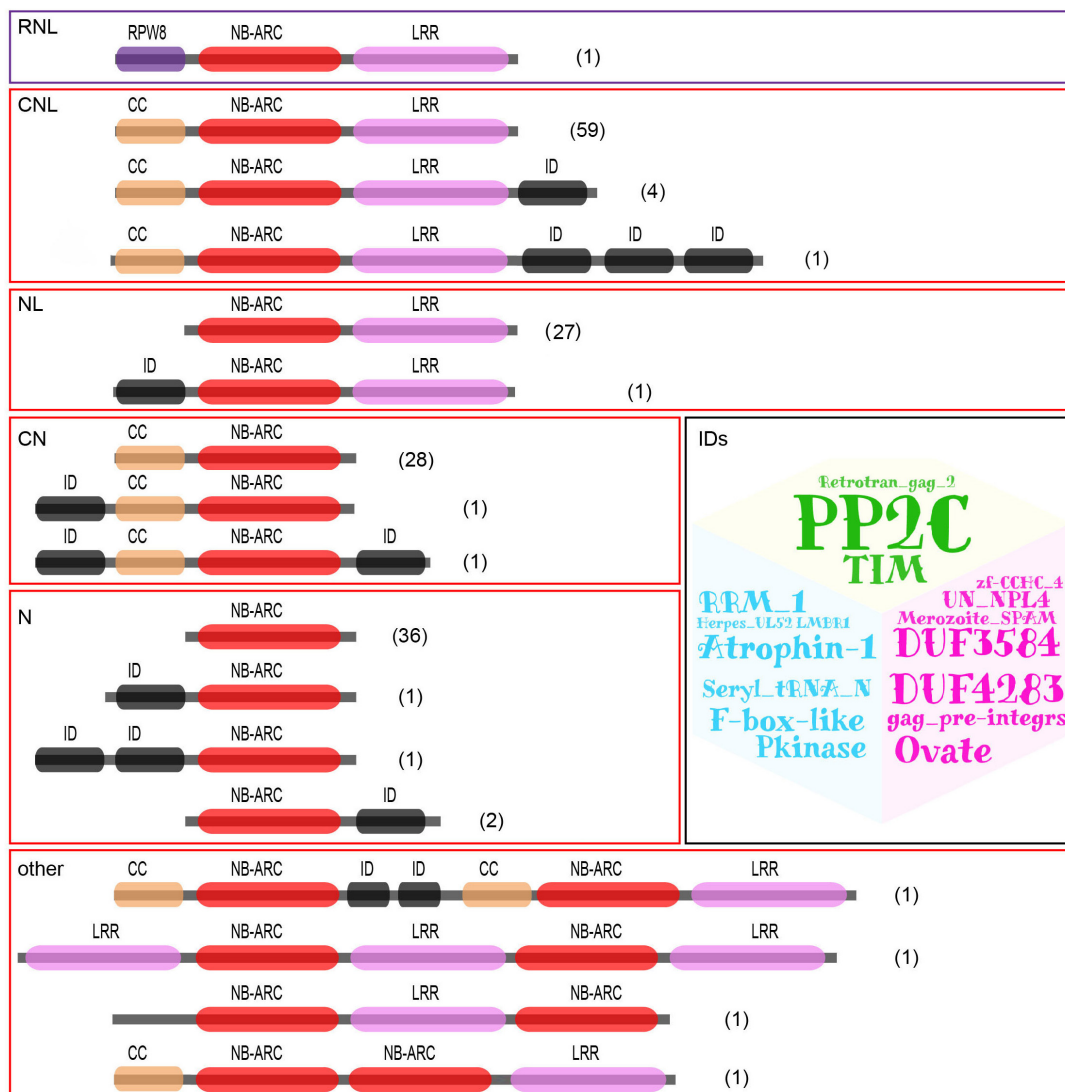


FIGURE 1 | Identification and classification of *D. rotundata* NBS-LRR genes based on protein domain structure. 167 NBS-LRR genes were identified from the *D. rotundata* genome and classified into six groups with different domain compositions and arrangements. The identified integrated domains (IDs) are also listed as a word cloud.

CC domain. Three other groups contain partial *CNL* genes, namely, *NL* (28 genes), *CN* (30 genes), and *N* (40 genes). These genes lack the N-terminal CC domain, the C-terminal LRR domain, or both, respectively. The remaining group of genes, namely, “others,” contains genes that also encode *CNL* proteins, yet have a complicated domain arrangement. Besides the characteristic domains regularly found in NBS-LRR proteins, 16 different integrated domains were detected to be encoded by 15 genes from the five *CNL* groups (Supplementary Table S1).

MEME analysis was performed on the amino acid sequence of NBS domain of *CNL* genes. The result showed that the highly conserved amino acid sequence of “GKTTLA,” “GLPL,” “DDVW,” and “TTR” at the four motifs P-loop, GLPL, Kinase-2 and RNBS-B are readily detected in *D. rotundata* *CNL* genes

(Supplementary Table S2). Furthermore, the “DDVW” region in the Kinase-2 motif is conserved in both *CNL* and *RNL* genes as has been reported in other angiosperms (Shao et al., 2016).

Chromosomal Distribution of *D. rotundata* NBS-LRR Genes

The 167 identified NBS-LRR genes were plotted against the 21 *D. rotundata* chromosomes based on their physical locations retrieved from the GFF3 file. NBS-LRR genes within an interval of less than 250 kb were treated as a cluster (Ameline-Torregrosa et al., 2008). The result showed that *D. rotundata* NBS-LRR genes are unevenly distributed on 17 of the 21 chromosomes (Figure 2). More than ten NBS-LRR genes were detected on chromosomes 2, 3, 7, 8, 13, and 16, whereas

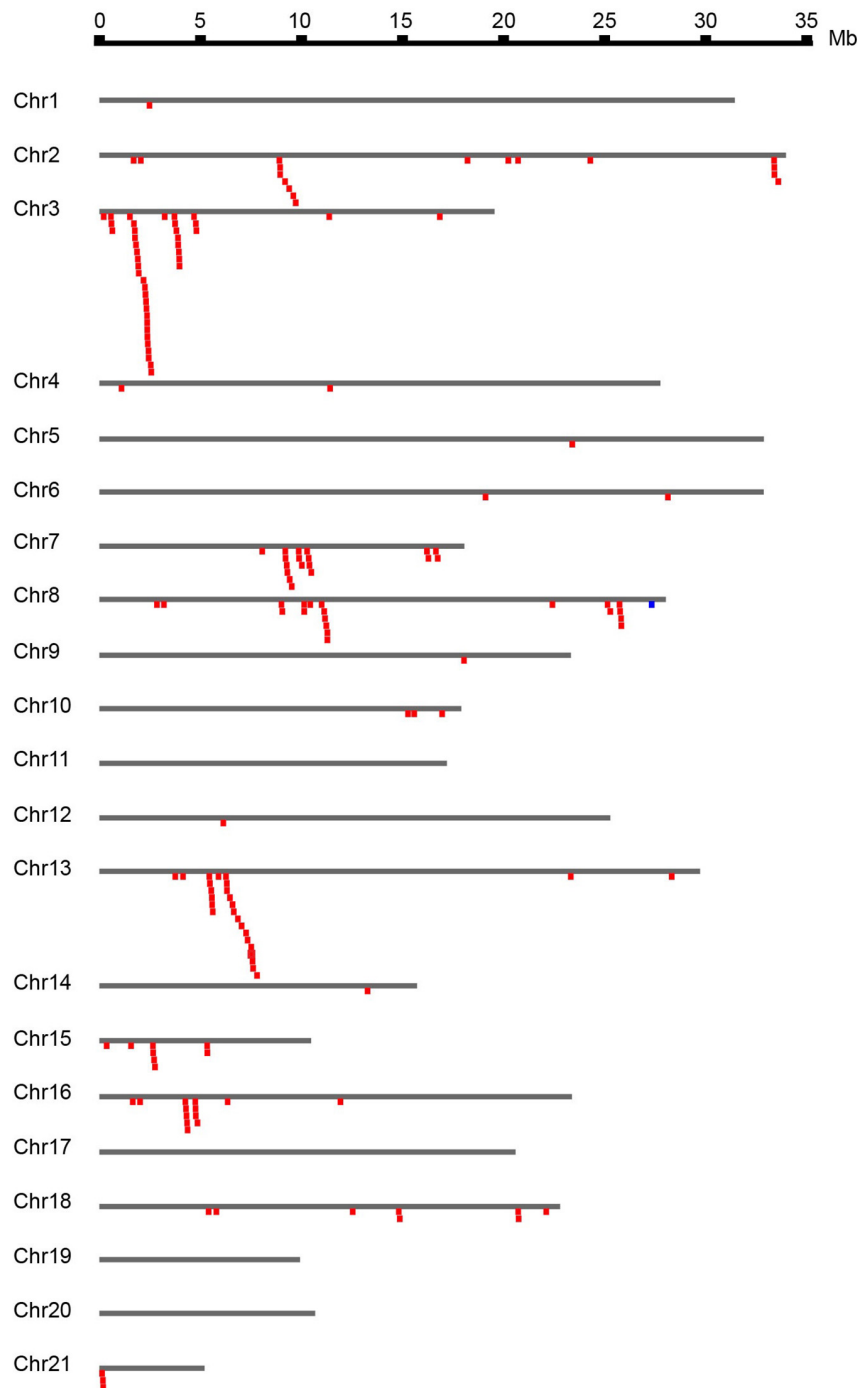


FIGURE 2 | Chromosomal distribution of *D. rotundata* NBS-LRR genes. The 167 identified NBS-LRR genes are plotted against the *D. rotundata* chromosomes based on their physical locations retrieved from the GFF3 file. NBS-LRR genes within an interval of less than 250 kb were treated as a cluster (Ameline-Torregrosa et al., 2008).

only one NBS-LRR gene was detected on chromosomes 1, 5, 9, 12, and 14. No NBS-LRR genes were detected on chromosomes 11, 17, 19, and 20. No significant correlation was detected between the chromosomal length and the NBS-LRR gene number.

Based on the physical locations, the NBS-LRR genes on the 17 chromosomes were classified into 68 loci, including 43 singletons and 25 multigene clusters. The result demonstrated that 124 NBS-LRR genes are present in the 25 clusters. On average, there are five genes per cluster. Among the 25 defined clusters, the

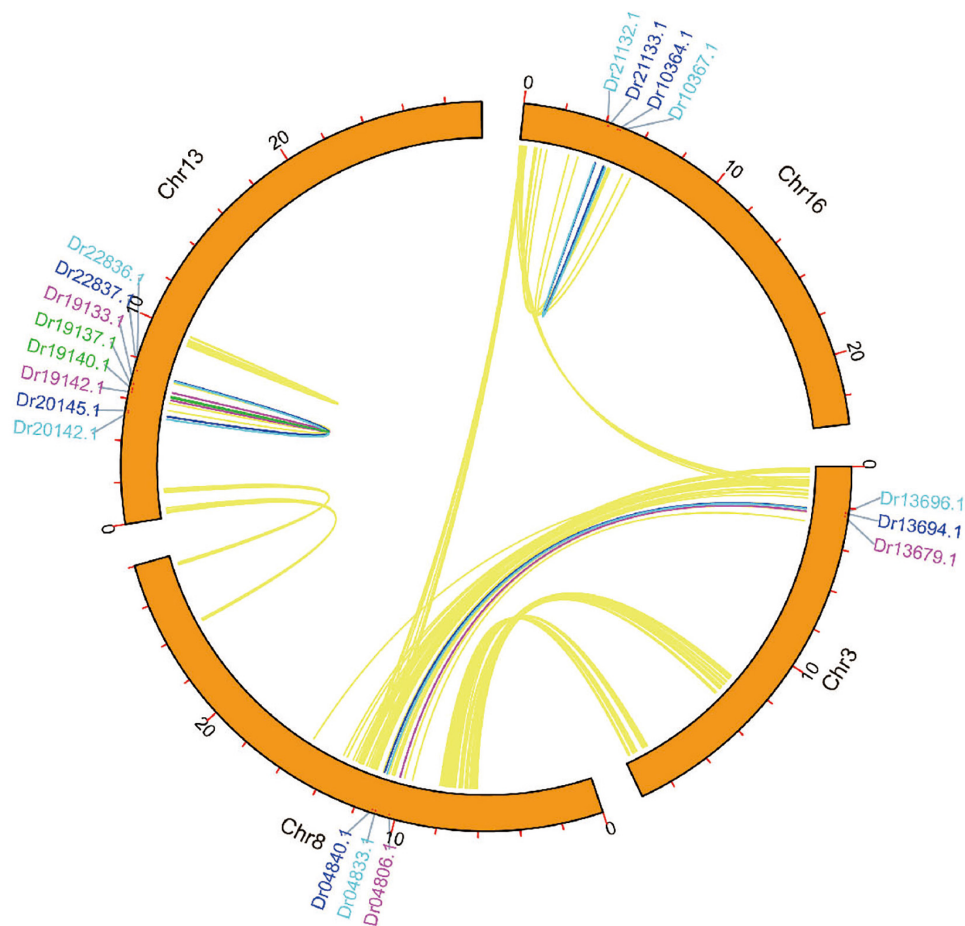


FIGURE 3 | Syntenic relationship of the 18 segmental-duplicated *NBS-LRR* genes.

smallest ones only have two adjacent genes, including loci 27 and 28 on chromosome 7, loci 31, 32, and 36 on chromosome 8, locus 55 on chromosome 15, and loci 65 and 66 on chromosome 18. The largest cluster was locus 12 on chromosome 3, which has 23 *NBS-LRR* genes.

Different Types of Gene Duplications Contributed to *NBS-LRR* Gene Expansion

A large number of *NBS-LRR* genes were present in clusters, suggesting that tandem duplication plays an important role in the *D. rotundata* *NBS-LRR* gene expansion. Thus, the output of different duplication types was detected. The result showed that 108 of the 167 genes were duplicated through tandem duplications, 18 resulted from segmental duplications, and 41 were from ectopic or dispersed duplication. It is interesting that over 10% of the *NBS-LRR* genes originated from segmental duplications, whereas no whole genome duplications have been detected for the *D. rotundata* genome (Tamiru et al., 2017). Further analysis revealed that the 18 segmental duplicated genes are related to three segmental duplication events (Figure 3). One

of them occurred between chromosomes 3 and 18, and resulted in duplication of three genes to form six. The remaining two events were intra-chromosomal small-scale inversions in chromosomes 13 and 16, resulting in the doubling of four and two ancestral *NBS-LRR* genes, respectively.

Phylogenetic and Ka/Ks Analysis of *NBS-LRR* Genes From *D. rotundata*

To trace the evolutionary history of *D. rotundata* *NBS-LRR* genes, phylogenetic analysis was performed by incorporating *NBS-LRR* genes from the other Dioscoreaceae species *Trichopus zeylanicus* (Supplementary Table S3) and CNL and RNL genes from a dicot species *Arabidopsis thaliana* (Zhang et al., 2016). *TNL* genes of *A. thaliana* were not included in the analysis, because no *NBS-LRR* genes of this subclass were found in *D. rotundata*. The phylogenetic result (Figure 4 and Supplementary Figure S1) showed that *RNL* genes from the three species form an independent clade with a high support value, which corresponds to the CNL-A clade of *A. thaliana* *NBS-LRR* phylogeny constructed by Meyers et al. (2003). The topology supports the ancient divergence of the *RNL* and *CNL*

subclasses documented by other studies (Shao et al., 2016, 2019). *RNL* genes from the three species further separated into two lineages, the *ADR1* and *NRG1* lineages. The *RNL* genes from *D. rotundata* and *T. zeylanicus* form a highly supported lineage with *Arabidopsis ADR1* genes, suggesting loss of *NRG1* genes in the two species. This is in accordance with previous reported loss of *NRG1* genes in other monocot species (Shao et al., 2016). *CNL* genes from the three species also form two well-supported large clades (Figure 4). *D. rotundata* and *T. zeylanicus* genes in the first clade cluster with *A. thaliana* *CNL-B* clade genes, whereas *D. rotundata* and *T. zeylanicus* genes in the second clade cluster with *A. thaliana* *CNL-C* and *D* genes (Meyers et al., 2003).

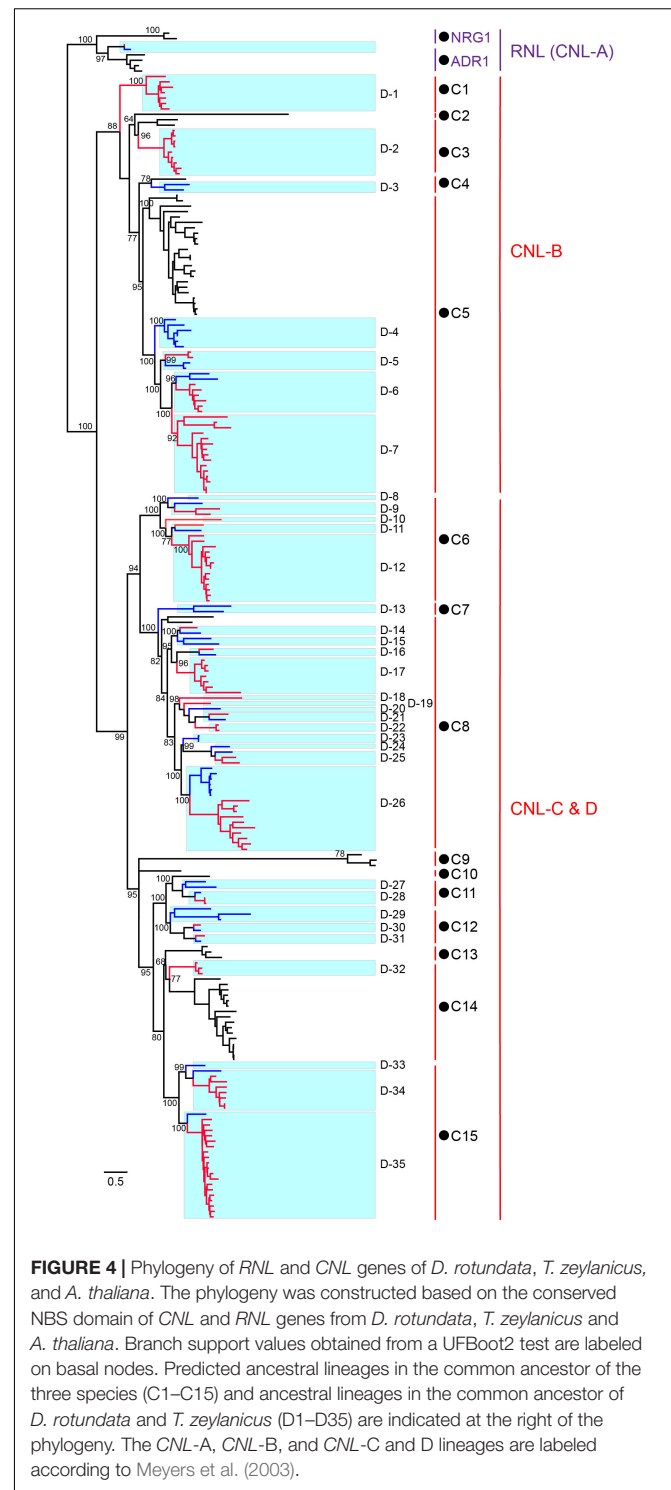
Reconciling the *NBS-LRR* phylogeny revealed that *CNL* genes from the two *Dioscoreaceae* species and *A. thaliana* were derived from 15 ancestral lineages of the progenitor before the divergence of monocot and dicot plants (Figure 4). Among the 15 ancestral lineages, six were inherited by both *A. thaliana* and the *Dioscoreaceae* species (lineages 3, 4, 5, 8, 11, and 14). The lineage 5 has expanded a lot in both *A. thaliana* and the *Dioscoreaceae* species. In comparing, the lineage 3 and lineage 8 were only expanded in the *Dioscoreaceae* species. The lineage 11 was conservatively evolved in all of the three species, with only one to three genes in each genome. It is also worth noting that lineage 11 includes a functional *A. thaliana* *R* gene against *Pseudomonas syringae*, *RPM1*. Among the remaining nine lineages, four of them were inherited by *A. thaliana*, whereas five were inherited by the two *Dioscoreaceae* species (Figure 4).

In total, 11 of the 15 ancestral *CNL* lineages that emerged in the common ancestor of *A. thaliana* and the *Dioscoreaceae* species were inherited by the *Dioscoreaceae* species. These ancestral *CNL* lineages further diverged into 35 sub-lineages (D-1 to D-35) in the common ancestor of the two *Dioscoreaceae* species (Figure 4). Among them, eight sub-lineages (D1, D2, D6, D7, D12, D17, D26, and D35) had experienced considerable duplications in *D. rotundata*, resulting in a large number of descendant genes in the modern genome. Overall, the phylogenetic analysis revealed that the 11 ancestral *CNL* lineages that present in the common ancestor of monocots and dicots, have experienced step-wise expansion during *D. rotundata* evolution. This contributed to the expansion of the *NBS-LRR* gene number in *D. rotundata* to over ten times that of its ancestor.

The non-synonymous substitution to synonymous substitution (*Ka/Ks*) ratio is an informative value of positive selection. To detect whether some *NBS-LRR* genes are under positive selection, *Ka/Ks* analysis was performed on *D. rotundata* *NBS-LRR* genes from all aforementioned sub-families. The result showed that *Ka/Ks* for all but four gene pairs were less than one, indicating the majority of duplicated genes underwent purifying selection (Supplementary Table S4).

Expression Profile of *NBS-LRR* Genes From *D. rotundata*

To obtain the expression pattern of *NBS-LRR* genes in *D. rotundata*, the transcriptome data of four *D. rotundata* tissues from the public database were analyzed. The result showed that most *NBS-LRR* genes are not expressed or are only expressed



at very low levels in all of the tissues studied (flower, leaf, tuber, and stem). However, the expression of some *NBS-LRR* genes could reach 100 fragments per kilobase million (FPKM; Figure 5 and Supplementary Table S5). Furthermore, the high expression of some *NBS-LRR* genes is often tissue specific. For example, among the five genes that were expressed at a level

of more than 100 FPKM, four were highly expressed only in the tuber tissue.

The expression of all detected *NBS-LRR* genes was compared among the four tissues. The average expression value of the 167 genes is 2.9, 6.6, 9.9, and 4.6 FPKM in the flower, leaf, tuber, and stem, respectively. It is obvious that the tuber and leaf have higher average *R* gene expression levels than the stem and flower, although this difference is not statistically significant (Figure 5). The highest expression value of each gene was detected among the four tissues. The result showed that among the 146 genes that were expressed in at least one tissue, 78 genes show the highest expression value in the tuber, 29 in the leaf, 28 in the stem, and 11 in the flower (Figure 5). This mirrors the situation of the average expression level of all *R* genes among the four tissues. Overall, the expression analysis indicated that *NBS-LRR* genes in *D. rotundata* are expressed at a low level, with some genes showing high expression in specific tissues.

DISCUSSION

NBS-LRR genes are the largest group among all plant disease resistance genes, and they play vital roles in plant defense against various pathogens. Defining a complete set of *NBS-LRR* genes in a species is not only helpful for obtaining new insights into the evolution of this important gene family but also practical for the identification and utilization of functional *R* genes from the species and its close relatives (Meyers et al., 2003; Zhang et al., 2015). Genome-wide identification and evolutionary analysis has been performed in many angiosperms in the past 20 years (Bai et al., 2002; Meyers et al., 2003; Shao et al., 2016, 2019; Zhang et al., 2016; Neupane et al., 2018a). Several evolutionary features have been documented, including frequent tandem duplication for gene expansion, cluster organization on the chromosome, rapid species-specific gene loss, and duplication (Meyers et al., 2003; Gu et al., 2015; Shao et al., 2016; Die et al., 2018; Song et al., 2019). However, most of the previous studies have concentrated on dicots, especially the rosoid lineage of the angiosperms. Only a few monocot species, mainly in the grass family, have been investigated. A recent study analyzed *NBS-LRR* genes in several genomes of the orchid family, which increased the catalog of analyzed lineages (Xue et al., 2020). In this study, this list was further expanded by analyzing *NBS-LRR* genes from *D. rotundata*, a species from an early diverged monocot lineage.

A comprehensive analysis of the 167 identified *NBS-LRR* genes clearly recovered previously documented evolutionary features of the *NBS-LRR* genes. The data revealed that 124 of the identified *NBS-LRR* genes are present within 25 clusters on the chromosomes. Furthermore, the cluster distribution of the *NBS-LRR* genes is consistent with their duplication mechanisms. The proportion of *NBS-LRR* genes that clustered reached 74%, which is higher than that of *A. thaliana* (Meyers et al., 2003). Segmental duplication of *NBS-LRR* genes is also frequently found in species that have recently experienced whole genome duplications (Shao et al., 2014). In this study, three segmental duplications involving 18 *NBS-LRR* genes were detected in *D. rotundata*, although no recent whole genome duplication has been recorded for this

species (Tamiru et al., 2017). This result suggested that small-scale segmental duplications also play a role in *NBS-LRR* gene expansion. Identification and phylogenetic analysis of *RNL* genes from *D. rotundata* supports the stance that the *NRG1* lineage has been lost in monocot lineages (Collier et al., 2011; Shao et al., 2016). Several recent studies have shown that many *TNL* proteins rely on *NRG1* to transduce immune signals (Peart et al., 2005; Qi et al., 2018; Castel et al., 2019; Wu et al., 2019). The functional codependence of *TNL* and *NRG1* was further strengthened by our observation that *NRG1* and *TNL* genes are co-absent in *D. rotundata*.

The arms race between *NBS-LRR* genes and plant pathogens drives rapid turnover of *NBS-LRR* profiles in a species. Therefore, conserved *NBS-LRR* lineages across different species are present at very low levels. The analysis of four legume species that diverged 54 million years ago revealed that over 94% (112 of 119) of ancestral *NBS-LRR* lineages experienced deletions or significant expansions during speciation. Meanwhile, only seven ancestral lineages were maintained in a conservative manner (Shao et al., 2014). In this study, the phylogenetic analysis of *NBS-LRR* genes from *D. rotundata* and *A. thaliana* revealed two ancestral *RNL* genes and 15 ancestral *CNL* genes between these two species that diverged over 100 million years ago. It is not surprising to see the loss of the *NRG1* lineage and the preservation of only one copy of the *ADR1* lineage *RNL* gene, because of their specific function in signal transduction rather than pathogen detection (Peart et al., 2005; Collier et al., 2011; Qi et al., 2018; Castel et al., 2019; Wu et al., 2019). Among the 15 ancestral *CNL* lineages, nine were lost in one of the two species. Only two of the remaining lineages were conservatively inherited by both species, whereas five lineages were expanded greatly in at least one species. It is interesting to find that the *A. thaliana* *RPM1* is located in one of the conservatively evolved ancestral lineages (lineage 11). *RPM1* has been proposed as an anciently originated *NBS-LRR* gene that defends against *P. syringae* (Mackey et al., 2002; Shao et al., 2016). *RPM1* recognizes infection of the pathogen by monitoring a host protein, *RIN4*. Although maintenance of *RPM1* requires a high fitness cost (Tian et al., 2003), this gene has not been erased during long-term evolution, suggesting the importance of this anciently originated *R* gene and its functional mechanism. The finding of two *RPM1* orthologs in *D. rotundata* suggests that a function and mechanism similar to that of *RPM1* may have been adopted by monocot plants as well. It would be very interesting if this could be validated, because no *NBS-LRR* genes other than *RNLs* have been evidenced to maintain their function for such a long evolutionary time.

NBS-LRR genes are invaluable resources for mining functional *R* genes. A previous study in *D. alata* has tried to isolate *NBS-LRR* genes against Anthracnose by PCR (Saranya et al., 2016). However, PCR analysis can only be designed to amplify sequences encoding the conserved *NBS* domain without a reference genome. The full list of 167 *NBS-LRR* genes obtained in this study may serve as templates for mining full length orthologous or homologous *NBS-LRR* genes from *D. alata* and other yam crops. The results from this study provide a fundamental resource for molecular breeding of *D. rotundata*

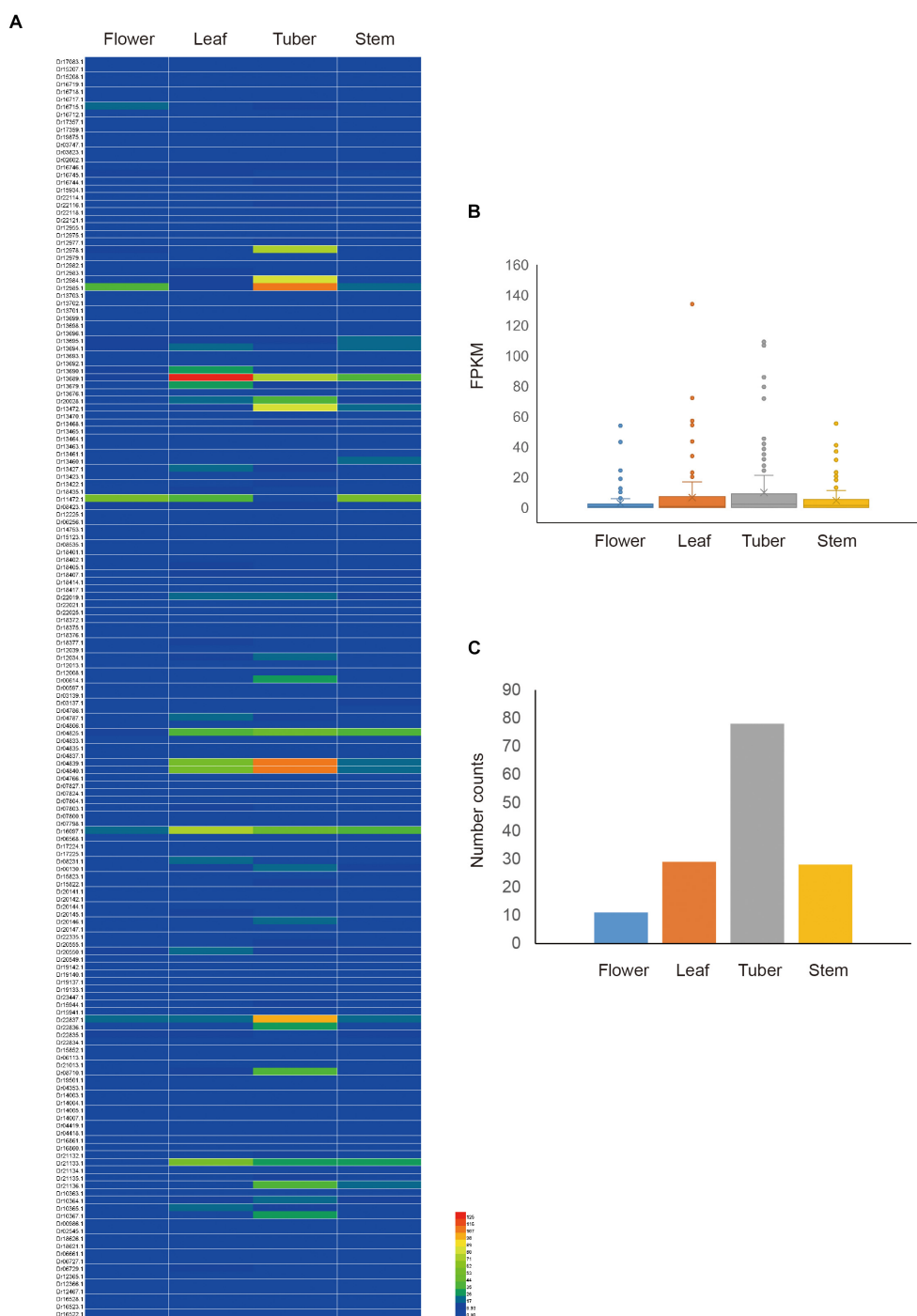


FIGURE 5 | Expression pattern of *D. rotundata* NBS-LRR genes. **(A)** Heatmap of the expression of 167 NBS-LRR genes in four different plant tissues. **(B)** Average expression of 167 NBS-LRR genes in the four tissues. **(C)** Distribution of the top expression tissue for the 167 NBS-LRR genes.

and its relatives that have not yet been sequenced. The clustering of NBS-LRR genes on chromosomes has been associated with generating high sequence diversity and functional genes against various pathogens (Kuang et al., 2004; Wroblewski et al., 2007). In soybean, 54 million years of evolution has enabled one ancestral gene to be tandemly duplicated into more than ten offspring on

chromosome 13, from which resistance has evolved to several different pathogens, including bacteria and different viruses (Ashfield et al., 2012; Shao et al., 2014; Nepal and Benson, 2015; Neupane et al., 2018b). Therefore, NBS-LRR clusters are excellent loci for mining functional *R* genes. In the present study, 25 clusters in *D. rotundata* aggregated 124 of the 167 NBS-LRR

genes. Six chromosomes were found to have *NBS-LRR* clusters possessing more than five genes each (Figure 2). These clusters may serve as candidates for mining functional *R* genes against *D. rotundata* pathogens. However, the role of singleton *NBS-LRR* loci should not be neglected. It will be helpful to link the *NBS-LRR* locus identified in this study when genetic mapping is used for *R* gene discovery in *D. rotundata*.

In summary, the present study identified a complete set of 167 *NBS-LRR* genes from the *D. rotundata* genome. The genomic organization and evolutionary pattern were comprehensively revealed by integrating different analysis tools. These results may serve as a fundamental resource for the molecular breeding of *D. rotundata*.

MATERIALS AND METHODS

Data Used in This Study

Genome sequence and annotation files of *D. rotundata* were downloaded from the Ensembl database¹. The raw RNA-seq data (accession numbers: DRX040448, Flower; DRX040449, Leaf; DRX040450, Tuber; and DRX040451, Stem) generated by Tamiru et al. (2017) was downloaded from the National Center for Biotechnology Information (NCBI) sequence read archive (SRA) database. *Arabidopsis thaliana* *NBS-LRR* genes were retrieved from our previous study (Zhang et al., 2016).

Identification of *NBS-LRR* Genes

BLAST and hidden Markov models search (HMMsearch) methods were used to identify *NBS-LRR* genes in the *D. rotundata* genome as described previously (Zhang et al., 2016). Briefly, the amino acid sequence of the NB-ARC domain (Pfam accession number: PF00931) was used as a query to search for *NBS-LRR* proteins using the BLASTp program of the NCBI BLAST software; the threshold expectation value was set to 1.0. Simultaneously, the protein sequences of *D. rotundata* were scanned by HMMsearch using the HMM profile of the NB-ARC domain as a query with an E-value setting of 1.0. Then, the results from the two methods were merged to produce the maximum number of *NBS-LRR* genes. In order to confirm the presence of the NBS domain, a round of HMMscan was performed for all the obtained hits against the Pfam-A database (E-value set to 0.0001). Genes without a conserved NBS domain were removed from the datasets. All of the non-redundant candidate sequences were compared with the NCBI Conserved Domains Database (CDD)² and the MARCOIL server³ to further verify the CC, TIR (Pfam accession number: PF01582), RPW8 (Pfam accession number: PF05659), LRR, and other integrated domains.

MEME analysis (Bailey et al., 2009) was performed to discover conserved motifs in the NBS domain of the identified *NBS-LRR* genes. The number of displayed motifs was set to 20 with all other parameters default settings as described by Nepal et al. (2017).

¹ http://ftp.ensemblgenomes.org/pub/plants/release-44/fasta/dioscorea_rotundata

² <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

³ https://bcf.isb-sib.ch/webmarcoil/webmarcoilC1.html?tdsourcetag=s_pcqq_aiomsg

Distribution of *NBS-LRR* Genes in Different Chromosomes

To determine the distribution of the *NBS-LRR* genes on the chromosomes of the *D. rotundata* genome, the GFF3 annotation file was parsed to extract the genomic locations of the *NBS-LRR* genes. A sliding window analysis was performed with a window size of 250 kb to identify the number of genes that appeared in a cluster on a chromosome as described by Ameline-Torregrosa et al. (2008). If two successive annotated *NBS-LRR* genes were located within 250 kb on a chromosome, they were considered as clustered.

Phylogenetic and Ka/Ks Analysis

Sequence alignment and phylogenetic analysis were performed as described by Xue et al. (2020). Briefly, amino acid sequences of the conserved NBS domain of the identified *NBS-LRR* genes were aligned using ClustalW with default options, and then manually corrected in MEGA 7.0 (Kumar et al., 2016). Too short or extremely divergent sequences were excluded from the analysis. Phylogenetic analysis was carried out with IQ-TREE using the maximum likelihood method (Nguyen et al., 2015) after selecting the best-fit model using ModelFinder (Kalyanamoorthy et al., 2017). Branch support values were estimated using UFBoot2 tests (Minh et al., 2013). Reconcile the phylogeny was performed as described in our previous studies (Shao et al., 2014; Zhang et al., 2016) to reconstruction the ancestral state of the *NBS-LRR* genes.

The Ka and Ks were calculated for gene pairs within each *NBS-LRR* subfamilies. The nucleotide coding sequences (CDSs) of each subfamily were aligned by MEGA 7.0 (Kumar et al., 2016) and the values of Ka, Ks, and Ka/Ks were calculated by DnaSP (Librado and Rozas, 2009).

Synteny and Gene Duplication Analysis

Pair-wise all-against-all BLAST was performed for the *D. rotundata* protein sequences. The obtained results and the GFF annotation file were then subjected to MCScanX for microsynteny detection and determination of the gene duplication type (Wang et al., 2012). Microsynteny relationships were displayed using TBtools⁴.

Gene Expression Analysis

To analyze the expression of *D. rotundata* *NBS-LRR* genes, the RNA-seq data of various tissues were downloaded from GenBank and checked with FastQC software⁵ to avoid containing adapter or low-quality reads. Clean reads from each sample were mapped to the reference genome of *D. rotundata* using TopHat with default settings (Trapnell et al., 2012; Kim et al., 2013). The mapping results were subjected to Cufflinks to assemble transcripts in each sample and then merged into one cohesive set using Cuffmerge. The expression of each gene was evaluated using Cuffdiff (Trapnell et al., 2012). All analyses by cufflinks were performed with default settings. A gene with the FPKM value larger than 100 was recognized as a high expression gene in the analysis.

⁴ <https://github.com/CJ-Chen/TBtools>

⁵ <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

Y-MZ and Y-YH conceived and designed the project. Y-MZ obtained and analyzed the data and wrote the manuscript. MC, LS, YW, JY, JL, and X-QS participated in the data analysis and discussion. Y-YH revised the manuscript. All authors contributed to discussion of the results, reviewed the manuscript and approved the final article.

FUNDING

This work was supported by the National Natural Science Founding of China (31500191) to Y-MZ and the Natural Science Founding of Jiangsu Province (BK20180316) to MC.

REFERENCES

- Ameline-Torregrosa, C., Wang, B. B., O'Blenness, M. S., Deshpande, S., Zhu, H., Roe, B., et al. (2008). Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol.* 146, 5–21. doi: 10.1104/pp.107.104588
- Amusa, N. (2001). Screening of cassava and yam cultivars for resistance to anthracnose using toxic metabolites of *Colletotrichum* species. *Mycopathologia* 150, 137–142. doi: 10.1023/A:1010955402457
- Andersen, E. J., Ali, S., Byamukama, E., Yen, Y., and Nepal, M. P. (2018). Disease resistance mechanisms in plants. *Genes* 9:E339. doi: 10.3390/genes9070339
- Ashfield, T., Egan, A. N., Pfeil, B. E., Chen, N. W., Podicheti, R., Ratnaparkhe, M. B., et al. (2012). Evolution of a complex disease resistance gene cluster in diploid phaseolus and tetraploid glycine. *Plant Physiol.* 159, 336–354. doi: 10.1104/pp.112.195040
- Bai, J., Pennill, A., Ning, J., Lee, S. W., Ramalingam, J., Webb, C. A., et al. (2002). Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Res.* 12, 1871–1884. doi: 10.1101/gr.454902
- 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
- Bhattacharjee, R., Nwadi, C. O., Saski, C. A., Paterne, A., Scheffler, B. E., Augusto, J., et al. (2018). An EST-SSR based genetic linkage map and identification of QTLs for anthracnose disease resistance in water yam (*Dioscorea alata* L.). *PLoS One* 13:e0197717. doi: 10.1371/journal.pone.0197717
- Castel, B., Ngou, P. M., Cevik, V., Redkar, A., Kim, D. S., Yang, Y., et al. (2019). Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. *New Phytol.* 222, 966–980. doi: 10.1111/nph.15659
- Collier, S. M., Hamel, L. P., and Moffett, P. (2011). Cell death mediated by the N terminal domains of a unique and highly conserved class of NB-LRR protein. *Mol. Plant Microbe Interact* 24, 918–931. doi: 10.1094/MPMI-03-11-0050
- Coyne, D. L., Tchabi, A., Baimey, H., Labuschagne, N., and Rotifa, I. (2006). Distribution and prevalence of nematodes (*Scutellonema bradys* and *Meloidogyne* spp.) on marketed yam (*Dioscorea* spp.) in West Africa. *Field Crop Res.* 96, 142–150. doi: 10.1016/j.fcr.2005.06.004
- Die, J. V., Castro, P., Millán, T., and Gil, J. (2018). Segmental and tandem duplications driving the recent NBS-LRR gene expansion in the asparagus genome. *Genes* 9:E568. doi: 10.3390/genes9120568

ACKNOWLEDGMENTS

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00484/full#supplementary-material>

FIGURE S1 | Detailed phylogeny of NBS-LRR genes from *Dioscorea rotundata*, *Trichopus zeylanicus* and *A. thaliana*.

TABLE S1 | Detailed features of identified NBS-LRR genes.

TABLE S2 | Discovered Motifs from NBS domain of *Dioscorea rotundata* NBS-LRR genes.

TABLE S3 | NBS-LRR genes identified from *Trichopus zeylanicus*.

TABLE S4 | Ka and Ks values of *Dioscorea rotundata* NBS-LRR pairs.

TABLE S5 | Expression of NBS-LRR genes in the four *Dioscorea rotundata* tissues.

- Gu, L., Si, W., Zhao, L., Yang, S., and Zhang, X. (2015). Dynamic evolution of NBS-LRR genes in bread wheat and its progenitors. *Mol. Genet. Genomics* 290, 727–738. doi: 10.1007/s00438-014-0948-8
- Kalyanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/Nmeth.4285
- Kim, D., Perte, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013, R36. doi: 10.1186/gb-2013-14-4-r36
- 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. *Plant Cell* 30, 285–299. doi: 10.1105/tpc.17.00579
- Kuang, H., Woo, S. S., Meyers, B. C., Nevo, E., and Michelmore, R. W. (2004). Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* 16, 2870–2894. doi: 10.1105/tpc.104.025502
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Li, P., Quan, X., Jia, G., Xiao, J., Cloutier, S., and You, F. M. (2016). RGAurgy: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genomics* 17:852. doi: 10.1186/s12864-016-3197-x
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Mackey, D., Holt, B. F., Wiig, A., and Dangel, J. L. (2002). RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell* 108, 743–754. doi: 10.1016/S0092-8674(02)00661-X
- 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
- Mignouna, H. D., Abang, M. M., Onasanya, A., Agindotan, B., and Asiedu, R. (2002). Identification and potential use of RAPD markers linked to Yam mosaic virus resistance in white yam (*Dioscorea rotundata*). *Ann. Appl. Biol.* 140, 163–169. doi: 10.1111/j.1744-7348.2002.tb00169.x

- Mignouna, H. D., Njukeng, P., Abang, M. M., and Asiedu, R. (2001). Inheritance of resistance to Yam mosaic virus, genus Potyvirus, in white yam (*Dioscorea rotundata*). *Theor. Appl. Genet.* 103, 1196–2000. doi: 10.1007/s001220100728
- Minh, B. Q., Nguyen, M. A. T., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024
- Muzac-Tucker, I., Asemota, H. N., and Ahmad, M. H. (1993). Biochemical composition and storage of Jamaican yams (*Dioscorea* sp.). *J. Sci. Food Agr.* 62, 219–224. doi: 10.1002/jsfa.2740620303
- Nepal, M. P., Andersen, E. J., Neupane, S., and Benson, B. V. (2017). Comparative genomics of Non-TNL disease resistance genes from six plant species. *Genes* 8:249. doi: 10.3390/genes8100249
- Nepal, M. P., and Benson, B. V. (2015). CNL disease resistance genes in soybean and their evolutionary divergence. *Evol. Bioinform Online* 11, 49–63. doi: 10.4137/EBO.S21782
- Neupane, S., Andersen, E. J., Neupane, A., and Nepal, M. P. (2018a). Genome-wide identification of NBS-encoding resistance genes in sunflower (*Helianthus annuus* L.). *Genes* 9:E384. doi: 10.3390/genes9080384
- Neupane, S., Ma, Q., Mathew, F. M., Varenhorst, A. J., Andersen, E. J., and Nepal, M. P. (2018b). Evolutionary divergence of TNL disease-resistant proteins in soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*). *Biochem. Genet.* 56, 397–422. doi: 10.1007/s10528-018-9851-z
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Oyelana, O. A., Durugbo, E. U., Olukanni, O. D., Ayodele, E. A., Aikulola, Z. O., and Adewole, A. I. (2011). Antimicrobial activity of ficus leaf extracts on some fungal and bacterial pathogens of *Dioscorea rotundata* from southwest nigeria. *J. Biol. Sci.* 11, 359–366. doi: 10.3923/jbs.2011.359.366
- Peart, J. R., Mestre, P., Lu, R., Malcuit, I., and Baulcombe, D. C. (2005). NRG1, a CC-NB-LRR protein, together with N, a TIR-NB-LRR protein, mediates resistance against tobacco mosaic virus. *Curr. Biol.* 15, 968–973. doi: 10.1016/j.cub.25005.04.053
- Qi, T., Seong, K., Thomazella, D. P. T., Kim, J. R., Pham, J., Seo, E., et al. (2018). NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant immunity in *Nicotiana benthamiana*. *Proc. Natl. Acad. Sci. U.S.A.* 115, E10979–E10987. doi: 10.1073/pnas.1814856115
- Salawu, M. B., Ibrahim, A. G., Lamidi, L. O., Salau, M. A., and Ogunleye, B. T. (2014). Economic analysis of yam processing in Oyo State. *Eur. J. Busin. Manag.* 39, 25–30.
- Saranya, G., Nath, V. S., Jeeva, M. L., Sheela, M. N., and Makesh Kumar, T. (2016). Mining of resistance gene analogues for anthracnose disease in greater yam (*Dioscorea alata* L.). *J. Root Crops* 42, 115–120.
- Scarcelli, N., Cubry, P., Akakpo, R., Thuillet, A. C., Obidiegwu, J., Baco, M. N., et al. (2019). Yam genomics supports West Africa as a major cradle of crop domestication. *Sci. Adv.* 5:eaaw1947. doi: 10.1126/sciadv.aaw1947
- Shao, Z. Q., Xue, J. Y., Wang, Q., Wang, B., and Chen, J. Q. (2019). Revisiting the origin of plant NBS-LRR Genes. *Trends Plant Sci.* 24, 9–12. doi: 10.1016/j.tplants.2018.10.015
- 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 genes reveal three anciently diverged classes with distinct evolutionary patterns. *Plant Physiol.* 170, 2095–2109. doi: 10.1104/pp.15.01487
- Shao, Z. Q., Zhang, Y. M., Hang, Y. Y., Xue, J. Y., Zhou, G. C., Wu, P., et al. (2014). Long-term evolution of nucleotide-binding site-leucine-rich repeat genes: understanding gained from and beyond the legume family. *Plant Physiol.* 166, 217–234. doi: 10.1104/pp.114.243626
- Song, H., Guo, Z., Hu, X., Qian, L., Miao, F., Zhang, X., et al. (2019). Evolutionary balance between LRR domain loss and young NBS-LRR genes production governs disease resistance in *Arachis hypogaea* cv. Tifrunner. *BMC Genomics* 20:844. doi: 10.1186/s12864-019-6212-1
- Tamiru, M., Natsume, S., Takagi, H., White, B., Yaegashi, H., Shimizu, M., et al. (2017). Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. *BMC Biol.* 15:86. doi: 10.1186/s12915-017-0419-x
- Tian, D., Traw, M. B., Chen, J. Q., Kreitman, M., and Bergelson, J. (2003). Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* 423, 74–77. doi: 10.1038/nature01588
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016
- 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.
- Wroblewski, T., Piskurewicz, U., Tomczak, A., Ochoa, O., and Micheltore, R. W. (2007). Silencing of the major family of NBS-LRR-encoding genes in lettuce results in the loss of multiple resistance specificities. *Plant J.* 51, 803–818. doi: 10.1111/j.1365-313X.2007.03182.x
- Wu, Z., Li, M., Dong, O. X., Xia, S., Liang, W., Bao, Y., et al. (2019). Differential regulation of TNL-mediated immune signaling by redundant helper CNLs. *New Phytol.* 222, 938–953. doi: 10.1111/nph.15665
- Xue, J. Y., Zhao, T., Liu, Y., Liu, Y., Zhang, Y. X., Zhang, G. Q., et al. (2020) Genome-wide analysis of the nucleotide binding site leucine-rich repeat genes of four orchids revealed extremely low numbers of disease resistance genes. *Front. Genet.* 10:1286. doi: 10.3389/fgene.2019.01286
- Zhang, X., Yang, S., Wang, J., Jia, Y., Huang, J., Tan, S., et al. (2015). A genome-wide survey reveals abundant rice blast R genes in resistant cultivars. *Plant J.* 84, 20–28. doi: 10.1111/tpl.12955
- Zhang, Y. M., Shao, Z. Q., Wang, Q., Hang, Y. Y., Xue, J. Y., Wang, B., et al. (2016). Uncovering the dynamic evolution of nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes in Brassicaceae. *J. Integr. Plant Biol.* 58, 13. doi: 10.1111/jipb.12365
- Zhang, Y. M., Xue, J. Y., Liu, L. W., Sun, X. Q., Zhou, G. C., Chen, M., et al. (2017). Divergence and conservative evolution of XTNX genes in land plants. *Front. Plant Sci.* 8:1844. doi: 10.3389/fpls.2017.01844

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, Chen, Sun, Wang, Yin, Liu, Sun and Hang. 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 and Evolutionary Analyses Reveal the Powdery Mildew Resistance Gene *Pm21* Undergoing Diversifying Selection

Huagang He^{1*†}, Jian Ji^{1†}, Hongjie Li², Juan Tong¹, Yongqiang Feng¹, Xiaolu Wang³, Ran Han³, Tongde Bie⁴, Cheng Liu^{3*} and Shanying Zhu^{1,5*}

OPEN ACCESS

Edited by:

Zhu-Qing Shao,
Nanjing University, China

Reviewed by:

Diaoguo An,
Institute of Genetics and
Developmental Biology(CAS), China
Zujun Yang,
University of Electronic Science and
Technology of China, China
Dale Zhang,
Henan University, China

*Correspondence:

Huagang He
hghe@mail.ujs.edu.cn
Cheng Liu
lch6688407@163.com
Shanying Zhu
zhushanying@mail.ujs.edu.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: 17 March 2020

Accepted: 20 April 2020

Published: 12 May 2020

Citation:

He H, Ji J, Li H, Tong J, Feng Y,
Wang X, Han R, Bie T, Liu C and
Zhu S (2020) Genetic Diversity and
Evolutionary Analyses Reveal the
Powdery Mildew Resistance Gene
Pm21 Undergoing Diversifying
Selection. *Front. Genet.* 11:489.
doi: 10.3389/fgene.2020.00489

¹ School of Food and Biological Engineering, Jiangsu University, Zhenjiang, China, ² National Engineering Laboratory for Crop Molecular Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, ³ Crop Research Institution, Shandong Academy of Agricultural Sciences, Jinan, China, ⁴ Yangzhou Academy of Agricultural Sciences, Yangzhou, China, ⁵ School of Environment, Jiangsu University, Zhenjiang, China

Wheat powdery mildew caused by *Blumeria graminis* f. sp. *tritici* (*Bgt*) is a devastating disease that threatens wheat production and yield worldwide. The powdery mildew resistance gene *Pm21*, originating from wheat wild relative *Dasypyrum villosum*, encodes a coiled-coil, nucleotide-binding site, leucine-rich repeat (CC-NBS-LRR) protein and confers broad-spectrum resistance to wheat powdery mildew. In the present study, we isolated 73 *Pm21* alleles from different powdery mildew-resistant *D. villosum* accessions, among which, 38 alleles were non-redundant. Sequence analysis identified seven minor insertion-deletion (InDel) polymorphisms and 400 single nucleotide polymorphisms (SNPs) among the 38 non-redundant *Pm21* alleles. The nucleotide diversity of the LRR domain was significantly higher than those of the CC and NB-ARC domains. Further evolutionary analysis indicated that the solvent-exposed LRR residues of *Pm21* alleles had undergone diversifying selection (dN/dS = 3.19734). In addition, eight LRR motifs and four amino acid sites in the LRR domain were also experienced positive selection, indicating that these motifs and sites play critical roles in resistance specificity. The phylogenetic tree showed that 38 *Pm21* alleles were divided into seven classes. Classes A (including original *Pm21*), B and C were the major classes, including 26 alleles (68.4%). We also identified three non-functional *Pm21* alleles from four susceptible homozygous *D. villosum* lines (DvSus-1 to DvSus-4) and two susceptible wheat-*D. villosum* chromosome addition lines (DA6V#1 and DA6V#3). The genetic variations of non-functional *Pm21* alleles involved point mutation, deletion and insertion, respectively. The results also showed that the non-functional *Pm21* alleles in the two chromosome addition lines both came from the susceptible donors of *D. villosum*. This study gives a new insight into the evolutionary characteristics of *Pm21* alleles and discusses how to sustainably utilize *Pm21* in wheat production. This study also reveals the sequence variants and origins of non-functional *Pm21* alleles in *D. villosum* populations.

Keywords: *Dasypyrum villosum*, *Pm21* allele, genetic diversity, evolutionary analysis, diversifying selection, wheat powdery mildew resistance

INTRODUCTION

Dasyphyrum villosum L. Candargy ($2n = 2x = 14$, VV), a diploid species native to the Mediterranean region, is an important wild resource for the improvement of common wheat (*Triticum aestivum* L., $2n = 6x = 42$, AABBDD). *D. villosum* possesses good resistance to multiple wheat diseases, such as wheat spindle streak mosaic disease, eyespot, take-all, stem rust, stripe rust, and powdery mildew (Li and Zhu, 1999; De Pace et al., 2011; Wang et al., 2017). Four powdery mildew resistance (*Pm*) genes, *Pm21* (Chen et al., 1995), *PmV* (Li et al., 2005), *Pm55* (Zhang et al., 2016), and *Pm62* (Zhang et al., 2018), have been found in *D. villosum*. Among them, both *Pm21* and *PmV* are located on the short arm of chromosome 6V (6VS) and confer immunity to powdery mildew at the whole growth stages of wheat. *Pm55* and *Pm62* are mapped to the short arm of chromosome 5V (5VS) and the long arm of chromosome 2V (2VL), respectively, which provide powdery mildew resistance at the adult-plant stage.

Pm21 was originally transferred from an accession of *D. villosum*, collected from Cambridge Botanical Garden, United Kingdom, to durum wheat (*T. turgidum* var. *durum* L.), and then a translocation line of wheat-*D. villosum* T6AL-6VS carrying *Pm21* was further developed (Chen et al., 1995). Using this translocation line as the powdery mildew resistance source, more than 20 varieties have been developed and released in the middle and lower reaches of the Yangtze River Valley and the southwest wheat-producing area, the most rampant areas of powdery mildew in China, where some *Pm* genes, such as *Pm2a* and *Pm4a*, are gradually losing their resistance (Bie et al., 2015a).

Undoubtedly, *Pm21* is a very valuable gene that confers highly effective resistance to tested isolates of *Blumeria graminis* f. sp. *tritici* (*Bgt*). However, no recombination occurs between the alien chromosome arm 6VS carrying *Pm21* and the wheat homoeologous chromosome arms, which limits the genetic mapping and the cloning of *Pm21* in the wheat backgrounds (Zhu et al., 2018). Recently, four seedling-susceptible *D. villosum* lines were identified from the natural populations. Based on the fine genetic map constructed, the gene *Pm21* was cloned and confirmed to encode a single coiled-coil, nucleotide-binding site, leucine-rich repeat (CC-NBS-LRR) protein (He et al., 2017, 2018).

In the present study, we isolated the *Pm21* alleles from different resistant *D. villosum* accessions and determined their genetic diversity, non-synonymous and synonymous substitution rates and positive selection sites. On the other hand, *D. villosum* germplasms susceptible to powdery mildew are rare, and only four susceptible *D. villosum* lines (DvSus-1 to DvSus-4) and two wheat-*D. villosum* chromosome 6V disomic addition lines (DA6V#1 and DA6V#3) were identified (Qi et al., 1998; Liu et al., 2011; He et al., 2017). Understanding the reason that these *D. villosum* germplasms keep or lose their resistance to powdery mildew will be useful to extend the effective duration of *Pm21* in agriculture. We also detected the sequence variations of *Pm21* alleles in the above germplasms for tracing their origins in natural population of *D. villosum*.

MATERIALS AND METHODS

Plant Materials

Dasyphyrum villosum accessions were gifted from Germplasm Resources Information Network (GRIN), GRIN Czech, Genebank Information System of the IPK Gatersleben (GBIS-IPK), and Nordic Genetic Resource Center (NordGen). The wheat-*D. villosum* chromosome 6V disomic addition lines DA6V#1 and DA6V#3 were provided by GRIN and Dr. Bernd Friebe (Kansas State University, Manhattan, KS, USA), respectively (Table S1). The *D. villosum* line DvRes-1 carries the original *Pm21* gene. DvRes-2 and DvRes-3 were derived from the powdery mildew resistant individuals of the accessions GRA961 and GRA1114, respectively. Lines DvSus-1 to DvSus-4 were derived from the susceptible individuals of the accessions GRA2738, GRA962, GRA1105, and PI 598390, respectively. The wheat variety (cv.) Yangmai 18 was a wheat-*D. villosum* translocation line that carries *Pm21*. The wheat cv. Yangmai 9 was susceptible to powdery mildew. Both of them were developed in Yangzhou Academy of Agricultural Sciences, Yangzhou, China. Plants were grown under a daily cycle of 16 h of light and 8 h of darkness at 24°C in a greenhouse.

Evaluation of Powdery Mildew Resistance

Blumeria graminis f. sp. *tritici* (*Bgt*) isolate YZ01 is a virulent isolate collected from Yangzhou region (Jiangsu Province, China). All plants, *D. villosum* accessions or lines and wheat varieties, were inoculated with *Bgt* isolate YZ01 at one-leaf stage (He et al., 2016). The powdery mildew responses of plants were evaluated at 8 d after inoculation.

Allelic Test

The susceptible homozygous *D. villosum* line DvSus-1 was used as female parent to cross with other susceptible lines, DvSus-2, DvSus-3, and DvSus-4, to produce three F₁ hybrids, DvSus-1/DvSus-2, DvSus-1/DvSus-3, and DvSus-1/DvSus-4, respectively. The wheat-*D. villosum* chromosome 6V disomic addition line DA6V#1 susceptible to powdery mildew was crossed with another susceptible chromosome 6V addition line DA6V#3 to result in F₁ hybrid DA6V#1/DA6V#3. All F₁ plants derived from different crosses were inoculated with *Bgt* isolate YZ01 at one-leaf stage for investigation of their responses.

DNA Isolation and Molecular Analysis of *Pm21* Alleles

Genomic DNA was extracted from leaves of one-leaf-stage plants by the TE-boiling method (He et al., 2017). The marker MBH1, developed from the promoter region of *Pm21* gene (Bie et al., 2015b), was used to detect genetic diversity of different *D. villosum* individuals. PCR amplification was carried out according to our previous description (He et al., 2017). PCR products with different sizes were T/A-cloned and sequenced.

Isolation of *Pm21* Alleles

Total RNA of different *D. villosum* accessions/lines and wheat materials was extracted from seedlings leaves using the TRIzol solution (Life Technologies, Carlsbad, California,

USA). About 2 μ g of total RNA was used for synthesis of cDNA using the PrimeScriptTM II 1st Strand cDNA Synthesis Kit (TaKaRa, Shiga, Japan) according to the manufacturer's guidelines. *Pm21* alleles were isolated from the cDNAs by PCR using the high fidelity PrimeSTAR Max Premix (TaKaRa, Shiga, Japan) and the primer pair (forward primer: 5'-TTACCCGGGCTCACCCGTTGGACTTGGACT-3'; reverse primer: 5'-CCCACTAGTCTCTCTTCGTTACATAATGTA GTGCCT-3'). PCR products were digested with *Sma*I and *Spe*I, inserted into pAHC25-MCS1 and sequenced. The genomic DNA of the alleles in the susceptible materials, DvSus-1 to DvSus-4, DA6V#1, and DA6V#3, were also isolated using PCR with LA *Taq* DNA polymerase (TaKaRa, Shiga, Japan) and the above primer pair. Each *Pm21* allele was amplified from its donor material by three independent PCR, followed by cloning and Sanger sequencing.

Sequence Data Analysis

Multiple alignment analysis was carried out using the CLUSTAL W tool (Thompson et al., 1994). Nucleotide diversity of *Pm21* alleles and their coding sequences of different domains or non-domain regions was analyzed using the MEGA7 software (Kumar et al., 2016) and assessed by Tajima's test of neutrality (Tajima, 1989). π meant the average number of nucleotide differences per site between two sequences. θ represented Watterson's nucleotide diversity estimator based on the value of π . Synonymous substitution rate (dS), non-synonymous substitution rate (dN), and natural selection for each codon were estimated by the HyPhy program in the MEGA7 software. Sequence logos of LRR motifs were created by the WebLogo tool (Crooks et al., 2004). For evolutionary analyses, all positions containing gaps were eliminated. So, there were a total of 2,718 positions in the final dataset. A phylogenetic tree based on the cDNA sequences of the *Pm21* alleles was constructed using the Neighbor-Joining method in the MEGA7 software (Kumar et al., 2016).

Accession Numbers

The accession number of *Pm21* gene in the GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) is MF370199. The *Pm21* alleles obtained have been deposited in the GenBank under the accession numbers MG831524–MG831526, MG831528–MG831561 and MH184801–MH184806.

RESULTS

Powdery Mildew Responses of Different Germplasms

The *D. villosum* accessions provided by different germplasm resource institutions were collected from the Mediterranean region, mainly from Greece and Italy (Figure 1; Table S1). A total of 62 accessions were used to detect the responses to *Bgt* isolate YZ01. All plants of the 58 accessions were immune to *Bgt* isolate YZ01, whereas in each of the other four accessions (GRA2738, GRA962, GRA1105, and PI 598390), several individuals (2–5%) were susceptible despite that most plants were resistant. The four susceptible homozygous lines derived from the above accessions were then designated as DvSus-1 to DvSus-4, respectively. The

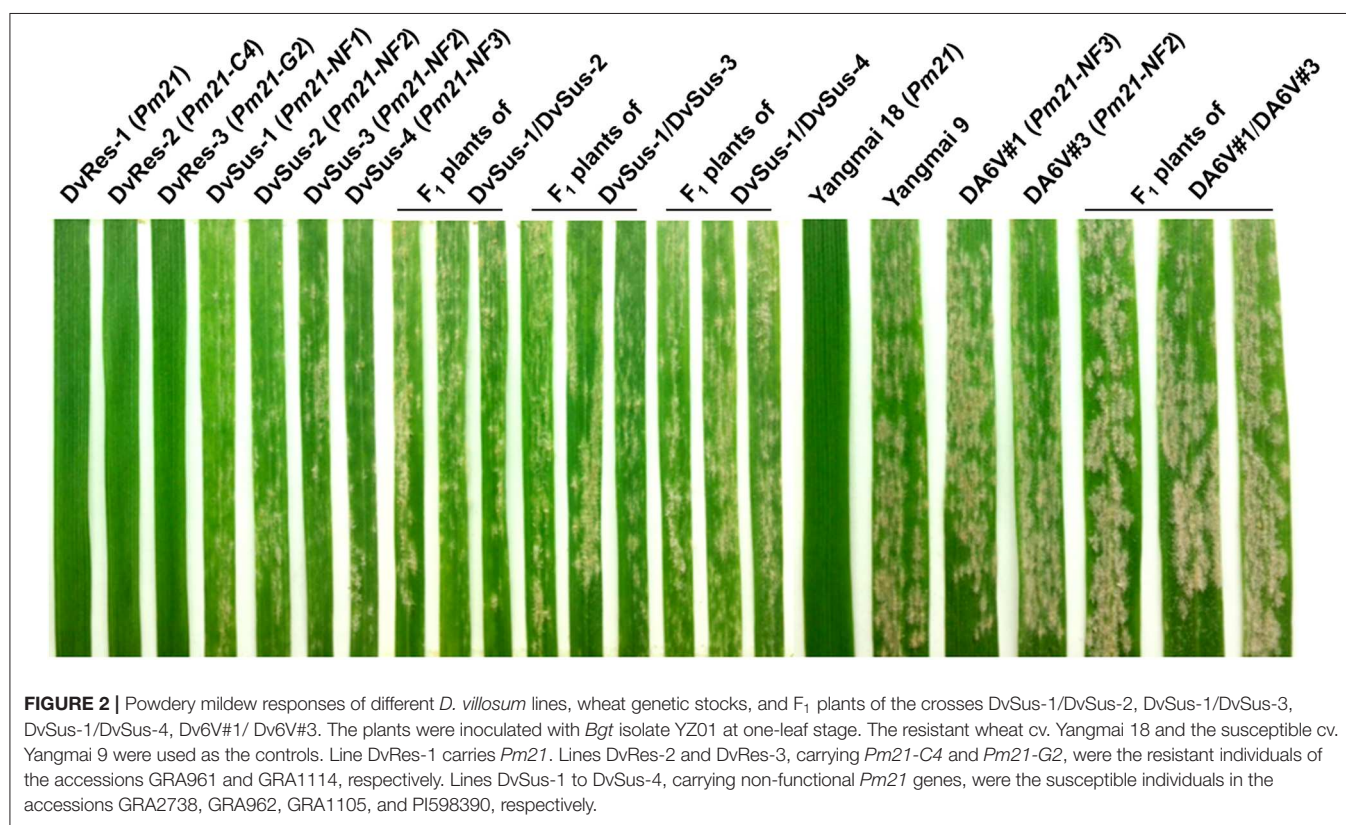
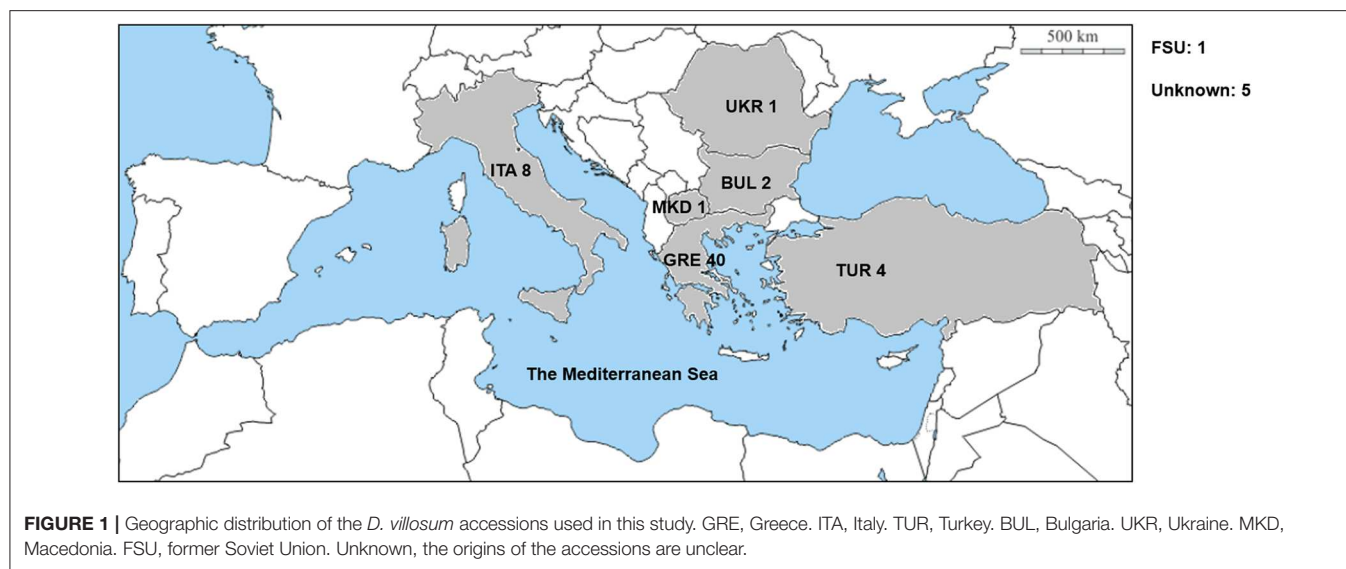
results also showed that the wheat-*D. villosum* chromosome 6V disomic addition lines DA6V#1 and DA6V#3 were susceptible to powdery mildew (Figure 2).

The powdery mildew responses of the F₁ plants derived from four different crosses, DvSus-1/DvSus-2, DvSus-1/DvSus-3, DvSus-1/DvSus-4, and DA6V#1/DA6V#3, were also assessed. The data showed that all the F₁ hybrids displayed high susceptibility to *Bgt* isolate YZ01 (Figure 2). It was indicated that there was no obvious allelic complementation in any of the above crosses. Therefore, it was suggested that the potential mutation(s), which led to susceptibility of the four *D. villosum* lines (DvSus-1, DvSus-2, DvSus-3, and DvSus-4) and the two wheat-*D. villosum* chromosome 6V disomic addition lines (DA6V#1 and DA6V#3), may all occur in the alleles of *Pm21*.

Molecular and Nucleotide Diversity of the *Pm21* Alleles

To understand the diversity at the *Pm21* loci, *MBH1*, designed based on the promoter sequence of *Pm21* (Bie et al., 2015b), was used to detect the resistant individuals from 62 different *D. villosum* accessions. The PCR products were sequenced and eight representative bands with different sizes, 271, 339, 340, 341, 342, 344, 396, and 467 bp, were found. This indicated that insertion-deletion (InDel) polymorphisms exist at the promoter regions of different *Pm21* alleles. Given that all *MBH1* sequences were isolated from resistant individuals, it was suggested that the variations in the promoter regions have no obviously adverse impact on the expression of *Pm21* alleles. In some individuals, two specific DNA bands were observed (Figures S1, S2), suggesting that these individuals might be heterozygous at the *Pm21* loci.

We then isolated *Pm21* alleles from the resistant individuals of 62 *D. villosum* accessions. Each of the individuals of 52 accessions had one copy of *Pm21* allele. However, due to open pollination of *D. villosum* species, each of the tested individuals of 9 accessions (PI 368886, W619414, W67270, GRA960, GRA1109, GRA1114, GRA2711, GRA2716, and 01C2300013) had two copies of *Pm21* alleles. In addition, three different alleles were, respectively, isolated from three individuals of the accession PI 251478. As a result, a total of 73 *Pm21* alleles were isolated in this study (Table S1). Among them, 38 alleles were non-redundant, sharing 91.7–100% identities with each other. In general, a total of seven InDels (Table S2), including three 3-bp insertions, one 30-bp insertion and three 3-bp deletions, and 400 single nucleotide polymorphism (SNP) sites were identified among these alleles. The 38 non-redundant *Pm21* alleles and their coding sequences of different domains were further used to determine the nucleotide diversity. The average pairwise nucleotide diversity π and Watterson's nucleotide diversity estimator θ of the full-length *Pm21* alleles were 0.039096 and 0.035027, respectively. Compared with the full-length alleles, the values of π and θ of the NB-ARC domain-encoding sequences were slightly lower ($\pi = 0.036868$ and $\theta = 0.034204$), whereas those of the CC domain-encoding sequences were significantly lower ($\pi = 0.013115$ and $\theta = 0.012973$) and those of the LRR domain-encoding sequences were obviously higher



($\pi = 0.051892$ and $\theta = 0.044652$). These results indicated that the CC domain was more conserved than other domains whereas the LRR domain was more variable. We also analyzed the π and θ values of Linker 1 and Linker 2, the regions between the CC and NB-ARC domains, and between the NB-ARC and LRR domains, respectively. The data showed that Linker 1 had no nucleotide diversity. Contrarily, Linker 2 had the highest nucleotide diversity ($\pi = 0.054507$ and $\theta = 0.054092$) in different

domains or regions of *Pm21* alleles (Figure 3; Table 1). Up to now, the function of Linker 2 is unclear yet. One reasonable explanation for its high variation is that Linker 2 may be an extension of the LRR domain.

Selection Pressure Analysis

To determine the potential evolutionary selection occurred in *Pm21* alleles, dN and dS rates were assessed using the HyPhy

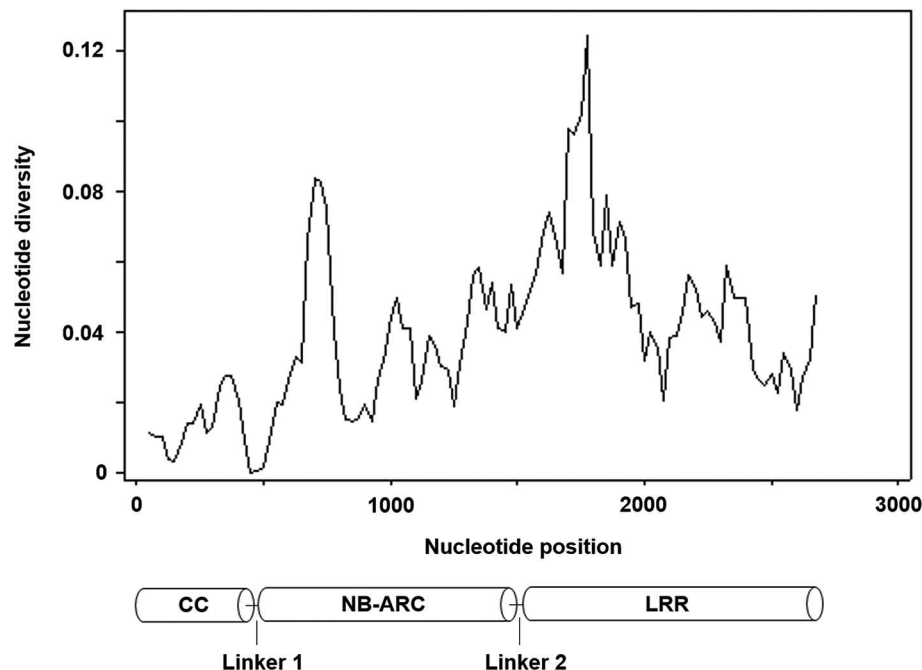


FIGURE 3 | Nucleotide diversity of 38 non-redundant *Pm21* alleles isolated from the resistant *D. villosum* accessions. All the positions containing gaps were eliminated. Therefore, there were a total of 2,718 positions in the final dataset. The predicted protein structure is shown at the bottom.

TABLE 1 | Nucleotide diversity of the *Pm21* alleles and their domains.

Gene or domain	Position (aa)	Position (bp)	N	S	π	θ	D
FL	1–906	1–2718	2,718	400	0.039096	0.035027	0.437596
CC	1–159	1–477	477	26	0.013115	0.012973	0.037612
Linker 1	160–168	478–504	27	0	0.000000	0.000000	n.a.
NB-ARC	169–502	505–1506	1,002	144	0.036868	0.034204	0.289608
Linker 2	503–524	1507–1572	63	15	0.054507	0.054092	0.024658
LRR	525–906	1573–2718	1,149	215	0.051892	0.044652	0.606923

FL, full-length *Pm21* alleles. Linker 1, linker between the CC and NB-ARC domains. Linker 2, linker between the NB-ARC and LRR domains. n, total number of sites. S, the number of segregating (polymorphic) sites. π , the average number of nucleotide differences per site between two sequences. θ , Watterson's nucleotide diversity estimator based on the value of π . D, Tajima's D statistics for neutrality test. n.a., not applicable.

program. The dN/dS ratio of full-length *Pm21*, CC-, NB-ARC-, and LRR-encoding sequences were 0.72046, 0.22671, 0.48723, and 1.15098, respectively, which suggested that the LRR domain might be under positive selection. The dN/dS ratio of the structural LRR residues and the solvent-exposed LRR residues, the two parts of the LRR domain, were 0.88106 and 3.19734, respectively (Table 2). This indicated diversifying selection acting on the solvent-exposed residues in the LRR domain of *Pm21* alleles.

The LRR domain of *Pm21* consists of 16 LRR motifs. The dN/dS ratios of 8 LRR motifs (LRR4–LRR7, LRR10, LRR11, LRR15, and LRR16) were greater than 1. Among them, the dN/dS ratio of LRR11 was 8.58259 and that of LRR16 was infinite because its dS value was zero (Figure 4; Table 2). These results indicated that the above 8 LRR motifs have undergone positive selection. In the LRR domain, four sites at the positions 628, 885, 903, and 905 were subject to positive selection, detected by four

different models (Felsenstein 1981 model, Hasegawa-kishino-Yano model, Tamura-Nei model, and General Time Reversible model). Position 628 lied in the LRR5 motif and positions 885, 903, and 905 were all located in the LRR16 motif (Figure 4; Table S3).

Phylogenetic Analysis and Classification of the *Pm21* Alleles

The phylogenetic tree for *Pm21* alleles showed that 38 non-redundant *Pm21* alleles were clustered into seven clades (Clade A to G). Among these clades, Clades A, B, and C were the major types in the *D. villosum* populations, which included 26 members, accounting for 68.4% (Figure 5).

According to the clades categorized in the phylogenetic tree, the *Pm21* alleles isolated from the resistant *D. villosum* accessions were correspondingly divided into seven classes (Class A to G).

TABLE 2 | dN, dS, and dN/dS ratio of *Pm21* alleles and their domains or motifs.

Gene, domain or motif	Position (bp)	dN	dS	dN/dS ratio
Full-length <i>Pm21</i>	1–2,718	0.27309	0.37905	0.72046
CC	1–477	0.08821	0.38906	0.22671
Linker 1	478–504	0.00000	0.00000	n.a.
NB-ARC	505–1,506	0.16928	0.34744	0.48723
Linker 2	1,507–1,572	0.37287	0.86044	0.43336
LRR	1,573–2,718	0.44728	0.38860	1.15098
Structural LRR	-	0.38044	0.43180	0.88106
Solvent-exposed LRR	-	0.69844	0.21844	3.19734
LRR1	1,573–1,635	0.27554	0.83026	0.33187
LRR2	1,636–1,707	0.33048	0.70341	0.46982
LRR3	1,708–1,785	0.71357	0.69042	1.03357
LRR4	1,786–1,851	0.46933	0.18153	2.58544
LRR5	1,852–1,917	0.62027	0.34232	1.81196
LRR6	1,918–1,980	0.29715	0.11953	2.48594
LRR7	1,981–2,049	0.41649	0.24384	1.70807
LRR8	2,050–2,127	0.19744	0.25129	0.78572
LRR9	2,128–2,208	0.43540	0.66865	0.65117
LRR10	2,209–2,277	0.53271	0.19266	2.76496
LRR11	2,278–2,349	0.48589	0.05661	8.58259
LRR12	2,350–2,418	0.46961	0.67420	0.69654
LRR13	2,421–2,487	0.26525	0.27089	0.97917
LRR14	2,488–2,562	0.19280	0.78675	0.24506
LRR15	2,563–2,652	0.29745	0.14091	2.11088
LRR16	2,653–2,718	1.26260	0.00000	Infinite

FL, full-length *Pm21* alleles. Linker 1, the linker between the CC and NB-ARC domains. Linker 2, the linker between the NB-ARC and LRR domains. Solvent-exposed LRR, the residue x in the LxxLxLxx motif. Structural LRR, other residues except the residue x in the LRR domain (Srichumpa et al., 2005). LRR1 to LRR16, 16 LRR motifs predicted in the LRR domain. n.a., not applicable.

Class A consisted of 9 alleles, *Pm21-A1* to *Pm21-A9*, whose open reading frames (ORFs) were 2,730 bp in length sharing the highest identities with *Pm21* (99.2% on average). Class B contained 10 alleles, *Pm21-B1* to *Pm21-B10*, most of which were 2,724 bp sharing 96.6% identity with *Pm21* on average. Class C harbored 7 alleles, *Pm21-C1* to *Pm21-C7*, with 2,730 bp in length and had 96.7% identity with *Pm21* on average. The remaining 12 alleles, sharing 92.1–97.0% identities with *Pm21*, were divided into four classes, Class D to G, whose obvious sequence characteristics was a 30-bp insertion compared with *Pm21* (Table 3).

Natural Variations of *Pm21* Alleles in Susceptible Germplasms

To test the rare natural variations leading to lose of resistance to powdery mildew, we isolated *Pm21* alleles from the susceptible *D. villosum* lines DvSus-1 to DvSus-4, derived from the accessions GRA2738, GRA962, GRA1105, and PI 598390, respectively. The non-functional allele *Pm21-NF1* isolated from the genome of DvSus-1 was 3,699 bp in length, whose ORF was 2,730 bp. Compared with *Pm21*, *Pm21-NF1* had 98 SNPs; however, compared with the 38 non-redundant alleles isolated from the resistant *D. villosum* accessions, *Pm21-NF1* only had

two specific variations. The first variation was a transversion G61T leading to the amino acid change A21S in the CC domain. The second variation was a transition A821G resulting in the change D274G (Figure S3A), corresponding to the latter aspartate (D) in kinase-2 motif (also called Walker B motif; consensus sequence: LLVLDDVW) in the NB-ARC domain. The latter D is considered to act as the catalytic site for ATP hydrolysis and activation of disease resistance protein (Meyers et al., 1999; Tameling et al., 2006). Here, bioinformatic analysis showed that the latter D was highly conserved in all the tested disease resistance proteins from *Arabidopsis thaliana*, barley (*Hordeum vulgare* L.) and wheat (Figure S4), suggesting that the amino acid change D274G might lead to loss-of-function of *Pm21-NF1*.

The genomic sequence of the non-functional allele *Pm21-NF2* isolated from the susceptible DvSus-2 was 3,698 bp in length, whose ORF contained a 1-bp deletion after position 876, leading to frame shift and resulting in a truncated protein (296 aa). The variations of *Pm21* alleles isolated from DvSus-3 and DA6V#3 were both identical to that of *Pm21-NF2*. In DvSus-4 and DA6V#1, the sequences of the alleles were identical (4,988 bp) and designated as *Pm21-NF3*. *Pm21-NF3* harbored an insertion of 1281 bp that caused a premature stop codon (Figure S3B) and led to loss of the last four LRR motifs. These results suggested that the non-functional *Pm21* alleles in DA6V#1 and DA6V#3 both directly originated from their *D. villosum* donors susceptible to powdery mildew.

Molecular Tracing of the Origins of Non-functional *Pm21* Alleles

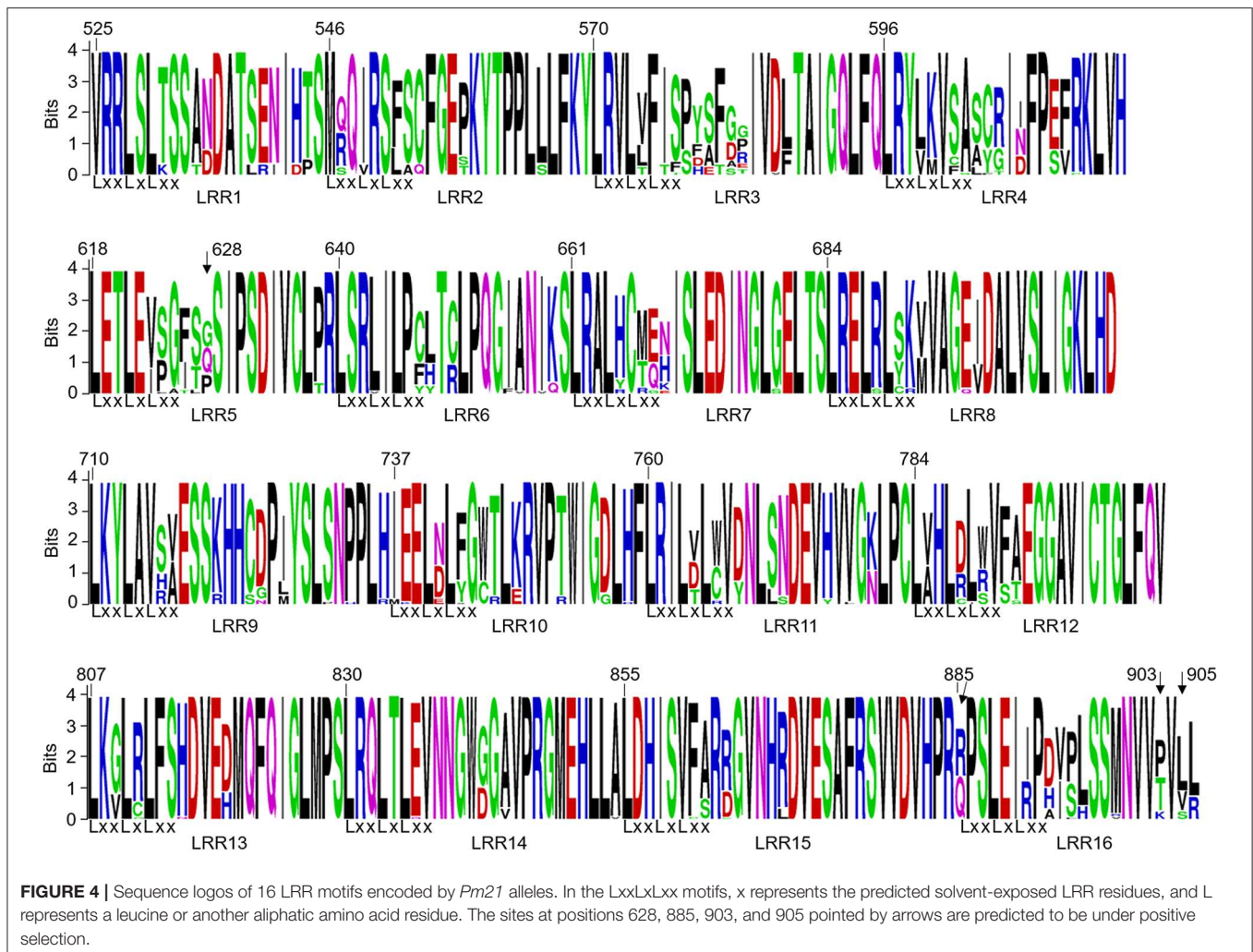
Phylogenetic analysis showed that DvSus-1, DvSus-2, DvSus-3, GRA961, and GRA1164 were clustered in Clade C (Figure 5). In contrast to the alleles, *Pm21-C4* in GRA961 and *Pm21-C1* in GRA1164, the non-functional allele *Pm21-NF1* in DvSus-1 had 8 and 10 SNPs, and *Pm21-NF2* in DvSus2/DvSus-3 had 1 and 3 SNPs, respectively (Figure S3B). This suggested that the non-functional allele *Pm21-NF2* originated from the allele *Pm21-C4* in the resistant accession GRA961 (Figures 5, 6; Table 3). In the tested accessions, the origin of *Pm21-NF1* could not be well-traced yet.

The data also indicated that lines DvSus-4, GRA1113, and GRA1114 were clustered in Clade G (Figure 5). Except the 1281-bp insertion, *Pm21-NF3* in DvSus-4 had no difference from *Pm21-G2* in GRA1114 (Figure S3C). This result revealed that the non-functional allele *Pm21-NF3* came from the variation of the allele *Pm21-G2* in the resistant accession GRA1114 (Figures 5, 6; Table 3).

DISCUSSION

Diversity, Classification and Geographic Distribution of *Pm21* Alleles

As a wild relative of wheat, *D. villosum* possesses several powdery mildew resistance genes that have important potential for controlling wheat powdery mildew disease (He et al., 2017). Among them, *Pm21* and *PmV*, located on chromosome 6VS derived from different *D. villosum* accessions, confer powdery



mildew resistance at whole-plant growth stages. It seems that *Pm21* and *PmV* may be allelic (Bie et al., 2015b). Both *Pm55* and *Pm62* confer resistance at adult-plant stage but not at the seedling stage (Zhang et al., 2016, 2018). In this study, *Bgt*-responses of all *D. villosum* accessions were detected at one-leaf stage, which could exclude the resistance conferred by *Pm55* and *Pm62*. Therefore, the seedling-resistance in these materials was considered to be provided by *Pm21* alleles.

Recently, the broad-spectrum powdery mildew resistance gene *Pm21* was isolated from *D. villosum* using the map-based cloning strategy (He et al., 2018). Based on the investigation of powdery mildew responses of different *D. villosum* accessions collected from the Mediterranean countries, we isolated 73 *Pm21*-like sequences from the resistant individuals. The previous work showed that *Pm21* is adjacent to another CC-NBS-LRR-encoding gene *DvRGA1* (He et al., 2018). Although *DvRGA1* is the highest matched gene of *Pm21* in Genbank database, they had only 72.7% nucleotide sequence identity. Here, the isolated *Pm21*-like genes shared 91.7–100% identities with each other, indicating that all the sequences are identical or allelic to *Pm21*. Of the 73 sequences, 38 were different from each other.

Compared with *Pm21*, the other 37 non-redundant alleles have seven InDels involved in 3-bp, 6-bp, 30-bp, 33-bp, or 36-bp, which make the alleles maintain correct ORFs and encode full-length proteins. The alleles also had many SNPs and the average pairwise nucleotide diversity of the LRR-encoding region was significantly higher than those of the CC- or NB-ARC-encoding regions. Compared with other domains, the LRR domain were supposed to have undergone faster evolution. Because all of the individuals containing these alleles were still effective against the highly virulent *Bgt* isolate YZ01, it was proposed that the wide variations of *Pm21* alleles have no obviously adverse effect on the disease resistance. However, whether they still keep broad-spectrum resistance remains to be disclosed.

Phylogenetic analysis identified seven independent clades that involved all the *Pm21* alleles. Among them, Classes A to C represented the three major classes. The functional *Pm21* gene was originally found in an accession provided by Cambridge Botanic Garden in the United Kingdom, but the exact collection site of this accession was unclear. *Pm21*, with the systemic name *Pm21-A1* here, belongs to Class A whose members were only found in the accessions of Greece or Turkey. In particular, among



Geographic distributions of different *Pm21* alleles were further investigated in this study. It is indicated that the *Pm21* alleles isolated from Greece *D. villosum* accessions had more genetic diversity and covered the most members of all the seven classes (Class A to G). In addition, *Pm21*-A8, *Pm21*-E2, and *Pm21*-F3 were only detected in Turkey accessions, and *Pm21*-B7 and *Pm21*-G2 were only detected in Italy accessions (**Table S1**). The characteristics of geographic distributions of the *Pm21* alleles may help to search the accessions carrying specific *Pm21* alleles as donors for future breeding purpose.

It has been believed that *D. villosum* resources are all resistant to wheat powdery mildew (Qi et al., 1998). In our previous work, four *D. villosum* lines DvSus-1 to DvSus-4 susceptible

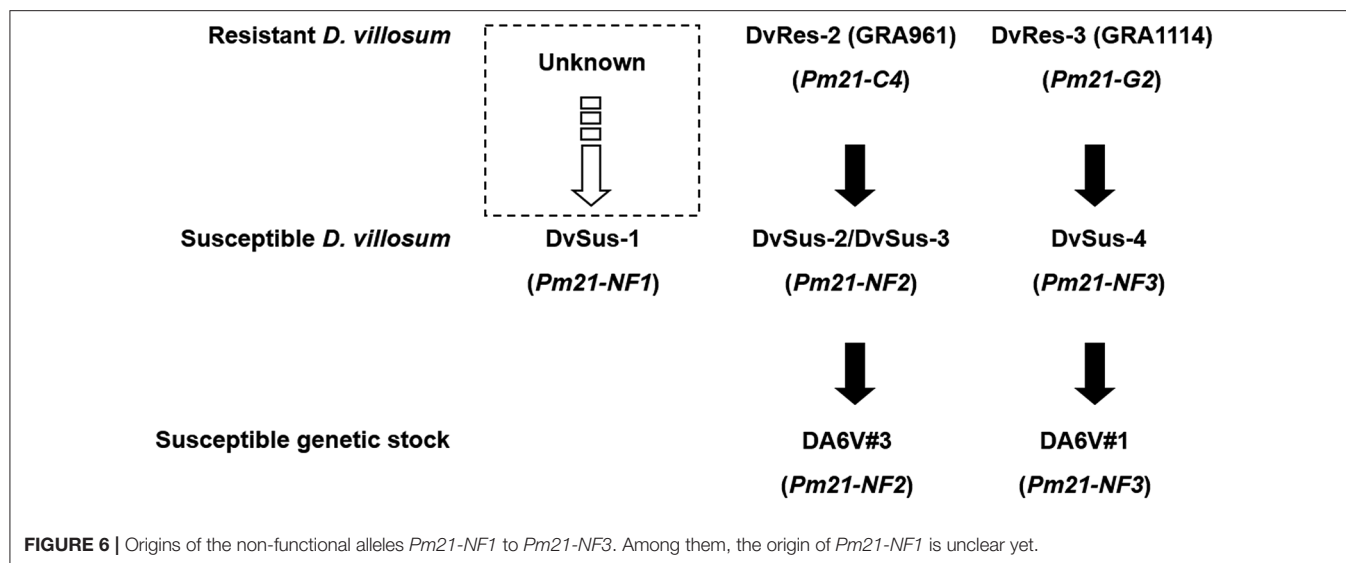
Previously, the wheat-*D. villosum* chromosome 6V disomic addition lines DA6V#1 and DA6V#3 were reported to be highly susceptible to powdery mildew (Qi et al., 1998; Liu et al., 2011). During the creation of the two addition lines, colchicine was used for chromosome doubling, which is proved to be an effective mutagen in fact (Gilbert and Patterson, 1965). So, researchers did not know if the susceptibilities of DA6V#1 and DA6V#3 came from colchicine treatment or the *D. villosum* donors. Since *Pm21* has been cloned, through sequencing of allele genes here, we demonstrated that *Pm21*

TABLE 3 | Classification of *Pm21* alleles isolated from resistant individuals of *D. villosum*.

Class	Allele	GenBank accession number	ORF length (bp)	InDel (compared with <i>Pm21</i>)	Identity with (%)		Occurrence in population
					<i>Pm21</i> on average	Class	
A	<i>Pm21</i> (<i>Pm21</i> -A1)	MF370199	2,730	-	99.2	98.0–100	7
	<i>Pm21</i> -A2	MG831538	2,730	-			1
	<i>Pm21</i> -A3	MG831528	2,730	-			2
	<i>Pm21</i> -A4	MG831529	2,730	-			1
	<i>Pm21</i> -A5	MG831531	2,730	-			1
	<i>Pm21</i> -A6	MG831536	2,730	-			1
	<i>Pm21</i> -A7	MG831539	2,730	-			1
	<i>Pm21</i> -A8	MG831537	2,730	-			1
	<i>Pm21</i> -A9	MG831533	2,730	-			1
B	<i>Pm21</i> -B1	MG831540	2,724	In-2, Del-1—Del-3	96.6	97.4–100	1
	<i>Pm21</i> -B2	MG831545	2,724	In-2, Del-1—Del-3			5
	<i>Pm21</i> -B3	MG831546	2,724	In-2, Del-1—Del-3			1
	<i>Pm21</i> -B4	MG831547	2,724	In-2, Del-1—Del-3			4
	<i>Pm21</i> -B5	MG831548	2,724	In-2, Del-1—Del-3			1
	<i>Pm21</i> -B6	MG831561	2,727	In-2, Del-1—Del-2			2
	<i>Pm21</i> -B7	MG831541	2,724	Del-2—Del-3			4
	<i>Pm21</i> -B8	MG831542	2,724	Del-2—Del-3			1
	<i>Pm21</i> -B9	MG831543	2,724	Del-2—Del-3			1
	<i>Pm21</i> -B10	MG831544	2,724	Del-2—Del-3			3
C	<i>Pm21</i> -C1	MG831524	2,730	-	96.7	97.9–99.9	4
	<i>Pm21</i> -C2	MG831525	2,730	-			1
	<i>Pm21</i> -C3	MG831526	2,730	-			4
	<i>Pm21</i> -C4	MG831530	2,730	-			2
	<i>Pm21</i> -C5	MG831532	2,730	-			1
	<i>Pm21</i> -C6	MG831534	2,730	-			1
	<i>Pm21</i> -C7	MG831535	2,730	-			1
D	<i>Pm21</i> -D1	MG831557	2,766	In-1–In-3	92.4	99.2–100	3
	<i>Pm21</i> -D2	MG831558	2,766	In-1–In-3			1
	<i>Pm21</i> -D3	MG831559	2,766	In-1–In-3			1
	<i>Pm21</i> -D4	MG831560	2,766	In-1–In-3			1
E	<i>Pm21</i> -E1	MG831550	2,760	In-1–In-2, Del-1	96.4	97.8	1
	<i>Pm21</i> -E2	MG831553	2,760	In-1–In-2, Del-1			1
F	<i>Pm21</i> -F1	MG831549	2,760	In-1–In-2, Del-1	95.8	99.3–100	4
	<i>Pm21</i> -F2	MG831551	2,760	In-1–In-2, Del-1			2
	<i>Pm21</i> -F3	MG831552	2,760	In-1–In-2, Del-1			1
	<i>Pm21</i> -F4	MG831554	2,760	In-1–In-2, Del-1			1
G	<i>Pm21</i> -G1	MG831555	2,763	In-1, In-4, Del-1	95.2	99.8	3
	<i>Pm21</i> -G2	MG831556	2,763	In-1, In-4, Del-1			1

alleles isolated from DA6V#1 and DA6V#3 had identical variations to *Pm21*-NF3 (DvSus-4) and *Pm21*-NF2 (DvSus-2 and DvSus-3), respectively. Therefore, it was suggested that the variations of the *Pm21* alleles from DA6V#1 and DA6V#3 both originated from their *D. villosum* donors, rather than colchicine treatment.

The non-functional alleles, *Pm21*-NF1, *Pm21*-NF2, and *Pm21*-NF3, were found in the accessions GRA2738, GRA962, PI 598390, respectively. In theory, their wild-type genes could be isolated from the corresponding accessions above. We tried to do so but not succeeded. The major reason may be that *D. villosum* is highly outcrossing which causes that the pollen



with a mutated gene is subject to separate from the one carrying a corresponding wild-type gene. Therefore, we attempted to trace the origins of the non-functional alleles through evolutionary analysis. The origins of the two non-functional alleles, *Pm21-NF2* and *Pm21-NF3*, were both traceable in the natural populations of *D. villosum*. Except the identified mutations, the sequences of *Pm21-NF2* and *Pm21-NF3* were entirely identical to those of *Pm21-C4* and *Pm21-G2* that were cloned from the resistant individuals of the accessions GRA961 and GRA1114, respectively. Hence, we concluded that the non-functional alleles *Pm21-NF2* and *Pm21-NF3* originated from *Pm21-C4* and *Pm21-G2*, respectively. However, the origin of *Pm21-NF1* remains unclear yet.

Diversifying Selection Acting on the Solvent-Exposed LRR Residues of *Pm21* Alleles

It was confirmed that the broad-spectrum resistance of *Pm21* is conferred by a single CC-NBS-LRR-encoding gene (He et al., 2018). However, it is believed that the resistance provided by such kind of genes is most likely race-specific, which is prone to be overcome by fast-evolving pathogens. For instance, *Pm8* from rye (*Secale cereale* L.), also encoding a CC-NBS-LRR protein, previously provided effective resistance to wheat powdery mildew (Hurni et al., 2013), has lost its resistance in most wheat producing regions with the worldwide utilization. In this study, the value of dN/dS (3.19734) significantly exceeded 1 in the solvent-exposed LRR residues, which is considered to take part in the specific recognition of pathogens (Meyers et al., 1999). This result suggested that the solvent-exposed LRR residues of *Pm21* have been undergone diversifying selection and may play critical roles in resistance specificity. This situation is similar to those of race-specific powdery mildew resistance gene *Pm3* from wheat (Srichumpa et al., 2005) and *Mla* from barley (Seeholzer et al., 2010). In several works, the researchers reported that the wheat varieties carrying

Pm21 could be infected by *Bgt* pathogens in different regions (Shi et al., 2009; Yang et al., 2009). Therefore, combined the data given by evolutionary analysis, it is speculated that *Pm21* may be a race-specific resistance gene although it still provides broad-spectrum resistance to the most *Bgt* isolates so far.

Since 1995 when the translocation line of wheat-*D. villosum* T6AL.6VS was released, many wheat varieties carrying *Pm21* have been commercialized in China, mainly in the middle and lower reaches of the Yangtze River Valley and the southwest wheat-producing regions, where *Bgt* pathogen is prevailing (Jiang et al., 2014; Bie et al., 2015a; Cheng et al., 2020). The long-time and wide-range application of *Pm21* in agriculture would accelerate the evolution of *Bgt* pathogens. Correspondingly, *Pm21* would face to an increasing risk of losing its resistance to powdery mildew. Consequently, it will be a great challenge to sustainably utilize the *Pm21* resistance in the future. In this study, a total of 38 non-redundant *Pm21* alleles were obtained, which allows to comparatively analyze their fine functions against *Bgt* pathogens in further researches. Utilization of different *Pm21* alleles with functional diversity would be a way to extend the lifespan of *Pm21* resistance in wheat production. The marker *MBH1*, which can reveal genetic diversity of *Pm21* alleles in some degree, will be a useful tool when transferring them from *D. villosum* into common wheat. Other reasonable means would be diversifying use of *Pm* genes in field, such as pyramiding other effective *Pm* gene(s) into *Pm21*-carrying varieties or exploring new *Pm* genes and developing wheat varieties carrying different *Pm* genes.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the GenBank, MG831524–MG831526, MG831528–MG831561, MH184801–MH184806.

AUTHOR CONTRIBUTIONS

HH, CL, and SZ conceived and designed the experiments. HH, JJ, JT, YF, XW, RH, and TB performed the experiments. HH, JJ, and HL analyzed the data and wrote the paper.

FUNDING

This study was supported by grants from Shandong Agricultural Seed Improvement Project (2019LZGC016), National Natural Science Foundation of China (31971874), Jiangsu Agricultural Science and Technology Innovation Fund [CX(19)2042], Priority Academic Program Development of Jiangsu Higher Education Institutions, Jiangsu Education Department and Taishan Scholars Project (tsqn201812123).

ACKNOWLEDGMENTS

This manuscript has been released as a pre-print at bioRxiv (Zhu and He, 2018). The authors are thankful to Germplasm Resources Information Network (GRIN), GRIN Czech, and Genebank Information System of the IPK Gatersleben (GBIS-IPK), Nordic Genetic Resource Center (NordGen) and Dr. Bernd Friebe (Kansas State University) for providing the *Dasypyrum villosum* accessions and wheat genetic stocks.

REFERENCES

- Bie, T., Zhao, R., Jiang, Z., Gao, D., Zhang, B., and He, H. (2015a). Efficient marker-assisted screening of structural changes involving *Haynaldia villosa* chromosome 6V using a double-distal-marker strategy. *Mol. Breeding* 35:34. doi: 10.1007/s11032-015-0211-y
- Bie, T., Zhao, R., Zhu, S., Chen, S., Cen, B., Zhang, B., et al. (2015b). Development and characterization of an efficient breeding-practical marker MBH1 simultaneously tagging Pm21 and PmV genes conferring resistance to wheat powdery mildew. *Mol. Breeding* 35:189. doi: 10.1007/s11032-015-0385-3
- Chen, P. D., Qi, L. L., Zhou, B., Zhang, S. Z., and Liu, D. J. (1995). Development and molecular cytogenetic analysis of wheat-*Haynaldia villosa* 6VS/6AL translocation lines specifying resistance to powdery mildew. *Theor. Appl. Genet.* 91, 1125–1128. doi: 10.1007/BF00223930
- Cheng, B., Ding, Y., Gao, X., Cao, N., Xin, Z., and Zhang, L. (2020). The diversity of powdery mildew resistance gene loci among wheat germplasm in Southwest China. *Cereal Res. Commun.* 48, 65–70. doi: 10.1007/s42976-020-00015-2
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- De Pace, C., Vaccino, P., Cionini, P. G., Pasquini, M., Bizzarri, M., and Qualset, C. O. (2011). “*Dasypyrum*,” in *Wild Crop Relatives: Genomic And Breeding Resources, Cereals* ed C. Kole (Berlin: Springer), 185–292. doi: 10.1007/978-3-642-14228-4_4
- Gilbert, S. K., and Patterson, F. L. (1965). Colchicine-induced mutants in Decatur barley. *Crop Sci.* 5, 44–47. doi: 10.2135/cropsci1965.0011183X000500010014x
- He, H., Ji, Y., Zhu, S., Li, B., Zhao, R., Jiang, Z., et al. (2017). Genetic, physical and comparative mapping of the powdery mildew resistance gene Pm21 originating from *Dasypyrum villosum*. *Front. Plant Sci.* 8:1914. doi: 10.3389/fpls.2017.01914
- He, H., Zhu, S., Jiang, Z., Ji, Y., Wang, F., Zhao, R., et al. (2016). Comparative mapping of powdery mildew resistance gene Pm21 and functional characterization of resistance-related genes in wheat. *Theor. Appl. Genet.* 129, 819–829. doi: 10.1007/s00122-016-2668-4

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00489/full#supplementary-material>

Figure S1 | Molecular analysis of the diversity of *D. villosum* by the marker MBH1 that was developed from the promoter region of Pm21. M, DNA marker DL2000. Line 1 to 24, PCR products obtained from resistant individuals of different *D. villosum* accessions.

Figure S2 | Multiple sequence alignment of different representative products PCR-amplified with the marker MBH1.

Figure S3 | Detection of mutations in the non-functional Pm21 alleles. (A) Mutations of Pm21-NF1 in DvSus-1 contrasted to Pm21. (B) Mutations of Pm21-NF2 in DvSus-2, DvSus-3, and DA6V#3 contrasted to Pm21-C4 in DvRes-2 (derived from GRA961). (C) Mutations of Pm21-NF3 in DvSus-4 and DA6V#1 in contrast to Pm21-G2 in DvRes-3 (derived from GRA1114). SNPs, tandem premature stop codons and insertion sequences are shown by arrows, underlines and brackets, respectively.

Figure S4 | Multiple sequence alignment of the surrounding sequences of kinase-2 motif (consensus sequence: LLVDDVW) of plant disease resistance proteins. The conserved second aspartate (D) of kinase-2 motif is marked by an arrow.

Table S1 | Pm21 alleles and the corresponding germplasms.

Table S2 | InDel polymorphisms in the Pm21 alleles. All InDels, compared with Pm21, occur after the positions showed in the brackets.

Table S3 | Amino acid sites under positive selection.

- He, H., Zhu, S., Zhao, R., Jiang, Z., Ji, Y., Ji, J., et al. (2018). Pm21, encoding a typical CC-NBS-LRR protein, confers broad-spectrum resistance to wheat powdery mildew disease. *Mol. Plant* 11, 879–882. doi: 10.1016/j.molp.2018.03.004
- Hurni, S., Brunner, S., Buchmann, G., Herren, G., Jordan, T., Krukowski, P., et al. (2013). Rye Pm8 and wheat Pm3 are orthologous genes and show evolutionary conservation of resistance function against powdery mildew. *Plant J.* 76, 957–969. doi: 10.1111/tj.12345
- Jiang, Z., Wang, Q., Wu, J., Xue, W., Zeng, Q., Huang, L., et al. (2014). Distribution of powdery mildew resistance gene Pm21 in Chinese winter wheat cultivars and breeding lines based on gene-specific marker. *Scientia Agric. Sin.* 47, 2078–2087. (in Chinese with English abstract)
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Li, H., Chen, X., Xin, Z., Ma, Y., Xu, H., Chen, X., et al. (2005). Development and identification of wheat-*Haynaldia villosa* T6DL6VS chromosome translocation lines conferring resistance to powdery mildew. *Plant Breed.* 124, 203–205. doi: 10.1111/j.1439-0523.2004.01062.x
- Li, H., and Zhu, Z. (1999). The potential value of *Dasypyrum villosum* in breeding and its chromosome manipulation. *Chin Bull Bot.* 16, 504–510. (in Chinese with English abstract)
- Liu, C., Qi, L., Liu, W., Zhao, W., Wilson, J., Friebe, B., et al. (2011). Development of a set of compensating *Triticum aestivum* – *Dasypyrum villosum* robertsonian translocation lines. *Genome* 54, 836–844. doi: 10.1139/g11-051
- Meyers, B. C., Dickerman, A. W., Michelsmore, R. W., Sivaramakrishnan, S., Sobral, B. W., and Yong, N. D. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* 20, 317–332. doi: 10.1046/j.1365-3113.1999.t01-1-00606.x
- Qi, L. L., Wang, S. L., Chen, P. D., Liu, D. J., and Gill, B. S. (1998). Identification and physical mapping of three *Haynaldia villosa* chromosome-6V deletion lines. *Theor. Appl. Genet.* 97, 1042–1046. doi: 10.1007/s001220050989
- Seeholzer, S., Tsuchimatsu, T., Jordan, T., Bieri, S., Pajonk, S., Yang, W., et al. (2010). Diversity at the Mla powdery mildew resistance locus from cultivated barley reveals sites of positive selection. *Mol. Plant Microbe Interact.* 23, 497–509. doi: 10.1094/MPMI-23-4-0497

- Shi, Y., Wang, B., Li, Q., Wu, X., Wang, F., Liu, H., et al. (2009). Analysis of the virulent genes of *Erysiphe graminis* f. sp. *tritici* and the resistance genes of wheat commercial cultivars in shaanxi province. *J. Triticeae Crop* 29, 706–711. (in Chinese with English abstract)
- Srichumpa, P., Brunner, S., Keller, B., and Yahiaoui, N. (2005). Allelic series of four powdery mildew resistance genes at the Pm3 locus in hexaploid bread wheat. *Plant Physiol.* 139, 885–895. doi: 10.1104/pp.105.062406
- Tajima, F. (1989). Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Tameling, W. I. L., Vossen, J. H., Albrecht, M., Lengauer, T., Berden, J. A., Michel, A., et al. (2006). Mutations in the NB-ARC domain of I-2 that impair ATP hydrolysis cause autoactivation. *Plant Physiol.* 140, 1233–1245. doi: 10.1104/pp.105.073510
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Wang, H., Dai, K., Xiao, J., Yuan, C., Zhao, R., Doležal, J., et al. (2017). Development of intron targeting (IT) markers specific for chromosome arm 4VS of *Haynaldia villosa* by chromosome sorting and next-generation sequencing. *BMC Genomics* 18:167. doi: 10.1186/s12864-017-3567-z
- Yang, L., Xiang, L., Zeng, F., Wang, H., Shi, W., and Yu, D. (2009). Virulence gene structure analysis of *Blumeria graminis* f. sp. *tritici* in Hubei. *Plant Protect.* 35, 76–79. (in Chinese with English abstract)
- Zhang, R., Fan, Y., Kong, L., Wang, Z., Wu, J., Xing, L., et al. (2018). Pm62, an adult-plant powdery mildew resistance gene introgressed from *Dasyphyrum villosum* chromosome arm 2VL into wheat. *Theor. Appl. Genet.* 131, 2613–2620. doi: 10.1007/s00122-018-3176-5
- Zhang, R., Sun, B., Chen, J., Cao, A., Xing, L., Feng, Y., et al. (2016). Pm55, a developmental-stage and tissue-specific powdery mildew resistance gene introgressed from *Dasyphyrum villosum* into common wheat. *Theor. Appl. Genet.* 129, 1975–1984. doi: 10.1007/s00122-016-2753-8
- Zhu, S., Ji, Y., Bie, T., Gao, A., and He, H. (2018). Fine physical bin mapping of the powdery mildew resistance gene Pm21 based on chromosomal structural variations in wheat. *Int. J. Mol. Sci.* 19:643. doi: 10.3390/ijms19020643
- Zhu, S., and He, H. (2018). Characteristics and origins of non-functional Pm21 alleles in *Dasyphyrum villosum* and wheat genetic stocks. *bioRxiv [Preprint]*. doi: 10.1101/332841

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 He, Ji, Li, Tong, Feng, Wang, Han, Bie, Liu 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.



Wheat Encodes Small, Secreted Proteins That Contribute to Resistance to Septoria Tritici Blotch

Binbin Zhou^{1†}, Harriet R. Benbow^{1†}, Ciarán J. Brennan¹,
Chanemougasoundharam Arunachalam¹, Sujit J. Karki², Ewen Mullins³,
Angela Feechan², James I. Burke² and Fiona M. Doohan^{1*}

¹ UCD School of Biology and Environmental Science, UCD Earth Institute, UCD O'Brien Centre for Science (East), University College Dublin, Dublin, Ireland, ² UCD School of Agriculture and Food Science, University College Dublin, Dublin, Ireland, ³ Department of Crop Science, Teagasc, Carlow, Ireland

OPEN ACCESS

Edited by:

Zhu-Qing Shao,
Nanjing University, China

Reviewed by:

Lili Huang,
Northwest A&F University, China
Mark Derbyshire,
Curtin University, Australia

*Correspondence:

Fiona M. Doohan
fiona.doohan@ucd.ie

[†] 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: 24 February 2020

Accepted: 16 April 2020

Published: 12 May 2020

Citation:

Zhou B, Benbow HR,
Brennan CJ, Arunachalam C,
Karki SJ, Mullins E, Feechan A,
Burke JI and Doohan FM (2020)
Wheat Encodes Small, Secreted
Proteins That Contribute
to Resistance to Septoria Tritici
Blotch. *Front. Genet.* 11:469.
doi: 10.3389/fgene.2020.00469

During plant–pathogen interactions, pathogens secrete many rapidly evolving, small secreted proteins (SSPs) that can modify plant defense and permit pathogens to colonize plant tissue. The fungal pathogen *Zymoseptoria tritici* is the causal agent of Septoria tritici blotch (STB), one of the most important foliar diseases of wheat, globally. *Z. tritici* is a strictly apoplastic pathogen that can secrete numerous proteins into the apoplast of wheat leaves to promote infection. We sought to determine if, during STB infection, wheat also secretes small proteins into the apoplast to mediate the recognition of pathogen proteins and/or induce defense responses. To explore this, we developed an SSP-discovery pipeline to identify small, secreted proteins from wheat genomic data. Using this pipeline, we identified 6,998 SSPs, representing 2.3% of all proteins encoded by the wheat genome. We then mined a microarray dataset, detailing a resistant and susceptible host response to STB, and identified 141 *Z. tritici*-responsive SSPs, representing 4.7% of all proteins encoded by *Z. tritici* – responsive genes. We demonstrate that a subset of these SSPs have a functional signal peptide and can interact with *Z. tritici* SSPs. Transiently silencing two of these wheat SSPs using virus-induced gene silencing (VIGS) shows an increase in susceptibility to STB, confirming their role in defense against *Z. tritici*.

Keywords: Septoria tritici blotch (STB), *Zymoseptoria tritici*, small secreted proteins (SSPs), wheat disease resistance, apoplastic proteins, protein secretion

INTRODUCTION

One of the most economically important species in the plant kingdom is bread wheat, *Triticum aestivum*. Wheat dominates the European arable sector, with ~150 million tons of wheat grown in the European Union annually (FAOSTAT 2019). While yields are generally high across the EU, wheat production is threatened by a range of pests and pathogens. One of the most important of these is Septoria tritici blotch, a foliar disease caused by the pathogenic fungus *Zymoseptoria tritici* (*Z. tritici*) (O'Driscoll et al., 2014). *Z. tritici* is a strictly apoplastic fungus, and is a host-specific pathogen of wheat. The high selection pressure within intensive agricultural systems [high fungicide usage and dense planting of STB-resistant varieties (Fones and Gurr, 2015)], combined with rapid evolution of the pathogen (Dooley, 2015), has led to the widespread occurrence of

Z. tritici populations that are resistant to fungicides, or can overcome resistance genes deployed in elite cultivars, or both (Cools and Fraaije, 2013; McDonald and Stukenbrock, 2016; Heick et al., 2017). There are two main phases of STB disease: the symptomless latent phase, during which hyphae of *Z. tritici* enter the leaf tissue via the stomata and begin to colonize the substomatal cavity (Kema et al., 1996), and the subsequent necrotrophic phase. The symptomless phase lasts ~12 days (dependent on wheat cultivar, *Z. tritici* isolate and environmental conditions) (Hehir et al., 2018), after which the fungus switches to a necrotrophic feeding habit and host tissue begins to die (Keon et al., 2007).

Plants have evolved a multi-layered immune system to recognize and defend themselves against invading pathogens such as *Z. tritici* (Jones and Dangl, 2006). The first layer of plant immunity is pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI). There is a growing body of evidence demonstrating that the apoplast, i.e., the space outside of the plasma membrane, serves as the front-line between the plant host and invading pathogens, and is spatially significant for PTI (Jashni et al., 2015; Wang and Wang, 2018; Schellenberger et al., 2019). Immune receptors on the plant cell surface [known as pattern-recognition receptors (PRRs)], typically with an external binding, lectin or lysin-motif (LysM) domain, play determinant roles during infection by detecting PAMPs; for example the Chitin Elicitor Binding Protein (CEBiP) and Chitin Elicitor Receptor Kinase1 (CERK1), which can recognize the fungal PAMP chitin in *Arabidopsis thaliana* (Miya et al., 2007; Desaki et al., 2018). These receptors activate downstream plant defense responses [encoded by pathogenesis-related (PR) genes], such as the production of reactive oxygen species (ROS), the activation of transcription factors, and the secretion of various pathogenesis-related (PR) proteins into the apoplast that can: hydrolyse glucans, chitin and polypeptides (Tian et al., 2004; Ilyas et al., 2015; Jashni et al., 2015; Ali et al., 2018), inhibit pathogen-secreted enzymes (Kim et al., 2009; Jashni et al., 2015; Rustgi et al., 2018), and phytochemically inhibit pathogen growth (Wirthmueller et al., 2013).

While PRRs recognize and play an important role in resistance to most non-adapted microbes, known as basal resistance (Couto and Zipfel, 2016), when adapted to their host, pathogens can deploy small secreted proteins (SPPs) that act as effectors to suppress or block PTI-induced defense pathways (Block et al., 2014). Hundreds of candidate *Z. tritici* effector genes have been identified via comparative genomics and transcriptomic analyses (Yang et al., 2013; Mirzadi Gohari, 2015; Rudd et al., 2015; Palma-Guerrero et al., 2016; Kettles et al., 2017; Plissonneau et al., 2018). Pathogen effectors are deployed in a spatial and time-dependant manner, depending on the stage of infection. In pathogenic bacteria, effectors are secreted directly out of bacterial cells and/or into the plant cells via multiple secretion systems. For example, the *Pseudomonas syringae* effector HopAO1 and *Ralstonia solanacearum* effector PopP2 are secreted directly into *A. thaliana* plant cells via the bacterial type-III secretion system, and suppress immune responses by targeting receptor kinases and multiple WRKY transcription factors (Macho et al., 2014; Le Roux et al., 2015). In fungi

and oomycetes, the effectors are secreted inside (cytoplasmic) or outside (apoplastic) plant cells via the general secretory pathway and through various feeding and infection structures, such as extracellular hyphae and haustoria (Petre and Kamoun, 2014; Wang et al., 2017). During *Z. tritici* infection of wheat, effector proteins are secreted in the apoplast of wheat plant cells, such as Mg3LysM (Marshall et al., 2011; Lee et al., 2014), which interferes with chitin-triggered immunity and helps establish the disease during the latent phase of infection. Although the causes for the rapid switch to necrotrophy in the *Z. tritici* life cycle are largely unknown, several *Z. tritici* effectors have been implicated in initiating the necrotrophic phase, such as MgNLP, ZtNIP1, and ZtNIP2 (Marshall et al., 2011; Ben et al., 2015).

In response to the secretion of effectors, plants have developed a second layer of immunity, in which host nuclear-binding leucine-rich repeat (NLR) proteins, typically characterized by an extracellular leucine-rich repeat (LRR) domain (Chiang and Coaker, 2015), recognize pathogen effectors. Recognition of effectors, leading to ETI, can elicit a hypersensitive response, often associated with salicylic acid (SA) signaling and systemic acquired resistance (SAR) (Kombrink and Schmelzer, 2001). Additionally, plant small secreted proteins have also been reported to play key roles in plant immunity (Lanver et al., 2017; Ziemann et al., 2018; Segonzac and Monaghan, 2019). The *A. thaliana* protein AtPep1 is a 23 AA long peptide that enhances plant resistance to various pathogens, including the bacterium *P. syringae*, the fungus *Botrytis cinerea*, and the oomycete *Phytophthora infestans* (Huffaker et al., 2006; Yamaguchi et al., 2010; Liu et al., 2013). In maize (*Zea mays*), the ortholog of AtPep1 (ZmPep1) was demonstrated to activate the production of jasmonic acid and induce multiple defense pathways to enhance resistance against the fungal pathogens *Cochliobolus heterostrophus* and *Colletotrichum graminicola* (Huffaker et al., 2011). Additionally in *Z. mays*, a 17 AA peptide, termed *Z. mays* immune signaling peptide 1 (Zip1), is a functional elicitor of SA signaling in maize (Ziemann et al., 2018). In the case of the wheat - *Z. tritici* interaction, wheat can secrete β -1,3-glucanase into the apoplast, which cleaves β -1,3-glucan in the *Z. tritici* cell wall to prevent colonization of *Z. tritici* (Shetty et al., 2009).

These findings have demonstrated that plant secreted proteins play significant roles in apoplastic immunity in plant-pathogen interactions, and that plant-encoded SSPs may be an important reservoir of potential STB-resistance genes for wheat. Using features typical of small secreted proteins, such as a protein length ≤ 250 amino acids and a secretion signal of an N-terminal signal peptide, we investigated the small secretome of wheat, to identify small secreted proteins from that may play a role in the wheat - *Z. tritici* interaction, and may interact with fungal SSPs that are also present in the apoplast during infection. The aims of this study were to determine: (1) if wheat-encoded SSPs are regulated during wheat-*Z. tritici* interactions, (2) whether some SSPs might be able to enhance wheat resistance to *Z. tritici* and (3) if yes, how molecular mechanisms for SSPs contribute to wheat resistance.

MATERIALS AND METHODS

Plant and Fungal Material

Wheat (*T. aestivum*) cultivars (cvs.) Stigg and Gallant were used in this study. The cv. Stigg [Pedigree: (BISCAY/LW-96-2930//TANKER)] is resistant to STB disease (Hehir et al., 2018) and cv. Gallant (Pedigree: TJB-268-175/HOBBIT) is susceptible to STB disease (Orton and Brown, 2016). Wheat seeds were incubated at 4°C for 5 days then subsequently transferred to a dark 19°C growth room for 3 days. Germinated seeds were transferred to 2 L trays filled with John Innes Compost No. 2 soil (Westland Horticulture, United Kingdom). Plants were grown under controlled conditions at 19°C with a 15/9 h light/dark cycle and the relative humidity was maintained at 80% using a Humidisk 10 humidifier (Carel, Italy). *Nicotiana benthamiana* seeds were incubated at 4°C for 3 days in a cold room. Then the seeds were transferred to a growth chamber at 22°C (day) to 19°C (night) with a 16/8 h light/dark for 5 weeks before infiltration for all experiments.

The *Z. tritici* isolate used in this study was a field isolate collected from the wheat cv. Cordiale in Cork, Ireland hereafter referred to as 'Cork Cordiale 4.' Glycerol stocks were provided by Dr. Thomas Welch (Teagasc, Crops Research Centre, Carlow, Ireland). *Z. tritici* was cultured by inoculating YPDA (10 g Yeast extract, 20 g Bacteriological peptone, 20 g D-Glucose, and 15 g Agar in 1 L water) plates with 50 µl of the glycerol stock to generate conidia. The petri dishes were transferred to a near-ultraviolet light incubator for 7 days at 20°C with a 12:12 h light/dark cycle. Plates were flooded with 3 mL sterile water and scraped with a sterile spreader to collect the *Z. tritici* spores (pycnidiospores). Spores were filtered through sterile cheesecloth and the concentrations were measured using a Glasstic hemocytometer (Kova International, United States). The final spore concentration was adjusted to 1×10^6 per ml and 0.02% Tween20 (Fisher Bioreagents, United States).

Identification and Characterization of Wheat Small Secreted Proteins

A bioinformatics pipeline was developed to automate the identification of wheat small, secreted proteins (TaSSPs), written in the Bash-command and R languages (Figure 1A). Briefly, the script takes gene IDs as an input and retrieves their corresponding protein sequences from the IWGSC refseq V1.1 protein annotation¹ (IWGSC, 2018), using the SAMtools fasta index function (Li et al., 2009; Li, 2011). The length of each query protein was retrieved, and the standalone SignalP V5.0 software (Armenteros et al., 2019), with default parameters, was used to detect the presence of a signal peptide and the location of their cleavage site in the protein sequences (Armenteros et al., 2019). The pipeline scripts are open access and can be accessed at https://github.com/hbenbow/SSP_pipeline.git. Using this pipeline, the entire wheat proteome was searched for SSPs, by splitting the reference protein annotation by chromosome and mining each chromosome individually. The resultant set of predicted SSPs

was further refined by identifying and removing proteins with any transmembrane helices, [using TMHMM v2.0 (Krogh et al., 2001)], and glycosylphosphatidylinositol (GPI)-anchors [using GPI-SOM (Fankhauser and Maser, 2005)]. TaSSP genes were annotated using Blast2GO (Conesa et al., 2005), and a Fisher's enrichment test was carried out between the TaSSP set and the whole genome to test if any GO terms were significantly enriched in the TaSSP set [note; only high-confidence TaSSP genes (i.e., 4,532 of the total 6,998) were included in this analysis, as only high confidence gene annotations were present in the reference Blast2GO annotation]. To further characterize the small secreted proteins, the signal peptide sequence was cut from the sequence FASTA file using the Bedtools getfasta algorithm (Quinlan and Hall, 2010), using a.BED file of coordinates from 1:n, where n = the cleavage site position defined by SignalP. The MEME-suite (Bailey et al., 2009) was used to identify motifs in the signal peptides, using the options -nmotifs 10 to find the top 10 motifs in the set of signal peptide sequences. Following MEME, MAST (Bailey and Gribskov, 1998) was used to align the motifs back to the sequences and identify which sequences contained which motifs. To identify signal peptide sequences that were similar to each other, the sequences were clustered with CD-HIT (Li et al., 2001), where sequences with >90% similarity to each other were clustered into groups. The distribution of clusters and cluster size was reported using the Perl script plot_len.pl from <https://github.com/weizhongli/cdhit>.

Identification and Validation of *Z. tritici*-Responsive SSPs

Zymoseptoria tritici-responsive SSPs were identified using the differentially expressed probe set from Brennan et al. (2020, in press); a microarray study, which assessed the transcriptome responses of winter wheat cvs. Stigg and Gallant to *Z. tritici* (isolate IPO323) at 4, 8, and 12 days post-inoculation (dpi), available at <https://doi.org/10.6084/m9.figshare.11882601.v1>. Microarray probe sequences were retrieved from Affymetrix². Probe sequences for every differentially expressed probe in each cultivar \times timepoint combination were BLASTn searched against the IWGSC v1.1 reference CDS annotation (IWGSC, 2018) using BLAST+. As the microarray probes could potentially hybridize to all three homoeologues of each wheat gene, a one-to-one search algorithm was not appropriate for identifying the full gene sequence of each microarray probe. Therefore, bespoke Bash and R scripts were created to identify the top three IWGSC hits for each microarray probe. The probe sequences were used as the query and the IWGSC reference was used as the search subject. The BLASTn short sequence algorithm was used with the parameters -max_target_seqs 1, -max_hsps 3 and -task blastn-short, to return a maximum of 3 high-scoring pairs. The BLASTn results were returned in tabular format (-outfmt 6). The output file was sorted first by query ID, then by (in this order): bitscore (descending), E-value (ascending) and percentage identity (descending). From this sorted file, the top three hits for each query sequence were retained. These

¹https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Annotations/v1.1/, accessed January 2020

²http://www.affymetrix.com/Auth/analysis/downloads/data/wheat.probe_fasta.zip, accessed January 2020

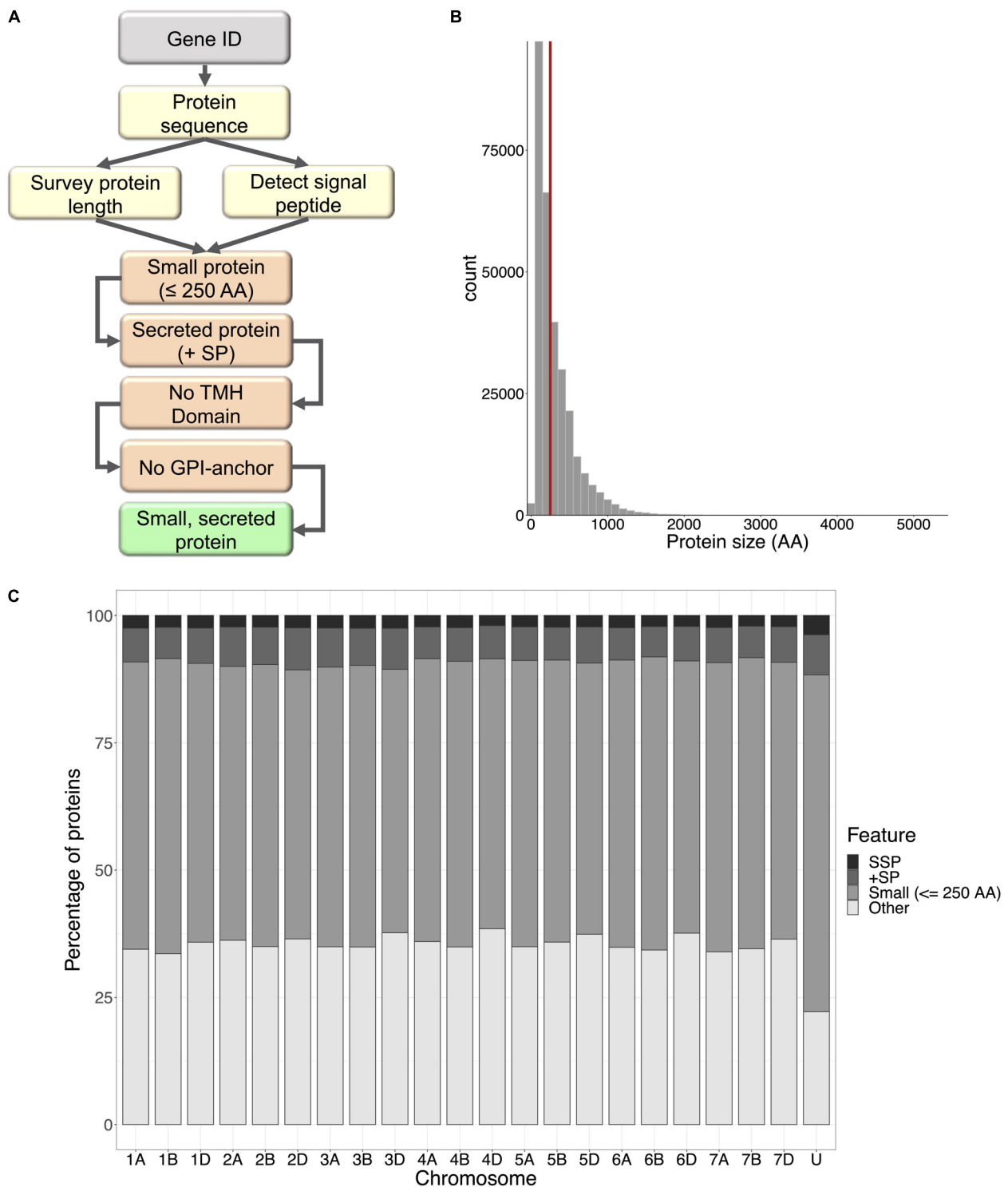


FIGURE 1 | Small secreted proteins of wheat. **(A)** The small secreted protein (SSP) discovery pipeline, written in Bash command and R languages. Protein sequences are retrieved for query gene IDs, surveyed for length and SignalP v5.0 was used to detect presence of a signal peptide. Small proteins with a signal peptide were filtered for transmembrane helix domains (TMH) and GPI-anchors. The resulting proteins were designated SSPs. **(B)** Frequency distribution of protein size for all proteins encoded within the wheat genome. The cut-off size for SSPs (250 amino acids) is shown by the red line; 58% of all wheat proteins were ≤ 250 AA. **(C)** The percentage of small proteins, proteins with a signal peptide, and SSPs with no TMH domains and GPI-anchors encoded by genes on each chromosome of the wheat genome.

scripts are open access and can be accessed at https://github.com/hbenbow/SSP_pipeline.git. *Z. tritici*-responsive genes were cross-referenced against the list of SSPs. To choose candidate *Z. tritici*-responsive SSPs, we focused on SSP genes with a high fold-change in cv. Stigg (resistant) compared to cv. Gallant (susceptible), and candidates were chosen for cloning and further study. *In silico* analysis of these genes was done using a BLASTx search to the NCBI non-redundant nucleotide database, using default parameters, and InterProScan. Both of these were performed as part of the OmicsBox desktop application.

Expression of candidate SSP genes was validated by qRT-PCR as per Brennan et al. (2020). Plants of cvs. Stigg and Gallant were grown as stated above, and at growth stage 21 (Zadocks et al., 1974), the third leaf was spray inoculated with 1 ml (1×10^6 spores) *Z. tritici* on both the adaxial and abaxial surface using a Hozelock 0.5 L hand-held mist sprayer (Hozelock LTD., United Kingdom). Control plants were inoculated with a solution of 0.02% Tween20. A total of three independent trials were conducted, each with four plants (2 per pot) per time point per cultivar per treatment. At 4, 8, and 12 dpi, the entire third leaf was excised and flash frozen in liquid nitrogen for RNA extraction.

RNA Extraction and cDNA Synthesis

Leaf tissue was ground in liquid nitrogen in a sterile mortar and pestle. Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, Germany) following the manufacturer's recommendations. gDNA was removed from RNA extraction samples using TURBO DNA-free™ Kit (Ambion, United States) in accordance with the manufacturer's protocol. RNA quality and integrity were checked using an ND-1000 Spectrophotometer NanoDrop (Thermo Scientific, United States) and it was visualized on a 1.5% agarose gel. DNA removal was validated by PCR using Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)-specific primers, which span an intron (Supplementary Table S1). Each PCR reaction contained 0.125 µl Ex Taq™, 2.5 µl 10X Ex Taq Buffer, 2 µl dNTP mixture, 2 µl treated RNA sample (or 2 µl gDNA 50 ng/µl serving as the positive control), 2 µl 5 µM Primer in 25 µl reaction volume, with following conditions: 1 cycle of 30 s at 98°C; 40 cycles of 5 s at 98°C and 20 s at 60°C; and a final cycle of 2 min at 72°C. PCR products were visualized using 1.5% agarose gel electrophoresis. Reverse transcription of total RNA was performed using SuperScript II Reverse Transcriptase (Invitrogen, United States) following the manufacturer's recommendations. Two cDNA samples were synthesized from each RNA sample.

Quantitative Real-Time PCR (qRT-PCR) Analysis

Quantitative real-time PCR (qRT-PCR) analysis was conducted using the Stratagene Mx3000™ Real-Time PCR (Stratagene, United States). Each reaction was performed with 1.25 µL of a 1:5 (V/V) dilution of cDNA, 0.2 µM of each of the primers (Supplementary Table S1) and 1X SYBR Premix Ex Taq (Takara, Japan) in a total reaction volume of 12.5 µL, with following conditions: 1 cycle of 1 min at 95°C; 40 cycles of 5 s at 95°C and 20 s at 60°C; and a final cycle of 1 min at 95°C,

30 s, at 55°C, and 30 s at 95°C for the dissociation curve. All real-time qRT-PCR analyses were conducted in duplicate. Two housekeeping genes were used as reference genes, α -tubulin and Glyceraldehyde phosphate dehydrogenase 2 (GAPDH2) (Supplementary Table S1).

The threshold cycle (Ct) values obtained by real-time RT-PCR were used to calculate the Δ Ct values for the formula $\Delta Ct = Ct_{(target\ gene)} - \mu[Ct_{(housekeeping\ genes)}]$. Relative expression was calculated using the formula $2^{-\Delta Ct}$ (Livak and Schmittgen, 2001). qRT-PCR was carried out for each of the three independent trials, with 2 reactions per cDNA, and 2 cDNAs per RNA extraction.

Cloning of Wheat Small Secreted Protein (TaSSP) Encoding Genes

The full length *TaSSP* genes were amplified from cDNA produced from *Z. tritici*-infected wheat leaf (cv. Stigg or Gallant) using primers matching the 5' and 3' UTR of the *TaSSPs* genes (Supplementary Table S1). The PCR reactions (50 µl) contained 0.25 µl Ex Taq™, 5 µl 10X Ex-Taq Buffer, 4 µl dNTP mixture, 2 µl treated cDNA, 4 µl primer (5 µM), and 3 technical replicates with the following program constituted 98°C for 30 s, 35 cycles of 98°C for 10 s, extension of 68°C for 1 min, with a final extension at 72°C for 5 min. The PCR product was cloned into pDONR207 (Invitrogen, United States) after the 2nd amplification using the attB1 and attB2 primers (Supplementary Table S1) and was subsequently introduced into different expression vectors by the Gateway cloning technology (Invitrogen, United States) for gene function analysis.

Developing a Sucrose Transport Protein Signal Sequence Trap System and Testing of TaSSPs Secretion

A yeast sucrose transport protein *SUC2* signal sequence trap system was developed and used to determine whether *TaSSP* proteins were secreted (the schematic diagram of the yeast secretion assay is shown in Supplementary Figure S1). The sucrose transport protein *SUC2* gene of *Saccharomyces cerevisiae* strain SEY6210 (ATCC: The Global Bioresource Center) was replaced by the tryptophan synthesis (*Trp1*) gene via homologous recombination (Horecka and Davis, 2014) using primers POP-IN-U2-F, POP-IN-D2-R, Pop-Trp-U2-F and Pop-Trp-D2-R (Supplementary Table S1). The *suc2* mutant yeast cells were selected on synthetic Trp dropout (-Trp) yeast media (Takara, Japan). To construct yeast expression vectors for the secretion assay, the full length and truncated (without signal peptide, *SUC2*^{22–511}) *SUC2* genes were cloned into the pGADT7 plasmid (Clontech, United States) using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB, United States) according to the manufacturer's instructions. A DNA fragment containing Gateway cassette, HA tag, Kex2 cleavage site (TCTCATGGTTCTTTGGATAAAAGAGAGGCTGA) and *SUC2*^{22–511} gene was synthesized by General Biosystems (United States) and ligated into the pGADT7 plasmid (Clontech, United States) to generate a Gateway compatible vector pGAD-GW-*SUC2*^{22–511} for *TaSSP* protein secretion assays in yeast. All yeast expression vectors sequences for secretion analysis are

presented in **Supplementary Figure S2**. To test the secretion of TaSSPs, the candidate TaSSP genes were cloned into the secretion vector pGAD-GW-SUC2^{22–511} using Gateway recombination cloning technology (Invitrogen, United States). The TaSSP genes were fused to the N-terminus of the SUC2^{22–511} gene. The vectors were transformed into the *suc2* mutant yeast following the yeast transformation protocols (Sigma-Aldrich). Yeast was spread on a synthetic Trp and Leu dropout (-TL) plates with sucrose (10 mM) as the sole carbon source. The Petri dishes were transferred to an incubator at 28°C for 3 days. If TaSSPs were secreted, the positive *suc2* mutant yeast transformants grew on the plate and were visible after 3 days (Plett et al., 2017). Four trials were conducted, in which nine independent yeast clones were grown and three technical reps from each yeast clone were tested. Three biological replicates were conducted per trial and yeast spotting on the media was performed using serial dilutions from an initial OD₆₀₀ of 1.0, 0.1, 0.01, and 0.001, respectively.

Yeast Two-Hybrid Analysis

The interaction between TaSSP proteins and *Z. tritici* SSPs (ZtSSPs) were assessed via yeast two-hybrid (Y2H) analysis. Twenty-seven non-annotated ZtSSPs were identified by filtering the publicly available secretome dataset from do Amaral et al. (2012) to identify *Z. tritici* small secreted proteins (ZtSSPs). The candidate sequences were filtered based on the following features: EST support, size (≤ 315 aa), presence of cysteine residue, presence of signal peptide using SignalP v5.0 (Armenteros et al., 2019) and lack of transmembrane domain predicted by TMHMM v2 (Krogh et al., 2001). Finally, putative secreted proteins with unknown functional conserved domains were selected using NCBI CDD (Conserved Domain Database) and the Pfam database (JGI Protein ID of predicted ZtSSPs is listed in **Supplementary Table S2**). Truncated TaSSP and ZtSSP genes (lacking their signal peptides) were amplified by PCR using gene-specific primers (**Supplementary Table S1**) and cloned into the vector pDONR207 using the Gateway cloning technology (Invitrogen, United States). Both TaSSPs and ZtSSPs were then recombined into bait and prey vectors derived from pGADT7 and pGBKT7 plasmids (Clontech, United States). The bait and prey vectors were transformed into a yeast strain (Y2H Gold, Clontech) and grown on Trp and Leu drop-out medium (-TL) at 28°C for 3 days. The yeast cells carrying both plasmids were selected on Trp/Leu/His/Ade drop-out medium (-TLHA). Three technical replicates were performed per TaSSP-ZtSSP combination. If TaSSPs interacted with ZtSSPs, the yeast can grow on -TLHA plates at approximately 3–7 days. Three trials were performed, in which nine independent yeast clones were grown and divided into three technical replicates to be tested.

Bimolecular Fluorescence Complementation

Bimolecular fluorescence complementation (BiFC) was used to validate the interactions between TaSSPs and ZtSSPs *in planta*. The relevant pDONR207 vectors encoding TaSSPs and ZtSSPs used for Y2H were recombined into the BiFC vectors pDEST-VYCE^{GW} and pDEST-VYNE^{GW} (Gehl et al., 2009) using Gateway cloning technology (Invitrogen, United States). This

generated constructs wherein proteins were fused to the YFP C-terminal (YFP^C) or N-terminal fragment (YFP^N). The vectors were then transformed into the *Agrobacterium tumefaciens* strains GV3101 by electroporation. The transformed GV3101 strains were cultured in LB liquid medium containing gentamicin (20 µg/ml), kanamycin (50 µg/ml), and rifampicin (50 µg/ml) at 28°C overnight. *A. tumefaciens* was harvested by centrifuge at 4000 rpm for 10 min and washed once with distilled water. The *A. tumefaciens* cells were resuspended in infiltration buffer (10 mM MES pH 5.6, 10 mM MgCl₂, and 150 µM acetosyringone) to an OD₆₀₀ = 0.5 and incubated in the dark for 2 h at room temperature. The leaves of 5 weeks old *N. benthamiana* plants were infiltrated using a 1 ml needleless syringe. Epidermal cells of leaves were assayed for YFP fluorescence using a Confocal Laser Scanning Microscope (Olympus fluoview FV1000) at 2 days post-infiltration. YFP fluorescence was excited at 515 nm and detected in the range between 530 and 630 nm. Three trials were conducted and within each trial, three independent leaves were analyzed per TaSSP-ZtSSP combination.

Agrobacterium-Mediated Expression of ZtSSPs

It is reported that some ZtSSPs can induce cell death in *N. benthamiana* leaves (Kettles et al., 2017). The ZtSSPs that interacted with TaSSPs were cloned into a high-level expression vector pEAQ-HT-DEST3 (Sainsbury et al., 2009) using Gateway cloning technology (Invitrogen, United States). The constructs were transformed into *A. tumefaciens* strains GV3101 by electroporation. Five week old *N. benthamiana* plants were infiltrated as described by Kettles et al. (2017). The only modification was that the GV3101 was finally resuspended at OD₆₀₀ = 1.0 before infiltration. The infiltrated leaves were observed for 10 days to check for cell death. Four independent leaves were analyzed per ZtSSP per trial, and three trials were conducted in total. As a negative control, GFP was infiltrated into four independent leaves per trial. To test if infiltration of co-expressed ZtSSP-TaSSP combinations affected the cell death phenotyping in *N. benthamiana* leaves, co-expressed proteins were infiltrated into *N. benthamiana* leaves as above, and six biological replicates (from three independent plants) were conducted.

Virus-Induced Gene Silencing (VIGS)

Virus-induced gene silencing (VIGS) was used to determine the impact of TaSSP genes on STB disease, based on the Barley Stripe Mosaic Virus (BSMV) method (Scofield et al., 2005; Gunupuru et al., 2015). Two gene fragments were used for VIGS of TaSSP6 and TaSSP7 genes and these were amplified from the CDS of TaSSP6 and TaSSP7 (VIGS primer sets are listed in **Supplementary Table S1**). VIGS target sequences were chosen to preferentially silence all three (A, B, and D genome) homoeologues (where present) using the publicly available online Wheat Ensembl database³. The PCR amplicons of silencing fragments were digested and ligated into BSMV-γ vectors using *NotI/PacI* (NEB, United States). They were named

³http://plants.ensembl.org/Triticum_aestivum/Info/Index

BSMV:TaSSP6-V1, BSMV:TaSSP6-V2, BSMV:TaSSP7-V1, and BSMV:TaSSP7-V2. Inserted gene fragments were confirmed by sequencing. In addition, a BSMV- γ vector with silencing fragments for phytoene desaturase (PDS) was used as a positive control and a BSMV- γ empty vector as a negative control. Plasmid linearization, *in vitro* transcription of RNA, and flag leaf inoculation with 1:1:1 mixtures of the *in vitro* transcripts of BSMV α , β , and γ RNA were done as previously described (Scofield et al., 2005). Plants were placed in low light conditions overnight and allowed to recover from mechanical stress; thereafter plants were returned to normal growth conditions. The *Z. tritici* inoculation was applied to the third and fourth leaves of wheat plants at 7 days post-VIGS constructs inoculation. The third leaf of each plant was taken for qRT-PCR validation at 8 days after *Z. tritici* inoculation and the fourth leaf was used for subsequent phenotyping. STB disease severity was assessed by scoring the percentage of leaf area bearing necrosis at 21 dpi. The leaves were then excised from the plant and placed in 100% humidity to promote pycnidia growth. For the VIGS experiment, each trial included 12 plants per treatment combination and the trial was replicated three times.

Statistical Analysis

All statistical analyses were performed in R v3.5.2. Data were checked for normality using a Shapiro–Wilk test. To adjust for variation between the trials for the gene expression time course, relative expression was adjusted to a percentage of relative gene expression in control plants of cv. Gallant at 4 DPI. Data were transformed using Johnson transformation and One-way ANOVA was used to calculate differences in gene expression between Cultivar + Treatment + Timepoint combinations. Tukey's HSD test was used for multiple comparisons of means. A Kruskal–Wallis test was used to analyze VIGS phenotype data with Dunn's *post hoc* test. VIGS qRT-PCR data for SSP6 were analyzed using a Kruskal–Wallis test and means were compared using Dunn's test, and SSP7 data were analyzed using One-way ANOVA with Tukey's HSD *post hoc* test.

RESULTS

Identification of Wheat Small Secreted Proteins (TaSSPs)

From across the wheat protein reference (298,774 proteins), the mean protein length was 311 AA, and the median was 218 AA (Figure 1B). The shortest proteins were 12 AA long, and the longest was 5,360 AA. We identified 166,086 small (≤ 250 AA) proteins, 20,763 proteins with a predicted signal peptide, and 8,467 (2.8% of the wheat proteome) proteins that were small and had a signal peptide (Figure 1C). From this set, 1,460 proteins had a predicted GPI-anchor, 12 contained more than one transmembrane helix (TMH) domain, and 3 had both GPI-anchor and TMH. This left a total of 6,998 unique wheat proteins that were classified as 'small, secreted proteins,' representing 2.3% of the wheat protein annotation. The SSPs are available in **Supplementary File 2**. The percentage of SSPs attributed to the 21 wheat chromosomes ranged from 1.9–2.4%, and 3.7% were attributed to chromosome 'Un' (a chromosome

reference that contains assembled sequence contigs that have not been unambiguously assigned to a chromosome) (Table 1). Of the annotated *TaSSP* genes, the most dominant biological process was the negative regulation of peptidase activity. The top biological processes also included the defense response, the negative regulation of catalytic activity, and lipid transport. The most common molecular functions were nutrient reservoir activity, manganese binding activity and enzyme inhibitor activity. The most common cellular component of the *TaSSPs* was the extracellular region, followed by the membrane and the apoplast. The Fisher's enrichment test revealed that 119 biological processes, 46 molecular functions, and 12 cellular components were significantly (FDR < 0.05) enriched (over-represented) in the *TaSSP* set, compared to the full wheat gene reference set. The most over-represented GO terms included the regulation of molecular functions, the negative regulation of catalytic activity, enzyme inhibitory activity, the defense response, peptidase inhibitory activity, the extracellular region and the apoplast, among others (Figure 2).

The MEME-suite was used to identify sequence motifs in the signal peptide sequences of the proteins, using the cut-off of 10 motifs. Ten motifs were discovered in the sequence data with an *E*-value < $1e^{-15}$. When the motif sequences were realigned back to the sequences, we profiled which signal peptide sequences contained which motif. Each motif represented an average of 41 signal peptide sequences, suggesting that there is a lot of sequence dissimilarity and there are no motifs or common features found between all, or most, of the sequences. We clustered sequences based on their sequence similarity, clustering signal peptides that were >90% similar to each other. Only clusters with more than 3 sequences were considered true clusters, as signal peptides from homoeologous proteins almost exclusively clustered together, contributing to a high number of clusters with 2 or 3 proteins. By comparing the motifs within the clusters, we identified some clusters of signal peptides that all contained the same motif, but most of the clusters were not represented by any of the 10 motifs found. Due to the large number of small clusters, it seems that the SP sequences are generally too different from each other to analyze, and each cluster is represented by a different sequence motif. Therefore, we conclude that there are no defining motif features within wheat SSPs.

TaSSP Gene Expression Was Induced During *Z. tritici* Infection

We mined microarray data to identify *TaSSPs* that were responsive to *Z. tritici* (isolate IPO323) at 4, 8, and 12 days post-inoculation in cultivars that were STB-resistant (cv. Stigg) or susceptible (cv. Gallant). From the microarray data, 5,163 genes in total, corresponding to 2,968 unique *Z. tritici*-responsive genes (some genes were differentially expressed at multiple timepoints or in both cultivars) were identified via the BLASTn algorithm corresponding Affymetrix probes to IWGSC refseq V1.1 genes IDs, and the proteins corresponding to these genes were retrieved from the protein annotation. In total, 198 SSPs were differentially expressed in the microarray study across the two cultivars and three timepoints (Table 2). These 198 genes/proteins correspond to 141 unique proteins, representing

TABLE 1 | The number of small, secreted proteins (SSPs) in the wheat genome.

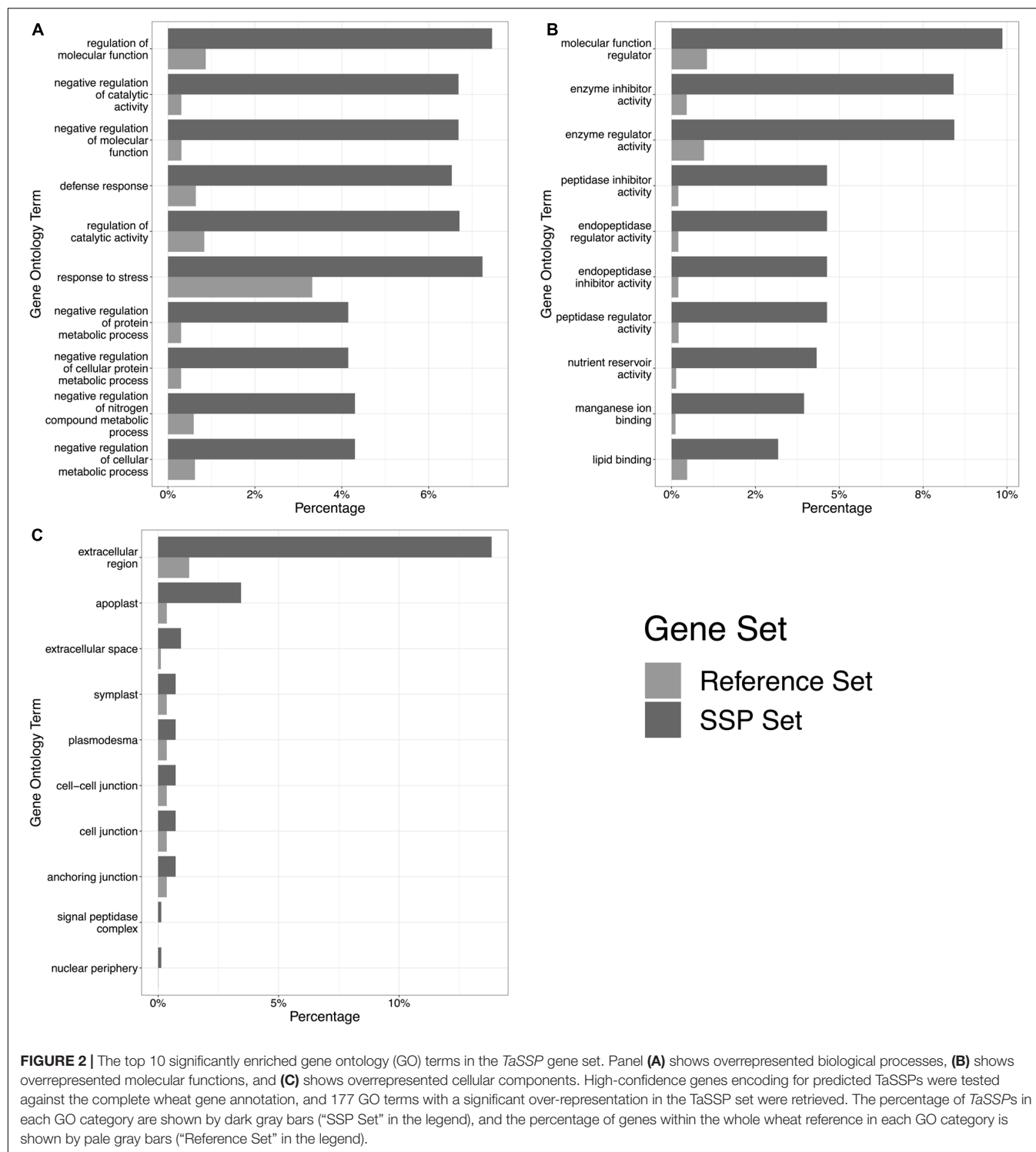
Chromosome	Whole genome			Filtered		SSPs ^e
	Total	Small ^a	+SP ^b	+GPI-anchor ^c	+TMH ^d	
1A	11947	6743	793	55	1	295
1B	14047	8144	864	70	0	322
1D	11653	6384	808	66	1	285
2A	15186	8172	1176	81	0	339
2B	17266	9567	1268	86	0	392
2D	14982	7920	1235	83	0	361
3A	14391	7909	1103	81	1	352
3B	16845	9322	1227	85	0	419
3D	13769	7129	1109	71	0	342
4A	13791	7664	860	75	0	307
4B	11409	6403	754	59	2	270
4D	9600	5091	626	53	0	189
5A	14389	8094	949	72	1	319
5B	15355	8509	989	59	1	352
5D	13814	7360	981	73	0	307
6A	11533	6510	731	37	0	275
6B	14248	8207	847	36	0	307
6D	10383	5554	703	29	0	221
7A	15146	8608	1043	77	1	355
7B	14760	8442	914	70	0	305
7D	14430	7847	1009	68	1	314
Un	9830	6507	774	74	3	370

^aSmall (≤ 250 amino acids) proteins. ^bThe number proteins with a signal peptide (predicted to be secreted). ^cSmall proteins, with a signal peptide and a predicted glycosylphosphatidylinositol anchor. ^dSmall proteins, with a signal peptide and predicted transmembrane helices (≥ 1). ^ePredicted small, secreted proteins (with no GPI-anchor or TMH).

4.7% of all differentially expressed genes/proteins. Of the 141 unique SSPs, 35 were uniquely differentially expressed in cv. Stigg, and 75 were uniquely differentially expressed in cv. Gallant. The remaining 27 SSPs were differentially expressed in both cultivars. The number of SSPs in the differentially expressed genes is significantly higher (χ^2 P -value < 0.05) than the total percentage of SSPs across the wheat proteome (2.3%), indicating enrichment of SSPs in the disease response of wheat. Across the two cultivars, Stigg (STB-resistant) and Gallant (STB-susceptible), there was little difference in the percentage of differentially expressed genes that encoded SSPs, although a slightly higher percentage of SSPs was detected in Gallant (5.3%) as compared to Stigg (4.7%). Of the cv. Gallant-specific SSPs (75 SSPs), the most abundant biological processes and molecular functions were lipid transport and binding, cell surface receptor signaling, redox homeostasis, electron transfer, and peptidase activity. From the cv. Stigg-specific SSPs, the most dominant biological processes and molecular functions were the negative regulation of endopeptidase activity, lipid transport, and metal ion binding. A higher percentage of the differentially expressed genes were up-regulated (6.5%) versus down-regulated (3.5%) across both cultivars. The most striking difference was the temporal difference in SSP expression: 7.2% of the differentially expressed genes (DEGs) at 4 dpi across both cultivars were SSPs, compared to 4.6% at 8 dpi and 3.3% at 12 dpi (Table 2). From the *Z. tritici*-responsive SSPs, we chose two for

further characterization based on their fold change in cv. Stigg compared to cv. Gallant; TaSSP6 and TaSSP7 (Table 3). TaSSP6 was represented by the microarray probe Ta.23397.1.S1_x_at, which was upregulated in cv. Stigg at 4 dpi by 4.1-fold, but was not differentially expressed in cv. Gallant at 4 dpi. This probe was upregulated in cv. Gallant at 8 dpi by 6.3-fold, but was not differentially expressed in cv. Stigg at 8 dpi. TaSSP7 was represented by the probe Ta.28289.2.S1_x_at, which was upregulated at 8 dpi in cv. Stigg by 1.7-fold, and at 12 dpi in cv. Stigg by 44.7-fold. TaSSP7 was not differentially expressed in cv. Gallant. Therefore, TaSSP6 is *Z. tritici*-responsive in both cvs. Stigg and Gallant, but was upregulated earlier in Stigg than Gallant, and TaSSP7 is specific to cv. Stigg (at least at the time points explored). TaSSP6 consists of three homoeologues on the group 2 chromosomes; the A homoeologue encodes one splice variant, the B homoeologue encodes three splice variants, and the D homoeologue encodes two splice variants. All six variants of TaSSP6 encode for small, secreted proteins. BLASTx of the TaSSP6 homoeologues and variants revealed that 6 of the TaSSP6 variants had significant homology to a glycine rich protein, and one (TaSSP6-D.1), had homology to a probable H/ACA ribonucleoprotein complex subunit 1 (Table 4). Four of the TaSSP6 variants had a hit to the PANTHER classification system⁴ domain PTHR37389, which

⁴<http://www.pantherdb.org/>



is an uncharacterized protein domain in wheat. However, in *Z. mays* and *Nicotiana tabacum* this domain ID is described as a glycine-rich protein domain. The only gene ontology (GO) terms associated with these genes were the biological process cell wall organization and the cellular components extracellular region and cell wall. *TaSSP7* consists of three

homoeologues on the group 3 chromosomes, each with one splice variant. All three *TaSSP7* genes encode small, secreted proteins. *TaSSP7-A.1* had significant homology to a papilin-like isoform, while *TaSSP7-B.2* had no BLASTx description, and *TaSSP7-D.1* had homology to a Kunitz/Bovine pancreatic trypsin inhibitor domain protein (Table 4). None of the *TaSSP7* homoeologues

had any domains, based on the InterProScan. The GO terms associated with the A and D-genome homoeologues were the biological processes chitin metabolic process, proteolysis, and negative regulation of peptidase activity, the molecular functions serine-type endopeptidase inhibitor activity, chitin binding, and peptidase activity, and the cellular component associated with these genes was the extracellular region. No GO terms were associated with the B-genome homoeologue. Although multiple GO terms were associated with TaSSP7, no protein domains were found within these sequences from the InterProScan search.

Both TaSSP6 and TaSSP7 were cloned from cv. Stigg. TaSSP6-2B from cv. Stigg has 97% identity to the Chinese Spring sequence, with 5 single nucleotide polymorphisms (SNPs) within the gene sequence between TaSSP6-2B from cv. Stigg and TaSSP6-2B from cv. Chinese Spring. TaSSP6-2B has 94.6% identity to TaSSP6-2D and 93.7% identity to SSP6-2A. TaSSP7-3A was cloned from cv. Stigg and has 97.5% identity to the cv. Chinese Spring reference sequence, with 7 SNPs in TaSSP7-3A between the two cultivars. TaSSP7-3A has 94.6% identity to cv. Chinese spring TaSSP7-3D, and 84.3% identity to TaSSP7-3B. qRT-PCR primers for each gene were designed to amplify all homoeologues of both TaSSP genes. These were used to assess the expression of TaSSP6 and TaSSP7 in response to another isolate of *Z. tritici* (Cork Cordiale 4) in wheat seedlings of cvs. Stigg and Gallant at 4, 8, and 12 dpi. qRT-PCR of TaSSP6 revealed an increase in expression in the *Z. tritici*-treated samples at 8 dpi. While this difference was significant in a *t*-test ($P < 0.05$), it was not significant once corrected for multiple comparisons (Tukey's *post hoc* test). qRT-PCR of TaSSP7 showed an increase in transcript abundance in *Z. tritici*-treated plants at 12 dpi, but this difference was not significant. Expression of both TaSSP6 and TaSSP7 was much greater in cv. Stigg than in cv. Gallant in both treated and control conditions (Supplementary Figure S3).

TaSSPs Have Functional Secretion Signals

To test the secretion of wheat TaSSPs, a complementation assay system in which the survival of the host depends on the secretion of the protein of interest was chosen. The sucrose transport protein (*SUC2*) gene of *S. cerevisiae* strain SEY6210 was completely knocked out to generate a *suc2* mutant yeast strain. Because the *suc2* gene of the yeast was replaced by tryptophan synthesis (*Trp1*) gene by homologous recombination, the *suc2* mutant yeast can grow on Trp dropout (-Trp) yeast media plate with glucose as a carbon supply, but it cannot grow on media containing sucrose as the sole energy source without the presence of a secreted SUC2 protein. A series of expression vectors were developed for the secretion assay in yeast (Supplementary Figure S2): the pGADT7 vector was used as a backbone expression vector containing a yeast alcohol dehydrogenase promoter (pADH1) to drive gene expression. The pGAD-SUC2^{Full length} was used as a positive control, and the pGAD-ΔSP:SUC2^{22–511} (without signal peptide) was used as a negative control. The Gateway-compatible pGAD-GW-SUC2^{22–511} vector was used for TaSSPs protein yeast

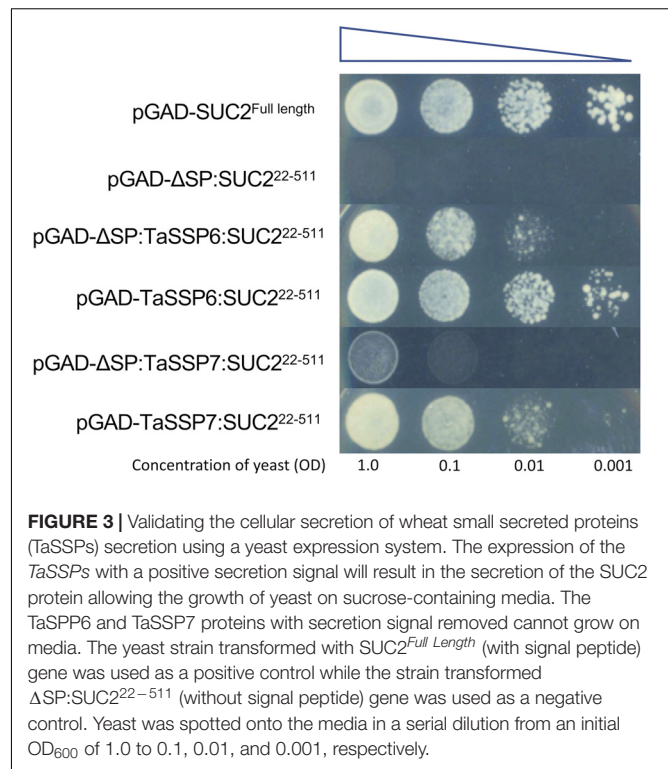


FIGURE 3 | Validating the cellular secretion of wheat small secreted proteins (TaSSPs) secretion using a yeast expression system. The expression of the TaSSPs with a positive secretion signal will result in the secretion of the SUC2 protein allowing the growth of yeast on sucrose-containing media. The TaSSP6 and TaSSP7 proteins with secretion signal removed cannot grow on media. The yeast strain transformed with SUC2^{Full length} (with signal peptide) gene was used as a positive control while the strain transformed ΔSP:SUC2^{22–511} (without signal peptide) gene was used as a negative control. Yeast was spotted onto the media in a serial dilution from an initial OD₆₀₀ of 1.0 to 0.1, 0.01, and 0.001, respectively.

secretion assay, wherein a linker (HA tag-Kex2 cleavage site) was added between Gateway Reading Frame Cassette and truncated SUC2 gene. The Kex2 cleavage site improved yeast secretion productivity and ensured fusion proteins did not affect the SUC2 activity. The TaSSPs genes were fused to the N-terminus of SUC2^{22–511} gene, and the recombinant genes were then expressed in the *suc2* mutant yeast strains (as were the positive and negative expression vectors pGAD-SUC2^{Full length} and pGAD-ΔSP:SUC2^{22–511}). The result showed that all *suc2* mutant yeast cells containing pGAD-TaSSPs:SUC2^{22–511} (full length TaSSPs, including their signal peptide) vector grew on Trp and Leu dropout (-TL) plates, using sucrose as the sole carbon source. When the signal peptide of TaSSP6 and TaSSP7 were deleted, the associated yeast cells either failed to grow (pGAD-ΔSP:TaSSP7:SUC2^{22–511}) or growth was diminished (pGAD-ΔSP:TaSSP6:SUC2^{22–511}). In conclusion, the TaSSPs have a functional secretion signal peptide that can complete the function of the signal peptide of SUC2 protein in yeast. Thus we concluded that these TaSSP proteins are secreted (Figure 3).

Silencing TaSSPs Enhances Wheat Susceptibility to *Z. tritici*

Virus induced gene silencing (VIGS) was used to study the function of TaSSP6 and TaSSP7, to determine if reducing transcript levels of TaSSP6 and TaSSP7 altered the phenotypic response to *Z. tritici* isolate Cork Cordiale 4 in the STB-resistant cv. Stigg. Two different VIGS fragments were designed to silence all homoeologues of each TaSSP gene.

TABLE 2 | The number of STB-responsive SSPs from a microarray of wheat cultivars Gallant (susceptible) and Stigg (resistant) at 4, 8, and 12 days post inoculation with *Zymoseptoria tritici*^a.

Cultivar	Regulation	Timepoint (DPI)	Number of proteins			GPI-Anchor ^e	Small and secreted ^f (%)
			DEGs ^b	Small ^c	+Signal peptide ^d		
Gallant	Up	4	33	6	15	0	4 (12.1)
		8	1063	312	121	2	32 (3)
		12	1321	366	250	13	63 (4.7)
	Down	4	51	11	13	2	3 (5.8)
		8	204	64	21	2	9 (4.4)
		12	600	164	62	3	14 (2.3)
Stigg	Up	4	37	12	5	0	4 (10.8)
		8	273	66	66	1	15 (5.5)
		12	1174	342	127	2	38 (3.2)
	Down	4	4	0	0	0	0 (0)
		8	122	38	17	2	7 (5.7)
		12	281	75	38	0	9 (3.2)

^aMicroarray probes were converted to genes using BLAST, and signal peptides were predicted with SignalP V5.0 standalone. ^bDEGs, Differentially expressed genes that corresponded to proteins. ^cSmall (≤ 250 amino acids) proteins responsive to *Z. tritici*. ^dThe number of the differentially expressed genes that code for a protein with a signal peptide (predicted to be secreted). ^eSmall, differentially expressed proteins, with a signal peptide and a predicted glycosylphosphatidylinositol anchor. ^fThe number and percentage of differentially expressed proteins that are predicted to be small, secreted proteins.

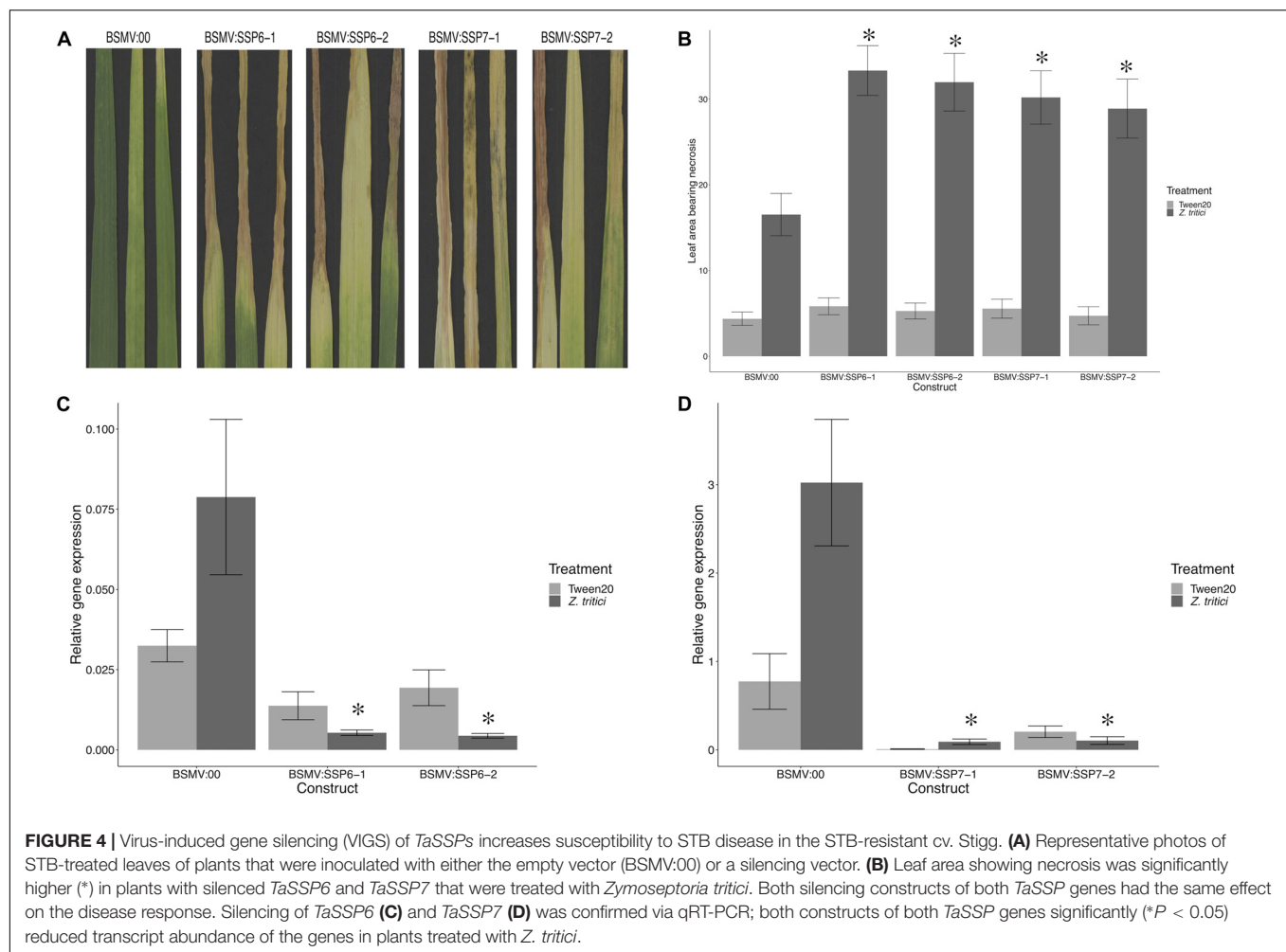
TABLE 3 | Characteristics of selected *Z. tritici* – responsive small, secreted proteins.

Affymetrix probe/ Gene ID ^a	Days post inoculation ^b	Fold change in cv. Stigg	% Identity/ E-value ^c	IWGSC ID	Length (AA)	P (SP) ^d	Cleavage site ^e	
							Position	Sequence
Ta.23397.1.S1_x_at (<i>TaSSP6</i>)	4	4.11	92.9/3.6E ⁻⁹²	TraesCS2A02G513900.1	164	0.99	27–28	TQA-KK
			94/3.1E ⁻⁸⁰	TraesCS2B02G542000.1	120	0.99	27–28	TQA-KK
			94/3.1E ⁻⁸⁰	TraesCS2B02G542000.2	118	0.99	27–28	TQA-KK
			94/3.1E ⁻⁸⁰	TraesCS2B02G542000.3	152	0.99	27–28	TQA-KK
			98.3/3.1E ⁻¹¹⁷	TraesCS2D02G515500.1	152	0.99	27–28	TQA-KK
			98.3/3.1E ⁻¹¹⁷	TraesCS2D02G515500.2	126	0.99	27–28	TQA-KK
Ta.28289.2.S1_x_at (<i>TaSSP7</i>)	12	44.7	92.3/2.9E ⁻¹¹	TraesCS3A02G095900.1	141	0.91	27–28	MMA-VT
			89.4/3.2E ⁻¹⁷	TraesCS3B02G111700.1	115	0.99	28–29	ADA-SA
			95.1/7.2E ⁻⁹	TraesCS3D02G096300.1	138	0.98	28–29	TDA-SA

^aID of the Affymetrix probe from the wheat 61K Affymetrix microarray. These probes were differentially expressed by *Z. tritici* treatment and encoded small, secreted proteins. ^bDays post inoculation (with *Z. tritici*). ^c% identity and E-value of the Affymetrix probe sequence to the IWGSC V1.1 reference gene annotation. ^dLikelihood probability of a signal peptide (based on SignalP v5.0). ^eThe cleavage site is the site of cleavage of the signal peptide from the nascent protein.

TABLE 4 | BLASTx hits and InterProScan domain IDs for the *TaSSP* genes.

Gene	BLASTx description	E-value of BLASTx hit	InterPro ID
<i>TaSSP6-A.1</i>	Glycine rich protein	8.20E ⁻¹⁷	–
<i>TaSSP6-B.1</i>	Glycine-rich cell wall structural protein precursor	3.60E ⁻¹⁷	PTHR37389
<i>TaSSP6-B.2</i>	Glycine-rich cell wall structural protein precursor	3.43E ⁻¹⁷	PTHR37389
<i>TaSSP6-B.3</i>	Glycine rich protein	3.16E ⁻¹⁷	–
<i>TaSSP6-D.1</i>	Probable H/ACA ribonucleoprotein complex subunit 1	2.43E ⁻⁷²	PTHR37389
<i>TaSSP6-D.2</i>	Glycine-rich cell wall structural protein precursor	2.19E ⁻²²	PTHR37389
<i>TaSSP7-A.1</i>	Papilin-like isoform X2	6.09E ⁻⁸⁷	–
<i>TaSSP7-B.1</i>	Unnamed protein product	5.55E ⁻⁷⁵	–
<i>TaSSP7-D.1</i>	Kunitz/Bovine pancreatic trypsin inhibitor domain protein	7.10E ⁻⁷⁹	–

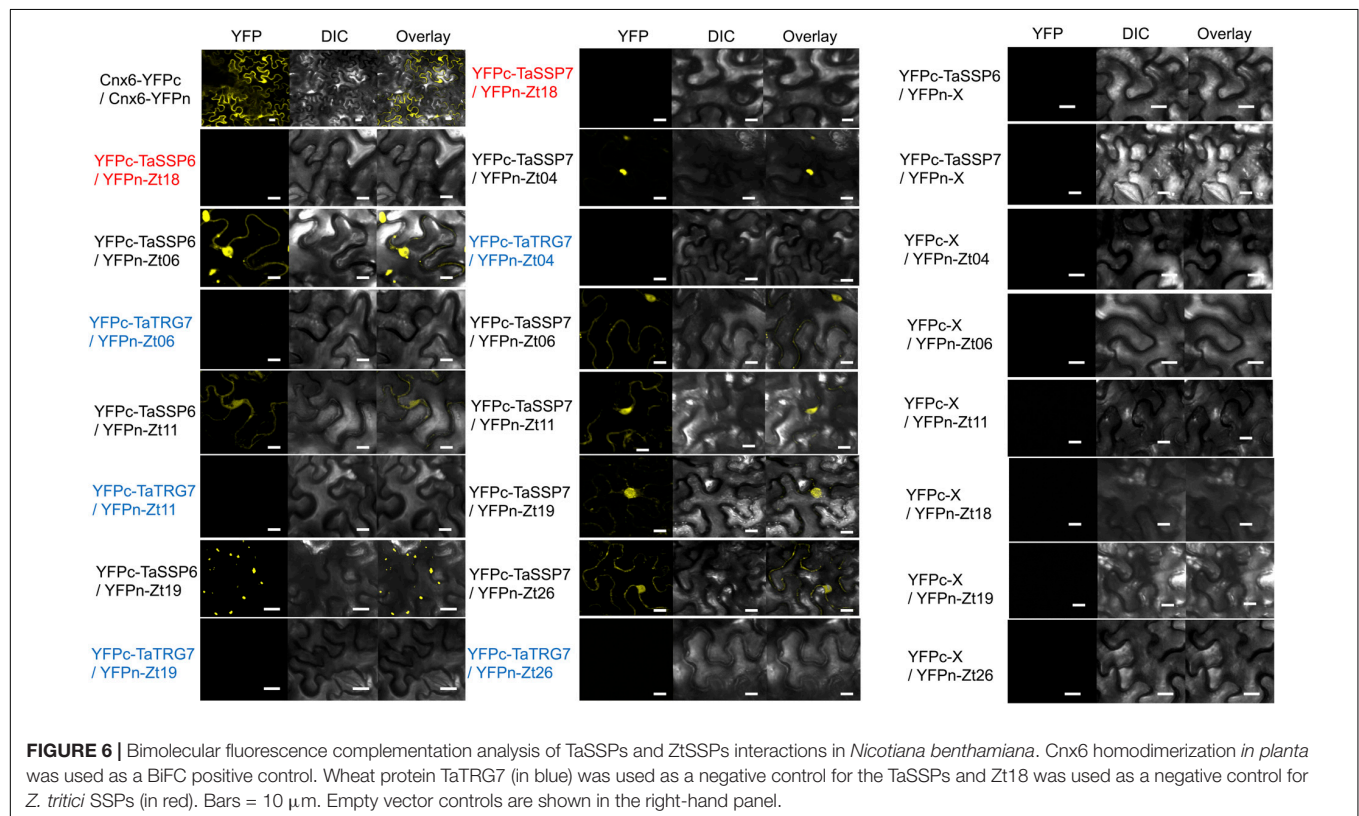
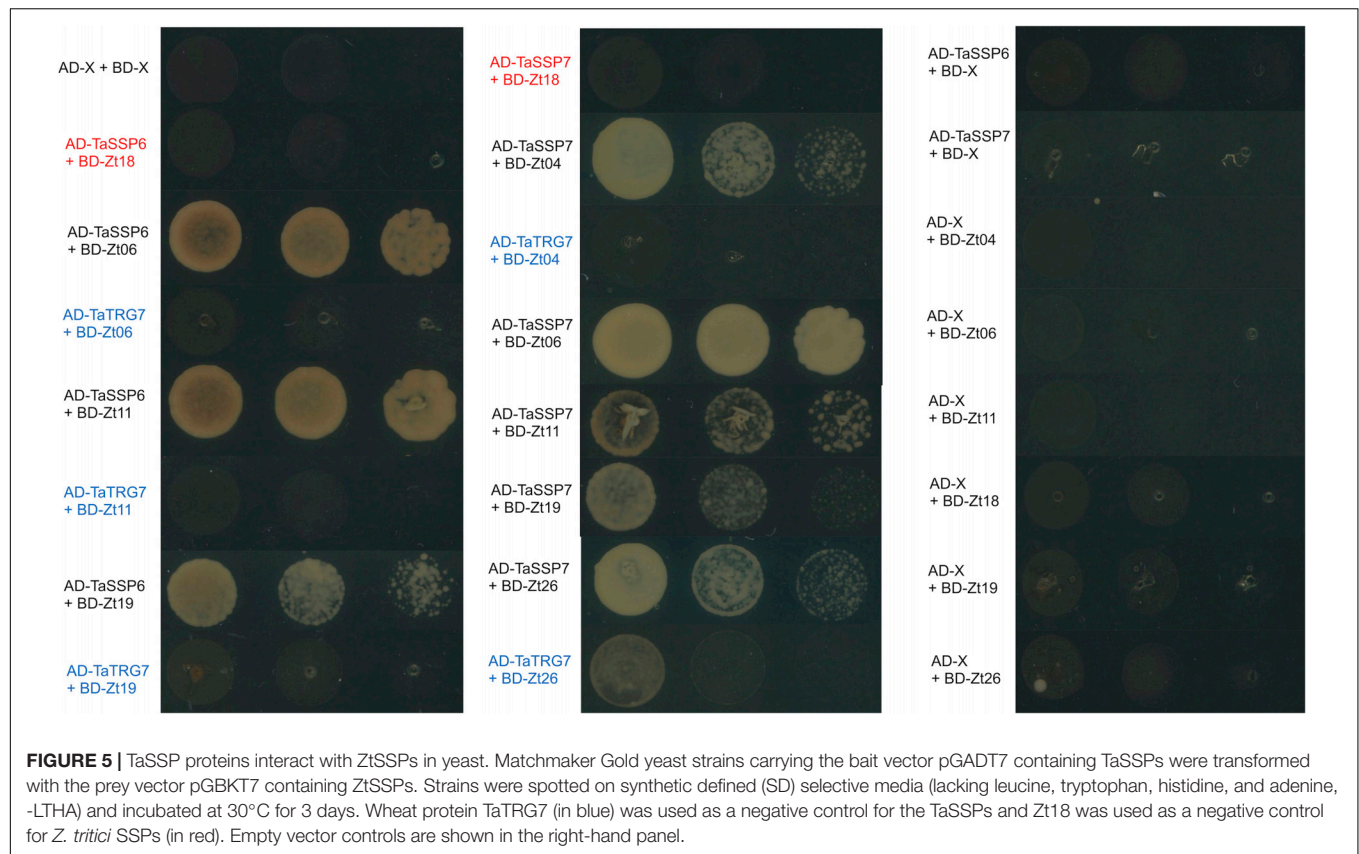


The phenotype of silenced plants was assessed by scoring the percentage of leaf area bearing necrosis at 21 dpi. Pycnidia coverage on the leaves after 10 days in 100% humidity confirmed that the STB disease developed as expected on the *Z. tritici*-treated leaves, and that no STB disease developed on the control leaves (**Supplementary Figure S4**). VIGS silencing of *TaSSP6* caused a significant ($P < 0.01$) increase in necrosis by 2-fold (construct BSMV:TaSSP6-V1) and 1.9-fold (construct BSMV:TaSSP6-V2). Silencing of *TaSSP7* caused a significant (P -value < 0.05) increase in necrosis by 1.8-fold (construct BSMV:TaSSP7-V1) and 1.7-fold (construct BSMV:TaSSP7-V2). There were no significant differences in disease levels between the two constructs for each gene (**Figures 4A,B**). The efficiency of *TaSSP* silencing were confirmed by qRT-PCR. *TaSSP6* and *TaSSP7* expression was induced by *Z. tritici* in the BSMV:00 plants at the timepoint analyzed, but this difference was not significant. In plants treated with *Z. tritici*, VIGS silencing of *TaSSP6* caused a significant ($P < 0.01$) decrease in transcript abundance of *TaSSP6* by 18-fold (construct BSMV:TaSSP6-V1) and 22-fold (construct BSMV:TaSSP6-V2) (**Figure 4C**). Silencing of *TaSSP7* caused a significant (P -value < 0.05) decrease in transcript

abundance of *TaSSP7* by 24-fold (construct BSMV:TaSSP7-V1) and 23-fold (construct BSMV:TaSSP7-V2) (**Figure 4D**). VIGS silencing of both *TaSSP* genes did not significantly reduce gene expression in the control plants (treated with 0.02% Tween20).

TaSSPs Interact With Fungal Small, Secreted Proteins

We hypothesized that one or more of *TaSSP6* and *TaSSP7* may interact with *ZtSSPs*. We used yeast two-hybrid (Y2H) analysis to identify *ZtSSPs* that can physically interact with *TaSSP* proteins. Using *TaSSP6* or *TaSSP7* as bait, we screened the interaction with 27 *ZtSSPs* (**Supplementary Table S2**) using a galactose-responsive transcription factor GAL4 (GAL4)-based yeast two-hybrid system. The results showed that *TaSSP6* could interact with three *ZtSSPs*, and *TaSSP7* could interact with five *ZtSSPs* in yeast. Three of the *ZtSSPs* were common to *TaSSP6* and *TaSSP7* (**Figure 5**). *Zt18* was used as a negative control for *ZtSSPs*, and the wheat STB-responsive non-secreted protein *TaTRG7* protein was used as a negative control for the *TaSSP* proteins, as it was previously demonstrated not to interact with these *ZtSSPs* (Brennan et al., 2020, in press).



Based on Y2H results, we deduced that TaSSP6 interacted with three separate ZtSSPs: Zt06, Zt11 and Zt19. TaSSP7 interacted with five ZtSSPs: Zt04, Zt06, Zt11, Zt19, and Zt26. We then investigated whether the TaSSPs and ZtSSPs could interact *in planta* using BiFC assays. The N-terminal part of YFP was fused to the N-terminal of TaSSPs (without the signal peptide) to create YFPn-TaSSP6 and YFPn-TaSSP7. The C-terminal part of YFP was fused to the N-terminal of the ZtSSPs (without the signal peptide) to create YFPc-Zt04, YFPc-Zt06, YFPc-Zt11, YFPc-Zt19, and YFPc-Zt26 fusion. YFPn-TaTRG7 and YFPc-Zt18 were used as negative controls. Using *Agrobacterium*-mediated transient co-expression in *N. benthamiana*, interactions between these fusion proteins were assayed for YFP fluorescence using a Confocal Laser Scanning Microscope at 2 days post-infiltration. We observed a strong YFP signal in the cytoplasm of leaf cells co-infiltrated with *A. tumefaciens*. Based on fluorescent signals we deduced that TaSSP6 interacted with ZtSSPs Zt06, Zt11, and Zt19, and TaSSP7 interacted with Zt04, Zt06, Zt11, Zt19, and Zt26 (**Figure 6**). Thus, BiFC confirmed that the interactions occurring in yeast also occurred *in planta*. We used a BLASTx search to the NCBI non-redundant protein database and InterProScan analysis to identify if any of the ZtSSP genes had any known domains or predicted function. None of Zt04, Zt06, Zt11, Zt19, or Zt26 had any known domains or predicted function from the BLASTx search, and none had any domains based on the InterProScan analysis.

Fungal SSPs That Interact With TaSSPs Induce Cell Death in the Non-host *N. benthamiana*

We tested the ability of ZtSSPs that interacted with TaSSPs to induce cell death in tobacco leaves. A high-level and long-lasting protein expression vector pEAQ-HT (Sainsbury et al., 2009) was used to express the six ZtSSPs via an *Agrobacterium*-mediated transient expression assays in *N. benthamiana* leaves. This system had been successfully used for ZtSSP expression in *N. benthamiana* (Kettles et al., 2017). The results showed that three of the ZtSSPs (Zt06, Zt11, and Zt19, all of which interact with both TaSSP6 and TaSSP7) induced cell death in *N. benthamiana* leaves (**Figure 7**). The other two ZtSSPs, Zt04 and Zt26, which interact with TaSSP7, did not induce cell death. The GFP alone was also transiently expressed as a negative control and no cell death was detected in GFP-expressing control leaves. These data thus showed that a subset of ZtSSPs that interact with TaSSP proteins can induce cell death in the non-host plant *N. benthamiana*. To test whether the interaction between TaSSPs and ZtSSPs affected cell death, the interacted TaSSP and ZtSSP combinations were co-expressed in *N. benthamiana* leaves, and similar phenotypes could be observed (**Supplementary Figure S5**), suggesting that the TaSSP-ZtSSP interaction did not affect the cell death phenotype *in planta*, at least in *N. benthamiana*. The TaSSP proteins did not induce a cell death phenotype when infiltrated into *N. benthamiana* leaves alone (**Supplementary Figure S5**).

DISCUSSION

The apoplastic space is one of the first sites of conflict between plant and pathogen. In the case of STB disease of wheat, the fungus is purely apoplastic, and the conflict between fungal effectors and plant proteins determines the outcome of the disease progression (Doehlemann and Hemetsberger, 2013). In this study, we identified numerous wheat small secreted proteins (SSPs) expressed during the interaction with the fungal pathogen *Z. tritici* and showed that SSPs can enhance wheat resistance to STB disease.

Until recently, large-scale or automated identification or characterization of gene families in wheat was difficult due to the lack of an annotated reference genome. However, since the release of the IWGSC refseq (IWGSC, 2018), we were able to automate the discovery of predicted SSPs from the protein annotation by writing a wrapper script to survey protein length and predict the presence of a signal peptide using SignalP v5.0 (Armenteros et al., 2019). By using this pipeline, we discovered that 58% of all wheat proteins were smaller than 250 amino acids in length, and a positive-skew in the distribution of protein length revealed that smaller proteins were more abundant in the genome than longer proteins. Plant-specific proteins are known to be on-average shorter than those of animals and fungi as they generally contain fewer exons than the genomes of the other eukaryotic kingdoms (Ramirez-Sanchez et al., 2016). Due to their lower capacity to house domains, and limited folding potential, small (<200 AA) proteins are usually limited in function (Chothia et al., 2003), but are known to be important for multiple biological processes, including the stress response (Storz et al., 2014).

We used SignalP to predict the presence of signal peptide sequences at the N terminal of the wheat proteins. Signal peptides are found on secreted as well as transmembrane proteins, and also in proteins within cellular organelles (Armenteros et al., 2019). Signal peptides were predicted in 7% of the wheat proteins, and all of these proteins were predicted to be secreted via the general secretory pathway; protein translocation across the endoplasmic reticulum membrane (Vitale and Denecke, 1999). Combining the secretion predictions with protein length, and filtering out proteins with any transmembrane domains and GPI anchors, we identified 6,998 proteins that were small, and had a secretion signal (SSPs). We used Blast2Go to functionally annotate the TaSSP genes and test for any enrichment of function. Many gene ontology terms were significantly enriched in the TaSSP gene set, including many GO terms that are important for the disease response. These included pathways and cellular components that have been characterized and implicated in the plant defense response, including peptidase inhibitor activity (Benbow et al., 2019), and the apoplastic space (Doehlemann and Hemetsberger, 2013; Jashni et al., 2015).

Using the *Z. tritici* microarray data set from Brennan et al. (2020), we identified *Z. tritici*-responsive SSPs. We found a significant enrichment of SSPs in the STB-response of wheat, indicating that SSPs are over-represented in the wheat disease response to *Z. tritici*. Additionally, we observed a temporal decrease in the number of SSPs that were



FIGURE 7 | ZtSSPs induced cell death in *N. benthamiana* leaves. The candidate ZtSSPs, which interacted with TaSSPs, were expressed in leaves of *N. benthamiana* by *Agrobacterium*-mediated expression. Three of the ZtSSPs induced cell death phenotypes. GFP was expressed as negative control. Leaves photographed at 7 days post infiltration (dpi).

differentially expressed in response to *Z. tritici*; from 8% at 4 dpi to 3.7% at 12 dpi. Four dpi is well within the latent phase of the disease, and is at the tail end of a peak in expression of *Z. tritici* effectors associated with the latent phase of the disease (Mirzadi Gohari et al., 2015). It seems that expression of the TaSSP genes studied here may be related to that of ZtSSPs. We hypothesize that many of these TaSSP genes evolved in response to pathogen attack, and as a mechanism for effector-triggered immunity in the wheat-*Z. tritici* pathosystem.

Of the *Z. tritici*-responsive SSPs, we chose to focus on two, TaSSP6 and TaSSP7, because they had a high fold change in the STB-resistant cv. Stigg, and were not differentially expressed in the susceptible cv. Gallant in response to isolate IPO323. The probes that were differentially expressed from the Brennan et al. (2020, in press) microarray study were used to identify genes from the IWGSC refseq v1.1 annotation, and we found that both genes were present as 3 homoeologues, and TaSSP6 could be alternatively spliced into five isoforms (1 × A, 3 × B, 2 × D). TaSSP6 is a putative glycine-rich protein, and four out of the six TaSSP6 variants had hits to the domain PTHR37389 – an uncharacterized domain in wheat. The PTHR37389 domain has 18 subfamilies, including glycine-rich protein-like and cold and drought regulated protein-like, suggesting that TaSSP7 contains a variant or subfamily of PTHR37389 that may be associated with the stress response. The GO terms associated with TaSSP6 were cell wall organization, the extracellular region, and the cell wall. The cell wall, and genes involved in cell wall reorganizing/cell wall remodeling have previously been associated with wheat resistance to STB, with increased activity of cell wall remodeling and reinforcement in a resistant wheat cultivar in response to *Z. tritici* (Yang et al., 2015). TaSSP7-A.1 has homology to a papilin-like isoform, TaSSP7-B.1 to an unnamed protein product, and TaSSP7-D.1 to a trypsin protease inhibitor domain protein, although no domains were found any of the three homoeologues. Both papilin-like proteins and trypsin protease inhibitors are serine-type endopeptidase inhibitors (serpins), that are known to play a role in the disease response in wheat and other plant species (Bhattacharjee et al., 2017; Bao et al., 2018; Benbow et al., 2019). However, none of the TaSSP7 homoeologues were characterized as wheat serpins in a recent genome wide characterization of the wheat serpin family (Benbow et al., 2019), indicating that although TaSSP7 may have partial homology

to serpin-like proteins, it doesn't contain any of the serpin protein domains.

As the microarray study provides no information on homoeologue specificity, we cannot be sure which homoeologue (and indeed isoform) is contributing to the differential expression of the probe. The independent time course generated for gene expression studies of the TaSSP genes used a different isolate of *Z. tritici*: the aggressive Irish field isolate Cork Cordiale 4 was used rather than the reference isolate IPO323. While we saw an increase in expression of TaSSP6 in cv. Stigg at 8 dpi, this difference was not significant, partly due to the large variance in gene expression in the leaves. Additionally, we saw no significant difference in TaSSP7 expression in our time course. We attribute this, in part, to a change in *Z. tritici* isolate used for the expression study.

In addition to their expression in response to *Z. tritici*, TaSSP6 and TaSSP7 were predicted to be secreted proteins, based on the presence of a signal peptide sequence at the N-terminus of the protein. To validate this prediction, a complementation-based secretion assay was used based on that of Plett et al. (2017). This system confirmed that both TaSSP6 and TaSSP7 have functional secretion signal peptides and are secreted. TaSSP7 that lacked its signal peptide was not secreted, indicating that the signal peptide is vital for secretion of TaSSP7. However, TaSSP6 was secreted (although to a lesser extent) once the signal peptide was removed. This phenomenon, known as leaderless secretion, is known to occur in bacteria (Bendtsen et al., 2005), and has been characterized in plants, where a normally non-secreted, cytoplasmic protein was secreted into the plant apoplast in response to SA signaling (Cheng et al., 2009). While TaSSP6 is ordinarily secreted and has a functional signal peptide, its secretion (at least in yeast) was improved by, but not wholly dependent on its signal peptide. Further study is warranted here to determine if TaSSP6 is secreted *in planta* without its signal peptide.

To test the function of TaSSP6 and TaSSP7, both genes were transiently silenced using the VIGS system. The expression of both TaSSP genes was induced by *Z. tritici* in the BSMV:00 (empty vector) plants, but the difference in gene expression between *Z. tritici* and Tween20 treated plants was not significant. This is likely due to the use of a different *Z. tritici* in both the temporal gene expression study and the VIGS, where we used the isolate Cork Cordiale 4 versus IPO323 that was used in the original microarray study. In plants silenced with either TaSSP gene, the

expression of the *TaSSP* gene was significantly reduced by the silencing construct, and a significant increase in STB disease was observed, demonstrating a ~2-fold increase in susceptibility of the STB-resistant cv. Stigg. It seems both genes, when silenced, give a similar phenotype and may serve to shorten the latent phase of cv. Stigg. Normally ~35 days long, Stigg's lengthy latent phase contributes to its exceptional STB resistance in the field (Hehir et al., 2018). With a somewhat elusive pedigree, Stigg's various chromosomal introgressions from wild wheat relatives are thought to contribute to its resistance, and the key players behind this characteristic are largely unknown. We suggest that, based on their expression and function, *TaSSP6* and *TaSSP7* are involved in the latent phase of infection and stave off disease by interacting with fungal small, secreted proteins. Supporting this is the confirmation that both *TaSSP6* and *TaSSP7* interact with ZtSSPs *in vitro* and *in planta*. Both *TaSSPs* interact with three common ZtSSPs: Zt06, Zt11, and Zt19. Interestingly, it was these three ZtSSPs that could also induce cell death in tobacco leaves. Although tobacco is a non-host to *Z. tritici*, these proteins were clearly recognized and responded to, suggesting the potential for activation of down-stream signaling cascades that can promote a hypersensitive response. However, co-expression of the *TaSSPs* with the ZtSSPs into the tobacco leaves did not alter the cell death phenotype, suggesting that the cell death phenotype depends on non-host resistance (Kettles et al., 2017). These ZtSSPs did not have any known domains, based on InterProScan, so we cannot hypothesize as to their function or method of interaction with *TaSSPs*, but as infection of wheat cv. Stigg with *Z. tritici* does not elicit a hypersensitive response, we must conclude that the elicitation of a host response by fungal SSPs is interfered with by host defense genes. In this case, we propose a role for *TaSSP6* and *TaSSP7* to stop, delay, or alter the effect of the ZtSSPs on the cytology of the host.

In summary, we present two novel wheat genes that encode for small, secreted proteins, and contribute to resistance to STB disease. The wheat proteins interact with fungal small, secreted proteins and this interaction may contribute the resistance phenotype observed in cv. Stigg. We hypothesize that these *TaSSP* proteins may be important for effector-triggered immunity in wheat infected with STB disease. This study gives insight into the complex mechanisms of the host-pathogen interaction in this economically important disease, and further characterization of wheat small, secreted proteins, especially those that are responsive to disease, may reveal insights into the evolution of effector-triggered immunity in plants.

DATA AVAILABILITY STATEMENT

The microarray data from Brennan et al. (2020) is available at doi: 10.6084/m9.figshare.11882601.v1. The raw data

REFERENCES

- Ali, S., Ganai, B. A., Kamili, A. N., Bhat, A. A., Mir, Z. A., Bhat, J. A., et al. (2018). Pathogenesis-related proteins and peptides as promising tools for engineering plants with multiple stress tolerance. *Microbiol. Res.* 212, 29–37. doi: 10.1016/j.micres.2018.04.008
- Armenteros, A. J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z
- 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 supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.
- BZ, HB, and FD designed the experiments. BZ, HB, CB, and CA carried out all experiments. BZ and HB analyzed the data. SK and AF identified SSPs from *Z. tritici*. BZ and HB wrote the manuscript. EM, JB, and FD reviewed the manuscript.

AUTHOR CONTRIBUTIONS

FUNDING

The authors would like to thank the Virtual Irish Centre for Crop Improvement (VICCI) project (14/S/819), CONSUS project (16/SPP/3296) and Science Foundation Ireland (14/1A/2508 and 15/CDA/3451). The authors would also like to thank Department of Agriculture, Food and the Marine Research Stimulus Project Wheat enhance (11/S/103), and the Cereal Improvement through Variety choice and understanding Yield Limitations (606 CYVIL) project (11/S/121) funded by DAFM in conjunction with Teagasc, AFBI, the John Innes Centre, Goldcrop, Seedtech, the HGCA, and Germinal seeds. This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 674964.

ACKNOWLEDGMENTS

We would like to thank Dr. Stephen Kildea and Thomas Welch of Teagasc Crops Research, Co. Carlow, Ireland for providing isolates of *Zymoseptoria tritici*. We would also like to thank Dr. Laurent Deslandes, Laboratoire des Interactions Plantes-Microorganismes (LIPM), INRA-CNRS (Toulouse, France) for providing *Agrobacterium tumefaciens* strain GV3101 and Plant Bioscience Limited (Norwich, United Kingdom) for supplying the pEAQ vectors. We would also like to acknowledge Dr. Alexandre Perochon, Dr. Ganesh Thapa, Brian Fagan, Jianguang Jia, and Liam Kavanagh from University College Dublin for technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00469/full#supplementary-material>

- searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335
- Bailey, T. L., and Gribskov, M. (1998). Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 14, 48–54. doi: 10.1093/bioinformatics/14.1.48
- Bao, J., Pan, G., Poncz, M., Wei, J., Ran, M., Zhou, Z., et al. (2018). Serpin functions in host-pathogen interactions. *PeerJ* 6:e4557. doi: 10.7717/peerj.4557
- Ben, M., Barek, S., Cordewener, J. H., Tabib Ghaffary, S. M., van der Lee, T. A., Liu, Z., et al. (2015). FPLC and liquid-chromatography mass spectrometry identify candidate necrosis-inducing proteins from culture filtrates of the fungal wheat pathogen *Zymoseptoria tritici*. *Fungal Genet. Biol.* 79, 54–62. doi: 10.1016/j.fgb.2015.03.015
- Benbow, H. R., Jermini, L. S., and Doohan, F. M. (2019). Serpins: genome-wide characterisation and expression analysis of the serine protease inhibitor family in *Triticum aestivum*. *G3* 9, 2709–2722. doi: 10.1534/g3.119.400444
- Bendtsen, J. D., Kiemer, L., Fausboll, A., and Brunak, S. (2005). Non-classical protein secretion in bacteria. *BMC Microbiol.* 5:58. doi: 10.1186/1471-2180-5-58
- Bhattacharjee, L., Singh, D., Gautam, J. K., and Nandi, A. K. (2017). *Arabidopsis thaliana* serpins AtSRP4 and AtSRP5 negatively regulate stress-induced cell death and effector-triggered immunity induced by bacterial effector AvrRpt2. *Physiol. Plant.* 159, 329–339. doi: 10.1111/ppl.12516
- Block, A., Toruño, T. Y., Elowsky, C. G., Zhang, C., Steinbrenner, J., Beynon, J., et al. (2014). The *Pseudomonas syringae* type III effector HopD1 suppresses effector-triggered immunity, localizes to the endoplasmic reticulum, and targets the *Arabidopsis* transcription factor NTL9. *New Phytol.* 201, 1358–1370. doi: 10.1111/nph.12626
- Brennan, C. J., Zhou, B., Benbow, H. R., Ajaz, S., Karki, S. J., Hehir, J. G., et al. (2020). Taxonomically restricted wheat genes interact with small secreted fungal proteins and enhance resistance to *Septoria tritici* blotch disease. *Front. Plant Sci.* 11:433. doi: 10.3389/fpls.2020.00433
- Cheng, F. Y., Zamski, E., Guo, W. W., Pharr, D. M., and Williamson, J. D. (2009). Salicylic acid stimulates secretion of the normally symplastic enzyme mannitol dehydrogenase: a possible defense against mannitol-secreting fungal pathogens. *Planta* 230, 1093–1103. doi: 10.1007/s00425-009-1006-3
- Chiang, Y.-H., and Coaker, G. (2015). Effector triggered immunity: NLR immune perception and downstream defense responses. *Arabidopsis Book* 13:e0183. doi: 10.1016/j.coi.2019.12.007
- Chothia, C., Gough, J., Vogel, C., and Teichmann, S. A. (2003). Evolution of the protein repertoire. *Science* 300, 1701–1703.
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., Robles, M., et al. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Cools, H. J., and Fraaije, B. A. (2013). Update on mechanisms of azole resistance in *Mycosphaerella graminicola* and implications for future control. *Pest Manag. Sci.* 69, 150–155. doi: 10.1002/ps.3348
- Couto, D., and Zipfel, C. (2016). Regulation of pattern recognition receptor signalling in plants. *Nat. Rev. Immunol.* 16, 537–552. doi: 10.1038/nri.2016.77
- Desaki, Y., Miyata, K., Suzuki, M., Shibuya, N., and Kaku, H. (2018). Plant immunity and symbiosis signaling mediated by LysM receptors. *Innate Immun.* 24, 92–100. doi: 10.1177/1753425917738885
- do Amaral, A. M., Antoniwi, J., Rudd, J. J., and Hammond-Kosack, K. E. (2012). Defining the predicted protein secretome of the fungal wheat leaf pathogen *Mycosphaerella graminicola*. *PLoS One* 7:e49904. doi: 10.1371/journal.pone.0049904
- Doehlemann, G., and Hemetsberger, C. (2013). Apoplastic immunity and its suppression by filamentous plant pathogens. *New Phytol.* 198, 1001–1016. doi: 10.1111/nph.12277
- Dooley, H. (2015). *Fungicide-Resistance Management Tactics: Impacts on Zymoseptoria tritici Populations*. Berkshire: University of Reading.
- Fankhauser, N., and Maser, P. (2005). Identification of GPI anchor attachment signals by a Kohonen self-organizing map. *Bioinformatics* 21, 1846–1852. doi: 10.1093/bioinformatics/bti299
- Fones, H., and Gurr, S. (2015). The impact of *Septoria tritici* blotch disease on wheat: an EU perspective. *Fungal Genet. Biol.* 79, 3–7. doi: 10.1016/j.fgb.2015.04.004
- Gehl, C., Waadt, R., Kudla, J., Mendel, R. R., and Hansch, R. (2009). New GATEWAY vectors for high throughput analyses of protein-protein interactions by bimolecular fluorescence complementation. *Mol. Plant* 2, 1051–1058. doi: 10.1093/mp/ssp040
- Gunupuru, L. R., Ali, S. S., and Doohan, F. M. (2015). Virus-induced gene silencing (VIGS) in barley seedling leaves. *Bio Protoc.* 5, 1506–1508. doi: 10.1104/pp.105.062810
- Hehir, J. G., Connolly, C., O'driscoll, A., Lynch, J. P., Spink, J., Brown, J. K. M., et al. (2018). Temporal and spatial field evaluations highlight the importance of the presymptomatic phase in supporting strong partial resistance in *Triticum aestivum* against *Zymoseptoria tritici*. *Plant Pathol.* 67, 573–583.
- Heick, T. M., Justesen, A. F., and Jorgensen, L. N. (2017). Resistance of wheat pathogen *Zymoseptoria tritici* to DMI and QoI fungicides in the Nordic-Baltic region - a status. *Eur. J. Plant Pathol.* 149, 669–682.
- Horecka, J., and Davis, R. W. (2014). The 50:50 method for PCR-based seamless genome editing in yeast. *Yeast* 31, 103–112. doi: 10.1002/yea.2992
- Huffaker, A., Dafoe, N. J., and Schmelz, E. A. (2011). ZmPep1, an ortholog of *Arabidopsis* elicitor peptide 1, regulates maize innate immunity and enhances disease resistance. *Plant Physiol.* 155, 1325–1338. doi: 10.1104/pp.110.166710
- Huffaker, A., Pearce, G., and Ryan, C. A. (2006). An endogenous peptide signal in *Arabidopsis* activates components of the innate immune response. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10098–10103. doi: 10.1073/pnas.0603727103
- Ilyas, M., Hörger, A. C., Bozkurt, T. O., van den Burg, H. A., Kaschani, F., Kaiser, M., et al. (2015). Functional divergence of two secreted immune proteases of tomato. *Curr. Biol.* 25, 2300–2306. doi: 10.1016/j.cub.2015.07.030
- IWGSC (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191
- Jashni, M. K., Mehrabi, R., Collemare, J., Mesarich, C. H., and de Wit, P. J. G. M. (2015). The battle in the apoplast: further insights into the roles of proteases and their inhibitors in plant-pathogen interactions. *Front. Plant Sci.* 6:584. doi: 10.3389/fpls.2015.00584
- Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329.
- Kema, G. H., Yu, D., Rijkenberg, F. H., Shaw, M. W., and Baayen, R. P. (1996). Histology of the pathogenesis of *Mycosphaerella graminicola* in wheat. *Phytopathology* 86, 777–786.
- Keon, J., Antoniwi, J., Carzaniga, R., Deller, S., Ward, J. L., Baker, J. M., et al. (2007). Transcriptional adaptation of *Mycosphaerella graminicola* to programmed cell death (PCD) of its susceptible wheat host. *Mol. Plant Microbe Interact.* 20, 178–193. doi: 10.1094/MPMI-20-2-0178
- Kettles, G. J., Bayon, C., Canning, G., Rudd, J. J., and Kanyuka, K. (2017). Apoplastic recognition of multiple candidate effectors from the wheat pathogen *Zymoseptoria tritici* in the nonhost plant *Nicotiana benthamiana*. *New Phytol.* 213, 338–350. doi: 10.1111/nph.14215
- Kim, J. Y., Park, S. C., Hwang, I., Cheong, H., Nah, J. W., Hahm, K. S., et al. (2009). Protease inhibitors from plants with antimicrobial activity. *Int. J. Mol. Sci.* 10, 2860–2872. doi: 10.3390/ijms10062860
- Kombrink, E., and Schmelzer, E. (2001). The hypersensitive response and its role in local and systemic disease resistance. *Eur. J. Plant Pathol.* 107, 69–78.
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Lanver, D., Tollot, M., Schweizer, G., Lo Presti, L., Reissmann, S., Ma, L. S., et al. (2017). *Ustilago maydis* effectors and their impact on virulence. *Nat. Rev. Microbiol.* 15, 409–421. doi: 10.1038/nrmicro.2017.33
- Le Roux, C., Huet, G., Jauneau, A., Camborde, L., Trémoussaygue, D., Kraut, A., et al. (2015). A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell* 161, 1074–1088. doi: 10.1016/j.cell.2015.04.025
- Lee, W. S., Rudd, J. J., Hammond-Kosack, K. E., and Kanyuka, K. (2014). *Mycosphaerella graminicola* LysM effector-mediated stealth pathogenesis subverts recognition through both CERK1 and CEBiP homologues in wheat. *Mol. Plant Microbe Interact.* 27, 236–243. doi: 10.1094/MPMI-07-13-0201-R
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509

- 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, W. Z., Jaroszowski, L., and Godzik, A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* 17, 282–283. doi: 10.1093/bioinformatics/17.3.282
- Liu, Z., Wu, Y., Yang, F., Zhang, Y., Chen, S., Xie, Q., et al. (2013). BIK1 interacts with PEPRs to mediate ethylene-induced immunity. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6205–6210. doi: 10.1073/pnas.1215543110
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Macho, A. P., Schwessinger, B., Ntoukakis, V., Brutus, A., Segonzac, C., Roy, S., et al. (2014). A bacterial tyrosine phosphatase inhibits plant pattern recognition receptor activation. *Science* 343, 1509–1512. doi: 10.1126/science.1248849
- Marshall, R., Kombrink, A., Motteram, J., Loza-Reyes, E., Lucas, J., Hammond-Kosack, K. E., et al. (2011). Analysis of two in planta expressed LysM effector homologs from the fungus *mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. *Plant Physiol.* 156, 756–769. doi: 10.1104/pp.111.176347
- McDonald, B. A., and Stukenbrock, E. H. (2016). Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20160026. doi: 10.1098/rstb.2016.0026
- Mirzadi Gohari, A. (2015). *Identification and Functional Characterization of Putative (a) Virulence Factors in the Fungal Wheat Pathogen Zymoseptoria tritici*. Wageningen: Wageningen University.
- Mirzadi Gohari, A., Ware, S. B., Wittenberg, A. H., Mehrabi, R., Ben, M., Barek, S., et al. (2015). Effector discovery in the fungal wheat pathogen *Zymoseptoria tritici*. *Mol. Plant Pathol.* 16, 931–945. doi: 10.1111/mpp.12251
- Miya, A., Albert, P., Shinya, T., Desaki, Y., Ichimura, K., Shirasu, K., et al. (2007). CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19613–19618. doi: 10.1073/pnas.0705147104
- O'Driscoll, A., Kildea, S., Doohan, F., Spink, J., and Mullins, E. (2014). The wheat–Septoria conflict: a new front opening up? *Trends Plant Sci.* 19, 602–610. doi: 10.1016/j.tplants.2014.04.011
- Orton, E. S., and Brown, J. K. (2016). Reduction of growth and reproduction of the biotrophic fungus *Blumeria graminis* in the presence of a necrotrophic pathogen. *Front. Plant Sci.* 7:742. doi: 10.3389/fpls.2016.00742
- Palma-Guerrero, J., Torriani, S. F., Zala, M., Carter, D., Courbot, M., Rudd, J. J., et al. (2016). Comparative transcriptomic analyses of *Zymoseptoria tritici* strains show complex lifestyle transitions and intraspecific variability in transcription profiles. *Mol. Plant Pathol.* 17, 845–859. doi: 10.1111/mpp.12333
- Petre, B., and Kamoun, S. (2014). How do filamentous pathogens deliver effector proteins into plant cells? *PLoS Biol.* 12:e1001801. doi: 10.1371/journal.pbio.1001801
- Plett, J. M., Yin, H., Mewalal, R., Hu, R., Li, T., Ranjan, P., et al. (2017). *Populus trichocarpa* encodes small, effector-like secreted proteins that are highly induced during mutualistic symbiosis. *Sci. Rep.* 7:382. doi: 10.1038/s41598-017-00400-8
- Plissonneau, C., Hartmann, F. E., and Croll, D. (2018). Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol.* 16:5. doi: 10.1186/s12915-017-0457-4
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Ramirez-Sanchez, O., Perez-Rodriguez, P., Delaye, L., and Tiessen, A. (2016). Plant proteins are smaller because they are encoded by fewer exons than animal proteins. *Genomics Proteomics Bioinformatics* 14, 357–370. doi: 10.1016/j.gpb.2016.06.003
- Rudd, J. J., Kanyuka, K., Hassani-Pak, K., Derbyshire, M., Andongabo, A., Devonshire, J., et al. (2015). Transcriptome and metabolite profiling of the infection cycle of *Zymoseptoria tritici* on wheat reveals a biphasic interaction with plant immunity involving differential pathogen chromosomal contributions and a variation on the hemibiotrophic lifestyle definition. *Plant Physiol.* 167, 1158–1185. doi: 10.1104/pp.114.255927
- Rustgi, S., Boex-Fontvieille, E., Reinbothe, C., von Wettstein, D., and Reinbothe, S. (2018). The complex world of plant protease inhibitors: insights into a Kunitz-type cysteine protease inhibitor of *Arabidopsis thaliana*. *Commun. Integr. Biol.* 11:e1368599. doi: 10.1080/19420889.2017.1368599
- Sainsbury, F., Thuenemann, E. C., and Lomonosoff, G. P. (2009). pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* 7, 682–693. doi: 10.1111/j.1467-7652.2009.00434.x
- Schellenberger, R., Touchard, M., Clément, C., Baillieul, F., Cordelier, S., Crouzet, J., et al. (2019). Apoplastic invasion patterns triggering plant immunity: plasma membrane sensing at the frontline. *Mol. Plant Pathol.* 20, 1602–1616. doi: 10.1111/mpp.12857
- Scofield, S. R., Huang, L., Brandt, A. S., and Gill, B. S. (2005). Development of a virus-induced gene-silencing system for hexaploid wheat and its use in functional analysis of the Lr21-mediated leaf rust resistance pathway. *Plant Physiol.* 138, 2165–2173. doi: 10.1104/pp.105.061861
- Segonzac, C., and Monaghan, J. (2019). Modulation of plant innate immune signaling by small peptides. *Curr. Opin. Plant Biol.* 51, 22–28. doi: 10.1016/j.pbi.2019.03.007
- Shetty, N. P., Jensen, J. D., Knudsen, A., Finnie, C., Geshi, N., Blennow, A., et al. (2009). Effects of β -1, 3-glucan from *Septoria tritici* on structural defence responses in wheat. *J. Exp. Bot.* 60, 4287–4300. doi: 10.1093/jxb/erp269
- Storz, G., Wolf, Y. I., and Ramamurthy, K. S. (2014). Small proteins can no longer be ignored. *Annu. Rev. Biochem.* 83, 753–777.
- Tian, M., Huitema, E., Da Cunha, L., Torto-Alalibo, T., and Kamoun, S. (2004). A Kazal-like extracellular serine protease inhibitor from *Phytophthora infestans* targets the tomato pathogenesis-related protease P69B. *J. Biol. Chem.* 279, 26370–26377. doi: 10.1074/jbc.M400941200
- Vitale, A., and Denecke, J. (1999). The endoplasmic reticulum - Gateway of the secretory pathway. *Plant Cell* 11, 615–628. doi: 10.1105/tpc.11.4.615
- Wang, S., Boevink, P. C., Welsh, L., Zhang, R., Whisson, S. C., Birch, P. R. J., et al. (2017). Delivery of cytoplasmic and apoplastic effectors from *Phytophthora infestans* haustoria by distinct secretion pathways. *New Phytol.* 216, 205–215. doi: 10.1111/nph.14696
- Wang, Y., and Wang, Y. (2018). Trick or treat: microbial pathogens evolved apoplastic effectors modulating plant susceptibility to infection. *Mol. Plant Microbe Interact.* 31, 6–12. doi: 10.1094/MPMI-07-17-0177-FI
- Wirthmueller, L., Maqbool, A., and Banfield, M. J. (2013). On the front line: structural insights into plant-pathogen interactions. *Nat. Rev. Microbiol.* 11, 761–776. doi: 10.1038/nrmicro3118
- Yamaguchi, Y., Huffaker, A., Bryan, A. C., Tax, F. E., and Ryan, C. A. (2010). PEPR2 is a second receptor for the Pep1 and Pep2 Peptides and Contributes to Defense Responses in *Arabidopsis*. *Plant Cell* 22, 508–522. doi: 10.1105/tpc.109.068874
- Yang, F., Li, W., Derbyshire, M., Larsen, M. R., Rudd, J. J., Palmisano, G., et al. (2015). Unraveling incompatibility between wheat and the fungal pathogen *Zymoseptoria tritici* through apoplastic proteomics. *BMC Genomics* 16:362. doi: 10.1186/s12864-015-1549-6
- Yang, F., Li, W., and Jørgensen, H. J. (2013). Transcriptional reprogramming of wheat and the hemibiotrophic pathogen *Septoria tritici* during two phases of the compatible interaction. *PLoS One* 8:e81606. doi: 10.1371/journal.pone.0081606
- Zadocks, J. C., Chang, T. T., and Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Res.* 14, 415–421.
- Ziemann, S., van der Linde, K., Lahrmann, U., Acar, B., Kaschani, F., Colby, T., et al. (2018). An apoplastic peptide activates salicylic acid signalling in maize. *Nat. Plants* 4, 172–180. doi: 10.1038/s41477-018-0116-y

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

Copyright © 2020 Zhou, Benbow, Brennan, Arunachalam, Karki, Mullins, Feechan, Burke and Doohan. 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.



Impact of DNA Demethylases on the DNA Methylation and Transcription of *Arabidopsis* NLR Genes

Weiwen Kong^{1,2*}, Xue Xia¹, Qianqian Wang³, Li-Wei Liu⁴, Shengwei Zhang¹, Li Ding¹, Aixin Liu⁵ and Honggui La^{3*}

¹ School of Horticulture and Plant Protection, Yangzhou University, Yangzhou, China, ² Joint International Research Laboratory of Agriculture and Agri-Product Safety of the Ministry of Education, Yangzhou University, Yangzhou, China, ³ College of Life Sciences, Nanjing Agricultural University, Nanjing, China, ⁴ State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing, China, ⁵ Department of Plant Protection, Shandong Agricultural University, Tai'an, China

OPEN ACCESS

Edited by:

Frank L. W. Takken,
University of Amsterdam, Netherlands

Reviewed by:

Thierry Halter,
INSERM U1024 Institut de biologie
de l'Ecole Normale Supérieure,
France

Changwei Shao,
Yellow Sea Fisheries Research
Institute (CAFS), China

*Correspondence:

Weiwen Kong
wwkong@yzu.edu.cn
Honggui La
hongguila@njau.edu.cn

Specialty section:

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

Received: 14 January 2020

Accepted: 14 April 2020

Published: 26 May 2020

Citation:

Kong W, Xia X, Wang Q, Liu L-W,
Zhang S, Ding L, Liu A and La H
(2020) Impact of DNA Demethylases
on the DNA Methylation
and Transcription of *Arabidopsis* NLR
Genes. *Front. Genet.* 11:460.
doi: 10.3389/fgene.2020.00460

Active DNA demethylation is an important epigenetic process that plays a key role in maintaining normal gene expression. In plants, active DNA demethylation is mediated by DNA demethylases, including ROS1, DME, DML2, and DML3. In this study, the available bisulfite sequencing and mRNA sequencing data from *ros1* and *ddd* mutants were analyzed to reveal how the active DNA demethylation process shapes the DNA methylation patterns of *Arabidopsis* nucleotide-binding leucine-rich repeat (NLR) genes, a class of important plant disease resistance genes. We demonstrate that the CG methylation levels of three NLR genes (*AT5G49140*, *AT5G35450*, and *AT5G36930*) are increased in the *ros1* mutants relative to the wild-type plants, whereas the CG methylation level of *AT2G17050* is decreased. We also observed increased CG methylation levels of *AT4G11170* and *AT5G47260* and decreased CG methylation levels of *AT5G38350* in *ddd* mutants. We further found that the expression of three NLR genes (*AT1G12280*, *AT1G61180*, and *AT4G19520*) was activated in both *ros1* and *ddd* mutants, whereas the expression of another three NLR genes (*AT1G58602*, *AT1G59620*, and *AT1G62630*) was repressed in these two mutants. Quantitative reverse transcriptase–polymerase chain reaction detection showed that the expression levels of *AT1G58602.1*, *AT4G19520.3*, *AT4G19520.4*, and *AT4G19520.5* were decreased in the *ros1* mutant; *AT3G50950.1* and *AT3G50950.2* in the *ddd* mutant were also decreased in expression compared to Col-0, whereas *AT1G57630.1*, *AT1G58602.2*, and *AT5G45510.1* were upregulated in the *ddd* mutant relative to Col-0. These results indicate that some NLR genes are regulated by DNA demethylases. Our study demonstrates that each DNA demethylase (ROS1, DML2, and DML3) exerts a specific effect on the DNA methylation of the NLR genes, and active DNA demethylation is part of the regulation of DNA methylation and transcriptional activity of some *Arabidopsis* NLR genes.

Keywords: nucleotide-binding leucine-rich repeat genes, DNA demethylases, cytosine methylation, active DNA demethylation, transcriptional regulation

INTRODUCTION

Cytosine DNA methylation is an important epigenetic mark (Johnson et al., 2002). It is observed on three sequence contexts, that is, CG, CHG, and CHH (where H represents A, C, or T), in the *Arabidopsis* genome (Chan et al., 2005). The regulation of gene expression by DNA methylation in plants has been discovered to play important roles in the cellular response to pathogen attacks (Downen et al., 2012; Yu et al., 2013; Le et al., 2014; Deleris et al., 2016). DNA methylation patterns in eukaryotes are shaped by DNA methylation and demethylation processes (Chan et al., 2005; Meyer, 2011; Zhang et al., 2018).

It has been demonstrated that a plant-specific pathway, RNA-directed DNA methylation (RdDM), mediates *de novo* cytosine methylation in three cytosine sequence contexts (Zhang and Zhu, 2011). More studies have revealed that two RdDM mechanisms, canonical and non-canonical RdDM pathways, establish DNA methylation in plants (Matzke and Mosher, 2014). In RdDM pathways, the *de novo* methyltransferase DRM1/2 plays key roles in sequence-specific cytosine methylation. Additionally, cytosine methylation has been determined to be established and maintained through several key methyltransferases in plants (Bender, 2004; Chan et al., 2005; Law and Jacobsen, 2010).

In *Arabidopsis* (*Arabidopsis thaliana*), active DNA demethylation is mediated by DNA glycosylase/lyases, that is, ROS1, DME, DML2, and DML3 (Choi et al., 2002; Penterman et al., 2007; Zhu, 2009). It is known that *Arabidopsis* ROS1 (repressor of silencing 1), a bifunctional DNA glycosylase/lyase, functions in repressing transcriptional gene silencing by the action of DNA demethylation (Gong et al., 2002). Mutations in ROS1 result in DNA hypermethylation and transcriptional silencing of specific genes (Penterman et al., 2007). It has been shown that hypermethylation is triggered in the promoters of some silenced loci in *ros1* mutants (Gong et al., 2002). *Arabidopsis* DME encodes a protein containing a DNA glycosylase domain and a nuclear localization domain, which is able to actively erase 5-methylcytosines by a base excision repair pathway (Choi et al., 2002; Morales-Ruiz et al., 2006). Another two DME paralogs, known as demeter-like proteins DML2 and DML3, were found in the genome of *Arabidopsis* (Choi et al., 2002; Penterman et al., 2007). DMEs function mainly in the central cells of female gametophytes, and they are vital for imprinted genes, for example, *MEA*, to be expressed in a maternal allele-specific pattern in the endosperm (Gehring et al., 2006; Bauer and Fischer, 2011). The other three demethylases, ROS1, DML2, and DML3, were shown to be largely active in *Arabidopsis* somatic cells (Gong et al., 2002; Ortega-Galisteo et al., 2008). It was found that approximately 180 discrete loci throughout the *Arabidopsis* genome were demethylated by DML enzymes, and more than 80% of these loci were located in genic regions (Penterman et al., 2007). Strikingly, the 5' and 3' ends of these regions were primarily targeted by the DML enzymes (Penterman et al., 2007). DML3 was also observed to demethylate preferentially symmetrical sequence contexts (CpG and CpHpG) (Ortega-Galisteo et al., 2008). *rdd* is a triple mutant with mutations in ROS1, DML2, and DML3 (Penterman et al., 2007). It was reported that many hypermethylated regions in

rdd do not overlap with those in *ros1* (Qian et al., 2012). This finding suggests that DML2 and DML3 have specific functions in contrast to ROS1. An earlier study demonstrated that after DNA demethylation occurred in *Arabidopsis*, activation of the defense response mediated by salicylic acid was observed, and bacterial pathogen multiplication was restricted (Yu et al., 2013). Another study revealed that stress-responsive genes in *Arabidopsis* can be modulated by DNA demethylases by targeting transposable elements within their promoters (Le et al., 2014). These results imply that active DNA demethylation is a factor that strongly affects disease resistance in plants.

Nucleotide-binding leucine-rich repeat (NLR) proteins, a class of immune receptors, play an important part in plant disease resistance. It was reported that approximately 150 typical *Arabidopsis* NLR genes were identified and characterized in ecotype Col-0 (Meyers et al., 2003). All the proteins were categorized into Toll/interleukin 1 receptor (TIR) or coiled-coil (CC) motif-containing NLR subfamilies, abbreviated as TNL and CNL, respectively (Meyers et al., 2003; McHale et al., 2006). Plant NLR genes have been well-known to play fundamental roles in disease resistance (Dangl and Jones, 2001). However, the transcriptional regulation of NLR genes has not been thoroughly elucidated, despite their importance in plant disease resistance. The expression levels of plant NLR genes may be regulated by diverse factors, including tissue types, developmental stages, environmental cues, and pathogen attacks (Yoshimura et al., 1998; Wang et al., 1999). A previous study revealed that most *Arabidopsis* NLR genes were expressed weakly, even with tissue-specific expression patterns (Tan et al., 2007). Some evidence has shown that small RNAs modulate the expression of plant NLR genes (Zhai et al., 2011; Li et al., 2012; Shivaprasad et al., 2012; Fei et al., 2013). Phased, secondary, small interfering RNAs (phasiRNAs), formerly known as *trans*-acting small interfering RNAs (tasiRNAs), are primed by miRNAs, a category of small RNAs. phasiRNAs and miRNAs were found to suppress the expression of tomato NLR genes (Shivaprasad et al., 2012). It was reported that an *Arabidopsis* NLR gene, *At4g11170*, temporarily named *resistance methylated gene 1* (*RMG1*) by the authors, is an outstanding RdDM target, and ROS1 is essential for its background expression and activated transcription (Yu et al., 2013). *RBA1*, encoding a TIR-containing, truncated NLR protein, is speculated to be regulated through cytosine methylation in the *Arabidopsis* Col-0 ecotype (Nishimura et al., 2017). In addition, new findings suggested that DNA methylation is involved in regulating the expression of some NLR genes in *Arabidopsis* and common bean (Kong et al., 2018; Richard et al., 2018).

A previous study demonstrated that single, double, and triple F2 mutants of *ROS1*, *DML2*, and *DML3* show no obvious morphological phenotypes under their growth conditions (Penterman et al., 2007; Ortega-Galisteo et al., 2008). However, developmental abnormalities were observed in some *ros1* mutants in later generations (Gong et al., 2002). Furthermore, the *ros1* mutant is sensitive to hydrogen peroxide and methyl methanesulfonate (Gong et al., 2002). Additionally, it was observed that a slightly increased bacterial growth occurred in the *ros1* mutant, but not in the *dml2* and *dml3* mutants with inoculation of *Pseudomonas syringae* pv. *tomato* strain

DC3000 (Yu et al., 2013). In the *rdd* mutant, an enhanced susceptibility was found to *Fusarium oxysporum* (Le et al., 2014). Another study showed that opposite phenotypes were observed in *Arabidopsis* hypomethylated mutants and hypermethylated mutants after infection with *Hyaloperonospora arabidopsidis* (Lopez Sanchez et al., 2016).

In this study, we used publicly available bisulfite sequencing (BS-Seq) data to identify *Arabidopsis* NLR genes that are targeted by demethylases, including ROS1, DML2, and DML3 in wild-type plants. We demonstrate that the CG methylation levels in the 5' upstream regions (UPRs) of 30 *Arabidopsis* NLR genes were increased in both the *ros1* and *rdd* mutant plants. Furthermore, we show that 32 *Arabidopsis* NLR genes were presumably regulated by both ROS1 and DML demethylases at the transcriptional level. In conclusion, our data indicate that active DNA demethylation by ROS1 and DML enzymes functions to protect *Arabidopsis* NLR genes from potentially deleterious methylation. The data also implicate ROS1 and DML demethylases in determining the DNA methylation profiles of *Arabidopsis* NLR genes. Additionally, we analyzed the available mRNA-Seq data from *Arabidopsis* *ros1*, *rdd* mutants, and their wild-type control plants. We found that mutations in DNA demethylases lead to changes in the transcriptional activities of some *Arabidopsis* NLR genes, suggesting that their expression is regulated by DNA demethylases.

MATERIALS AND METHODS

Retrieval of *Arabidopsis* BS-Seq and mRNA-Seq Data

The *Arabidopsis* BS-Seq data used in this study were retrieved from the Gene Expression Omnibus (GEO) database.¹ The GEO accession numbers for the data are GSM1859474 (SRR2179846, SRR2179847, SRR2179848, and SRR2179849)/GSM1859475 (SRR2179850, SRR2179851, SRR2179852, and SRR2179853) (wild-type/*ros1* mutant) and GSM819122/GSM819123/GSM819128/GSM819129 (wild-type/*rdd* mutant). The mRNA-Seq data from the wild-type, *ros1*, and *rdd* mutants were downloaded from the GEO database. Their GEO accession numbers are GSM1585887/GSM1585888/GSM1585889/GSM1585899/GSM1585900/GSM1585901 (wild-type/*ros1* mutant). The *rdd* mRNA-Seq data were retrieved from the NCBI SRA database,² whose accession numbers are SRR013411/SRR013412/SRR013413/SRR013414/SRR013415/SRR013416/SRR013426/SRR013427/SRR013428/SRR013429 (wild-type/*rdd* mutant).

Processing of *Arabidopsis* BS-Seq Data

The SRA-formatted BS-Seq data were changed into the FASTQ format, and their sequencing quality was then evaluated. The adapters for sequencing were removed, and the low-quality bases were deleted. The clean BS-Seq reads were mapped to

the TAIR10 genome (v36) with Bismark (v0.16.3) (Krueger and Andrews, 2011), allowing one base mismatch, and the unique paired-end reads were obtained for next analysis. To ensure dependable sequencing sites, cytosines covered by at least four reads were selected.

Methylation Analysis of *Arabidopsis* NLR Genes

Arabidopsis typical NLR genes encoding both NB and LRR domains were selected for further analysis (Meyers et al., 2003). The gene body region (GBR) (transcribed region) covers the genomic region from the transcription start site to the end site. The chromosomal coordinates of *Arabidopsis* NLR GBRs and 200- and 500-bp regions upstream of the transcription start sites were determined with the TAIR10 annotation file³ by custom Perl scripts (Supplementary Table S1). The cytosine methylation levels were calculated as described previously (Kong et al., 2018).

Processing of *Arabidopsis* mRNA-Seq Data

Possible adaptor sequences were cleaned from all the sequences before the reads were mapped to the *Arabidopsis* reference genome sequence, and the reads for which more than 50% of the bases had a low-quality value (≤ 5) were discarded. Then, the filtered reads were mapped through TopHat (v. 2.1.1) (Trapnell et al., 2012) to the TAIR10 genome sequence. The abundance of the *Arabidopsis* gene transcripts was determined and normalized with FPKM, that is, the expected fragments per kilobase of a transcript per million fragments sequenced, by Cufflinks software (v.2.2.1) (Trapnell et al., 2010, 2012).

HTSeq⁴ was used to measure the raw counts for all *Arabidopsis* genes determined through the TAIR10 annotation for coding genes (Anders et al., 2015). Then, the Cuffdiff program in the Cufflinks package (v2.2.1) was adopted to generate the differential expression data from these counts. The differentially expressed genes in each compared group were identified by the cutoff value of a more than twofold change and an adjusted *p*-value or FDR (false discovery rate) threshold ≤ 0.05 .

RNA Isolation and Real-Time Polymerase Chain Reaction Analysis

Total *Arabidopsis* RNAs were extracted from 2-week-old seedlings by TRIpure reagent (Aidlab Biotech, Beijing, China), and the possible contaminating DNAs were digested with DNase I (TransGen, Beijing, China). Two micrograms of total RNA was used for first-strand cDNA synthesis with the PrimeScript RT reagent kit (Takara, Dalian, China) according to the manufacturer's instructions. The cDNA reaction mixtures were then diluted fivefold. In a 20- μ L polymerase chain reaction (PCR) mixture, 1 μ L of the diluted cDNA solution was pipetted into a tube as the template. *Arabidopsis* ACTIN2 was used as an internal control. Program Premier 3 (Koressaar and Remm, 2007; Untergasser et al., 2012) was used to

¹<https://www.ncbi.nlm.nih.gov/gds>

²<https://www.ncbi.nlm.nih.gov/sra>

³<https://www.arabidopsis.org>

⁴https://htseq.readthedocs.io/en/release_0.10.0/

design the quantitative reverse transcriptase (qRT)–PCR primers (**Supplementary Table S2**). Quantitative reverse transcriptase–PCR was performed using the ABI 7500 Real Time PCR System (ABI, Carlsbad, CA, United States) with TransStart Top Green qPCR SuperMix (TransGen, Beijing, China). Three independent PCR analyses were carried out. The relative transcript levels were determined by the comparative threshold cycle (Ct) method (Relative Quantification Getting Started Guide; ABI). The mean fold changes were calculated using Livak's $2^{-\Delta\Delta(Ct)}$ method (Livak and Schmittgen, 2001).

RESULTS

Set of *Arabidopsis* NLR Genes Targeted by DNA Demethylases

DNA methylation occurring in the UPRs and within the transcribed gene bodies was observed in the majority of NLR genes in wild-type *Arabidopsis* plants, and the average methylation level of CG sequence contexts was greatly higher than that of CHG and CHH sequence contexts (Kong et al., 2018). In this study, we examined the DNA methylation status of NLR genes in the *ros1* and *rdd* mutant backgrounds by analyzing the BS-Seq data available from both the mutants and the corresponding wild-type controls (**Supplementary Tables S3, S4**).

Our results demonstrated that for CG, CHG, and CHH sequence contexts, the average methylation levels in the 200- and 500-bp regions lying immediately upstream of transcriptional starting sites and of the entire transcribed gene bodies of the 144 *Arabidopsis* NLR genes were, in most situations, increased in *ros1* and *rdd* mutants relative to wild-type controls, indicating that the NLR genes in general are the targets of DNA demethylases (i.e., ROS1, DML2, and/or DML3) (**Figure 1**). In addition, the average methylation level of CG sequence contexts of the *Arabidopsis* NLR genes was clearly higher than the levels of CHG and CHH sequence contexts in both the *ros1* and *rdd* mutants (**Figure 1**).

Because the average methylation level of the CG sequence contexts was significantly higher than those of the CHG and CHH sequence contexts in the *ros1* and *rdd* mutants, the 144 *Arabidopsis* NLR genes were classified into two groups on the basis of their CG methylation levels: group 1 (>0.1), whose methylation level is greater than 0.1, and group 2 (<0.1), whose methylation level is less than 0.1. The results demonstrated that the CG methylation levels of these NLR genes in *ros1* and *rdd* at the 200-bp UPR, 500-bp UPR, and GBR are all increased because the proportions of group 1 in both mutants at the three regions increase consistently compared to those in wild-type controls. For example, the proportions of group 1 at 500-bp UPR in *ros1* and *rdd* were 22 and 24% versus 15 and 14% in the corresponding wild-type controls (**Figure 2**). By comparison, the proportions of group two at three such regions in both mutants were decreased overall (**Figure 2**). It is worth noting that the increase of proportions at 500-bp UPR is more dramatic in both *ros1* and *rdd* mutants than at the other two regions in their respective wild-type controls (**Figure 2**). Thus, these data collectively suggest that the mutations of the DNA demethylases

generally lead to hypermethylation at the 200-bp UPRs, 500-bp UPRs, and GBRs of these NLR genes, and the 500-bp UPRs gain a higher level of methylation than the other two regions.

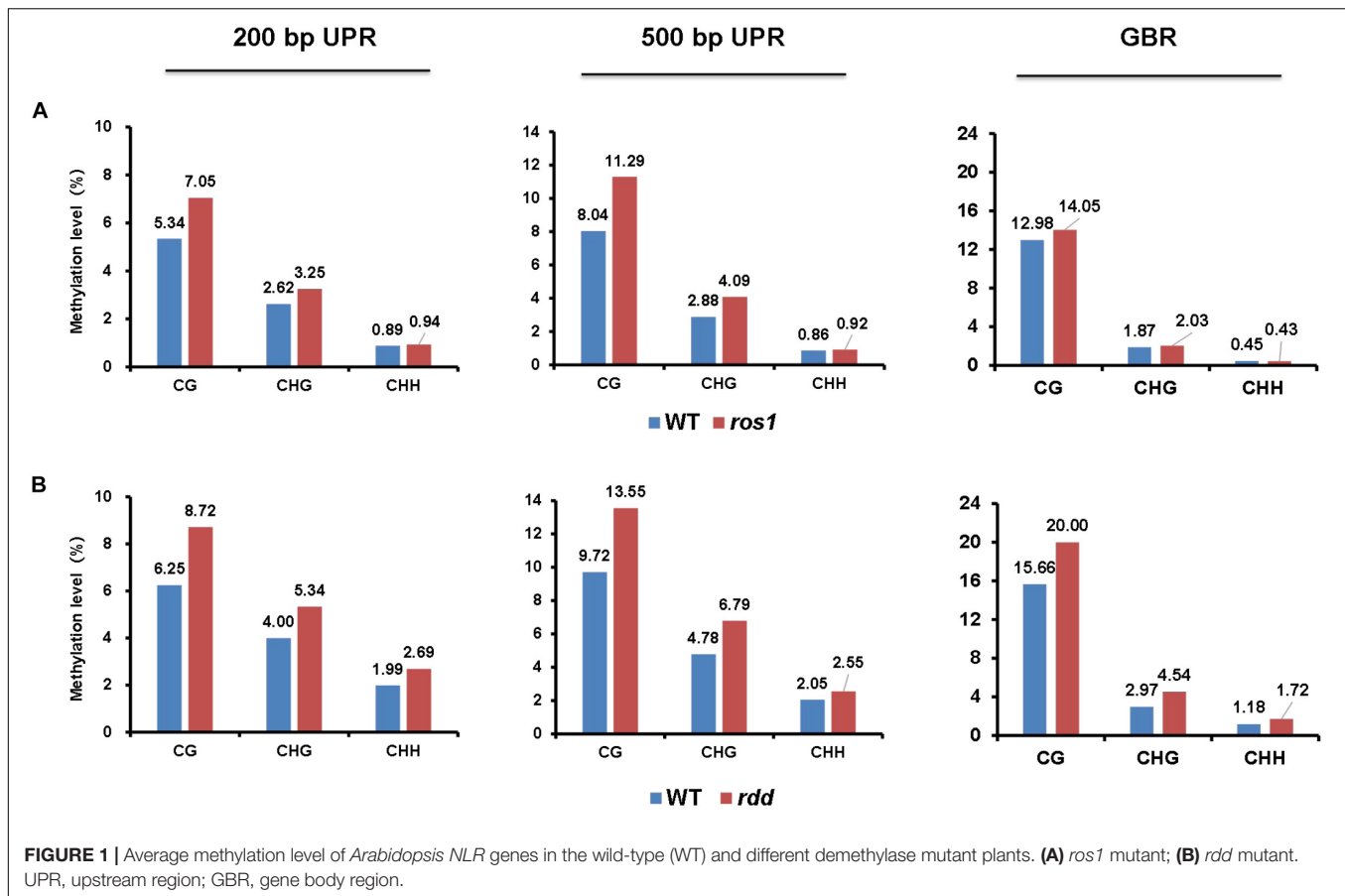
To further determine which members of the NLR genes have undergone evident changes in DNA methylation level in *ros1* and *rdd* mutants, the CG methylation levels of all the NLR genes were analyzed. Our results demonstrated that there are 10 NLR genes in which CG methylation levels at the 200-bp UPRs are significantly different between *ros1* and wild-type plants (**Supplementary Table S5**); eight of them show at least a 10% increase in DNA methylation level in the *ros1* mutant compared with the wild-type control. For *AT5G49140*, *AT5G35450*, and *AT5G36930*, their CG methylation levels were more than 50% higher in the *ros1* mutant relative to the wild-type control (**Figure 3** and **Supplementary Table S5**). In contrast, two genes, that is, *AT4G09430* and *AT2G17050*, exhibited decreased CG methylation levels in the *ros1* mutant, and notably, the proportion of CG methylation of *AT2G17050* decreased from 77.27% to zero (**Supplementary Table S5**).

For 500-bp UPRs, all 18 examined genes but one (*AT1G12280*) in *ros1* showed no less than 10% increase in CG methylation level compared to the wild-type control (**Supplementary Table S5**). Two genes, *AT5G35450* and *AT5G49140* in *ros1*, display a methylation increase of greater than 40% relative to wild type (**Supplementary Table S5** and **Figure 3**). It should be noted that there are four genes (i.e., *AT5G49140*, *AT4G27190*, *AT1G31540*, and *AT1G59780*) with no methylation at such regions in the wild-type control, showing increased methylation at least 20% in the *ros1* mutant. For the methylation status of GBRs, it appears that there are no obvious differences in DNA methylation levels between *ros1* and the wild-type control because the maximum difference is less than 9% as exemplified by *AT5G35450*, suggesting that the transcribed gene bodies of such NLRs are not the main targets of ROS1 (**Supplementary Table S5**).

In the *rdd* mutant, the CG methylation levels in the 200-bp UPRs of eight of nine *Arabidopsis* NLR genes increased by more than 10% (**Supplementary Table S6**). It is worth noting that three genes (*AT5G47260*, *AT4G11170*, and *AT5G45240*) have notably low levels of DNA methylation in the wild-type control, whereas they show a substantial increase of more than 40% in methylation levels in the *rdd* mutant (**Supplementary Table S6** and **Figure 4**).

Within the 500-bp UPRs, there were 23 of 24 NLR genes, which all exhibited a growth of 10% in methylation levels, and five of these genes displayed a 30% increase in methylation levels in the *rdd* mutant compared to the wild-type control. In contrast, the methylation level of *AT2G17050* was reduced by approximately 57% in *rdd* (**Supplementary Table S6**). For GBRs, 23 of 43 examined NLR genes showed an increase in methylation levels by more than 10% in *rdd* compared to the wild-type control (**Supplementary Table S6** and **Figure 4**). Notably, the increase in DNA methylation levels is generally larger within the 500-bp UPRs than the GBRs (**Figures 5A,B**). These data collectively indicate that triple mutations of ROS1, DML2, and DML3 lead to DNA hypermethylation within the promoters, as well as gene bodies of some specific NLR genes.

Our above analysis also revealed that the DNA methylation levels of different NLR genes in the same mutant were highly

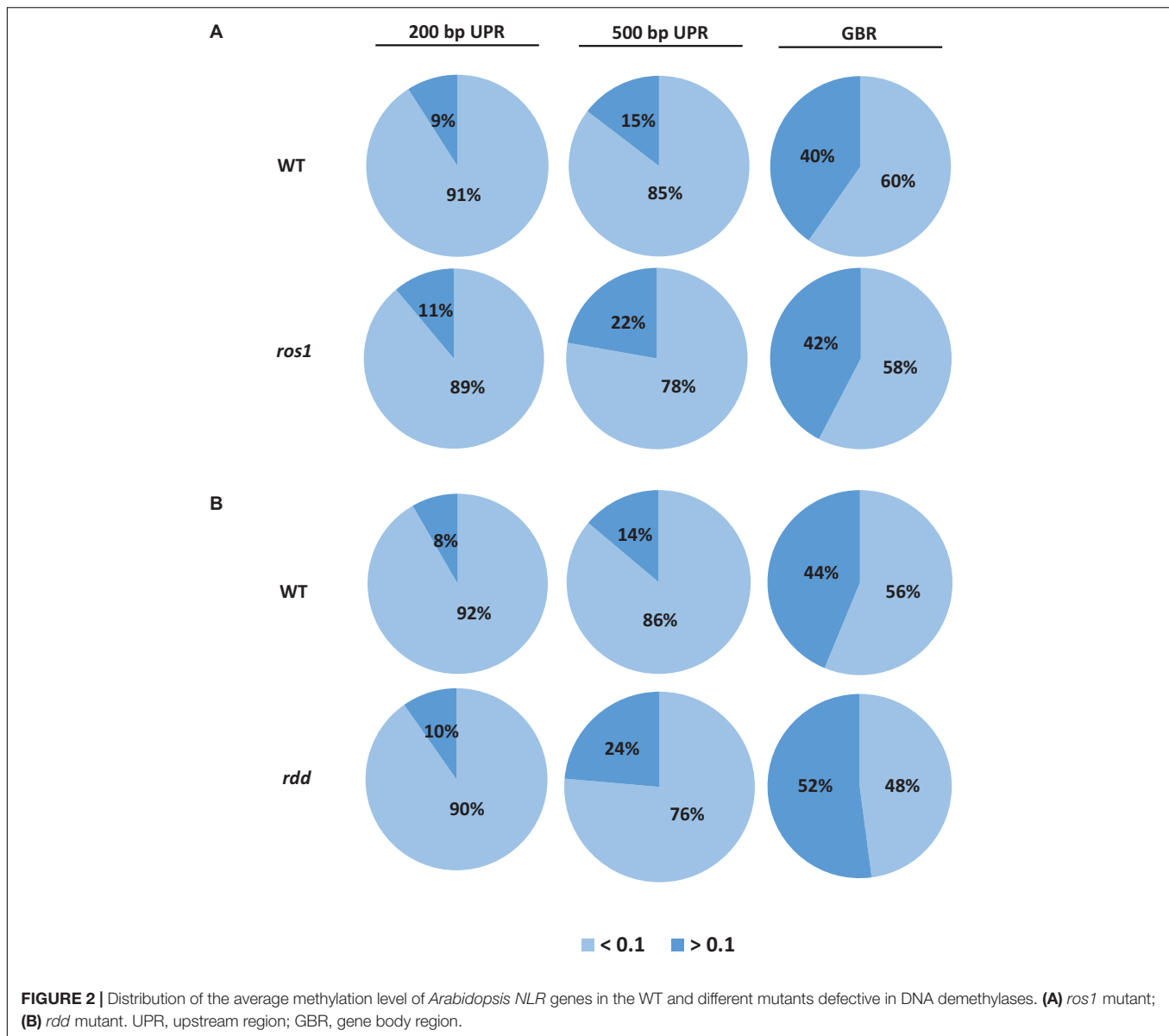


different (Figures 5A,B). For instance, in *ros1*, no DNA methylation was observed in the *AT1G58807*, *AT1G59124*, and *AT1G59218* genes (Supplementary Table S3). However, *AT1G27180* showed a lower level of DNA methylation in the *ros1* mutant than the corresponding wild-type control (Supplementary Table S3). Other NLR genes may be heavily methylated in *ros1* mutants. For example, *AT1G58602* had a CG methylation level of 52.23% and a CHG methylation level of 27.14% in its transcribed gene body in *ros1*; *AT3G46710* had 87.25% of CG methylation level and 15.04% of CHG methylation level in the upstream 500-bp region; *AT4G09360* had a CG methylation level of 86.25% and a CHG methylation level of 21.43% in the upstream 500-bp region, and its CG and CHG methylation levels within the transcribed region were 77.79 and 41.23%, respectively; in the upstream 500-bp region of *AT4G19520*, the CG methylation level was as high as 87.25%, and the CHG methylation level was 15.04%; for *AT5G36930*, the CG and CHG methylation levels in the upstream 500-bp region were as high as 88.03 and 55.63%, respectively (Supplementary Table S3). In the *ros1* mutant, unexpectedly, in the 200-bp UPR of *AT2G17050*, the CG methylation level was 77.27% lower than that of the wild-type control (Supplementary Table S3).

In the *rdd* mutant, a similar methylation profile exists (Figure 5B). Three genes, *AT1G58807*, *AT1G59124*, and *AT1G59218*, did not exhibit DNA methylation, whereas

AT1G12210, *AT1G27180*, *AT1G56540*, and *AT5G46260* were less highly methylated, but *AT3G46710*, *AT4G09360*, *AT4G19520*, and *AT5G36930* were highly methylated at their CG and CHG sites (Supplementary Table S4). In this mutant, compared to the wild-type plants, the methylation levels of CG, CHG, and CHH sites in the upstream 200-bp region of *AT5G45240* were increased by 42.47, 29.56, and 16.66%, respectively, whereas the CG methylation levels in the upstream 200-bp regions of *AT4G11170* and *AT5G47260* were 59.71 and 62.19% higher than those of the wild-type plants, respectively (Supplementary Table S4).

DNA demethylases play an important role in inhibiting the hypermethylation of endogenous genes in plants. However, this study demonstrated that some *Arabidopsis* NLR genes show high DNA methylation not only in *ros1* and *rdd* mutants but also in wild-type plants (Figures 5A,B). The DNA methylation of these genes was found to be similar between the wild-type and mutant plants. For example, *AT4G09360* and *AT5G47280* were highly methylated in the UPRs and transcribed gene bodies in both the *ros1* mutants and the wild-type plants (Figure 6 and Supplementary Table S3). In the UPRs of *AT4G19500* and *AT4G19510*, three cytosine sequence contexts were highly modified by DNA methylation in the wild-type and *ros1* mutant plants, and CG methylation was observed within their transcribed regions (Figure 6 and Supplementary Table S3). The other two genes, *AT2G17060* and *AT4G09430*, were



hypermethylated primarily in their UPRs in the *ros1* mutants and wild-type plants (Figure 6 and Supplementary Table S3). In the wild-type and *rdd* mutant, *AT4G09360*, *AT5G47260*, and *AT5G47280* were heavily methylated in the three cytosine sequence contexts of the upstream and transcribed regions (Figure 7 and Supplementary Table S4); *AT5G36930* was also clearly modified by DNA methylation, and the three cytosine sequence contexts of its UPR were significantly modified by DNA methylation, but CG methylation was mainly found within its transcribed region (Figure 7 and Supplementary Table S4). Interestingly, *AT4G09360* and *AT5G47280* were hypermethylated in both *ros1* and *rdd*, as well as their respective wild-type plants (Figures 6, 7 and Supplementary Tables S3, S4). In addition, *AT4G19500* and *AT4G19510* were the same as these two genes, but their methylation levels were considerably lower in the extent of modification (Figures 6, 7 and Supplementary Tables S3, S4).

The maintenance of heavy DNA methylation within these genes in wild-type plants suggests that DNA demethylases have little effect on them and that hypermethylation plays a critical role in their functions.

Transcriptional Activities of *Arabidopsis* NLR Genes in Wild-Type Plants and Various DNA Demethylase Mutants

It has been reported that there is a close relationship between DNA methylation and the transcriptional activity of a gene (Zilberman et al., 2007). The expression of NLR genes in *Arabidopsis* and soybean has also been shown to be regulated by their DNA methylation levels (Kong et al., 2018; Richard et al., 2018). To determine whether the mutations of the DNA demethylases affect the transcriptional activities of the

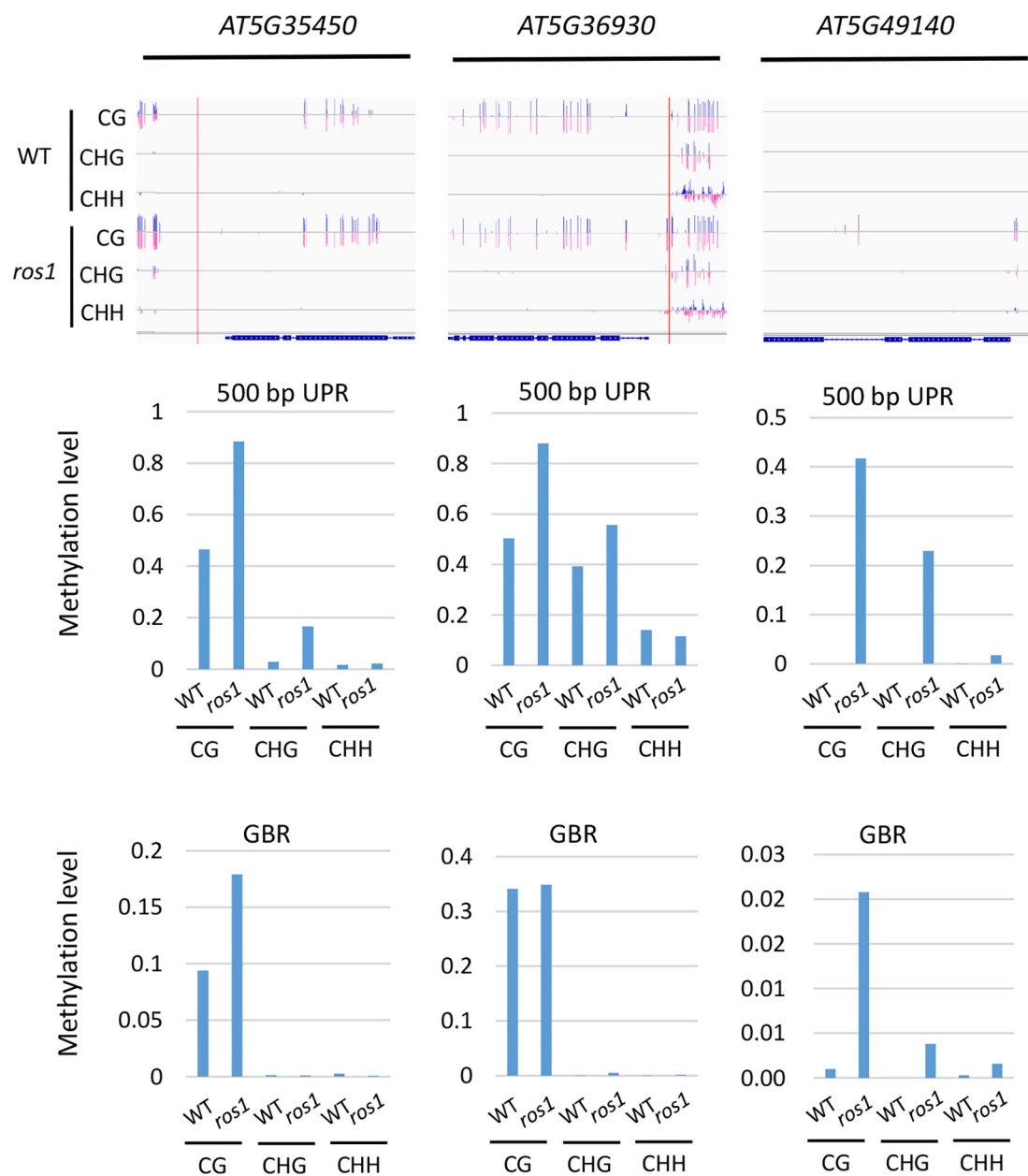


FIGURE 3 | Methylated *NLR* genes from *Arabidopsis* WT (Col-0) and *ros1* mutants. For each *NLR* gene, the DNA methylation patterns in the WT and *ros1* mutant are shown in the upper panel, and its DNA methylation levels are illustrated in the lower panels. The red rectangle indicates the 500-bp UPR.

Arabidopsis *NLR* genes, this study analyzed the available mRNA-Seq data from the *Arabidopsis* *ros1* and *rdd* mutants and their respective wild-type controls to examine the transcriptional activities of the *Arabidopsis* *NLR* genes.

An overall analysis of the transcriptional level of the *NLR* genes in the wild-type and mutants indicated that the expression levels of most *NLR* genes were very low in both wild-type plants and the mutants (Figure 8). However, most of the *NLR* genes with relatively high transcriptional activity in the wild-type plants showed a slightly higher expression level after the

mutation of *ROS1* (Figure 8A), whereas most of those *NLR* genes with relatively high expression levels in the wild-type plants demonstrated reduced expression in the *rdd* mutant (Figure 8B). Specifically, our analysis revealed that there are 43 transcribed *NLR* genes with the value of at least one FPKM in the wild-type plants or *ros1* mutants, and their ratios of FPKM values in *ros1* to the wild-type plants are ≥ 1.1 or ≤ 0.9 (Supplementary Table S7). Among these genes, the FPKM values of 38 *NLR* genes increased, and those of five *NLR* genes decreased, in *ros1* relative to the wild-type plants (Supplementary Table S7). It is

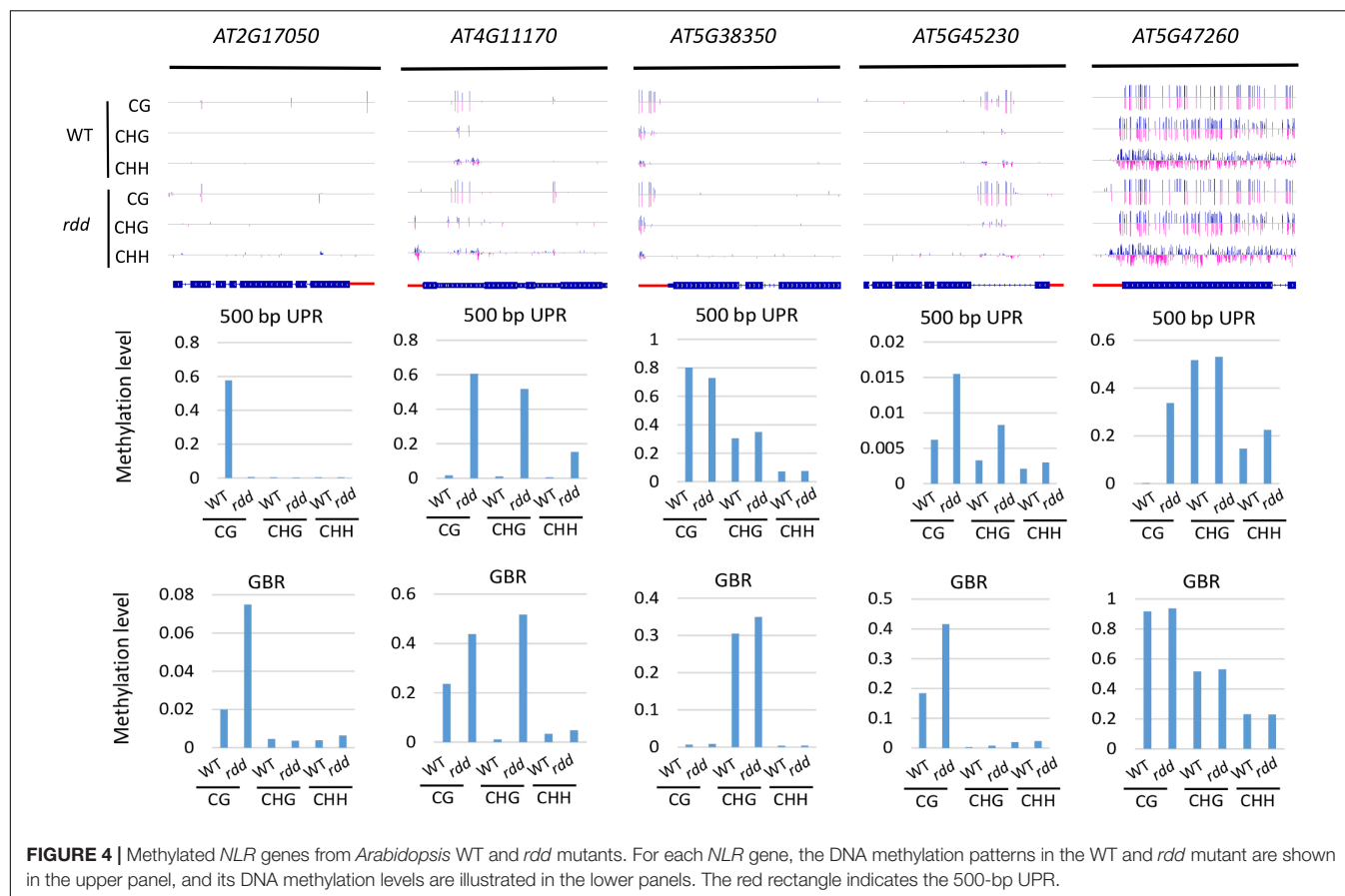


FIGURE 4 | Methylated *NLR* genes from *Arabidopsis* WT and *rdd* mutants. For each *NLR* gene, the DNA methylation patterns in the WT and *rdd* mutant are shown in the upper panel, and its DNA methylation levels are illustrated in the lower panels. The red rectangle indicates the 500-bp UPR.

noted that the FPKM value of *AT4G19520* in *ros1* was even 1.97 times the value of the gene in the wild-type plants, suggesting that the mutation of *ROS1* contributes to the transcription of such genes (Supplementary Table S7). However, there are five genes (i.e., *AT1G58602*, *AT1G10920*, *AT1G63750*, *AT1G62630*, and *AT1G59620*) that all have FPKM values of less than 1.0, indicating that these five genes are downregulated in the *ros1* mutant (Supplementary Table S7).

In the wild-type or *rdd* mutant plants, 64 *NLR* genes were found to be expressed with the value of at least one FPKM, and the ratios of the FPKM values were ≥ 1.1 or ≤ 0.9 (Supplementary Table S8). Only the ratios of FPKM values of *AT1G12280*, *AT1G61180*, and *AT4G19520* were over 1.1 in the *rdd* mutants, whereas the ratios of the other 61 *NLR* genes were all less than 0.9 (Supplementary Table S8). We also observed that the change in the transcriptional level of some *NLR* genes was inconsistent in *ros1* and *rdd* mutants; however, the transcriptional levels of *AT1G12280*, *AT1G61180*, and *AT4G19520* were higher in both *ros1* and *rdd* mutants than in the wild-type plants. In contrast, the transcriptional levels of *AT1G58602*, *AT1G59620*, and *AT1G62630* were lower in both *ros1* and *rdd* mutants than in the wild-type plants (Supplementary Table S9). This finding suggests that the transcriptional activities of these genes were likely to be regulated by DNA demethylases.

We identified the differentially expressed *NLR* genes between *ros1* or *rdd* and wild-type plants by analyzing their mRNA-Seq

data and then verified some identified *NLR* genes using real-time qRT-PCR. Five selected *NLR* genes were confirmed to be differentially expressed between the mutants and the wild-type plants (Figure 9). The expression levels of 10 transcripts encoded by these five *NLR* genes were detected in *ros1* and *rdd* mutants. The results demonstrated that the expression levels of *AT1G58602.1*, *AT4G19520.3*, *AT4G19520.4*, and *AT4G19520.5* were reduced in the *ros1* mutant relative to Col-0 (Figure 8). Among these genes, *AT4G19520.5* expression was notably reduced in the *ros1* mutant (Figure 8). In *rdd* mutants, *AT3G50950.1* and *AT3G50950.2* were detected to be reduced in expression compared with Col-0 (Figure 8). In contrast, in *rdd* mutants, *AT1G57630.1*, *AT1G58602.2*, and *AT5G45510.1* were upregulated relative to Col-0 (Figure 8). Thus, some *NLR* genes are suggested to be regulated by DNA demethylases.

DISCUSSION

Methylation Patterns of Some *NLR* Genes in *Arabidopsis* Are Shaped by Both DNA Methyltransferases and Demethylases

The DNA methylation patterns of some plant genes can be established and maintained by DNA methyltransferases, whereas

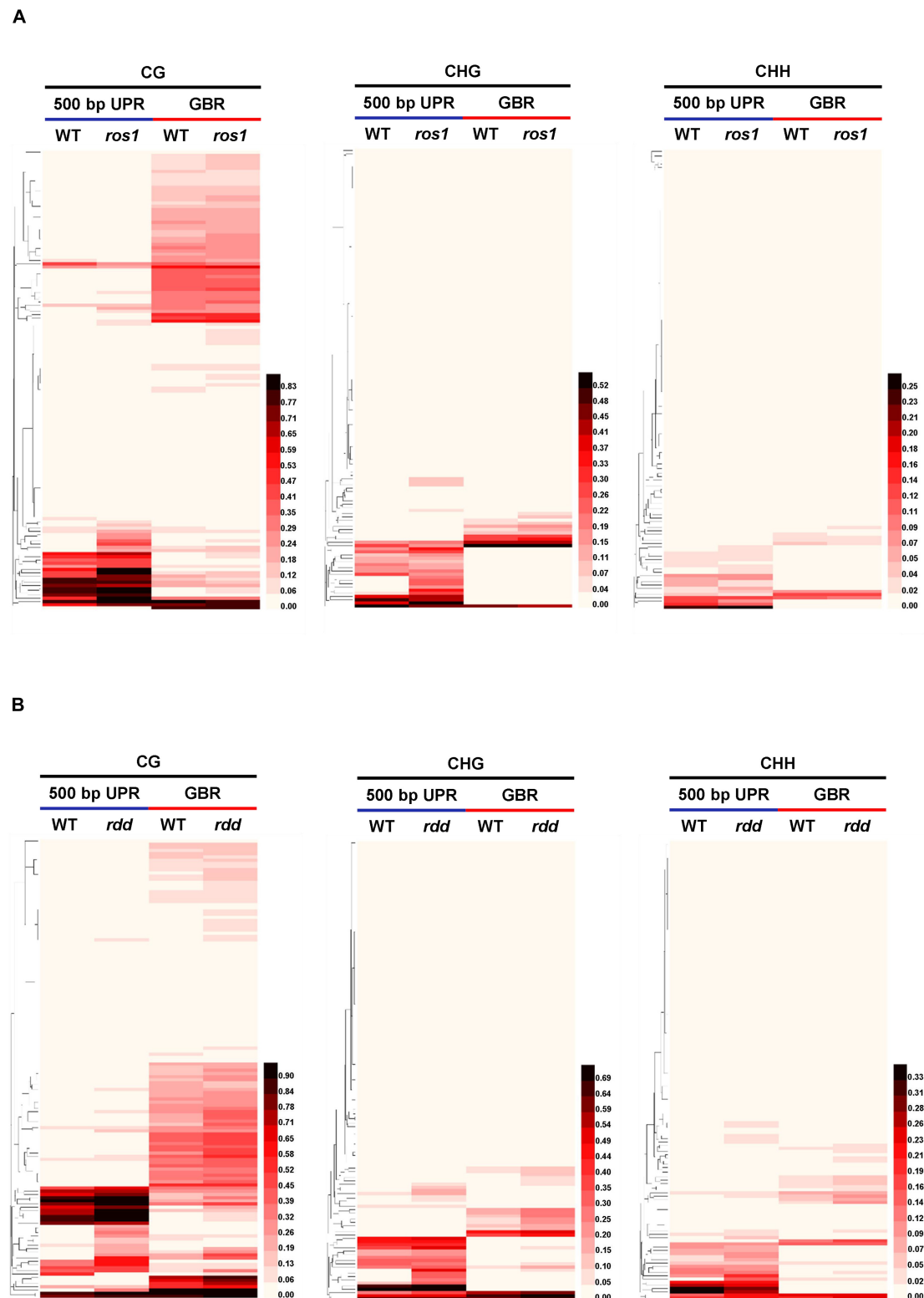


FIGURE 5 | Heatmap of the methylation levels of 144 *Arabidopsis* NLR genes in diverse genotypes. **(A)** NLR genes in wild type and *ros1* mutant; **(B)** NLR genes in wild type and *rdd* mutant.

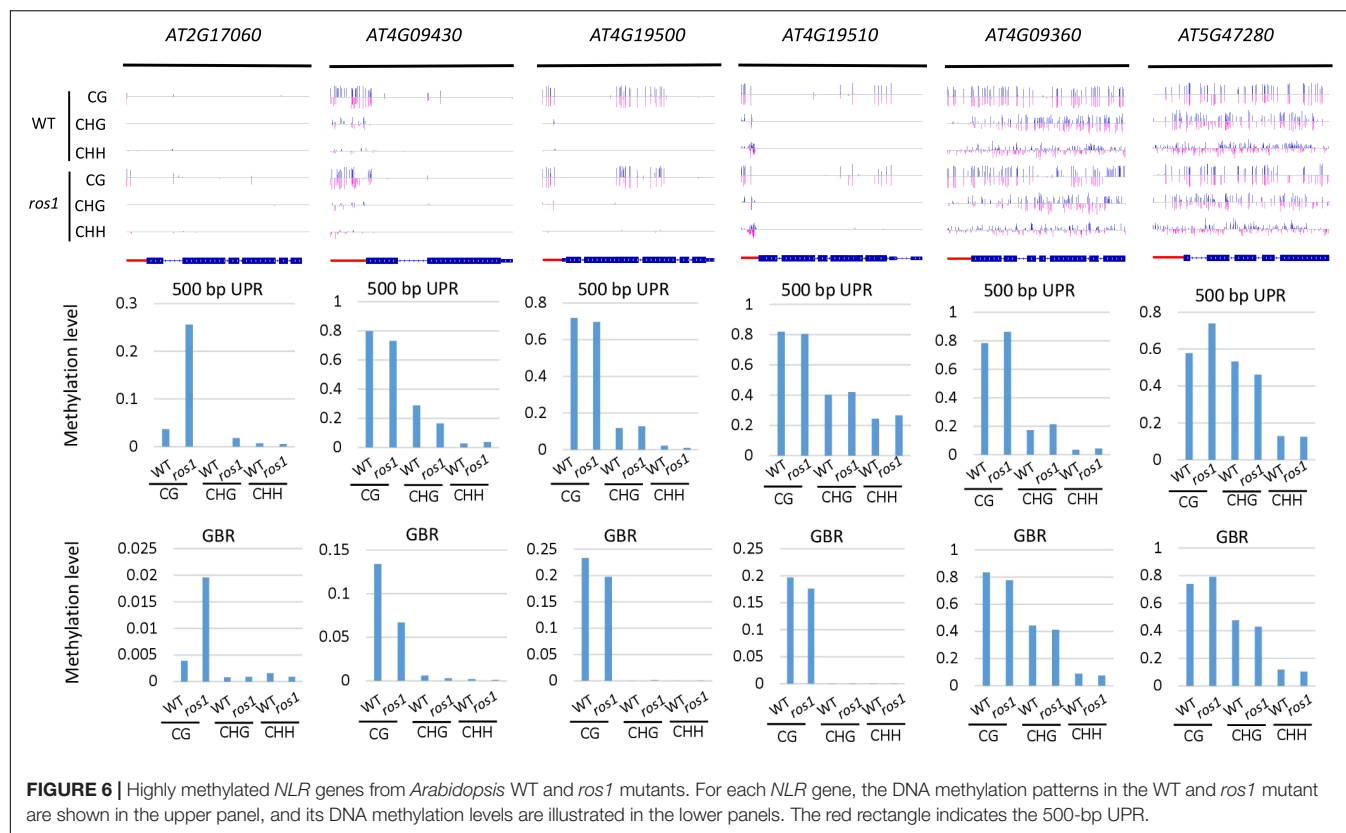


FIGURE 6 | Highly methylated *NLR* genes from *Arabidopsis* WT and *ros1* mutants. For each *NLR* gene, the DNA methylation patterns in the WT and *ros1* mutant are shown in the upper panel, and its DNA methylation levels are illustrated in the lower panels. The red rectangle indicates the 500-bp UPR.

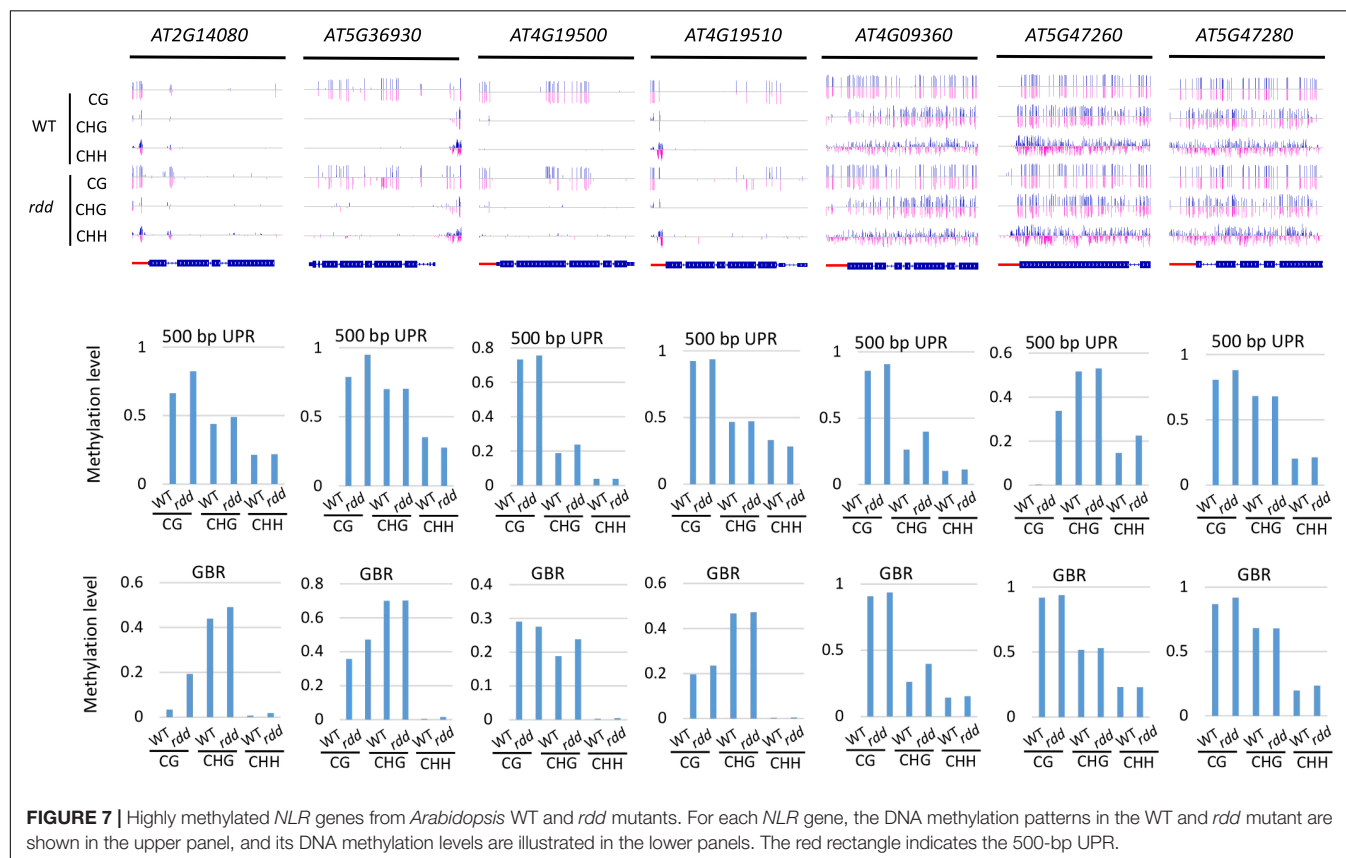
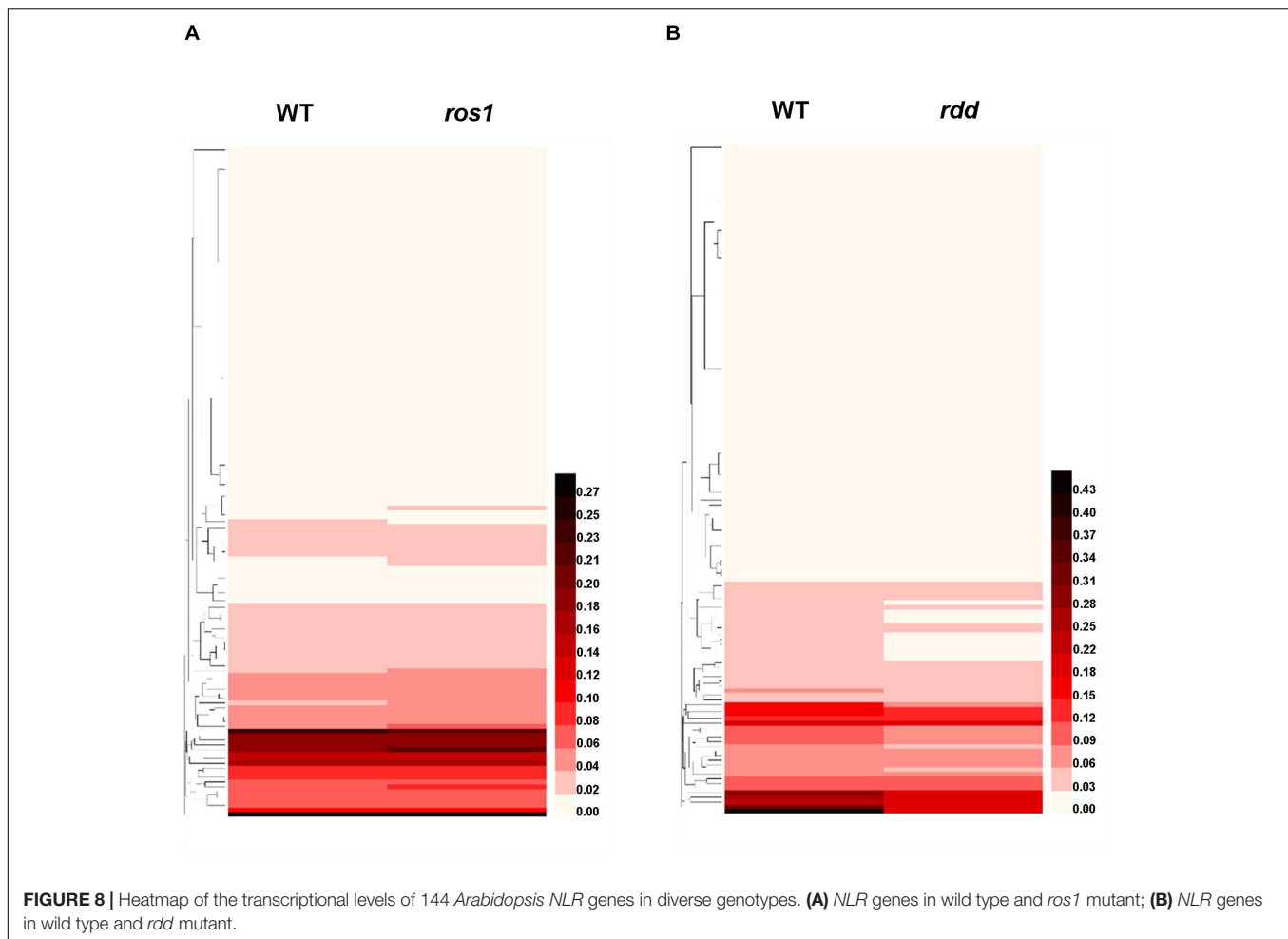


FIGURE 7 | Highly methylated *NLR* genes from *Arabidopsis* WT and *rdd* mutants. For each *NLR* gene, the DNA methylation patterns in the WT and *rdd* mutant are shown in the upper panel, and its DNA methylation levels are illustrated in the lower panels. The red rectangle indicates the 500-bp UPR.



those of others are jointly shaped by both methyltransferases and demethylases (Zhu, 2009). Our analyses revealed that the average methylation levels of the CG, CHG, and CHH sequence contexts in the 500-bp UPRs and transcribed gene bodies of *Arabidopsis* NLR genes varied in different DNA demethylase mutants. In *ros1* and *rdd* mutants, the average methylation levels of the three cytosine sequence contexts within the NLR genes were increased, but to a different extent. The average CG methylation levels within the NLR genes were higher than the average CHG and CHH methylation levels in the *ros1* and *rdd* mutants. It has been shown that most of the CG sites of some transposons and other genes are highly methylated in wild-type plants, whereas many CHG and CHH sites of the transposons and genes are methylated slightly or are even completely unmethylated; however, in *ros1*, these CHG and CHH sites are heavily methylated (Zhu et al., 2007). Similar to this result, higher levels of CHG and CHH methylation were observed within the NLR genes in *ros1* and *rdd* mutants than in the wild-type plants. Further analysis revealed increased NLR genes with a CG methylation level higher than 10% and decreased NLR genes with a CG methylation level of less than 10% in *ros1* and *rdd* mutants.

Among all the known demethylases in *Arabidopsis*, ROS1 is regarded as the predominant DNA demethylase in vegetative

tissues (Tang et al., 2016). However, mutations of DML2 and/or DML3 were observed to cause the hypermethylation of unmethylated or weakly methylated cytosine residues in wild-type plants (Ortega-Galisteo et al., 2008). Additionally, the heavily methylated cytosines in wild-type plants were shown to be hypomethylated in the *dml2* and/or *dml3* mutants (Ortega-Galisteo et al., 2008). Additionally, most of the hypermethylated loci in *ros1-4* were found to overlap with those in the *rdd* mutant (Qian et al., 2012). Thus, ROS1, DML2, and DML3 have their own distinct targets, although they overlap at some loci. We found that eight NLR genes (*AT1G56540*, *AT3G04220*, *AT4G09430*, *AT5G40100*, *AT5G45510*, *AT5G46260*, *AT5G47280*, and *AT5G49140*) showed elevated or reduced methylation levels by at least 10% within their 200- or 500-bp UPRs in *ros1* mutants but not *rdd* mutants (Supplementary Table S10). Similar changes were also observed in the other eleven NLR genes (*AT1G61180*, *AT3G07040*, *AT3G46530*, *AT4G16960*, *AT4G33300*, *AT5G38350*, *AT5G40060*, *AT5G45240*, *AT5G46490*, *AT5G47250*, and *AT5G47260*) in the *rdd* mutant (Supplementary Table S10). On the other hand, there are 14 NLR genes whose alterations in CG methylation level by at least 10% within the UPRs occurred in *ros1* mutants, as well as *rdd* mutants (Supplementary Table S10). Within the transcribed regions, 15 NLR genes displayed an

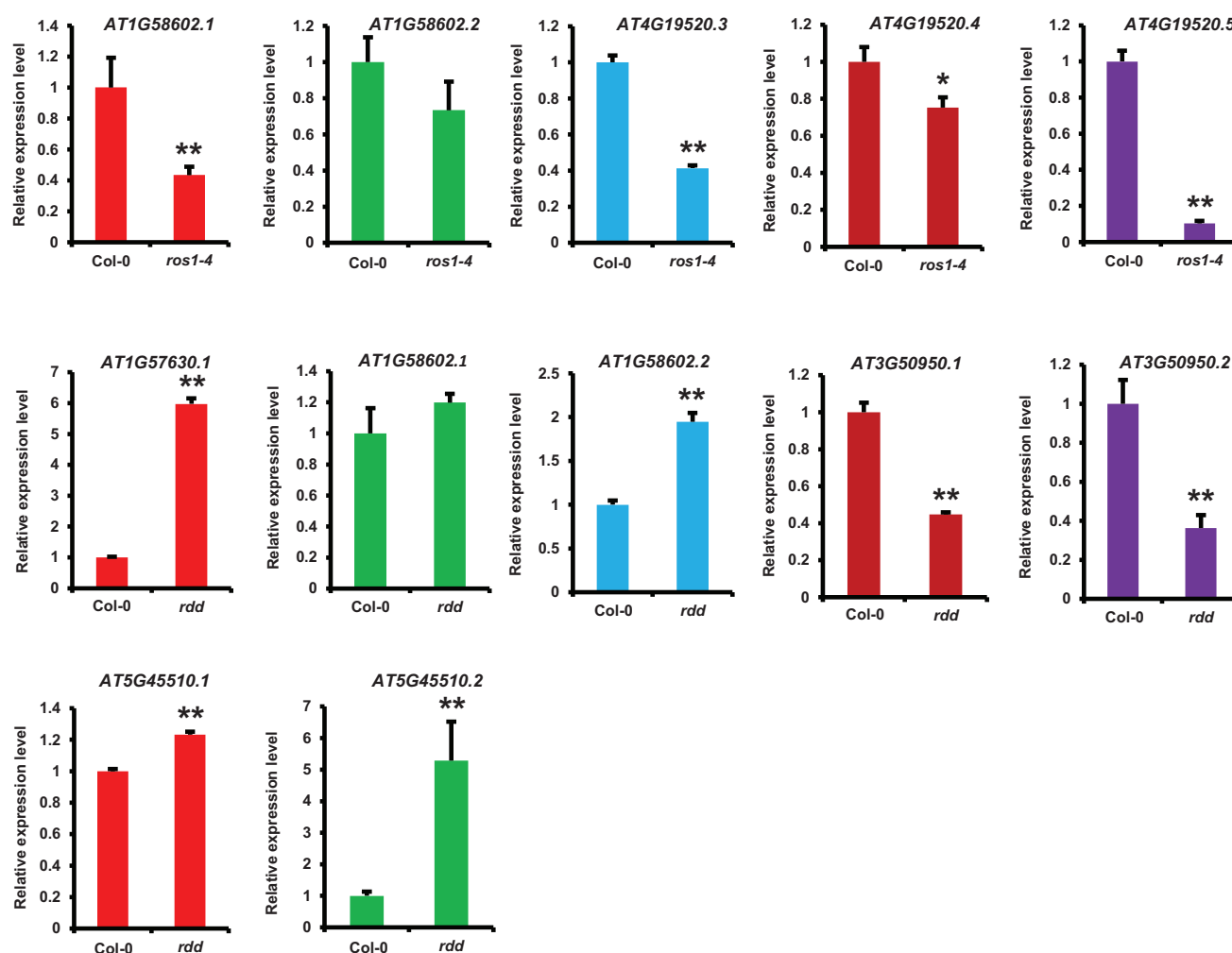


FIGURE 9 | Detection of the expression of *NLR* genes in different *Arabidopsis* genotypes by qRT-PCR. *ACTIN2* was used as an internal control. Error bars represent the standard deviation from three technical replicates. One-way analysis of variance (ANOVA) was applied to examine gene expression differences between the WT and each mutant. *Statistical significance ($p \leq 0.05$) and **high statistical significance ($p \leq 0.01$).

increased CG methylation level by at least 5% in the *ros1* mutant (Supplementary Table S11). However, 41 *NLR* genes showed an increase in CG methylation level by at least 5% in the *rdd* mutant, and 23 of them showed an increase up to greater than 10% (Supplementary Table S11). Among these genes, five *NLR* genes, which are *AT5G45230*, *AT4G09430*, *AT4G08450*, *AT1G53350*, and *AT5G05400*, displayed altered methylation in both *ros1* and *rdd* mutants (Supplementary Table S11). Additionally, *AT4G09360* and *AT4G09430* showed decreased CG methylation in the *ros1* mutant; *AT3G04220* and *AT4G19050* showed a decrease in the *rdd* mutant (Supplementary Table S11). Therefore, each DNA demethylase exerts a specific effect on the DNA methylation of the *NLR* genes. Similarly, the methylation levels within 7 of 14 loci in each single mutant were observed to be considerably less than in the *rdd* triple mutant, indicating that all the DML enzymes jointly demethylate these loci, whereas some other loci were found to be demethylated by a single DML (Penterman et al., 2007). Hence, these three

glycosylases function with partial redundancy. In this study, *ROS1* mutation does not cause an increase in DNA methylation at all *NLR* genes, even hypomethylation at some *NLR* genes can be observed, also suggesting that *DML2* and/or *DML3* are able to compensate for *ROS1* loss at some targets. It has been reported that the DNA methylation patterns of many *Arabidopsis* *NLR* genes are regulated by different DNA methyltransferases (Kong et al., 2018). Taken together, these results indicate that the methylation patterns of many *NLR* genes in *Arabidopsis* are regulated not only by DNA methyltransferases but also by DNA demethylases.

DNA Demethylases Mediate the Transcriptional Activities of *NLR* Genes in *Arabidopsis thaliana*

It has been revealed that the DNA methylation levels of some genes in *Arabidopsis* are closely related to their transcriptional

activities (Penterman et al., 2007). The mutation of ROS1 leads to increased DNA methylation and decreased expression in some *Arabidopsis* genomic loci (Zhu et al., 2007). Another study has shown that *Arabidopsis* DNA demethylases, including ROS1, DML2, and DML3, are able to modulate the transcriptional activity of many stress response genes, and these stress response genes are repressed in the *rdd* mutant (Le et al., 2014). In this study, we show that the transcriptional levels of some NLR genes are higher in different mutants defective in DNA demethylase than in the wild-type controls, whereas the levels of other NLR genes are lower in diverse DNA demethylase-defective mutants than in their wild-type controls.

We found that 28 NLR genes were upregulated in *ros1* but downregulated in *rdd* mutants in comparison to the wild-type controls, three NLR genes were upregulated in both *ros1* and *rdd* mutants, and one NLR gene (*AT1G62630*) was downregulated in both *ros1* and *rdd* mutants (Supplementary Table S12). We also observed that nine NLR genes were upregulated and two NLR genes (*AT1G10920* and *AT1G63750*) were downregulated only in *ros1* but *rdd* mutants (Supplementary Table S12). In addition, we discovered that 32 NLR genes were repressed in the *rdd* mutant (Supplementary Table S12). The *rdd* mutant was shown to exhibit increased susceptibility to *F. oxysporum*, and the transcriptional activities of *AT1G58602* and *AT4G09420* were detected to be downregulated (Le et al., 2014). Thus, the three demethylases may play partially redundant roles, and DML2 and/or DML3 can partially compensate some NLR genes for the loss of function of ROS1. On the other hand, the transcriptional activities of many NLR genes in *Arabidopsis* are mediated by different DNA demethylases, and the transcriptional activity varied among different NLR genes when the DNA demethylases were mutated.

Our qRT-PCR results further confirmed that some transcripts encoded by *Arabidopsis* NLR genes were increased or decreased at the transcriptional level in the mutants defective in DNA demethylases. Therefore, it is important and meaningful to reveal the mechanisms by which DNA demethylases modulate the expression of *Arabidopsis* NLR genes.

Relationships Between Methylation and Transcription of *Arabidopsis* NLR Genes

It was reported that only 182 genes demonstrated altered methylation (Penterman et al., 2007), and 167 genes presented differential expression (Lister et al., 2008) in the *rdd* mutant compared to wild-type plants. Therefore, changes in DNA methylation or gene expression are limited in the *rdd* mutant compared to wild-type plants. In another study, 348 genes were observed to be differentially expressed (Le et al., 2014). In their studies, the differentially expressed genes seldom overlapped with the differentially methylated genes (Le et al., 2014). We also found little overlap in a few NLR genes. For instance, three NLR genes (*AT1G31540*, *AT5G35450*, and *AT5G44870*) showed increased CG methylation within 500-bp UPRs and elevated transcriptional activity in the *ros1* mutant compared to wild-type plants, whereas *AT1G12280* showed decreased CG methylation and elevated expression

when ROS1 was mutated (Supplementary Table S13). In the *rdd* mutant, 11 NLR genes showed CG hypermethylation within the 500-bp UPRs, nine of which were downregulated compared to wild-type plants, whereas *AT1G12280* and *AT1G61180* were upregulated (Supplementary Table S14), suggesting that a close link exists between CG hypermethylation within UPRs and the expression of these NLR genes in the *rdd* mutant. Interestingly, a similar link occurs between CG hypermethylation within gene transcribed regions and the differential expression of 11 NLR genes in the *rdd* mutant (Supplementary Table S14). Of the 11 NLR genes, with the exception of *AT4G19520*, 10 were downregulated in the *rdd* mutant compared to wild-type plants. It is worth noting that *AT4G33300*, *AT5G36930*, and *AT5G44870* showed increased CG methylation within 500-bp UPRs and GBRs and downregulated expression in the *rdd* mutant compared to wild-type plants (Supplementary Table S14), indicating a negative connection between CG hypermethylation and their downregulated expression. Nevertheless, many NLR genes have no direct link between their changes in methylation status and transcriptional activity. A previous study also suggested the regulation of defense genes by DNA methylation not only based on *cis*-acting modes but also in *trans*, as well as the global influence of DNA demethylation on the activation of the defense-associated transcriptome through primarily *trans*-regulatory mechanisms (Lopez Sanchez et al., 2016).

CONCLUSION

In this study, we show that some *Arabidopsis* NLR genes can be demethylated by ROS1, DML2, and DML3 within their upstream and transcribed regions. We revealed that the loss of functions of the demethylases leads to obvious changes in DNA methylation levels within some members of *Arabidopsis* NLR genes. We found that demethylases have no effects on the DNA methylation status of some *Arabidopsis* NLR genes. We demonstrated that some *Arabidopsis* NLR genes were regulated by the DNA demethylases ROS1, DML2, and/or DML3. This study will provide a reference for future research into the expression of *Arabidopsis* NLR genes.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

WK conceived the project, designed study, interpreted the data, and wrote the manuscript. WK, HL, and AL supervised the study design. HL provided the plant materials and edited the manuscript. XX, L-WL, SZ, and LD conducted the bioinformatic analyses of the DNA methylome, transcriptome and statistical

analyses of the experimental data. QW carried out the qRT-PCR assays. All authors approved the final manuscript.

FUNDING

This study was funded by the National Natural Science Foundation of China (31970582), and the National Natural Science Foundation of China (31771427) to HL. This study

was also funded by the Shandong Provincial Key Laboratory of Agricultural Microbiology Open Fund (SDKL2017015) to WK.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00460/full#supplementary-material>

REFERENCES

- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Bauer, M. J., and Fischer, R. L. (2011). Genome demethylation and imprinting in the endosperm. *Curr. Opin. Plant Biol.* 14, 162–167. doi: 10.1016/j.pbi.2011.02.006
- Bender, J. (2004). DNA methylation and epigenetics. *Annu. Rev. Plant Biol.* 55, 41–68. doi: 10.1146/annurev.arplant.55.031903.141641
- Chan, S. W. L., Henderson, I. R., and Jacobsen, S. E. (2005). Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat. Rev. Genet.* 6, 351–360. doi: 10.1038/nrg1601
- Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J. J., Goldberg, R. B., et al. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell* 110, 33–42. doi: 10.1016/S0092-8674(02)00807-3
- Dangl, J. L., and Jones, J. D. G. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411, 826–833. doi: 10.1038/35081161
- Deleris, A., Halter, T., and Navarro, L. (2016). DNA methylation and demethylation in plant immunity. *Annu. Rev. Phytopathol.* 54, 579–603. doi: 10.1146/annurev-phyto-080615-100308
- Downen, R. H., Pelizzola, M., Schmitz, R. J., Lister, R., Downen, J. M., Nery, J. R., et al. (2012). Widespread dynamic DNA methylation in response to biotic stress. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2183–E2191. doi: 10.1073/pnas.1209329109
- Fei, Q. L., Xia, R., and Meyers, B. C. (2013). Phased secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell* 25, 2400–2415. doi: 10.1105/tpc.113.114652
- Gehring, M., Huh, J. H., Hsieh, T. F., Penterman, J., Choi, Y., Harada, J. J., et al. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* 124, 495–506. doi: 10.1016/j.cell.2005.12.034
- Gong, Z., Morales-Ruiz, T., Ariza, R. R., David, L., and Zhu, J.-K. (2002). ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell* 111, 803–814. doi: 10.1016/S0092-8674(02)01133-9
- Johnson, L. M., Cao, X., and Jacobsen, S. E. (2002). Interplay between two epigenetic marks: DNA Methylation and Histone H3 Lysine 9 methylation. *Curr. Biol.* 12, 1360–1367. doi: 10.1016/S0960-9822(02)00976-4
- Kong, W., Li, B., Wang, B., Wang, B., Duan, X., Ding, L., et al. (2018). Analysis of the DNA methylation patterns and transcriptional regulation of the NB-LRR-encoding gene family in *Arabidopsis thaliana*. *Plant Mol. Biol.* 96, 563–575. doi: 10.1007/s11103-018-0715-z
- Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291. doi: 10.1093/bioinformatics/btm091
- Krueger, F., and Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572. doi: 10.1093/bioinformatics/btr167
- Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220. doi: 10.1038/nrg2719
- Le, T. N., Schumann, U., Smith, N. A., Tiwari, S., Au, P. C., Zhu, Q. H., et al. (2014). DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in *Arabidopsis*. *Genome Biol.* 15:458. doi: 10.1186/s13059-014-0458-3
- Li, F., Pignatta, D., Bendix, C., Brunkard, J. O., Cohn, M. M., Tung, J., et al. (2012). MicroRNA regulation of plant innate immune receptors. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1790–1795. doi: 10.1073/pnas.1118282109
- Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., et al. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133, 523–536. doi: 10.1016/j.cell.2008.03.029
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Lopez Sanchez, A., Stassen, J. H., Furci, L., Smith, L. M., and Ton, J. (2016). The role of DNA (de)methylation in immune responsiveness of *Arabidopsis*. *Plant J.* 88, 361–374. doi: 10.1111/tpj.13252
- Matzke, M. A., and Mosher, R. A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* 15, 394–408. doi: 10.1038/nrg3683
- McHale, L., Tan, X., Koehl, P., and Michelmore, R. W. (2006). Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 7:212. doi: 10.1186/gb-2006-7-4-212
- Meyer, P. (2011). DNA methylation systems and targets in plants. *FEBS Lett.* 585, 2008–2015. doi: 10.1016/j.febslet.2010.08.017
- 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
- Morales-Ruiz, T., Ortega-Galisteo, A. P., Ponferrada-Marin, M. I., Martinez-Macias, M. I., Ariza, R. R., and Roldan-Arjona, T. (2006). DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6853–6858. doi: 10.1073/pnas.0601109103
- Nishimura, M. T., Anderson, R. G., Cherkis, K. A., Law, T. F., Liu, Q. L., Machius, M., et al. (2017). TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 114, E2053–E2062. doi: 10.1073/pnas.1620973114
- Ortega-Galisteo, A. P., Morales-Ruiz, T., Ariza, R. R., and Roldan-Arjona, T. (2008). *Arabidopsis* DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Mol. Biol.* 67, 671–681. doi: 10.1007/s11103-008-9346-0
- Penterman, J., Zilberman, D., Huh, J. H., Ballinger, T., Henikoff, S., and Fischer, R. L. (2007). DNA demethylation in the *Arabidopsis* genome. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6752–6757. doi: 10.1073/pnas.0701861104
- Qian, W., Miki, D., Zhang, H., Liu, Y., Zhang, X., Tang, K., et al. (2012). A histone acetyltransferase regulates active DNA demethylation in *Arabidopsis*. *Science* 336, 1445–1448. doi: 10.1126/science.1219416
- Richard, M. M. S., Gratiot, A., Thareau, V., Kim, K. D., Balzergue, S., Joets, J., et al. (2018). Genomic and epigenomic immunity in common bean: the unusual features of NB-LRR gene family. *DNA Res.* 25, 161–172. doi: 10.1093/dnares/dsx046
- Shivaprasad, P. V., Chen, H. M., Patel, K., Bond, D. M., Santos, B. A., and Baulcombe, D. C. (2012). A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell* 24, 859–874. doi: 10.1105/tpc.111.095380
- Tan, X., Meyers, B. C., Kozik, A., AlWest, M., Morgante, M., Clair, D. A. S., et al. (2007). Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in *Arabidopsis*. *BMC Plant Biol.* 7:56. doi: 10.1186/1471-2229-7-56
- Tang, K., Lang, X., Zhang, H., and Zhu, J. K. (2016). The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. *Nat. Plants* 2:16169. doi: 10.1038/nplants.2016.169

- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., and Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., and van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3 - new capabilities and interfaces. *Nucleic Acids Res.* 40:e115. doi: 10.1093/nar/gks596
- Wang, Z.-X., Yano, M., Yamanouchi, U., Iwamoto, M., Monna, L., Hayasaka, H., et al. (1999). The Pib gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. *Plant J.* 19, 55–64. doi: 10.1046/j.1365-3113X.1999.00498.x
- Yoshimura, S., Yamanouchi, U., Katayose, Y., Toki, S., Wang, Z.-X., Kono, I., et al. (1998). Expression of Xa1, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *Proc. Natl. Acad. Sci. U.S.A.* 95, 1663–1668. doi: 10.1073/pnas.95.4.1663
- Yu, A., Lepere, G., Jay, F., Wang, J., Bapaume, L., Wang, Y., et al. (2013). Dynamics and biological relevance of DNA demethylation in *Arabidopsis* antibacterial defense. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2389–2394. doi: 10.1073/pnas.1211757110
- Zhai, J. X., Jeong, D. H., De Paoli, E., Park, S., Rosen, B. D., Li, Y. P., et al. (2011). MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev.* 25, 2540–2553. doi: 10.1101/gad.177527.111
- Zhang, H., Lang, Z., and Zhu, J. K. (2018). Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* 19, 489–506. doi: 10.1038/s41580-018-0016-z
- Zhang, H., and Zhu, J. K. (2011). RNA-directed DNA methylation. *Curr. Opin. Plant Biol.* 14, 142–147. doi: 10.1016/j.pbi.2011.02.003
- Zhu, J., Kapoor, A., Sridhar, V. V., Agius, F., and Zhu, J. K. (2007). The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in *Arabidopsis*. *Curr. Biol.* 17, 54–59. doi: 10.1016/j.cub.2006.10.059
- Zhu, J. K. (2009). Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* 43, 143–166. doi: 10.1146/annurev-genet-102108-134205
- Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* 39, 61–69. doi: 10.1038/ng1929

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 Kong, Xia, Wang, Liu, Zhang, Ding, Liu and La. 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.



Bulked Segregant RNA-Seq Reveals Distinct Expression Profiling in Chinese Wheat Cultivar Jimai 23 Responding to Powdery Mildew

Tong Zhu^{1†}, Liru Wu^{1†}, Huagang He^{2†}, Jiancheng Song¹, Mengshu Jia¹, Liancheng Liu³, Xiaolu Wang⁴, Ran Han⁴, Liping Niu⁵, Wenxiao Du¹, Xu Zhang¹, Wenrui Wang¹, Xiao Liang¹, Haosheng Li⁴, Jianjun Liu⁴, Hongxing Xu^{6*}, Cheng Liu^{4*} and Pengtao Ma^{1*}

¹ School of Life Sciences, Yantai University, Yantai, China, ² School of Food and Biological Engineering, Jiangsu University, Zhenjiang, China, ³ Beijing Biomimics Technology Company Limited, Beijing, China, ⁴ Crop Research Institute, Shandong Academy of Agricultural Sciences, Jinan, China, ⁵ State Key Laboratory of Hybrid Rice, College of Life Sciences, Wuhan University, Wuhan, China, ⁶ State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, Kaifeng, China

OPEN ACCESS

Edited by:

Zhu-Qing Shao,
Nanjing University, China

Reviewed by:

Zhang Ruiqi,
Nanjing Agricultural University, China
Lihui Li,
Chinese Academy of Agricultural
Sciences, China

*Correspondence:

Hongxing Xu
hongxingxu@vip.henu.edu.cn
Cheng Liu
lch6688407@163.com
Pengtao Ma
ptma@ytu.edu.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: 14 January 2020

Accepted: 16 April 2020

Published: 27 May 2020

Citation:

Zhu T, Wu L, He H, Song J, Jia M,
Liu L, Wang X, Han R, Niu L, Du W,
Zhang X, Wang W, Liang X, Li H,
Liu J, Xu H, Liu C and Ma P (2020)
Bulked Segregant RNA-Seq Reveals
Distinct Expression Profiling
in Chinese Wheat Cultivar Jimai 23
Responding to Powdery Mildew.
Front. Genet. 11:474.
doi: 10.3389/fgene.2020.00474

Wheat powdery mildew, caused by *Blumeria graminis* f. sp. *tritici* (*Bgt*), is one of the most destructive fungal diseases threatening global wheat production. Host resistance is well known to be the most efficient method to control this disease. However, the molecular mechanism of wheat powdery mildew resistance (*Pm*) is still unclear. To analyze the molecular mechanism of *Pm*, we used the resistant wheat cultivar Jimai 23 to investigate its potential resistance components and profiled its expression in response to powdery mildew infection using bulked segregant RNA-Seq (BSR-Seq). We showed that the *Pm* of Jimai 23 was provided by a single dominant gene, tentatively designated *PmJM23*, and assigned it to the documented *Pm2* region of chromosome 5DS. 3,816 consistently different SNPs were called between resistant and susceptible parents and the bulked pools derived from the combinations between the resistant parent Jimai23 and the susceptible parent Tainong18. 58 of the SNPs were assigned to the candidate region of *PmJM23*. Subsequently, 3,803 differentially expressed genes (DEGs) between parents and bulks were analyzed by GO, COG and KEGG pathway enrichment. The temporal expression patterns of associated genes following *Bgt* inoculation were further determined by RT-qPCR. Expression of six disease-related genes was investigated during *Bgt* infection and might serve as valuable genetic resources for the improvement of durable resistance to *Bgt*.

Keywords: wheat, powdery mildew, BSR-Seq, expression profiling, DEG

INTRODUCTION

Bread wheat (*Triticum aestivum* L.) is one of the most important and widely planted crops worldwide. Producing a high and stable yield of wheat, however, is constantly challenged by a series of diseases. Wheat powdery mildew, caused by *Blumeria graminis* f. sp. *tritici* (*Bgt*), is a devastating disease, reaching epidemic levels in maritime or semi-continental climates (Morgounov et al., 2012; Mehta et al., 2018). Infection by powdery mildew will lead to a 10–15% yield reduction in general, but which sometimes can be as high as 62% in severely infected fields (Singh et al., 2016). Apart

from yield reduction, decreases in quality are also recognized and have been commonly reported (Pál et al., 2013). In China, the epidemic area of wheat powdery mildew has been around six million hectares for the last 10 years.¹ To control this disease and prevent epidemics, fungicides are often used, but drug resistance of *Bgt* is increasingly serious due to pathogenic variation (Manoharachary and Kunwar, 2014). Alongside drug resistance, environmental pollution and cost caused by the use of pesticides cannot be ignored (Saharan et al., 2019). Compared to the use of fungicides, the breeding and use of resistant cultivars is considered to be the most effective and environmentally friendly means of preventing disease epidemics (Selter et al., 2014).

To improve the resistance of wheat to powdery mildew, abundant resistance resources and identification of genes are essential. Up to now, 88 formally designated powdery mildew resistance (*Pm*) genes/alleles have been identified at 66 loci (*Pm1*-*Pm66*) (McIntosh et al., 2019; Li et al., 2020a). In addition, more than 30 temporarily designated *Pm* genes have also been reported and assigned to their corresponding wheat chromosomes (Li et al., 2020b).² Most of these genes were derived from common wheat and its relatives (Ma et al., 2018, 2019; McIntosh et al., 2019).

There are two types of resistance to wheat powdery mildew: qualitative and quantitative. Qualitative resistance is common in many of the reported *Pm* genes. These genes follow Mendel's law of segregation. In contrast, several of the resistance genes are quantitatively inherited, including *Pm38* (Spielmeyer et al., 2008), *Pm39* (Lillemo et al., 2008), *Pm46* (Herrera-Foessel et al., 2014), and *Pm54* (Hao et al., 2015). Qualitative resistance is often defeated after extended periods in production whereas quantitative resistance is rarely overcome. The underlying molecular mechanism of disease resistance needs to be clarified to support the rational use of the *Pm* genes. The mechanism for powdery mildew infection has been reported in grapevines (Fung et al., 2008), barley (Eckey et al., 2004), and *Arabidopsis thaliana* (Fauteux et al., 2006) using vitis GeneChip, cDNA-AFLP, and cDNA microarrays, respectively. But wheat powdery mildew is different from that of the plants mentioned above and relatively little is known regarding the molecular mechanism of powdery mildew resistance in wheat. Only individual genes, including NAC (NAM ATAF1/2 CUC2) and MYB (V-myb avian myeloblastosis viral oncogene homolog) transcription factors, have been analyzed and shown to play a role in the resistance process (Zhou et al., 2018; Zheng et al., 2019).

Jimai 23 is an elite wheat cultivar released in the Shandong province of China. It shows high level resistance to powdery mildew over its entire life cycle. In this study we confirmed that a *Pm2* allele confers the powdery mildew resistance in Jimai 23. Although a *Pm2*-related gene has been cloned, the *Pm2* locus has been shown to be a complex locus (Sánchez-Martín et al., 2016; Jin et al., 2018; Ma et al., 2018; Chen et al., 2019), and the resistance mechanism of this locus is even less clear during *Bgt* invasion. In order to dissect the composition of the *Pm2* locus and profile the expression of powdery mildew resistance genes in

Jimai 23 when subjected to *Bgt* invasion we: (1) confirmed the candidate interval of the *Pm* gene by distribution of differential SNPs; (2) identified and classified differentially expressed genes (DEGs) at the whole-genome scale; and (3) selected several key genes mediating powdery mildew resistance in Jimai 23 and profiled their expression following *Bgt* inoculation. To achieve this, we used RNA sequencing (RNA-seq), bulked segregant RNA-seq (BSR-seq) and reverse transcriptase quantitative PCR (RT-qPCR). RNA-seq is an effective and low-cost method to comprehensively assess the gene expression profiles of the *Pm* genes after inoculation by *Bgt* isolates. This is because RNA-seq is not dependent on pre-existing databases of expressed genes and can provide an unbiased view of gene expression profiling (Pearce et al., 2015; Pankiewicz et al., 2016). BSR-seq, which is a combination of RNA-seq and bulked segregant analysis (BSA), is an efficient method for both differential gene expression profiling and rapid gene/QTL mapping (Wang et al., 2017; Hao et al., 2019). Especially for the crop species with complex genomes, BSR-seq can also break through the adverse effects of the genome sequences and help to obtain relatively advantageous dissection (Liu et al., 2012; Trick et al., 2012; Wang et al., 2013, 2017, 2018; Ramirez-Gonzalez et al., 2015; Zhang et al., 2017).

MATERIALS AND METHODS

Plant Materials

The wheat cultivar Jimai 23, bred by the Crop Research Institute, Shandong Academy of Agricultural Sciences, was used as the resistant parent against powdery mildew in this study. The susceptible wheat cultivar Tainong 18 was crossed with Jimai 23 to produce F_1 , F_2 , and $F_{2:3}$ populations for genetic analysis and BSR-Seq analysis. Wheat cultivar Huixianhong, which we have previously shown to be susceptible to a range of *Bgt* isolates (Ma et al., 2018), was used as the susceptible control in phenotypic assessment experiments.

Preparation of Samples for BSR-Seq

Jimai 23, Tainong 18, together with their derived F_1 , F_2 , and $F_{2:3}$ progeny, were inoculated with *Bgt* isolate YT01 for phenotypic assessment. At the seedling stage, five seeds were sown in each cell using 128-cell rectangular trays in a growth chamber which was set at 20°C with a daily photoperiod of 14 h. For Jimai 23, Tainong 18 and their derived F_1 hybrids, ten seeds of each one were sown in six cells. For the F_2 population and $F_{2:3}$ families, 300 seeds and 200 families (20 seeds for each $F_{2:3}$ family) were sown for genetic analysis and preparation of the samples for BSR-Seq. The susceptible control, Huixianhong, was planted randomly in each tray. When the test seedlings had grown to the one leaf stage, they were inoculated with fresh conidiospores previously increased on Huixianhong seedlings, and immediately incubated in a dark and 100% humidity space at 18°C for 24 h, and then the growth chamber was set at 20°C with a daily photoperiod of 14 h. Over the next 2 days, the inoculation was conducted twice before dark. When the spores were fully developed on the first leaf of the susceptible control Huixianhong, which was at about 10–14 days

¹<https://www.natesc.org.cn/>

²<https://shigen.nig.ac.jp/wheat/komugi/genes/symbolClassList.jsp>

after inoculation, infection types (ITs) were scored using the 0–4 scale described by An et al. (2013), in which ITs 0, 0, 1, and 2 are regarded as resistant, and ITs 3 and 4 as susceptible. Three parallel experiments were conducted using the same procedure to confirm the phenotypic data.

When the spores were fully developed on the first leaf of susceptible control Huixianhong, Total RNA of Jimai 23, Tainong 18, and 40 homozygous resistant and susceptible $F_{2:3}$ families was isolated from the first leaf tissues using the Spectrum Plant Total RNA kit (Sigma-Aldrich) following the manufacturer's protocol. Resistant and susceptible RNA bulks were constructed by separately mixing equal amounts of mRNA from the 40 homozygous resistant and susceptible $F_{2:3}$ families, respectively.

Library Construction and RNA Sequencing

The RNA integrity of Jimai 23, Tainong 18, and the resistant and susceptible bulks was evaluated using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States). Samples with an RNA Integrity Number (RIN) ≥ 7 were regarded as meeting the sequencing standard and cDNA libraries were constructed using TruSeq Stranded mRNA LTSample Prep Kit (Illumina, San Diego, CA, United States) according to the manufacturer's instructions. The quality of the cDNA libraries was again assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States), and the acceptable cDNA libraries were sequenced on the Illumina HiSeq sequencing platform (Illumina HiSeq4000) by Beijing Biomix Technology Company Limited (Beijing, China). The sequencing indicator was set as 10 Gb clean data for parents and 20 Gb clean data for the two bulks. After the sequencing of the cDNA libraries, raw data was filtered, and joint sequences and poor-quality reads were eliminated to obtain high-quality clean data. The clean data were then assembled using the reference genome of Chinese Spring (The International Wheat Genome Sequencing Consortium, 2018) (v1.0) for subsequent SNP calling, differential gene expression, and GO and KEGG pathway analyses in Cloud Platform developed by Beijing Biomix Technology Company Limited.

SNP Calling and BSR Association Mapping

The reads of Jimai 23, Tainong 18, resistant and susceptible bulks were aligned with the Chinese Spring reference genome (v1.0) using the STAR (v2.3.0e) software. SNP calling was done following the reference flowchart aimed at RNA-Seq by the software GATK (v3.1-1). SNP index values in the two bulks was calculated using MutMap method (Abe et al., 2012) with SNPs in the susceptible parent as a reference. Subsequently, a Δ SNP index between resistant and susceptible parents and bulks for each SNP was calculated (Takagi et al., 2013) using the following formula:

$$\Delta\text{SNP_index} = (\text{SNP_index of resistance parent/bulk}) - (\text{SNP_index of susceptible parent/bulk}).$$

The average value of Δ SNP index in each window was calculated by sliding the window using a 5-Mb size as a step. The threshold for SNP screening was set as a test of 100,000 permutations in coupling with 99% confidence

(Takagi et al., 2013). Candidate regions with higher confidence (99%) and SNPs with larger than the threshold Δ SNP index value (set as 0.75) in candidate regions were considered to be candidate loci related to powdery mildew resistance.

DEGs Analysis

After clean reads were mapped to the reference genome, the expression level was calculated using FPKM (Fragments per kilo bases of exon per million fragments mapped) (Garber et al., 2011). DEGs were detected using Fold Change ≥ 2 and FDR (False discovery rate) < 0.01 as standard by the software EBSec. Statistical significance of DEGs was determined using a combination of multiple tests and false discovery rate (FDR) (Reiner et al., 2003). Statistics and clustering analysis of DEGs between parents and bulks were done to present the expression pattern genome-wide, including the candidate interval.

Functional Annotation and Enrichment Analysis of the DEGs

Functional annotation of the DEGs was performed using the IWGSC (International Wheat Genome Sequencing Consortium) database (v1.0). GO, COG, and KEGG pathway enrichment analyses were performed using an R package for DEGs (Yu et al., 2012). For GO analysis, GO Term Finder was used to describe the biological functions of a gene expression product (Boyle et al., 2004). For COG analysis, the Unigene sequences were aligned to the COG database to predict possible functions, and to determine the gene function distribution characteristics.³ For KEGG pathway analysis, the KEGG database was used to blast against the metabolic pathway.⁴

Sample Preparation and RT-qPCR

RT-qPCR was performed to profile the expression of the DEGs in the targeted interval that may be related to the powdery mildew resistance in Jimai 23. In addition, several genes that were not differentially expressed, but which are potentially related to pathogen invasion, were selected for further investigation because the DEG analysis only targeted one time point after inoculation. Seedlings of Jimai 23 and Tainong 18 were inoculated with the *Bgt* isolate YT01 at the one-leaf stage. The first leaf of Jimai 23 and Tainong 18 seedlings was sampled 3, 6, 12, 24, 36, 48, and 72 h after inoculation. Three parallel experiments were set up at this stage. The leaves were immediately frozen in liquid nitrogen and ground to a fine powder in a pestle and mortar. RNA was extracted using the Spectrum Plant Total RNA kit (Sigma-Aldrich) following the manufacturer's recommendations and quantified by measuring absorbance at the wavelengths of 260 and 280 nm using a Nano Drop 1000 spectrophotometer (Thermo Scientific). Finally, the RNA was DNase treated with Promega DNase I prior to cDNA synthesis.

One microgram of total RNA was used for cDNA synthesis using Invitrogen SuperScript-II reverse transcriptase following the manufacturer's guidelines. RT-qPCR analysis was performed as described previously (He et al., 2016, 2018), using SYBR

³<http://www.ncbi.nlm.nih.gov/COG>

⁴<https://www.kegg.jp/kegg/pathway.html>

green master mix (Applied Biosystems) with a Rotor-Gene-Q (Qiagen). Amplification was followed by melt curve analysis. The $2^{-\Delta\Delta C_t}$ method was used for relative quantification (He et al., 2018). To detect transcript levels, primers for specific genes were designed based on the coding sequences of the selected genes (Supplementary Table S1). Oligonucleotides amplifying *ACTIN* were used for normalization.

RESULTS

Powdery Mildew Resistance Evaluation and Genetic Analysis

When inoculated with *Bgt* isolate YT01, Jimai 23 showed no visible symptoms on the first leaf (IT 0), while Tainong 18

showed abundant sporulation with more than 80% of the leaf area covered with aerial hyphae. It was scored IT 4, susceptible. All the nine F_1 plants of Jimai 23 \times Tainong 18 were germinated and all showed resistance to YT01 with IT scores of 0. The 271 generated F_2 plants segregated into 198 resistant and 73 susceptible ones, which is consistent with the theoretical ratio for monogenic segregation ($\chi^2 = 0.54$; $P = 0.46$). Two hundred $F_{2:3}$ families further confirmed the ratio for monogenic segregation, with segregation ratio of 47 homozygous resistant (RR): 103 segregating (Rr): 50 homozygous susceptible (rr) families ($\chi^2 = 0.27$; $P = 0.87$) after YT01 infection. We therefore concluded that the seedling-resistance to *Bgt* isolate YT01 in Jimai 23 is controlled by a single dominant mendelian factor from the level of genetic analysis, tentatively designated *PmJM23*. Subsequently, mRNA from 40 randomly selected homozygous

TABLE 1 | Data summary of RNA-seq for Jimai 23, Tainong 18 and their derived resistant and susceptible bulks obtained from 40 homozygous resistant and 40 homozygous susceptible $F_{2:3}$.

Samples	Total base pairs	Clean reads	GC content of clean reads (%)	Clean reads Q30 (%)	Genome map rate (%)
Jimai 23	13,244,236,200	44,147,454	56.44%	94.50%	89.79%
Tainong 18	12,632,022,300	42,106,741	50.89%	94.75%	56.42%
Resistance bulk	53,859,298,800	179,530,996	51.83%	95.67%	88.57%
Susceptible bulk	37,628,667,600	125,428,892	48.92%	94.58%	50.05%

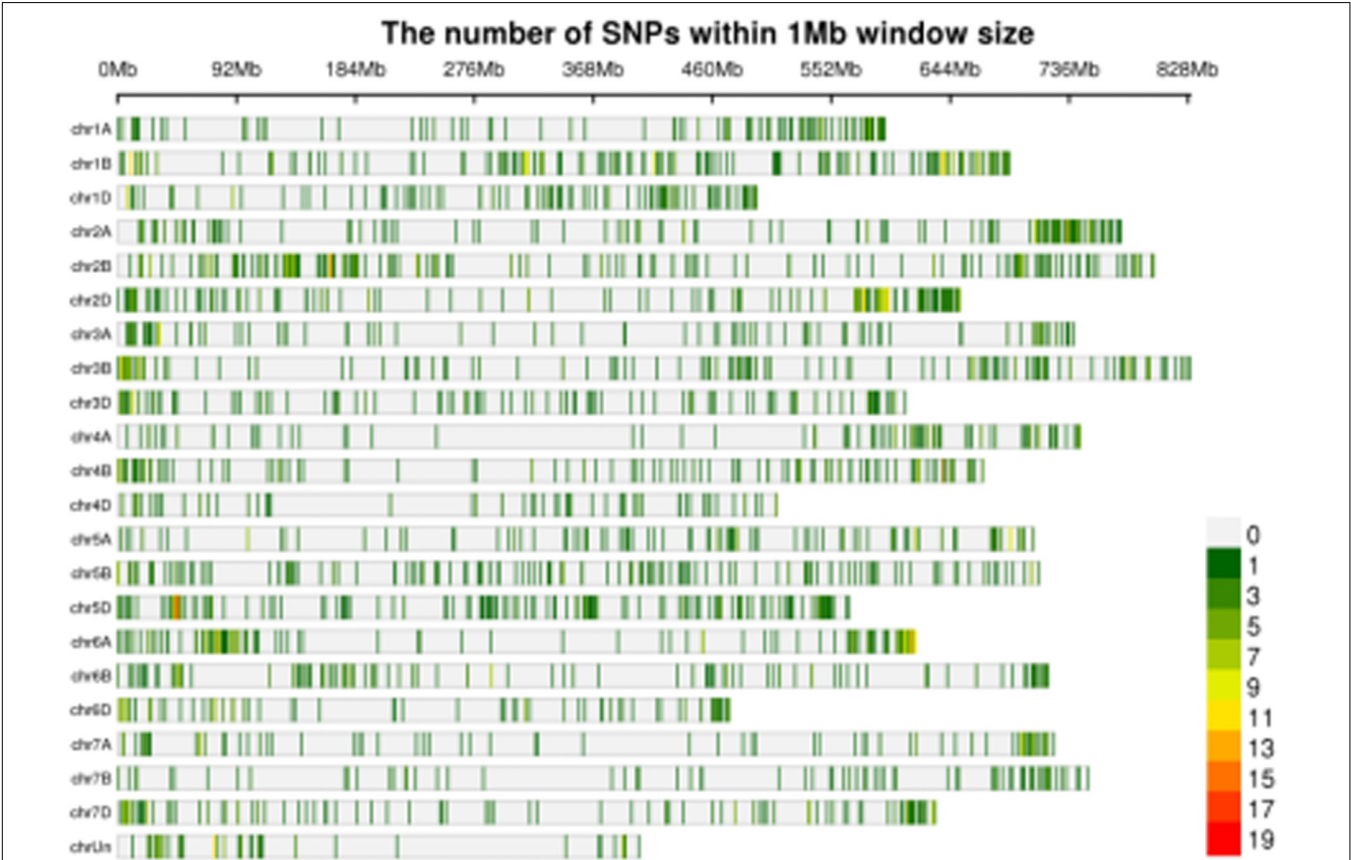
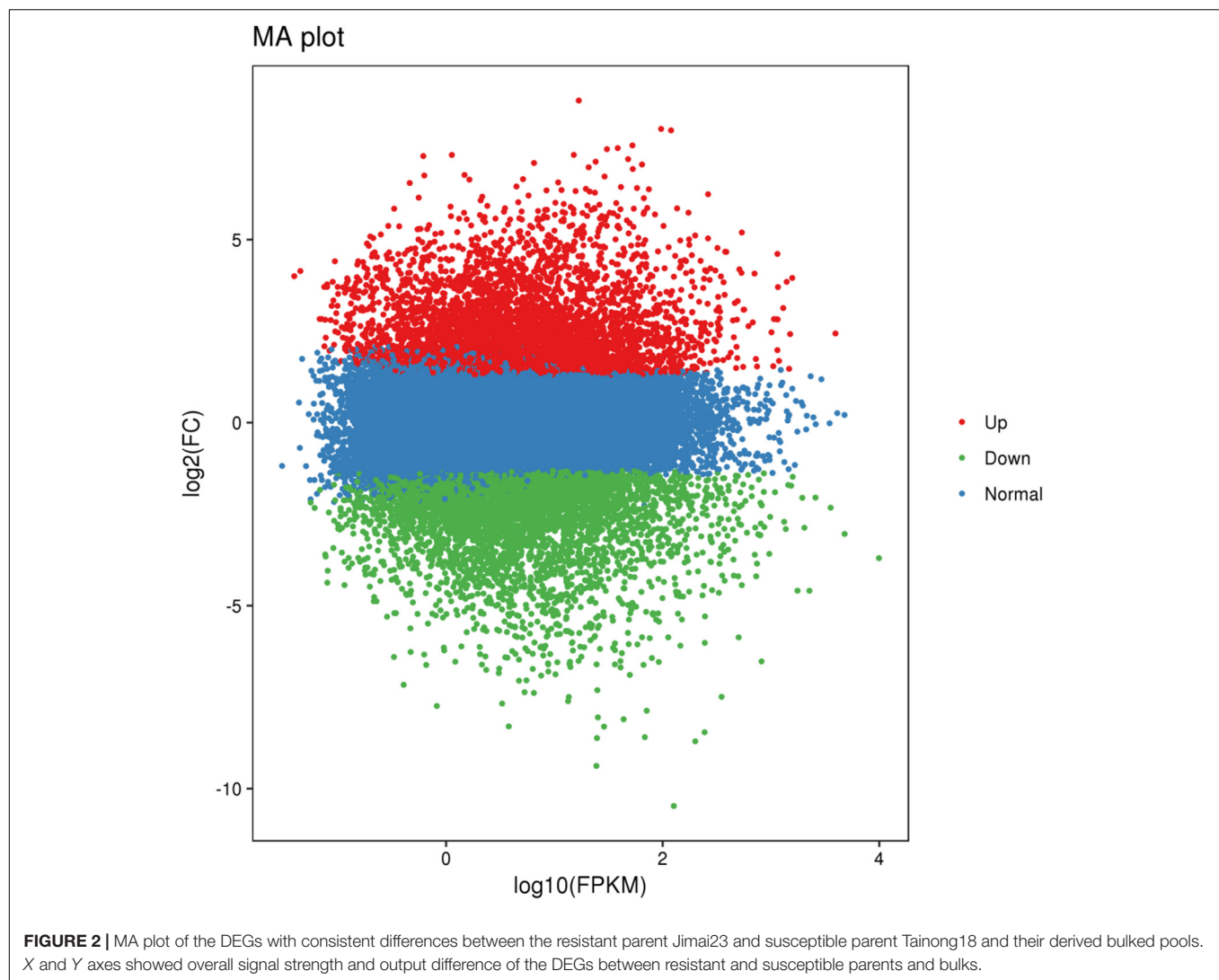


FIGURE 1 | Distribution of the SNPs with consistent differences between the resistant parent Jimai23 and susceptible parent Tainong18 and their derived bulked pools on 21 chromosomes based on Δ SNP index value. Density indicating meter is shown as the color scale in the bottom right corner.



resistant and susceptible $F_{2:3}$ families was isolated and pooled to make the resistant and susceptible bulks for subsequent RNA-Seq.

Clean Data, Quality Control, and Sequence Alignments

After filtering low-quality reads and adaptors, there was more than 12 Gb clean data for each of the parents, Jimai 23 and Tainong 18, and more than 37 Gb clean data for each of the resistant and susceptible bulks. The data size greatly exceeded the transcript size of the wheat genome and it was, therefore, assumed to cover most transcribed genes within the wheat genome. The percentage of clean reads with a Q30 was greater than 94% for all four of the samples, and the GC content ranged from 48.92 to 56.44%. After aligning the four sets of clean reads to the reference genome (IWGSC v1.0) individually, the percentage of reads mapping to the reference genome ranged from 50.05 to 89.79% (Table 1). In summary, the sequencing quality of the four samples was high and suitable for subsequent analysis.

SNP Calling and Confirmation of Candidate Interval

In order to confirm the candidate region within the wheat genome that was responding to powdery mildew infection, differential SNPs from the transcriptome data were screened. A total of 63,875 (Jimai 23), 237,919 (Tainong 18), 64,093 (Resistance bulk), and 77,438 (Susceptible bulk) SNPs were detected from the respective clean data. The three steps for SNP filtering were then carried out. Firstly, SNPs with support degree less than three were filtered; secondly, SNPs that were inconsistent between parent and corresponding bulk were filtered; thirdly, SNPs that were consistent between resistant and susceptible bulks were filtered. Finally, 3,816 SNPs with consistent differences between resistant and susceptible parents and bulks were obtained for subsequent Δ SNP index analysis (Figure 1 and Supplementary Table S2).

Using 99% confidence as the threshold, only one putative candidate region near the end of chromosome arm 5DS was identified (Figure 1). Multiple *Pm2* alleles have been reported in

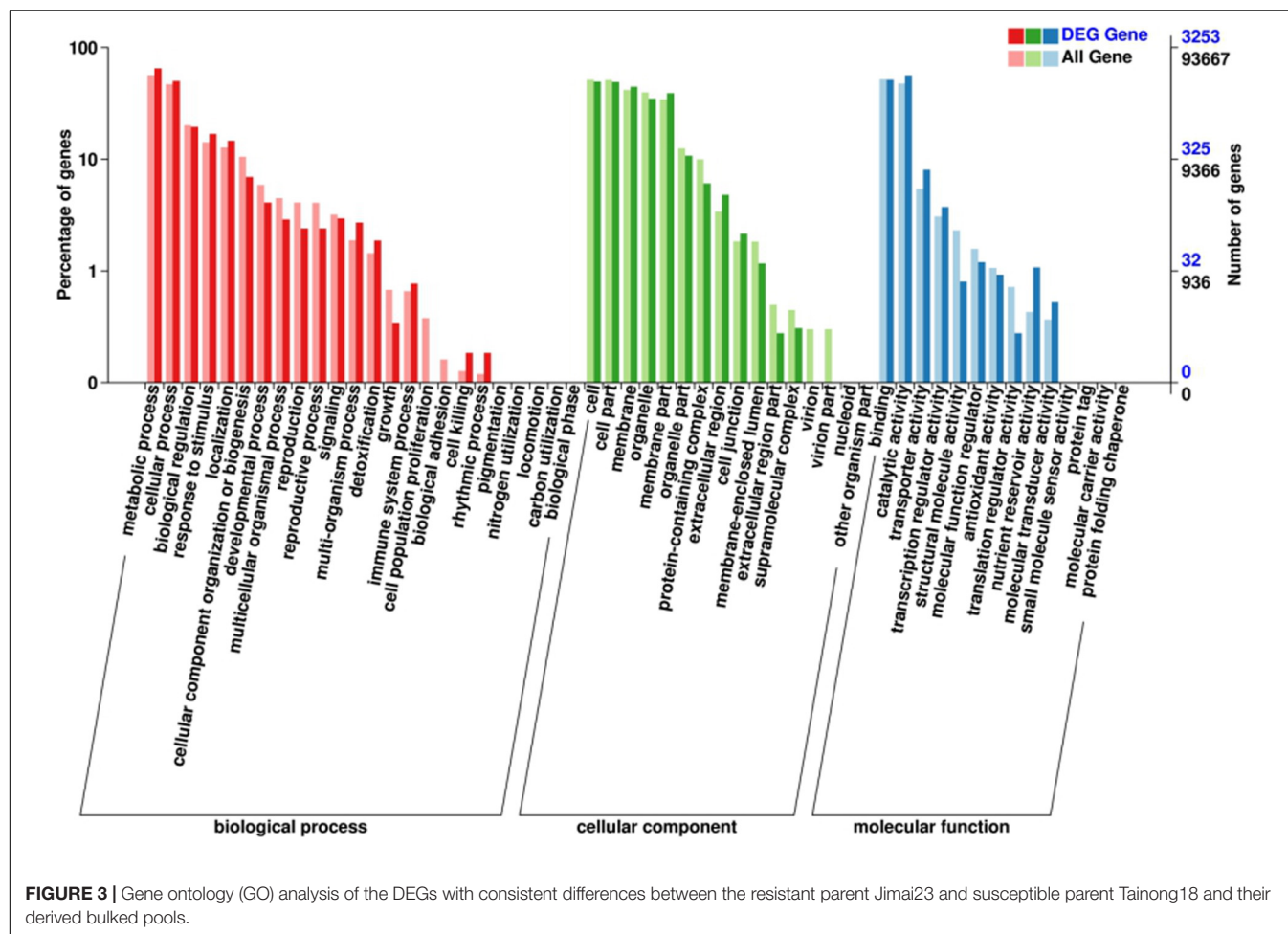


FIGURE 3 | Gene ontology (GO) analysis of the DEGs with consistent differences between the resistant parent Jimai23 and susceptible parent Tainong18 and their derived bulked pools.

this region (Ma et al., 2018). To confirm this locus, *Cfd81*, the universal detection marker for different *Pm2* alleles, was used to genotype $F_{2:3}$ families of Jimai 23 \times Tainong 18. The result showed that *Cfd81* co-segregated with the *Pm* gene in Jimai 23, indicating that the *Pm* in Jimai 23 was also likely controlled by a *Pm2* allele. In this interval, more than 50 SNPs with consistent differences between resistance and susceptible parents and bulks revealed high confidence. These SNPs were used as the reference for the DEG analysis (Supplementary Table S2).

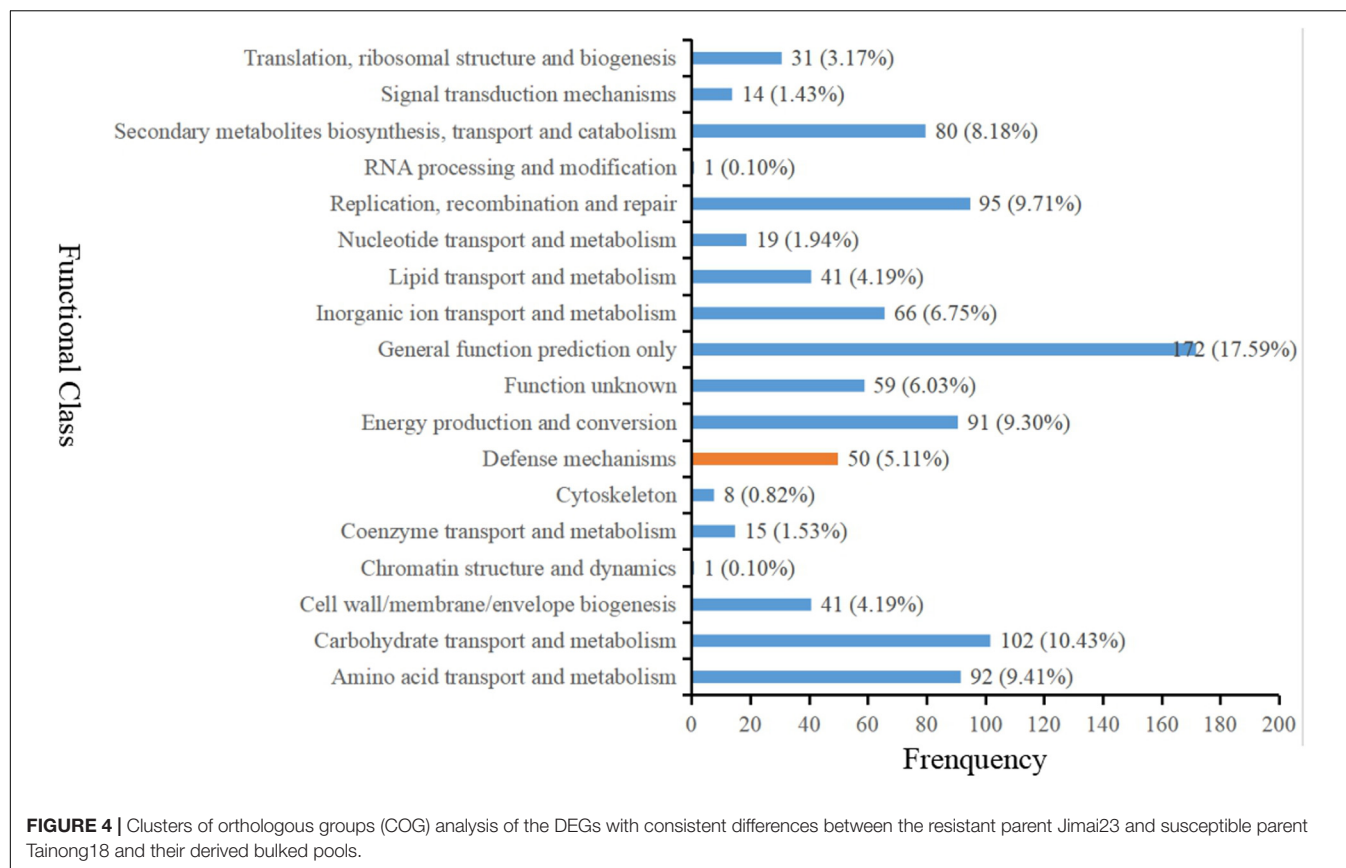
Discovery and Classification of DEGs

Using BSR-Seq, a total of 124,200 genes was identified from the parents and bulks. Among them, 12,361 DEGs were detected between Jimai 23 and Tainong 18, of which, 5,822 DEGs were down-regulated and 6,539 DEGs were up-regulated, using the expression index of Tainong 18 as a standard. Furthermore, 9,162 DEGs were detected between the resistance and susceptible bulks, of which 4,606 and 4,496 DEGs were downregulated and up-regulated, respectively. For further screening, 3,803 DEGs showed consistent expression difference between parents and bulks (Figure 2 and Supplementary Table S3). Combined with the candidate interval analysis, only 16 DEGs were located in this interval (Supplementary Table S1). These genes were considered

to be prime candidates in the resistance response of Jimai 23 to powdery mildew.

GO, COG, and KEGG Pathway Significant Enrichment Analyses of DEGs

Gene ontology (GO) analysis was firstly performed on the DEGs that showed consistent expression difference between parents and bulks via differential expression analysis. These DEGs were mainly involved in biological processes. These included: metabolic processes, cellular processes, biological regulation and response to stimulus; cell components including cell, cell part, membrane and organelle; and molecular functions comprising of binding catalytic and activity (Figure 3). However, the results of the GO analysis describe only the main processes exhibited after *Bgt* infection. Although the 'response to stimuli' process was significantly enriched and may directly participate in disease defense, no DEGs known to relate to defense mechanism(s) were detected. Therefore, clusters of orthologous groups (COG) analysis was performed using the same EDGs above. The data showed that the DEGs were mainly involved in transport and metabolic processes, such as amino acid and carbohydrate transport and metabolism, and DNA duplication, transcription, recombination and repair (Figure 4). However, a few DEGs



were directly involved in plant defense, but accounting for only 1.44% of the DEGs. These results indicated that more genes related to biological metabolism and synthesis were activated to participate in biological defense rather than defense-related genes themselves. In other words, activation of defense mechanisms needs the support of biosynthesis and metabolism.

To further investigate the signal transduction pathway(s) that the DEGs may be involved in, significance enrichment analysis for KEGG pathway was performed on the DEGs that showed consistent expression differences between parents and bulks in the differential expression analysis. Hundred and three significantly enriched ($Q \leq 0.05$) pathways involving 50 categories in cellular processes, environmental processing, genetic information processing, metabolism and organismal system were found (Figure 5). Among them, one plant-pathogen interaction pathway was enriched, and 14 DEGs were present in this pathway. These genes are a resource for further molecular studies into the plant response to powdery mildew (Figure 6).

RT-qPCR Verification for the Disease-Resistance Related Genes in Jimai 23

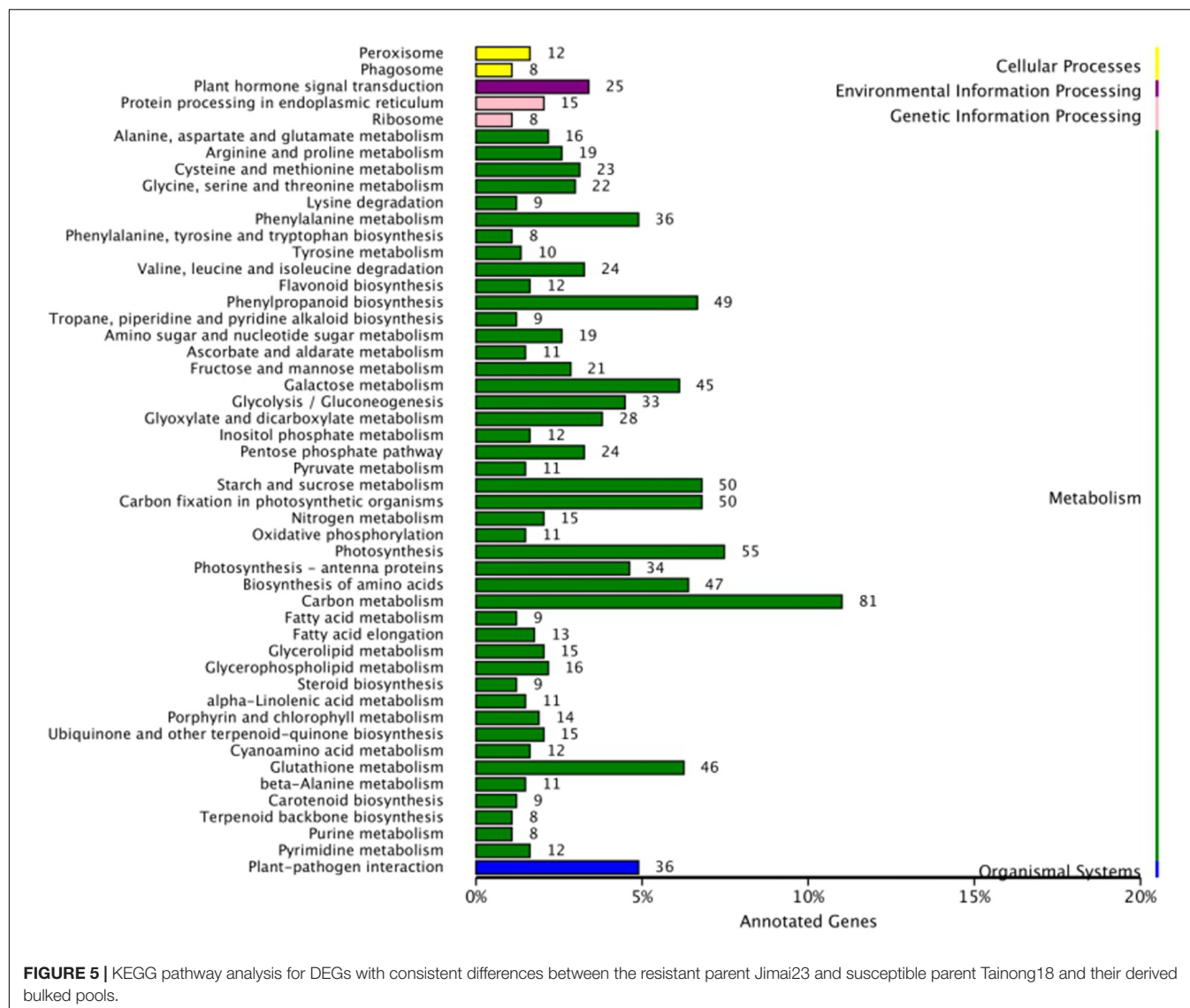
To profile the expression of the disease resistance-related genes in Jimai 23, we monitored the transcriptional level of 21 potential target genes (including 16 DEGs in the candidate interval) (Supplementary Table S1) at different stages after inoculation

with *Bgt* isolate YT01. Six of the target genes showed significant differences between the resistant Jimai 23 and the susceptible Tainong 18 in the time course analysis following *Bgt* inoculation.

The transcriptional levels of two genes [*TraesCS5D01G018000* encoding an early-responsive to dehydration (ERD) stress family protein and *TraesCS5D01G117600* encoding a 70 kDa heat shock protein (HSP70)] were rapidly up-regulated in Jimai 23 but not in Tainong 18 within 0–6 h after inoculation (Figures 7A,B). Their expression was highly induced only in Tainong 18 after 36 h (Figures 7A,B). *TraesCS5D01G104700* and *TraesCS5D01G105200* encode a reticulocyte-binding 2-a-like protein and a kinesin-related protein, respectively. The transcriptional levels of these two genes were initially both up-regulated in Jimai 23 and Tainong 18, but then elevated expression was maintained only in Tainong 18 (Figures 7C,D). *TraesCS5D01G099200* (encoding a S-adenosyl homocysteine deaminase-like protein) and *TraesCS5D01G111400* (encoding a dipeptidyl peptidase protein) were only up-regulated in Tainong 18 during infection.

DISCUSSION

Jimai 23 is an elite wheat cultivar released in the Shandong Province of China that is highly resistant to powdery mildew. In this study, through BSR-Seq combined genetic analysis, a dominant allele of *Pm2* was confirmed to control the resistance

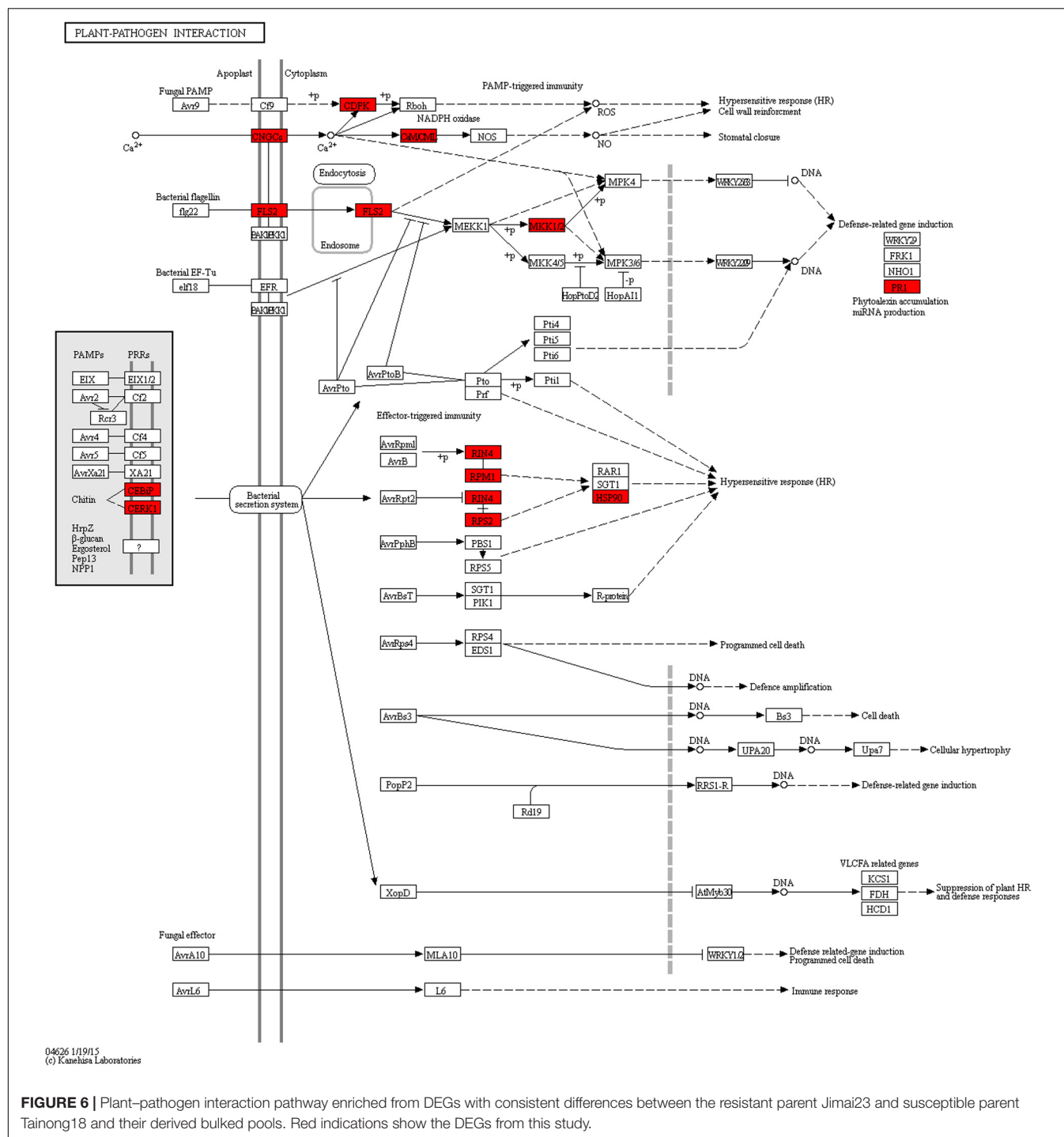


in Jimai 23. In our previous reports, a series of *Pm2* alleles with different reaction patterns to *Bgt* isolates have been identified in various wheat genotypes (Ma et al., 2014, 2015a,b,c, 2016, 2018; Xu et al., 2015; Jin et al., 2018). However, these *Pm2* alleles were only characterized at the genetic level. Even though a *Pm2*-related gene was cloned using mutant chromosome sequencing and candidate gene analysis using the reference genome of Chinese Spring (Sánchez-Martín et al., 2016; Chen et al., 2019), there are still aspects of this locus that require further investigation. For instance, while the homologous sequences of different *Pm2* alleles are identical to each other, these alleles have significantly different reaction patterns to *Bgt* isolates with different virulence spectra, something which cannot be explained by the background difference of the wheat genotypes (Jin et al., 2018; Ma et al., 2018).

Meanwhile, we determined the expression levels of 21 potential target genes in the *Pm2* candidate interval, but only six genes were differently expressed between Jimai 23 and Tainong 18. This further suggest that the *Pm2* locus in Jimai 23 is most

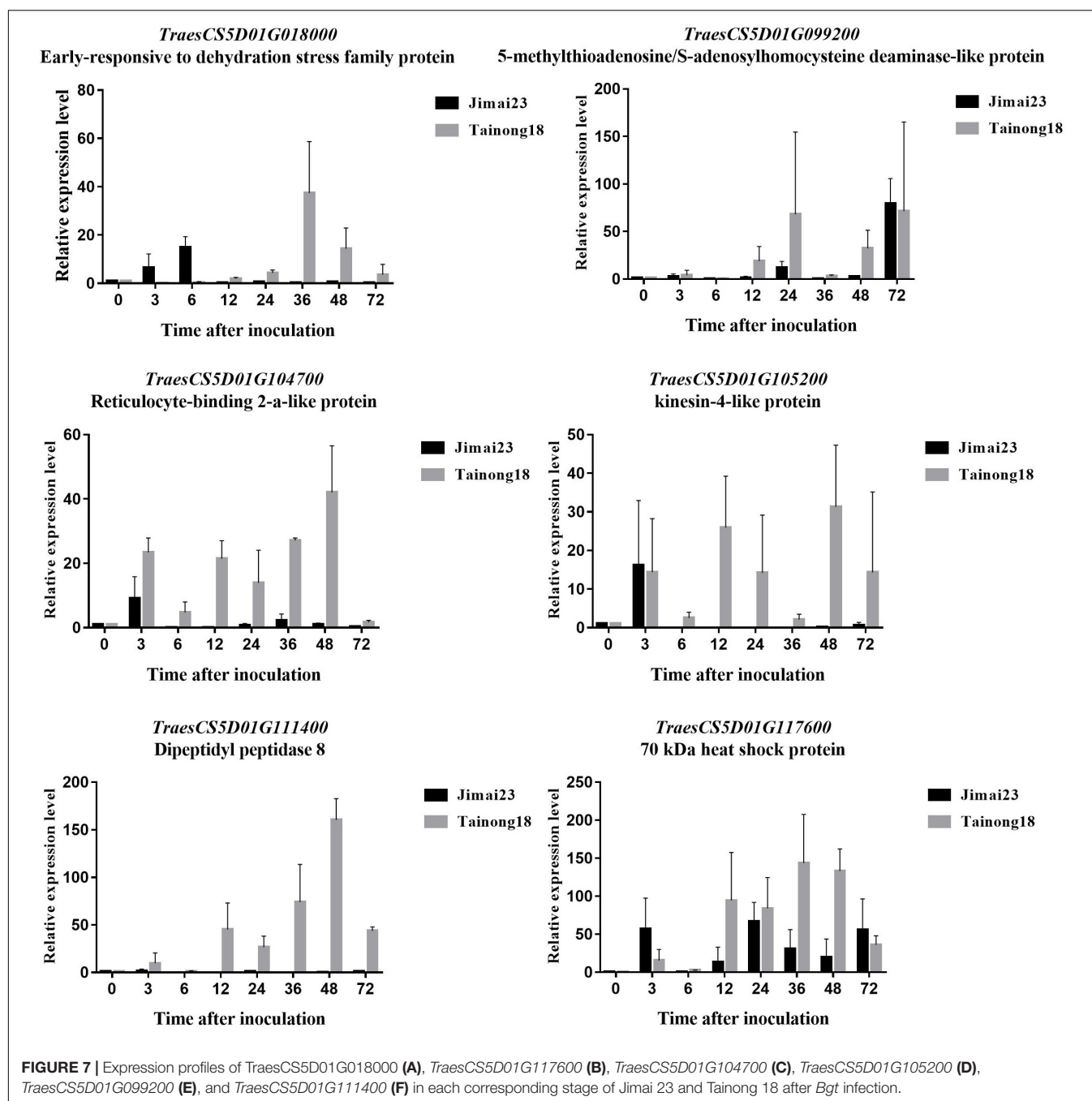
likely a more complex and larger interval compared to that shown within the reference genome of common wheat Chinese Spring, and that the powdery mildew resistance is most likely conferred by multiple genes from the level of gene analysis. So, in-depth study on the composition of the *Pm2* locus and dissection of its molecular mechanism is imperative. Surrounding the problems above, an expression profiling dissection was conducted on Jimai 23 using BSR-Seq in the present study, which is a high efficiency and low-cost means to investigate the overall expression profile of resistance-related genes (Hao et al., 2019). A large number of DEGs, including the target genes in the candidate interval which are important for defense against *Bgt* invasion, were identified and analyzed using GO, COG, and KEGG enrichment.

Plant resistance is a complex process in the course of host-pathogen interaction (Hikichi, 2016). From the host's perspective, a mass of genes will be activated in response to the intrusion of external pathogens. Entry of the pathogen could be prevented at different layers, such as cell wall, plasma membrane and various



enzymes in cytoplasm (Braeken et al., 2008; Soto et al., 2011; Siddiqui et al., 2012). However, there are relatively few studies of the regulatory mechanism within wheat in response to powdery mildew, and, so far, only the functional mechanism of individual genes has been analyzed during powdery mildew infection, including *MYB* and *NAC* transcription factors (Zhou et al., 2018; Zheng et al., 2019). From the BSR-Seq analysis, we selected 21 target genes for consideration at different stages of pathogen

invasion. These genes included structural protein, translocator, regulatory protein and stress response protein (**Supplementary Table S1**), indicating an overall response model from structure change, biosynthesis and transport, and direct stress response after *Bgt* invasion. It should also be pointed out that the sampling for BSR-Seq was only from the stage at which disease symptoms were visible, and did not represent the early stages of infection, so these genes were selected not solely based on differential



expression but also on the functional analysis of the genes in the candidate interval. We then profiled the expression of the six of the 21 selected genes following pathogen inoculation to analyze their response to powdery mildew in resistant versus susceptible cultivars.

TraesCS5D01G018000 encoded an ERD protein that is related to plant adaptation to stress conditions. The ERD families have been reported to provide enhanced drought and salt tolerance and respond to abscisic acid treatment in *Arabidopsis*, sugarcane and maize (Liu et al., 2009; Rai et al., 2016; Devi et al., 2019), suggesting the key role of the ERD protein families is in

plant stress tolerance or resistance. The *TraesCS5D01G117600*-encoded HSP70 is one of the members of the HSP70 family, which are well known as stress responsive molecular chaperones involved in the correct folding of newly synthesized proteins. Although they were first identified in the heat stress response, the heat shock proteins have also been reported to play key roles in innate immunity responses, and to be essential for the functioning of other resistance proteins (Neckers and Tatu, 2008). For wheat, preliminary proteomic analysis also revealed that HSP70 may be involved in regulation of resistance to powdery mildew (Mandal et al., 2014), which aligns with our findings.

These two genes were rapidly up-regulated in Jimai 23 at the early stage following *Bgt* inoculation, whereas in Tainong 18, they were induced only after 36h. We suggest that early defense activation of the two genes in Jimai 23 may be key to its resistance to powdery mildew. In Tainong 18, *Bgt* may have broken through any early defense barriers and the elevated expression at the later stage in Tainong 18 was no longer effective, leading to the sensitivity of Tainong 18 to powdery mildew.

In addition to the key defensive proteins mentioned above, we also selected two genes (*TraesCS5D01G104700* and *TraesCS5D01G105200*) encoding a structural protein and kinesin, respectively, for investigating cell structure changes after *Bgt* infection. The expression of these two genes were both up-regulated in Jimai 23 and Tainong 18 rapidly after *Bgt* inoculation, but remained elevated only in susceptible Tainong 18, suggesting that the attack of *Bgt* was blocked in Jimai 23 whereas it was aggravated in Tainong 18 within 6 h after inoculation, which is consistent with the expression of *TraesCS5D01G018000* and *TraesCS5D01G117600*. One reasonable explanation is that these transcriptional changes may lead structural change of the cell and be a consequence of the resistance, but not its cause. Changes in other structural proteins was also reported in wheat line N0308 after infection by *Bgt* using proteomic analysis (Mandal et al., 2014).

TraesCS5D01G099200 encoded an S-adenosyl homocysteine deaminase-like protein which is an intracellular oxidation-reduction enzyme (Meisel et al., 1980). *TraesCS5D01G111400* is a dipeptidyl peptidase (Park et al., 2006). These two genes have been reported to be involved in the plant resistance pathway within the cell, with the S-adenosyl homocysteine deaminase-like protein mediating disease resistance through the methionine cycle (Meisel et al., 1980; Park et al., 2006; Mäkinen and De, 2019), and dipeptidyl peptidase serving as a bio-marker of disease (Yazbeck et al., 2018). The transcription of these two genes was only up-regulated in Tainong 18 after *Bgt* infection, suggesting the defense system of susceptible Tainong 18 was vulnerable. We suggest that the activation of these two genes was not observed following *Bgt* inoculation of Jimai 23, as the pathogen did not invade its cells.

CONCLUSION

We investigated the holistic expression profile responding to powdery mildew in the resistant wheat cultivar Jimai 23, and six key genes potentially involved in the resistance process were further investigated following *Bgt* inoculation. Compared with previous reports, where the focus was mainly on the mechanism of a single gene, our study provided a perspective of overall

expression profiling, which can facilitate dissection of resistance pathways and accelerate improvement of durable resistance. The selection of potential key genes in the present study mainly focused on the candidate interval of *PmJM23*. In the future, we will select more genes in other intervals, especially from the plant pathogen interaction pathway, for a deeper dissection of the resistance mechanism.

DATA AVAILABILITY STATEMENT

We have uploaded the sequencing data to NCBI and the BioProject ID is PRJNA625022.

AUTHOR CONTRIBUTIONS

PM, CL, and HX conceived the research. TZ, LW, and MJ performed the experiments. HH, JS, LL, HL, JL, LN, WD, XW, and RH analyzed the data. XZ, WW, and XL performed the phenotypic assessment. TZ and PM wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was financially supported by the Shandong Agricultural Seed Improvement Project (2019LZGC016), the Key Research and Development Program of Yantai City (2019YT06000470), the Taishan Scholars Project (tsqn201812123), the Natural Science Foundation of China (31971874 and 31671771), the Jiangsu Agricultural Science and Technology Innovation Fund (CX(19)2042), and the Priority Academic Program Development of Jiangsu Higher Education Institutions, Jiangsu Education Department.

ACKNOWLEDGMENTS

We are grateful to Prof. Paula Jameson, University of Canterbury, financially supported by “Double Hundred” Plan for Foreign Experts in Shandong Province, China, for constructive comments and English language editing of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00474/full#supplementary-material>

REFERENCES

- Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., and Kanzaki, H. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* 30, 174–178. doi: 10.1038/nbt.2095
- An, D. G., Zheng, Q., Zhou, Y. L., Ma, P. T., Lv, Z. L., Li, L. H., et al. (2013). Molecular cytogenetic characterization of a new wheat-rye 4R chromosome translocation line resistant to powdery mildew. *Chromosome Res.* 21, 419–432. doi: 10.1007/s10577-013-9366-8
- Boyle, E. I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J. M., et al. (2004). GO: termfinder—open source software for accessing gene ontology information and finding significantly enriched gene ontology terms associated with a list of genes. *Bioinformatics* 20, 3710–3715. doi: 10.1093/bioinformatics/bt4456

- Braeken, K., Daniels, R., Ndayizeye, M., Vanderleyden, J., and Michiels, J. (2008). "Quorum sensing in bacteria-plant interactions," in *Molecular Mechanisms of Plant and Microbe Coexistence. Soil Biology*, Vol. 15, eds C. S. Nautiyal, and P. Dion (Berlin: Springer).
- Chen, F., Jia, H. Y., Zhang, X. J., Qiao, L. Y., Li, X., Zheng, J., et al. (2019). Positional cloning of *PmCH1357* reveals the origin and allelic variation of the *Pm2* gene for powdery mildew resistance in wheat. *Crop J.* 7, 771–783. doi: 10.1016/j.cj.2019.08.004
- Devi, K., Prathima, P. T., Gomathi, R., Manimekalai, R., Lakshmi, K., and Selvi, A. (2019). Gene expression profiling in sugarcane genotypes during drought stress and rehydration. *Sugar Tech.* 21, 717–733. doi: 10.1007/s12355-018-0687-y
- Eckey, C., Korell, M., Leib, K., Biedenkopf, D., Jansen, C., Langen, G., et al. (2004). Identification of powdery mildew-induced barley genes by cDNA-AFLP: functional assessment of an early expressed MAP kinase. *Plant Mol. Biol.* 55, 1–15. doi: 10.1007/s11103-004-0275-2
- Fauteux, F., Chain, F., Belzile, F., Menzies, J. G., and Belanger, R. R. (2006). The protective role of silicon in the *Arabidopsis*-powdery mildew pathosystem. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17554–17559. doi: 10.1073/pnas.0606330103
- Fung, R. W., Gonzalo, M., Fekete, C., Kovacs, L. G., He, Y., Marsh, E., et al. (2008). Powdery mildew induces defense-oriented reprogramming of the transcriptome in a susceptible but not in a resistant grapevine. *Plant Physiol.* 2008, 236–249. doi: 10.1104/pp.107.108712
- Garber, M., Grabherr, M. G., Guttman, M., and Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* 8, 469–477. doi: 10.1038/nmeth.1613
- Hao, Y. F., Parks, R., Cowger, C., Chen, Z. B., Wang, Y. Y., Bland, D., et al. (2015). Molecular characterization of a new powdery mildew resistance gene *Pm54* in soft red winter wheat. *Theor. Appl. Genet.* 128, 465–476. doi: 10.1007/s00122-014-2445-1
- Hao, Z. M., Geng, M. M., Hao, Y. R., Zhang, Y., Zhang, L. J., Wen, S. M., et al. (2019). Screening for differential expression of genes for resistance to Sitodiplosis mosellana in bread wheat via BSR-seq analysis. *Theor. Appl. Genet.* 132, 3201–3221. doi: 10.1007/s00122-019-03419-9
- He, H. G., Shanying Zhu, S. Y., Zhao, R. H., Jiang, Z. N., Ji, Y. Y., Ji, J., et al. (2018). *Pm21*, encoding a typical CC-NBS-LRR protein, confers broad-spectrum resistance to wheat powdery mildew disease. *Mol. Plant* 11, 879–882. doi: 10.1016/j.molp.2018.03.004
- He, H. G., Zhu, S. Y., Jiang, Z. N., Ji, Y. Y., Wang, F., Zhao, R. H., et al. (2016). Comparative mapping of powdery mildew resistance gene *Pm21* and functional characterization of resistance-related genes in wheat. *Theor. Appl. Genet.* 129, 819–829. doi: 10.1007/s00122-016-2668-4
- Herrera-Foessel, S. A., Singh, R. P., Lillemo, M., Huerta-Espino, J., Bhavani, S., Singh, S., et al. (2014). *Lr67/Yr46* confers adult plant resistance to stem rust and powdery mildew in wheat. *Theor. Appl. Genet.* 127, 781–789. doi: 10.1007/s00122-013-2256-9
- Hikichi, Y. (2016). Interactions between plant pathogenic bacteria and host plants during the establishment of susceptibility. *J. Gen. Plant Pathol.* 82, 326–331. doi: 10.1007/s10327-016-0680-9
- Jin, Y. L., Xu, H. X., Ma, P. T., Fu, X. Y., Song, L. P., Xu, Y. F., et al. (2018). Characterization of a new *Pm2* allele associated with broad-spectrum powdery mildew resistance in wheat line Subtil. *Sci. Rep.* 8:475. doi: 10.1038/s41598-017-18827-4
- Li, H. H., Dong, Z. J., Ma, C., Xia, Q., Tian, X. B., Sehgal, S., et al. (2020a). A spontaneous wheat-Aegilops longissima translocation carrying *Pm66* confers resistance to powdery mildew. *Theor. Appl. Genet.* 133, 1149–1159. doi: 10.1007/s00122-020-03538-8
- Li, Y. H., Shi, X. H., Hu, J. H., Wu, P. P., Qiu, D., Qu, Y. F., et al. (2020b). Identification of a recessive gene *PmQ* conferring resistance to powdery mildew in wheat landrace Qingxinmai using BSR-Seq analysis. *Plant Dis.* 104, 743–751. doi: 10.1094/PDIS-08-19-1745-RE
- Lillemo, M., Asalf, B., Singh, R. P., Huerta-Espino, J., Chen, X. M., He, Z. H., et al. (2008). The adult plant rust resistance loci *Lr34/Yr18* and *Lr46/Yr29* are important determinants of partial resistance to powdery mildew in bread wheat line Saar. *Theor. Appl. Genet.* 116, 1155–1166. doi: 10.1007/s00122-008-0743-1
- Liu, S., Yeh, C. T., Tang, H. M., Nettleton, D., and Schnable, P. S. (2012). Gene mapping via bulked segregant RNA-seq (BSR-seq). *PLoS One* 7:e36406. doi: 10.1371/journal.pone.0036406
- Liu, Y. H., Li, H. Y., Shi, Y. S., Song, Y. C., Wang, T. Y., and Li, Y. (2009). A maize early responsive to dehydration gene, *ZmERD4*, provides enhanced drought and salt tolerance in *Arabidopsis*. *Plant Mol. Biol. Rep.* 27:542. doi: 10.1007/s11105-009-0119-y
- Ma, P. T., Xu, H. X., Li, L. H., Zhang, H. X., Han, G. H., Xu, Y. F., et al. (2016). Characterization of a new *Pm2* allele conferring powdery mildew resistance in the wheat germplasm line FG-1. *Front. Plant Sci.* 7:546. doi: 10.3389/fpls.2016.00546
- Ma, P. T., Xu, H. X., Luo, Q. L., Qie, Y. M., Zhou, Y. L., Xu, Y. F., et al. (2014). Inheritance and genetic mapping of a gene for seedling resistance to powdery mildew in wheat line X3986-2. *Euphytica* 200, 149–157. doi: 10.1007/s10681
- Ma, P. T., Xu, H. X., Xu, Y. F., Li, L. H., Qie, Y. M., Luo, Q. L., et al. (2015a). Molecular mapping of a new powdery mildew resistance gene *Pm2b* in Chinese breeding line KM2939. *Theor. Appl. Genet.* 128, 613–622. doi: 10.1007/s00122-015-2457-5
- Ma, P. T., Xu, H. X., Zhang, H. X., Li, L. H., Xu, Y. F., Zhang, X. T., et al. (2015b). The gene *PmWf1* is a new member of the complex *Pm2* locus conferring unique powdery mildew resistance in wheat breeding line Wanfengjian 34. *Mol. Breed.* 35:210. doi: 10.1007/s11032-015-0403-5
- Ma, P. T., Zhang, H. X., Xu, H. X., Xu, Y. F., Cao, Y. W., Zhang, X. T., et al. (2015c). The gene *PmYB* confers broad-spectrum powdery mildew resistance in the multi-allelic *Pm2* chromosome region of the Chinese wheat cultivar YingBo 700. *Mol. Breed.* 35:124. doi: 10.1007/s11032-015-0320-7
- Ma, P. T., Xu, H. X., Xu, Y. F., Song, L. P., Liang, S. S., Sheng, Y., et al. (2018). Characterization of a powdery mildew resistance gene in wheat breeding line 10V-2 and its application in marker-assisted selection. *Plant Dis.* 102, 925–931. doi: 10.1094/PDIS-02-17-0199-RE
- Ma, P. T., Han, G. H., Zheng, Q., Liu, S. Y., Han, F. P., Wang, J., et al. (2019). Development of novel wheat-rye chromosome 4R translocations and assignment of their powdery mildew resistance. *Plant Dis.* 104. doi: 10.1094/PDIS-01-19-0160-RE
- Mäkinen, K., and De, S. (2019). The significance of methionine cycle enzymes in plant virus infections. *Curr. Opin. Plant Biol.* 50, 67–75. doi: 10.1016/j.pbi.2019.03.002
- Mandal, M. S. N., Fu, Y., Zhang, S., and Ji, W. (2014). Proteomic analysis of the defense response of wheat to the powdery mildew fungus *Blumeria graminis* f. sp. tritici. *Protein J.* 33, 513–524. doi: 10.1007/s10930-014-9583-9
- Manoharachary, C., and Kunwar, I. K. (2014). "Host-pathogen interaction, plant diseases, disease management strategies, and future challenges," in *Future Challenges in Crop Protection Against Fungal Pathogens. Fungal Biology*, eds A. Goyal, and C. Manoharachary (New York, NY: Springer).
- McIntosh, R. A., Dubcovsky, J., Rogers, W. J., Xia, X. C., and Raupp, W. J. (2019). "Catalogue of gene symbols for wheat: 2019 supplement," in *Annual Wheat Newsletter*, ed. W. J. Raupp (Manhattan, NY: Kansas State University), 98–113.
- Mehta, A., Basandrai, A. K., Banyal, D. K., and Basandrai, D. (2018). Effect of weather parameters on powdery mildew development of wheat at different location in himachal pradesh. *Indian Phytopathol.* 71, 349–353. doi: 10.1007/s42360-018-0062-7
- Meisel, A. D., Natarajan, C., Sterba, G., and Diamond, H. S. (1980). "Cyclic nucleotide levels and mechanism of inhibition of leucocyte function by adenosine deaminase inhibition," in *Purine Metabolism in Man-III. Advances in Experimental Medicine and Biology*, Vol. 122B, eds A. Rapado, R. W. E. Watts, and C. H. M. M. De Bruyn (Boston, MA: Springer). doi: 10.1007/978-1-4684-8559-2_41
- Morgounov, A., Tufan, H. A., Sharma, R., Akin, B., Bagci, A., Braun, H. J., et al. (2012). Global incidence of wheat rusts and powdery mildew during 1969-2010 and durability of resistance of winter wheat variety Bezostaya 1. *Eur. J. Plant Pathol.* 132, 323–340. doi: 10.1007/s10658-011-9879-y
- Neckers, L., and Tatu, U. (2008). Molecular chaperones in pathogen virulence emerging new targets for therapy. *Cell Host Microbe* 4, 519–527. doi: 10.1016/j.chom.2008.10.011
- Pál, M., Kovács, V., Vida, G., Szalai, G., and Janda, T. (2013). Changes induced by powdery mildew in the salicylic acid and polyamine contents and the antioxidant enzyme activities of wheat lines. *Eur. J. Plant Pathol.* 135, 35–47. doi: 10.1007/s10658-012-0063-9
- Pankiewicz, V. C. S., Camilios-Neto, D., Bonato, P., Balsanelli, E., Tadra-Sfeir, M. Z., Faoro, H., et al. (2016). RNA-seq transcriptional profiling of *Herbaspirillum*

- seropedicae colonizing wheat (*Triticum aestivum*) roots. *Plant Mol. Biol.* 90, 589–603. doi: 10.1007/s11103-016-0430-6
- Park, J., Ajami, K., Yu, D. M. T., and Gorrell, M. D. (2006). “Dipeptidyl peptidase 8 has post-proline dipeptidyl aminopeptidase and prolyl endopeptidase activities,” in *Dipeptidyl Aminopeptidases. Advances in Experimental Medicine and Biology*, Vol. 575, eds U. Lendeckel, D. Reinhold, and U. Bank (Boston, MA: Springer). doi: 10.1007/0-387-32824-6_10
- Pearce, S., Vazquez-Gross, H., Herin, S. Y., Hane, D., Wang, Y., Gu, Y. Q., et al. (2015). WheatExp: an RNA-seq expression database for polyploid wheat. *BMC Plant Biol.* 15:299. doi: 10.1186/s12870-015-0692-1
- Rai, A. N., Tamirisa, S., Rao, K. V., Kumar, V., and Suprasanna, P. (2016). RETRACTED ARTICLE: brassica RNA binding protein ERD4 is involved in conferring salt, drought tolerance and enhancing plant growth in *Arabidopsis*. *Plant Mol. Biol.* 90, 375–387. doi: 10.1007/s11103-015-0423-x
- Ramirez-Gonzalez, R. H., Segovia, V., Bird, N., Fenwick, P., Holdgate, S., Berry, S., et al. (2015). RNA-seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. *Plant Biotechnol. J.* 13, 613–624. doi: 10.1111/pbi.12281
- Reiner, A., Yekutieli, D., and Benjamini, Y. (2003). Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 19, 368–375. doi: 10.1093/bioinformatics/btf877
- Saharan, G. S., Mehta, N. K., and Meena, P. D. (2019). “Host resistance,” in *Powdery Mildew Disease of Crucifers: Biology, Ecology and Disease Management* (Singapore: Springer Nature Singapore Pvt. Ltd).
- Sánchez-Martin, J., Steuernage, B., Ghosh, S., Herren, G., Hurni, S., Adamski, N., et al. (2016). Rapid gene isolation in barley and wheat by mutant chromosome sequencing. *Genome Biol.* 17:221. doi: 10.1186/s13059-016-1082-1
- Selter, L., Shatalina, M., Singla, J., and Keller, B. (2014). “Identification and implementation of resistance: genomics-assisted use of genetic resources for breeding against powdery mildew and stagonospora nodorum blotch in wheat,” in *Genomics of Plant Genetic Resources*, eds R. Tuberosa, A. Graner, and E. Frison (Dordrecht: Springer).
- Siddiqui, Z. A., Nisha, R., Singh, N., and Alam, S. (2012). “Interactions of plant-parasitic nematodes and plant-pathogenic bacteria,” in *Bacteria in Agrobiolgy: Plant Probiotics*, ed. D. Maheshwari (Berlin: Springer).
- Singh, R. P., Singh, P. K., Rutkoski, J., Hodson, D. P., He, X. Y., Jørgensen, L. N., et al. (2016). Disease impact on wheat yield potential and prospects of genetic control. *Annu. Rev. Phytopathol.* 54, 303–322. doi: 10.1146/annurev-phyto-080615-095835
- Soto, M. J., Nogales, J., Pérez-Mendoza, D., Gallegos, M. T., Olivares, J., and Sanjuán, J. (2011). Pathogenic and mutualistic plant-bacteria interactions: ever increasing similarities. *Cent. Eur. J. Biol.* 6, 911–917. doi: 10.2478/s11535-011-0069-x
- Spielmeier, W., Singh, R. P., McFadden, H., Wellings, C. R., Huerta-Espino, J., Kong, X., et al. (2008). Fine scale genetic and physical mapping using interstitial deletion mutants of *Lr34/Yr18*: a disease resistance locus effective against multiple pathogens in wheat. *Theor. Appl. Genet.* 116, 481–490. doi: 10.1007/s00122-007-0684-0
- Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoaka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi: 10.1111/tpj.12105
- The International Wheat Genome Sequencing Consortium (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191
- Trick, M., Adamski, N. M., Mugford, S. G., Jiang, C. C., Febrer, M., and Uauy, C. (2012). Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol.* 12:14. doi: 10.1186/1471-2229-12-14
- Wang, R., Sun, L., Bao, L., Zhang, J., Jiang, Y., Yao, J., et al. (2013). Bulk segregant RNA-seq reveals expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish. *BMC Genomics* 14:929. doi: 10.1186/1471-2164-14-929
- Wang, Y., Xie, J. Z., Zhang, H. Z., Guo, B. M., Ning, S. Z., Chen, Y. X., et al. (2017). Mapping stripe rust resistance gene *YrZH22* in Chinese wheat cultivar Zhoumai 22 by bulked segregant RNA-Seq (BSR-Seq) and comparative genomics analyses. *Theor. Appl. Genet.* 130, 2191–2201. doi: 10.1007/s00122-017-2950-0
- Wang, Y., Zhang, H., Xie, J., Guo, B., Chen, Y., Zhang, H., et al. (2018). Mapping stripe rust resistance genes by BSRseq: *YrMM58* and *YrHY1* on chromosome 2AS in Chinese wheat lines Mengmai 58 and Huaiyang 1 are *Yr17*. *Crop J.* 6, 91–98. doi: 10.1016/j.cj.2017.03.002
- Xu, H. X., Yi, Y. J., Ma, P. T., Qie, Y. M., Fu, X. Y., Xu, Y. F., et al. (2015). Molecular tagging of a new broad-spectrum powdery mildew resistance allele *Pm2c* in Chinese wheat landrace Niamai. *Theor. Appl. Genet.* 128, 2077–2084. doi: 10.1007/s00122-015-2568-z
- Yazbeck, R., Jaenisch, S. E., and Abbott, C. A. (2018). Potential disease biomarkers: dipeptidyl peptidase 4 and fibroblast activation protein. *Protoplasma* 255, 375–386. doi: 10.1007/s00709-017-1129-5
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Zhang, H. Z., Xie, J. Z., Cheng, Y. X., Liu, X., Wang, Y., Yan, S. H., et al. (2017). Mapping stripe rust resistance gene *YrZM103* in wheat cultivar Zhengmai 103 by BSR-seq. *Acta Agron. Sin.* 43, 1643–1649. doi: 10.3724/SP.J.1006.2017.01643
- Zheng, H. Y., Dong, L. L., Han, X. Y., Jin, H. B., Yin, C. C., Han, Y. L., et al. (2019). The *TuMYB46L-TuACO3* module regulates ethylene biosynthesis in einkorn wheat defense to powdery mildew. *New Phytol.* 225, 2526–2541. doi: 10.1111/nph.16305
- Zhou, W. H., Qian, C., Li, R. C., Zhou, S., Zhang, R. Q., Xiao, J., et al. (2018). TaNAC6s are involved in the basal and broad-spectrum resistance to powdery mildew in wheat. *Plant Sci.* 277, 218–228. doi: 10.1016/j.plantsci.2018.09.014

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 Zhu, Wu, He, Song, Jia, Liu, Wang, Han, Niu, Du, Zhang, Wang, Liang, Li, Liu, Xu, Liu and Ma. 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.



Patterns of Sequence and Expression Diversification Associate Members of the *PADRE* Gene Family With Response to Fungal Pathogens

Marie Didelon, Mehdi Khafif, Laurence Godiard, Adelin Barbacci and Sylvain Raffaele*

Université de Toulouse, Laboratoire des Interactions Plantes Micro-organismes (LIPM), Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE) – Centre National de la Recherche Scientifique (CNRS), Castanet-Tolosan, France

OPEN ACCESS

Edited by:

Takaki Maekawa,
Max Planck Institute for Plant
Breeding Research, Germany

Reviewed by:

Thomas Griebel,
Freie Universität Berlin, Germany
Bilal Okmen,
University of Cologne, Germany
Andrea Ghelfi,
Kazusa DNA Research Institute,
Japan

*Correspondence:

Sylvain Raffaele
sylvain.raffaele@inrae.fr

Specialty section:

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

Received: 15 January 2020

Accepted: 20 April 2020

Published: 29 May 2020

Citation:

Didelon M, Khafif M, Godiard L,
Barbacci A and Raffaele S (2020)
Patterns of Sequence and Expression
Diversification Associate Members
of the *PADRE* Gene Family With
Response to Fungal Pathogens.
Front. Genet. 11:491.
doi: 10.3389/fgene.2020.00491

Pathogen infection triggers extensive reprogramming of the plant transcriptome, including numerous genes the function of which is unknown. Due to their wide taxonomic distribution, genes encoding proteins with Domains of Unknown Function (DUFs) activated upon pathogen challenge likely play important roles in disease. In *Arabidopsis thaliana*, we identified thirteen genes harboring a DUF4228 domain in the top 10% most induced genes after infection by the fungal pathogen *Sclerotinia sclerotiorum*. Based on functional information collected through homology and contextual searches, we propose to refer to this domain as the pathogen and abiotic stress response, cadmium tolerance, disordered region-containing (*PADRE*) domain. Genome-wide and phylogenetic analyses indicated that *PADRE* is specific to plants and diversified into 10 subfamilies early in the evolution of Angiosperms. *PADRE* typically occurs in small single-domain proteins with a bipartite architecture. *PADRE* N-terminus harbors conserved sequence motifs, while its C-terminus includes an intrinsically disordered region with multiple phosphorylation sites. A pangenomic survey of *PADRE* genes expression upon *S. sclerotiorum* inoculation in *Arabidopsis*, castor bean, and tomato indicated consistent expression across species within phylogenetic groups. Multi-stress expression profiling and co-expression network analyses associated At*PADRE* genes with the induction of anthocyanin biosynthesis and responses to chitin and to hypoxia. Our analyses reveal patterns of sequence and expression diversification consistent with the evolution of a role in disease resistance for an uncharacterized family of plant genes. These findings highlight *PADRE* genes as prime candidates for the functional dissection of mechanisms underlying plant disease resistance to fungi.

Keywords: plant disease resistance, diversification, DUF4228, intrinsic disorder, pathogenesis-related, gene expression profiling

INTRODUCTION

Wild plants and crops suffer from recurrent attacks by pathogenic microbes, threatening biodiversity and food production. Molecular and genetic studies revealed that plants possess an elaborate immune system able to detect pathogens and activate genetic pathways to mount effective defense responses (Dodds and Rathjen, 2010). Specific defense responses allow plants

to cope with microbial pathogens of diverse lifestyles and genotypes that target diverse plant organs (Glazebrook, 2005). In most cases, the activation of plant responses requires extensive transcriptional reprogramming, covering for instance up to 25% of the whole genome in *Arabidopsis thaliana* (Eulgem, 2005). In nature, one of the most frequent forms of plant immunity is designated as quantitative disease resistance (QDR) (Poland et al., 2009; Roux et al., 2014). QDR leads to a full continuum of disease resistance phenotypes in natural plant populations, from very susceptible to largely resistant, and generally involves a large number of genetic loci. Every gene adds a small contribution to form the overall resistance (Roux et al., 2014). Current knowledge of the molecular bases of QDR in plants remains very incomplete, but a few general properties have emerged. First, the molecular functions of QDR genes are very diverse, including for instance transporters (Krattinger et al., 2009), kinases (Derbyshire et al., 2019), proteases (Badet et al., 2017), and genes of unknown function (Fukuoka et al., 2009). Second, the function of QDR genes may not be limited to disease resistance and can include activity in cell morphology (Rajarammohan et al., 2018; Badet et al., 2019), metabolism (Rajarammohan et al., 2018), or embryogenesis (Derbyshire et al., 2019) in certain contexts. Third, QDR responses to a given pathogen species may involve hundreds or even thousands of genes (Corwin et al., 2016; Fordyce et al., 2018). Therefore, pathogen infection triggers extensive reprogramming of the plant transcriptome, including numerous genes the molecular function of which is currently unknown.

Recent progress in high-throughput omics techniques enabled the determination of the sequence of genes and proteins at an unprecedented pace. Homology relationships allow to rapidly transfer functional information from one sequence to another but suffer limitations (Pearson and Sierk, 2005), and our capacity to generate new sequences far exceeds our ability to interpret them. Sequence conservation across large evolutionary distances can identify previously unknown functional domains in proteins, such as in the case of the VAST domain (PF16016) (Khafif et al., 2014, 2017; Gatta et al., 2015). The Protein Family Database (Pfam) gathers protein families by their homology of sequence (El-Gebali et al., 2019). In 2019, the latest Pfam release (32.0) counted 17,929 entries, 3,961 (22%) of them being Domains of Unknown Function (DUFs). DUFs are protein families for which no member has an experimentally characterized function. Systematic structural analyses of DUF proteins revealed that a significant part of DUF proteins likely originate from extreme diversification and neofunctionalization of known protein domains (Jaroszewski et al., 2009). Due to their wide taxonomic distribution and their evolutionary sequence conservation, many DUFs are expected to compose essential proteins (Goodacre et al., 2013). Widely distributed genes encoding proteins with DUFs activated upon pathogen challenge are promising sources of new insight into the evolution and molecular mechanisms of plant disease resistance.

Sclerotinia sclerotiorum is a devastating fungal plant pathogen from the Ascomycota division with a necrotrophic lifestyle. It is responsible for the white and stem mold diseases on more than 400 plant species, including crops of high agricultural value like

sunflower, soybean, rapeseed, and tomato, among others (Boland and Hall, 1994; Hegedus and Rimmer, 2005). The host range of *S. sclerotiorum* also includes plants from the *Brassicaceae* family, such as the plant model *A. thaliana*. Resistance to *S. sclerotiorum* is typically quantitative with no complete resistance (Perchepped et al., 2010; Mbengue et al., 2016). The molecular bases of QDR to *S. sclerotiorum* are beginning to be elucidated, notably thanks to studies on *A. thaliana*, but remain very patchy (Mbengue et al., 2016). Global gene expression profiling by RNA sequencing revealed 4,703 *A. thaliana* genes significantly induced upon leaf inoculation with *S. sclerotiorum* (Badet et al., 2017), including several genes harboring a DUF4228 domain. Here, we took a survey of DUF4228 homologs across the plant kingdom and identified a few experimental insights into the function of these genes. We propose to refer to this domain as the pathogen and abiotic stress response, cadmium tolerance, disordered region-containing (PADRE) domain to facilitate future reference. We used phylogenetic analyses to document the extent diversity of PADRE sequences and infer scenarios for their evolution. PADRE proteins lack sequence similarity to characterized proteins but harbor a bipartite architecture with conserved motifs in the N-terminal region and a C-terminal region rich in phosphorylated residues and predicted to be intrinsically disordered. Pangenomic expression profiling in thale cress (*A. thaliana*), tomato (*Solanum lycopersicum*), and castor bean (*Ricinus communis*) plants inoculated by *S. sclerotiorum* identified groups of PADRE genes that respond to this fungal pathogen in a consistent manner across species. Finally, *AtPADRE* gene expression upon diverse stress treatments and co-expression network reconstruction suggests that several PADRE genes could function synergistically in plant defense. Our study reveals that responsiveness to fungal pathogen attack is conserved at the interspecific level in groups of PADRE genes and provides insights into the evolutionary history and functional diversification in this poorly characterized plant gene family.

RESULTS

Genes From the DUF4228 Family Are Over-Represented Among Genes Induced Upon *S. sclerotiorum* Inoculation

To get insights into plant processes activated during colonization by the fungal pathogen *S. sclerotiorum*, we analyzed RNA-Seq data for *A. thaliana* plants inoculated by *S. sclerotiorum*. Specifically, we focused on protein domains overrepresented among plant genes differentially expressed upon inoculation. To this end, we exploited the RNA sequencing data generated in Badet et al. (2017) (GSE106811). Differential expression analysis identified 4,703 genes significantly induced (\log_2 fold change (LFC) > 1.5, adjusted *p*-value (padj) < 0.01) and 5,812 genes significantly down-regulated (LFC < -1.5, padj < 0.01) in *A. thaliana* during infection by *S. sclerotiorum*. We annotated genes by their protein domains using the Pfam database. Using a proportion Z-test (*p*-value < 0.01), we counted 53 protein

domains significantly overrepresented among induced genes with at least 10 occurrences in *A. thaliana* genome (**Supplementary Table S1** and **Figure 1**). The ubiquitin-like domain PF14560 showed the strongest enrichment in induced genes (induced/total ratio = 0.83, p -val = 4.41×10^{-5}), and the protein kinase domain PF00069 had the most significant enrichment in induced genes (ratio 0.25, p -val 5.02×10^{-10}). Gene Ontology terms associated with the 74 overrepresented protein domains included defense mechanism and immune response in 64.5% of cases. For instance, 27 out of 72 genes harboring a WRKY domain (PF03106) were induced upon infection by *S. sclerotiorum* (ratio = 0.375, p -val 1.43×10^{-4}). Other protein domains enriched in induced genes included ubiquitin-like domains (PF10302, PF14560, PF11976, and PF00240), transport-related domains (PF01105, PF08449, and PF03105), calcium binding (PF14658), and heat shock response (PF00011) (**Supplementary Table S1**).

One domain enriched in induced genes had no known molecular function and was identified as Domain of Unknown Function DUF4228 (ratio 0.5, p -val 3.86×10^{-4}). We identified 28 genes with a DUF4228 domain in the genome of *A. thaliana* (hmmscan e -value < 1×10^{-10} , **Supplementary Table S2**), 19 of them being differentially expressed upon infection by *S. sclerotiorum* (14 induced and 5 down-regulated). The DUF4228 gene AT5G37840 was induced over 1,000 times (LFC 10.22, p -val 1.45×10^{-47}), and 13 genes harboring a DUF4228 domain were in the top 10% most induced genes after infection by *S. sclerotiorum* in *A. thaliana* (LFC > 4.04, **Supplementary Table S2**). Because of their dramatic induction pattern and

although uncharacterized to date, some DUF4228 genes could function in plant defense responses.

Taxonomic Distribution of the DUF4228/PADRE Domain

To document the taxonomic distribution of the DUF4228 domain across the tree of life, we performed a HMM search against the RefProt database of UniProtKB with an alignment of *A. thaliana* DUF4228 proteins as input (**Supplementary Datasheet S1**). We retrieved 3647 hits distributed in 98 species. As recently reported (Yang et al., 2020), DUF4228 appeared restricted to plants, including mosses, liverworts, and monocot and dicot species. The average size of DUF4228 domains detected in these proteins was 149.7 ± 40.3 amino acids, for proteins of 159.7 ± 37.6 amino acids long (**Figure 2A**). In good agreement, only 8.4% of proteins harboring a DUF4228 domain were multi-domain proteins. To identify the complete repertoire of DUF4228 in plant proteomes, we performed a HMM search against the Phytozome 12.1 database. Out of the 64 plant proteomes available at the time of our analysis, only the seven Chlorophyte proteomes did not show a single DUF4228 domain, indicating that the emergence of the DUF4228 domain occurred at least 450 million years ago (**Figure 2B**). Next, we used Timetree to relate the number of DUF4228-containing proteins with time of speciation in 49 plant species. In embryophytes, the number of DUF4228-containing proteins ranged from three (*Selaginella moellendorffii*) to 81 (*Glycine max*) (**Figure 2B**). A majority of embryophytes (28/45) had between 20 and 40 DUF4228-containing proteins, and there was no striking expansion of DUF4228 in a specific plant lineage. Recent whole genome duplication events were often associated with expanded DUF4228 repertoires, such as in *Brassica rapa*, *Malus domestica*, *G. max*, *Zea mays*, and *Musa acuminata*. Overall, the size of the DUF4228 family was well correlated ($R^2 = 0.5066$) with the total number of genes per genome across embryophyte species (**Figure 2C**).

Through homology and keyword searches, we found experimental insights into function for DUF4228 proteins. The *A. thaliana* AT4G37240 protein was identified as interacting with calmodulin proteins CAM4, 6, 7, 8, and 9 (Popescu et al., 2007). In addition, *A. thaliana* AT1G66480 was identified as interacting with Arabidopsis Response Regulator 14 (ARR14) in a yeast two-hybrid screen (Dortay et al., 2008). However, these protein-protein interactions have not been validated by independent approaches. In *Nicotiana tabacum*, the homolog of *A. thaliana* AT1G76600 was found responsive to tobacco mosaic virus and wounding and the corresponding protein designated as Pathogenesis-related protein of 23kDa (NtPRp23) (Akiyama et al., 2005). Its ortholog in *N. sylvestris* (LOC104235934) conferred tolerance to cadmium when expressed in yeast (Zhang et al., 2016). Recent work by Yang et al. (2020) revealed that several DUF4228 genes are responsive to drought, cold, or salt abiotic stress. Our analyses reported in this study indicated that several DUF4228 genes are responsive to infection by the fungal pathogen *S. sclerotiorum* and that *A. thaliana* DUF4228 proteins harbor intrinsically disordered regions. Based on this partial functional information and to facilitate further

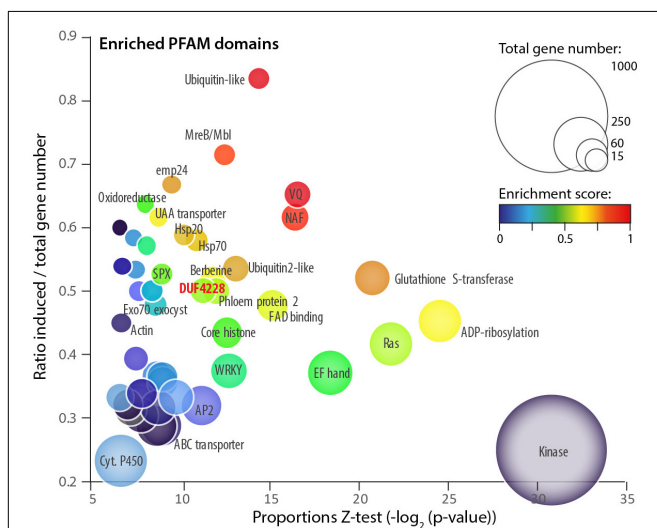
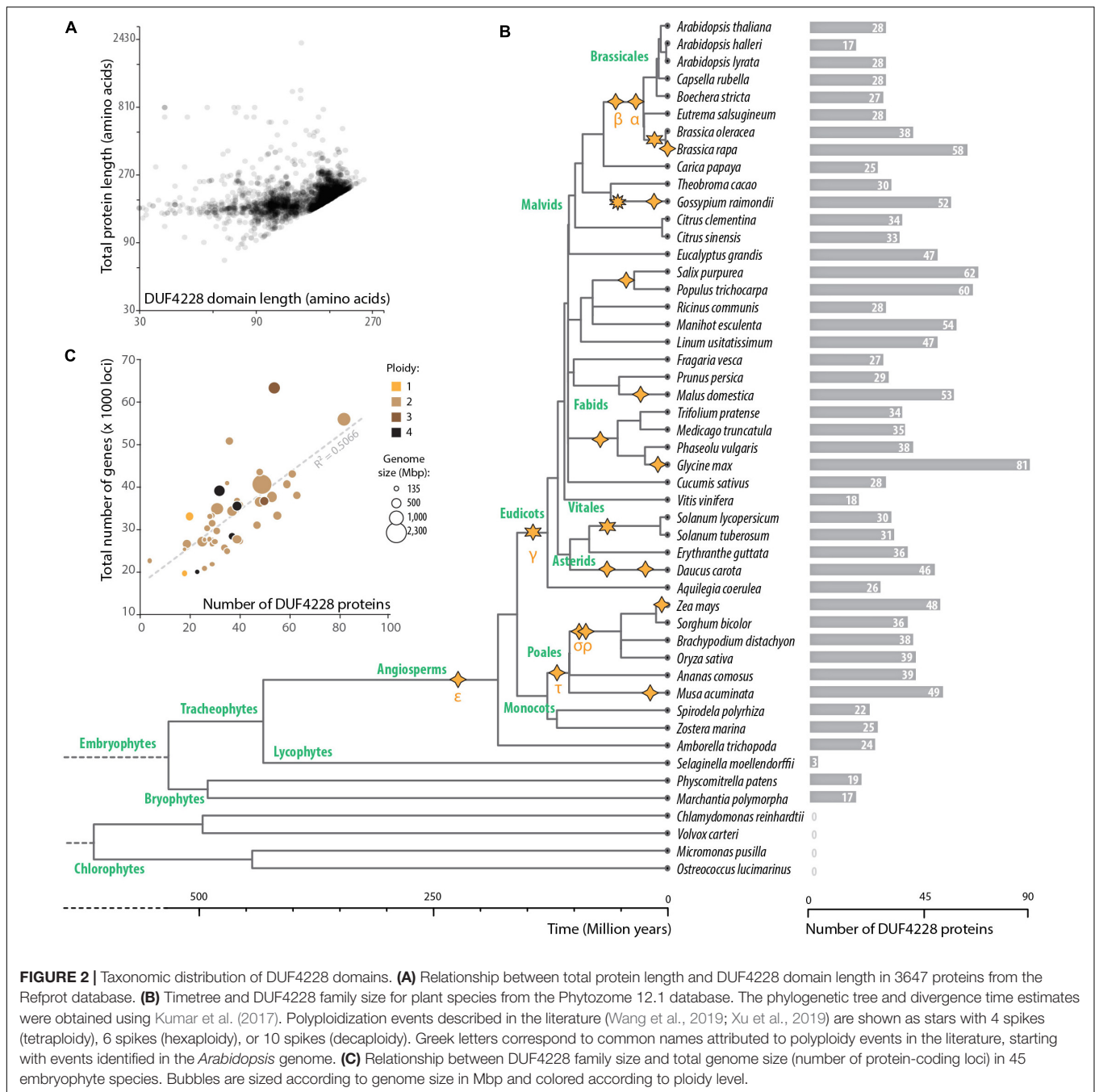


FIGURE 1 | Protein domains enriched among *A. thaliana* genes upregulated upon *S. sclerotiorum* inoculation. Each bubble shows one of 54 PFAM domains significantly enriched (proportion test p -value < 0.01) in induced genes (LFC > 1.5, p -value < 0.01). Bubbles are sized according to the total number of genes containing the domain in the *A. thaliana* genome. Enrichment is shown as the p -value of a proportion Z-test for enrichment (X-axis), the ratio between the number of induced/total genes (Y-axis) and a composite enrichment score (color scale, see the section “Materials and Methods”). The DUF4228 domain is labeled in bold red. Associated raw data corresponds to *Arabidopsis thaliana* samples from GEO accession number GSE106811.

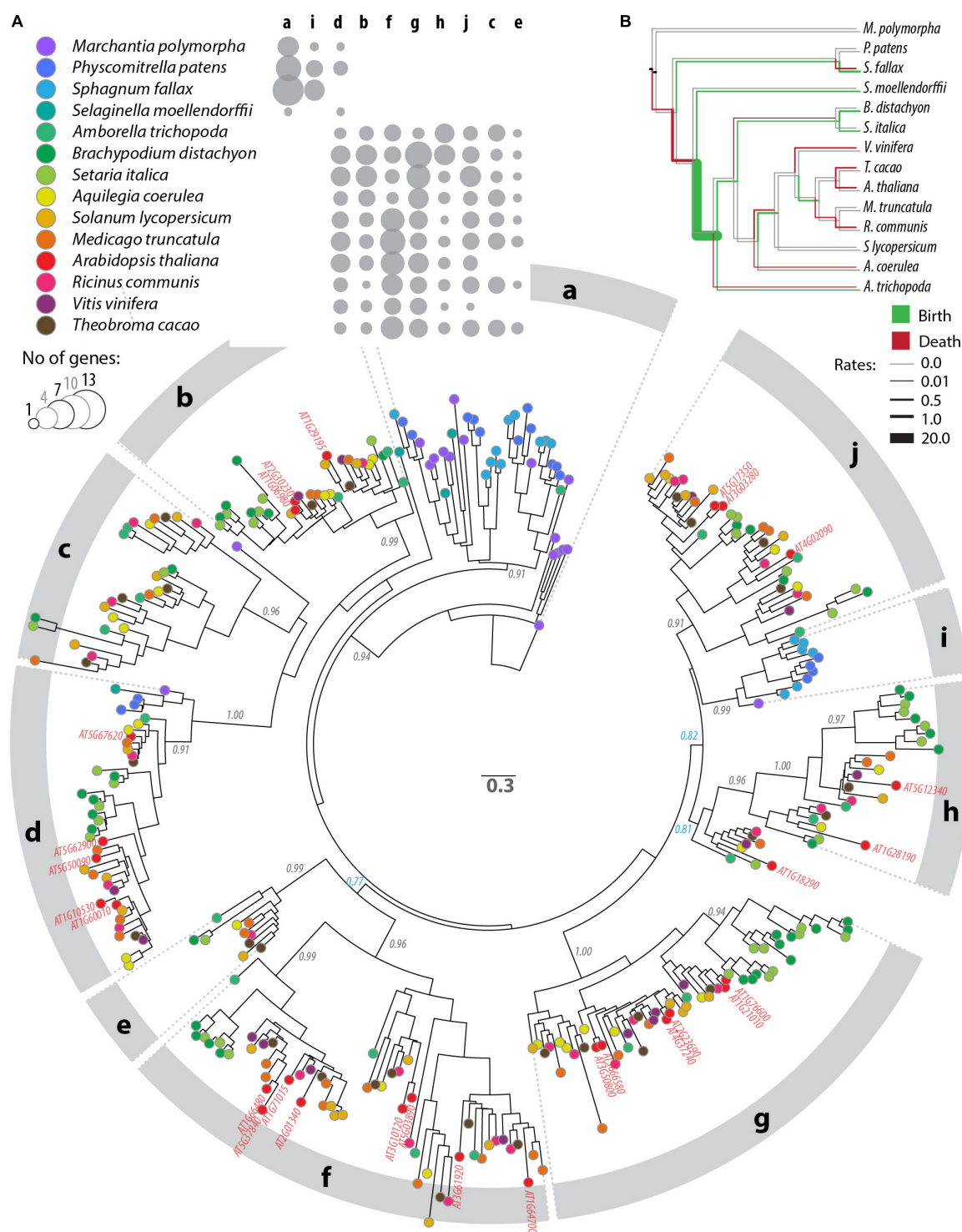


reference, we propose to refer to this family as the pathogen and abiotic stress response, cadmium tolerance, disordered region-containing (PADRE) family.

Sequence Diversification of the PADRE Domain

To analyze patterns of sequence diversification among PADRE proteins, we selected 13 plant genomes representative of the major Embryophyta lineages and constructed a phylogenetic tree of PADRE proteins from these species (see the section

“Materials and Methods”). For this, we generated a multiple protein alignment including 344 sequences and 116 informative sites located within the PADRE domain (**Supplementary Datasheet S2, S3**). We used maximum likelihood methods to represent phylogenetic relationships between these 354 PADRE domains as a tree (**Figure 3A** and **Supplementary Datasheet S4**). PADRE proteins classified into 10 monophyletic groups (a to j) supported by posterior probabilities ≥ 0.90 and encompassing 10 (clade e) to 55 (clade g) proteins. PADRE sequences diversified strongly since the divergence between Lycophytes and Angiosperms: groups a and i were restricted to Bryophytes



and Lycophytes, while groups b, c, e, f, g, h, and j were restricted to Angiosperms. Group d was represented in all species analyzed except *Sphagnum fallax*. Groups b, c, e, f, g, h, and j were represented in all Angiosperm species analyzed, with the exception of groups c and e that were absent from *Arabidopsis thaliana* and *Vitis vinifera*. This suggests that seven PADRE groups existed in the common Angiosperm ancestor and that groups c and e were lost in *A. thaliana* and *V. vinifera*. The number of PADRE groups expanded more rapidly in Angiosperms (reaching 6 and 8 distinct clades per species) than in Bryophytes and Lycophytes (2 or 3 clades per species), indicative of strong diversification of PADRE genes early in the evolution of Angiosperms. To estimate rates of domain birth and death in the PADRE family, we analyzed the species distribution of PADRE phylogenetic group with BadiRate (Librado et al., 2012) (Figure 3B). This revealed two major domain gain events during the emergence of Angiosperms and of core Eudicots and several lineage-specific gain events. Loss events mostly corresponded to the emergence of Tracheophytes and to terminal branches in the Fabids and Malvids clades.

PADRE Is a Bipartite Domain Including Disordered and Phosphorylated C-Termini

PADRE proteins do not display clear homology to functionally characterized proteins. In our alignment of PADRE proteins from 13 Tracheophyta species, sequence conservation appeared limited to four short motifs of 10 amino acids or less (Figure 4A). These conserved motifs correspond to motifs 1, 3, and 6 identified by Yang et al. (2020). As noted by Yang et al. (2020), additional short sequence motifs were restricted to specific PADRE groups. To get insights into PADRE protein sequence signatures and their potential functional implications, we scanned *A. thaliana* PADRE proteins with the ELM, PhosPhat, PrDOS, and Grantham Polarity calculation tools. First, we used the eukaryotic linear motif (ELM) resource to identify motifs similar to known functional sites in proteins (Gouw et al., 2018) (Figure 4B). Among motifs identified robustly in multiple AtPADRE proteins was an N-myristoylation motif, corresponding to the well-conserved GNXXX motif found at the very N-terminus of PADRE proteins. *In vitro* myristoylation provided experimental for N-myristoylation of AT4G37240 (group G) and AT1G10530 (group D) (Boisson et al., 2003, unpublished result available¹). Furthermore, AT1G21010 (group G) and ATGG17350 (group J) were identified in plasma-membrane fractions as expected if N-myristoylated (Majeran et al., 2018). The conserved LXXG motif of PADRE proteins overlapped with a WH2 motif for interaction with actin (LIG_Actin_WH2). The conserved YFLLP motif overlapped with a Tyrosine-based signal for interaction with the adaptor protein complex (TRG_ENDOCYTIC_2), a LIR motif for binding to the autophagy protein Atg8 (LIG_LIR_Gen_1), and a protein phosphatase interacting motif (DOC_PP1_RVXF_1). Basic nuclear localization signals were detected at the C-terminus

of several PADRE proteins. The C-terminal WRPXLXXIXE motif overlapped with an APCC-binding Destruction motif required for targeting to ubiquitin-mediated proteasome-dependent degradation (DEG_APCC_DBOX_1). ELM also detected numerous putative phosphorylation sites at the C-terminus of PADRE proteins. We took advantage of the PhosPhat 4.0 database to search for experimentally determined phospho-peptides in PADRE proteins (Durek et al., 2009). We retrieved phospho-peptides from seven AtPADRE proteins from group B (AT1G06980), D (AT1G60010), F (AT1G64700, AT1G66480, AT2G01340, AT5G37840), and G (AT1G76600) (Figure 4C). A large majority of the phosphorylated residues resided in the C-terminal half of the PADRE domain.

We used the PrDOS server (Ishida and Kinoshita, 2007) to predict natively disordered regions in the 28 AtPADRE proteins (Figure 4D). All AtPADRE proteins showed a relatively consistent pattern of disorder probability, indicating a short (~5 amino acids) N-terminal disordered region followed by an ordered region of ~60 amino acids and a C-terminal half with high probability of intrinsic disorder. To test whether the structural state of PADRE regions was associated with contrasted amino acid usage at the N and C terminus, we calculated the Grantham Polarity index (Grantham, 1974) along 28 AtPADRE proteins (Figure 4E). In average, the PADRE C-terminal region harbors more polar residues (average index 9.02) than the N-terminal region (average index 8.39).

Responsiveness to *S. sclerotiorum* Varies Across PADRE Phylogenetic Groups

The clear delineation of phylogenetic groups in the PADRE family and recent investigations of AtPADRE gene expression upon abiotic stress (Yang et al., 2020) suggested that PADRE genes could have acquired several distinct functions over evolution. Here, we set to investigate whether responsiveness to the fungal pathogen *S. sclerotiorum* contrasts across PADRE phylogenetic groups and whether responsiveness to fungal infection is consistent across plant species. To this end, we analyzed the expression of the PADRE gene repertoire of *A. thaliana*, *Solanum lycopersicum*, and *Ricinus communis* by RNA-sequencing in leaves of healthy plants and upon inoculation by *S. sclerotiorum*. We detected the expression of 74 PADRE genes, including 28 AtPADRE, 26 SiPADRE, and 23 RcPADRE (Figure 5 and Supplementary Table S3). Among them, 31 were significantly induced, 11 were significantly down-regulated, and 32 were not differentially expressed. PADRE genes highly expressed in healthy leaves were frequent in groups d and h and a subgroup of group f. Group h was very homogenous with all four genes significantly induced upon *S. sclerotiorum* infection, and groups j and f included mostly induced genes, with group d that included a majority of down-regulated genes. To determine whether PADRE gene induction differs significantly between phylogenetic groups or between species, we performed ANOVA on expression LFC for the 74 PADRE genes. In a one-way ANOVA, the phylogenetic group effect was found highly significant (p -value 0.0052) while the species was not significant (p -value 0.68). In a two-way

¹<https://www.i2bc.paris-saclay.fr/maturation/Myristoylome.html>

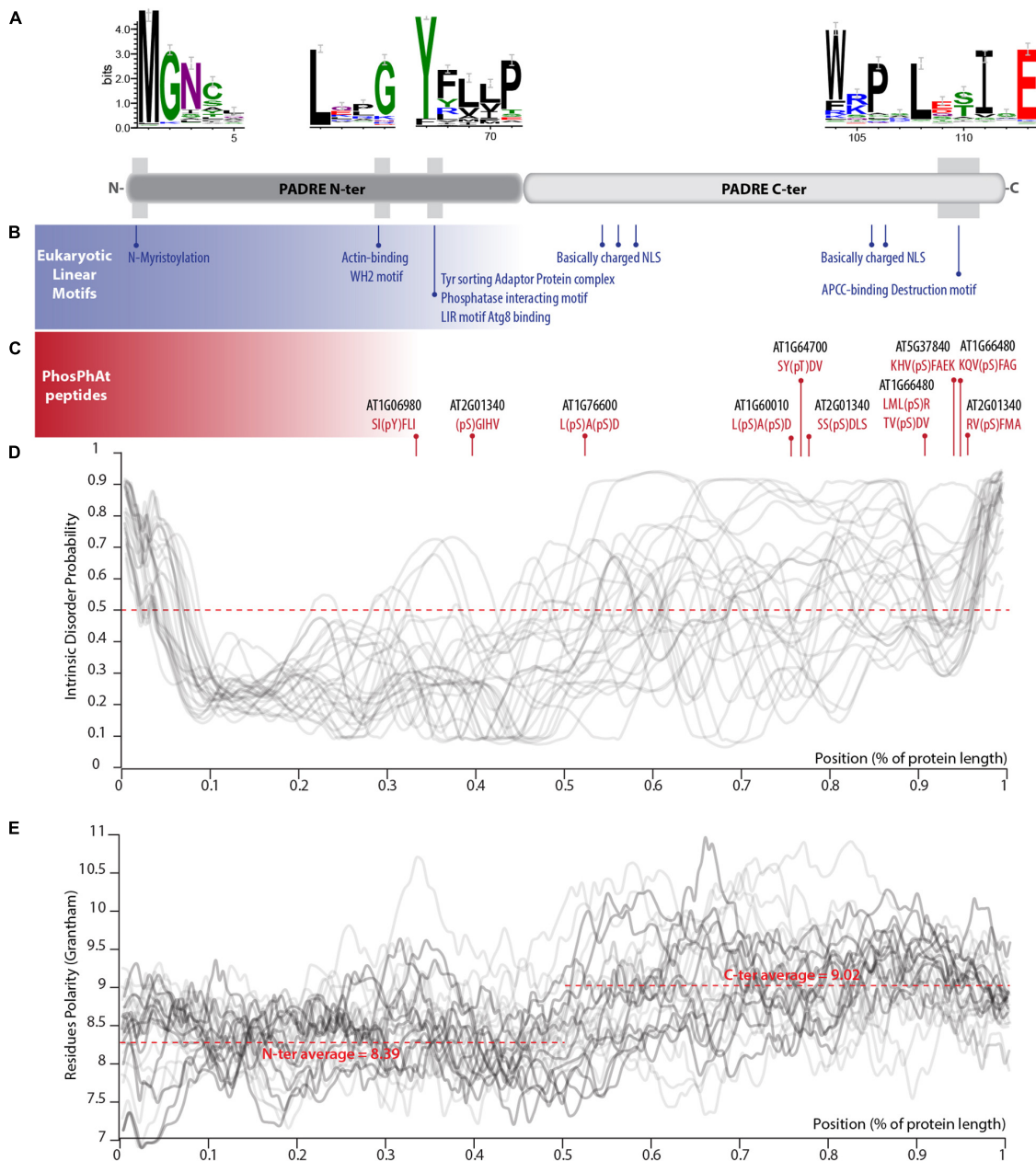
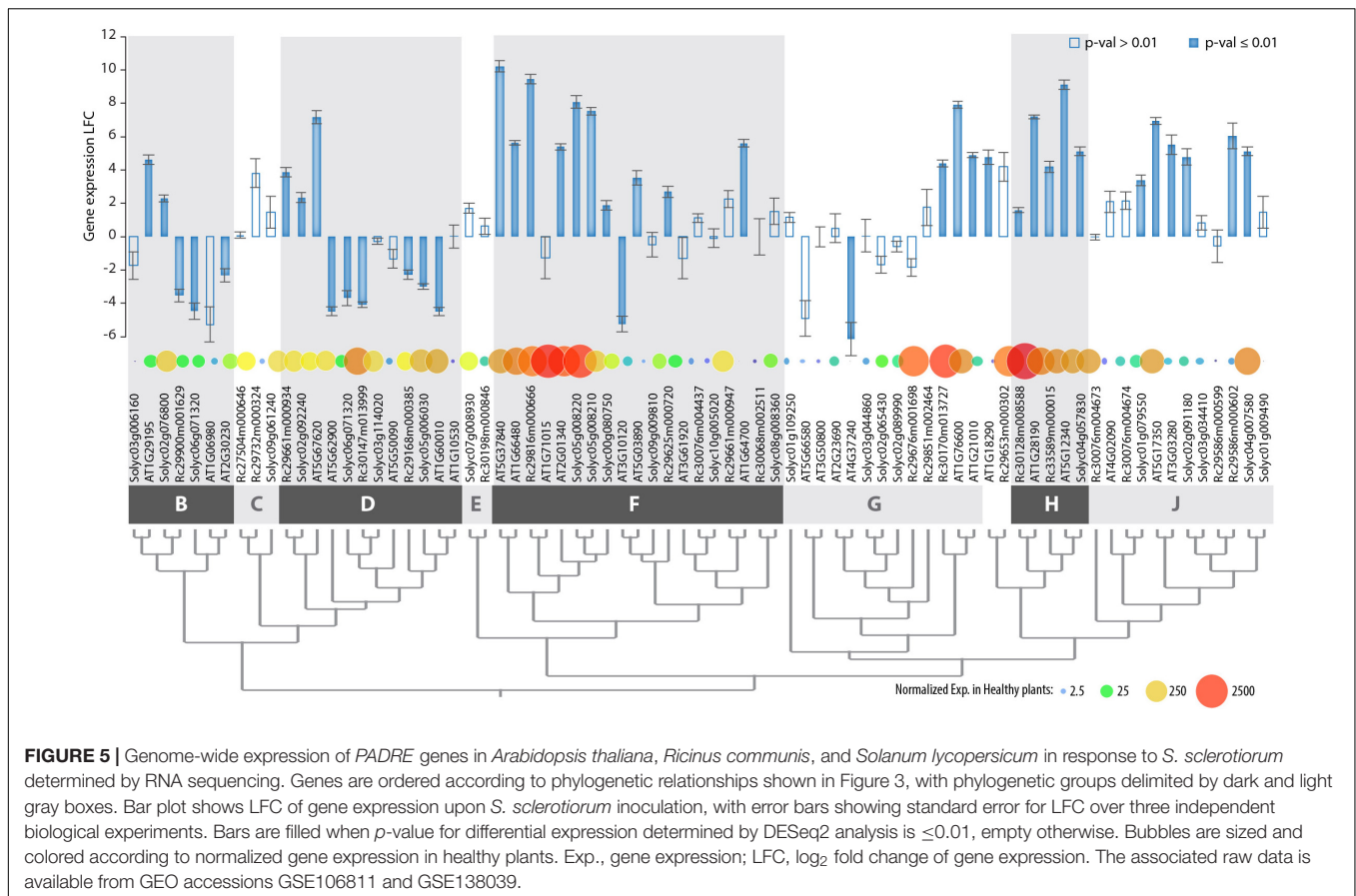


FIGURE 4 | Primary sequence features of PADRE proteins. **(A)** Conserved motifs detected in the alignment of 344 PADRE proteins from 13 Embryophytes. The relative position of these motifs is shown as a light gray shaded area on a diagram representing the PADRE protein structure. **(B)** Eukaryotic linear motifs identified in multiple AtPADRE proteins positioned along the protein diagram shown in **(A)**. **(C)** Phosphorylated residues determined experimentally registered in the PhosPhAt database, positioned along the protein diagram shown in **(A)**. **(D)** Intrinsic disorder probability **(D)** and Grantham residue polarity **(E)** along the 28 AtPADRE proteins. NLS, nuclear localization signal.

ANOVA, the phylogenetic group effect was significant (p -value 0.018); the species effect and group \times species interaction effect was not significant (p -value 0.72 and 0.99, respectively). We conclude that *PADRE* gene expression upon *S. sclerotiorum* inoculation differs between phylogenetic groups in a consistent manner across plant species.

Through a synteny analysis, we identified five pairs (AT1G06980/AT2G30230, AT1G10530/AT1G60010, AT1G210

10/AT1G76600, AT3G03280/AT5G17350, and AT3G10120/AT5G03890) and two quartets AT1G71015/AT2G01340/AT1G66480/AT5G37840, AT2G23690/AT4G37240/AT5G66580/AT3G50800) of *PADRE* genes associated as paralogs in the *A. thaliana* genome. In all instances, groups of paralogs belonged to the same phylogenetic group. The divergence in expression of the *PADRE* genes was limited within most of the paralog groups (**Supplementary Figure S1**). Only the pair of paralogs



AT3G10120-AT5G03890 (group f) showed significant divergence in expression (LFC-5.23 and 3.53, respectively), possibly due to their low basal level of expression. Altogether, our gene expression analysis suggests that responsiveness to pathogens was acquired early in the evolution of Angiosperms by specific groups or subgroups of *PADRE* genes.

A. thaliana PADRE Genes Respond to Multiple Stress and Associate With Plant Defense Ontologies

The finding that some *AtPADRE* genes are responsive to abiotic stresses (Yang et al., 2020) prompted us to investigate their expression under a range of biotic stresses. For this, we analyzed RNA sequencing data available in the Gene Expression Omnibus database (Figure 6A), collecting expression data for *A. thaliana* inoculated by the fungal pathogens *S. sclerotiorum* (Badet et al., 2017), *Botrytis cinerea* (Liu et al., 2015), *Alternaria brassicicola* (Rausch, 2016), and *Verticillium dahliae* (Scholz et al., 2018), the bacterial pathogen *Pseudomonas syringae* pv. *tomato* (*Pst*) DC3000 (Mine et al., 2018) and DC3000 expressing the effector *AvrRps4* (Bhandari et al., 2019), the Cabbage Leaf Curl Virus (CaLCuV) (Zorzatto et al., 2015), and the nematode *Heterodera schachtii* (Shanks et al., 2016). To serve as a reference, we also analyzed RNA sequencing data for plants inoculated with the endophytic fungus *Colletotrichum*

tofieldiae (Hacquard et al., 2016) and submitted to heat stress (Albhlal et al., 2018), cold stress (Zuther et al., 2019), and UV-B treatment (Tavridou et al., 2020). The analysis of *PADRE* gene differential expression revealed a cluster of six *PADRE* genes induced by multiple pathogens: *AT1G28190*, *AT5G12340*, *AT1G76600*, *AT1G21010*, *AT5G37840*, and *AT2G01340* are significantly induced in response to *S. sclerotiorum*, *B. cinerea*, *Pst* DC3000 *AvrRPS4*, and *V. dahliae*. Out of these six genes, five are also induced upon infection by *A. brassicicola*, three under heat stress. Three of them are down-regulated in root response to the non-pathogenic fungus *C. tofieldiae*. By contrast, a cluster of five *PADRE* genes (*AT4G37240*, *AT1G60010*, *AT1G06980*, *AT2G23690*, and *AT5G66580*) was down-regulated in response to pathogens and heat stress. The response of *PADRE* genes to heat stress shared more similarities with their response to pathogens than to other abiotic stimuli. *PADRE* genes were not responsive to all signals. Indeed, only *AT2G01340* was differentially expressed in response to the nematode *H. schachtii*, *AT3G50800* and *AT5G62900* upon infection by the virus CaLCuV, *AT1G76600*, *AT1G21010*, *AT2G01340*, and *AT3G10120* in response to *C. tofieldiae*, and *AT4G37240*, *AT3G61920*, *AT2G23690*, and *AT1G76600* to UV-B. To test for the relationship between phylogenetic clades and the response of *PADRE* genes to diverse stresses, we performed a two-way ANOVA. We found a significant effect of the phylogenetic group (*p*-value 0.042) and type of stress (*p*-value $4.15 \cdot 10^{-5}$) on *PADRE*

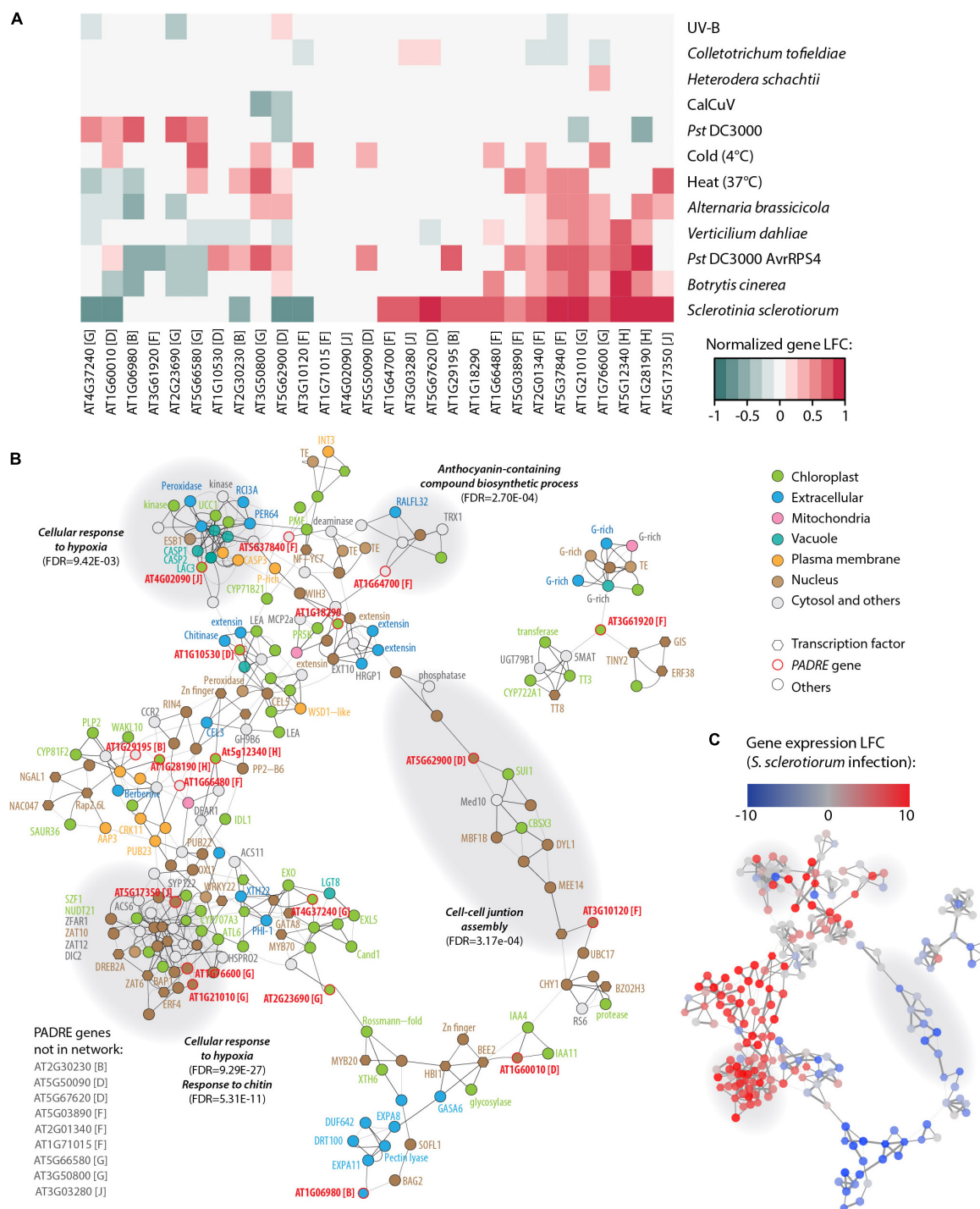


FIGURE 6 | Response to multiple pathogens and co-expression network for *AtPADRE* genes. **(A)** Expression profiles of *A. thaliana* *PADRE* genes under multiple biotic and abiotic stresses deduced from published RNA sequencing data. The expression levels of genes were normalized using min-max feature scaling to fit within the $[-1; 1]$ range for all experiments. Non-significant LFCs are displayed as 0. The phylogenetic group of *AtPADRE* genes is given between square brackets. The associated raw data is available from GEO accessions GSE132169, GSE70094, GSE72548, GSE56922, GSE88798, GSE112225, GSE85653, GSE83478, GSE104590, GSE116269, GSE66290, and GSE106811. *Pst*, *Pseudomonas syringae* pv. *tomato*. **(B)** Co-expression network for *AtPADRE* genes deduced from an experiment of 14,668 microarrays. Nodes are color-coded according to subcellular localization predicted by WOLF-Psort, shown as hexagons for transcription factors and as circles otherwise. *AtPADRE* genes are outlined and labeled in red. Edge widths are scaled according to a mutual rank score index for co-expression. Gray-shaded areas show a subnetwork identified by network modularity analysis, with associated specific gene ontologies labeled in bold italics. **(C)** The same co-expression network as in **(B)** with nodes color-coded according to LFC of gene expression upon infection by *S. sclerotiorum* determined by RNA sequencing (GEO accession GSE106811). LFC, \log_2 fold change of gene expression.

gene expression. Using a Tukey HSD test, we found that the phylogenetic group effect is due to contrasted expression patterns of genes from groups h and d (p -value 0.022). A Tukey HSD test on the stress variable indicated that the stress effect is due to *S. sclerotiorum* infection triggering PADRE gene expression significantly different from every other stresses (p -value < 0.05), except for *Pst* DC3000 *AvrRPS4* infection and heat stress (p -values 0.361 and 0.097, respectively). Therefore, we detected an association (i) between two PADRE phylogenetic groups and responsiveness to multiple stresses and (ii) between the expression of PADRE genes and infection by *S. sclerotiorum*. In addition, visual inspection of **Figure 6A** suggested that *S. sclerotiorum*, *B. cinerea*, *Pst* DC3000 *AvrRPS4*, *V. dahliae*, *A. brassicicola*, and heat stress induced similar transcriptional responses in PADRE genes.

To get insights into genes functioning in the same processes as PADRE genes, we retrieved a co-expression network for *AtPADRE* genes from the ATTED-II database covering 14,668 microarray samples (Obayashi et al., 2018) (**Supplementary Datasheet S5 and Figure 6B**). The network was composed of 225 nodes and 523 undirected edges, including 19 *AtPADRE* genes. We mapped LFC of gene expression upon *S. sclerotiorum* inoculation obtained from our RNA sequencing analysis onto this network, revealing one major sector including predominantly highly induced genes and another sector including mostly down-regulated genes (**Figure 6C**). To emphasize biological processes involving *AtPADRE* genes, we performed a modularity analysis based on the network topology (Blondel et al., 2008) to compute subnetworks and test next if every subnetwork corresponded to gene ontology. The modularity analysis identified 12 subnetworks, four of which were significantly associated with a specific biological function (**Figures 6B,C**). A subnetwork, strongly overexpressed during *S. sclerotiorum* infection (**Figure 6C**), was involved in the perception of the fungus cell wall and in the response to chitin (FDR = 5.3E-11). Response to hypoxia (FDR = 9.45E-3, FDR = 9.29E-27) was overexpressed during infection by *S. sclerotiorum* whereas genes involved in the cell-cell junction assembly were downregulated (**Figure 6C**). The subnetwork grouping genes associated with the biosynthesis of anthocyanin (FDR = 2.7E-4), secondary metabolites with antifungal activity (Kumar Sudheeran et al., 2019), appeared overexpressed during infection. To test further the role played by *AtPADRE* genes in the topology of the network, we computed the local centrality (or degree) of every gene (**Supplementary Figure S2**). Despite the high centrality of the *At5g17350* gene, centralities of *AtPADRE* genes did not differ significantly from other genes of the network (mean degree *AtPadre* 4.63, others 4.65, Wilcoxon's test p -value = 0.53).

DISCUSSION

Plant genomes harbor a remarkably large number of gene families that are not found in other life kingdoms, several of which function in cell signaling (Yamasaki et al., 2013) and defense (Tenhaken et al., 2005; Raffaele et al., 2007; Weidenbach et al., 2016). Through contextual searches, we identified experimental

evidence that DUF4228 genes are involved in pathogen and abiotic stress response, cadmium tolerance, disordered region-containing family (Zhang et al., 2016; Yang et al., 2020), and we propose to refer to this family as the PADRE family to reflect these functional and architectural information. Naming domains based on the first functional clues is unlikely to reflect all or the most prominent function of gene families but can foster further research on the function of these genes (Doerks et al., 2000; Habermann, 2004; Tenhaken et al., 2005). Some PADRE genes are responsive to environmental stimuli such as wounding and viruses (Akiyama et al., 2005), drought, cold, and salt (Yang et al., 2020), pointing toward yet uncharacterized molecular functions.

Insights Into the Evolutionary History of the PADRE Domain

We identified 344 high-quality PADRE protein sequences across 13 plant genomes and used this information in a phylogenetic analysis to explore the dynamics of the PADRE domain evolution. Analysis of the extent diversity of PADRE proteins suggests that they originated before the divergence between Bryophyta and Tracheophyta, like an estimated ~50% of plant-specific domains (Kersting et al., 2012). We classified PADRE proteins into 10 phylogenetic groups, corresponding approximately to subdivisions of the three groups proposed by Yang et al. (2020). In our analysis, the phylogenetic signal was too weak to infer a common ancestor to several groups and combine them with confidence. The BadiRate analysis highlighted a strong radiation of the PADRE domain at the base of the Angiosperms, around ~350 to 175 million years ago. It should be noted that our dataset does not include sequences from the Pinophyta and Pteridophyta lineages, so that the burst of PADRE diversification may date back to the divergence of these groups or to the Angiosperm most recent common ancestor. The recent duplication of PADRE genes from groups b, d, f, g, and j in *A. thaliana* is well supported by the phylogeny and synteny analysis and consistent with Yang et al. (2020). Recent duplications in these groups are also likely in *M. truncatula*, *S. lycopersicum*, and *A. coerulea*. However, PADRE domain births remained limited or null within the core Eudicot clade, where domain loss seemed predominant. This could indicate selection toward some degree of functional specialization in the PADRE family, favoring the expansion of a few clades to the detriment of the overall domain diversity. Our pangenomic expression analysis supported somewhat consistent patterns of PADRE gene expression upon *S. sclerotiorum* inoculation within phylogenetic groups and across species. This suggests that responsiveness to fungal infection was acquired by PADRE groups f, g, h, and j early in the evolution of core Eudicots. Nevertheless, there was striking contrast in expression within groups f and g, which may indicate some degree of neo- or subfunctionalization.

A Probable Bipartite Architecture With Structured and Disordered Regions

Sequence analysis pointed toward a bipartite architecture for the PADRE domain, with a combination of structured and intrinsically disordered regions. Intrinsically disordered

TABLE 1 | List of experiments used to analyze the response of *A. thaliana* PADRE to multiple biotic and abiotic stresses, with the corresponding GEO accession numbers.

Plants and treatment	Plant age	Time post treatment	GEO accession
<i>A. thaliana</i> and <i>S. lycopersicum</i> inoculated by <i>S. sclerotiorum</i>	4 weeks	48 h	GSE106811
<i>R. communis</i> inoculated by <i>S. sclerotiorum</i>	4 weeks	50 h	GSE138039
<i>A. thaliana</i> inoculated by <i>Botrytis cinerea</i>	4 weeks	14 h	GSE66290
<i>A. thaliana</i> inoculated by <i>Alternaria brassicicola</i>	NA	NA	GSE83478
<i>A. thaliana</i> inoculated by <i>Verticillium dahlia</i>	14 days	24 h	GSE104590
<i>A. thaliana</i> inoculated by <i>Colletotrichum tofieldiae</i>	10 days	10 days	GSE70094
<i>A. thaliana</i> inoculated by <i>Pseudomonas syringae</i> pv. <i>tomato</i> DC3000	30 days	24 h	GSE88798
<i>A. thaliana</i> inoculated by <i>Pseudomonas syringae</i> pv. <i>tomato</i> DC3000 AvrRPS4	4 weeks	24 h	GSE116269
<i>A. thaliana</i> inoculated by <i>Heterodera schachtii</i>	10 days	10 days	GSE72548
<i>A. thaliana</i> inoculated by CaLCuV	7 leaf stage	20 days	GSE56922
Heat-treated (37°C) <i>A. thaliana</i> Col-0	5 weeks	30 min	GSE85653
Cold-treated (4°C) <i>A. thaliana</i> Col-0	28 days	3 days	GSE112225
UV-B-treated <i>A. thaliana</i> Col-0	4 days	6 h	GSE132169

NA, not available.

regions (IDRs) are flexible protein regions lacking a stable 3D fold in solution, which may transition to an ordered state upon binding to natural ligands (Uversky, 2013). Proteins with IDRs are abundant in eukaryotic genomes and are depleted in hydrophobic residues and enriched in polar and charged residues. We found higher amino acid polarity at the C-terminus of PADRE proteins, in agreement with high disorder probability in this region. The peculiar composition and folding properties of IDRs confer specific functional properties (Sun et al., 2013; Uversky, 2013). First, IDRs are generally able to establish protein-protein interactions with multiple partners and are commonly found in hub proteins in eukaryotic networks. One paradigmatic example in plant immunity is RPM1-interacting protein 4 (RIN4) which interacts with multiple plant resistance proteins and bacterial effectors (Sun et al., 2014). In line with this property, PADRE proteins were shown experimentally to interact with calmodulins (Popescu et al., 2007) and response regulators (Dortay et al., 2008). Our co-expression network also suggests a high degree of connectivity for PADRE genes. Screening for protein-protein interactions involving PADRE proteins should prove an insightful avenue for future research. Second, IDRs are highly accessible regions and can therefore undergo complex regulations by post-translational modifications. For instance, Remorins are plant-specific proteins with a role in plant immunity (Raffaele et al., 2009; Bozkurt et al., 2014) containing structured and disordered regions, with their IDRs harboring multiple phosphorylation sites (Marín and Ott, 2012; Marín et al., 2012; Perraki et al., 2018). RIN4 also undergoes multiple post-translational modifications and regulation by proteolysis (Toruño et al., 2019). Similarly, we identified multiple phosphorylated residues in the C-terminal region of PADRE proteins, as well as degradation signals. Third, the ability to undergo a disorder-to-order transition can confer transient functionality to IDRs, such as membrane binding in Remorins (Perraki et al., 2012), cytotoxic activity of *Bordetella* CyaA toxin (O'Brien et al., 2018), and protein complex formation by cAMP response element-binding (CREB)

protein (Arai et al., 2015). We could then speculate that every PADRE protein could adopt several functions according to their cellular environment.

Toward a Functional Understanding of PADRE Family

We report the significant induction of 31 PADRE genes upon inoculation by *S. sclerotiorum*, including 14 *AtPADRE*, 7 *RcPADRE*, and 10 *SIPADRE* genes. Radiation of the PADRE family into 10 phylogenetic groups could provide the basis for some degree of functional diversification. In line with hypothesis, Yang et al. (2020) identified 3 *AtPADRE* genes induced upon osmotic stress, 4 upon salt stress, and 5 upon cold stress. Our work revealed an intrinsically disordered region in PADRE proteins, suggesting that PADRE gene function could be context-dependent. This could explain why Yang et al. (2020) found several *AtPADRE* genes mis-regulated by salt while none were significantly responding to NaCl in the RNA sequencing dataset we analyzed (Suzuki et al., 2016). The identification of multiple subcellular localization signals in PADRE proteins (N-myristoylation, NLS, endocytic vesicles) prevents predictions regarding the site of PADRE action. The use of fluorescent protein reporter fusions in a structure-function analysis will be required to this end. This approach shall be challenging given the presence of targeting signals at both ends of the PADRE domain. We found six *AtPADRE* genes induced upon inoculation by several fungal pathogens with a necrotrophic lifestyle (*S. sclerotiorum*, *B. cinerea*, and *A. brassicicola*), a bacterial pathogen (*P. syringae* pv. *tomato*), and a hemibiotrophic root-infecting fungus (*V. dahliae*), indicating that pathogens are very potent inducers of *AtPADRE* genes. The PADRE co-expression network included several important players in plant immunity such as the syntaxin SYP122 (Zhang et al., 2007), the C2-domain protein BAP1 (Yang et al., 2006), the patatin-like protein 2 PLP2 (La Camera et al., 2009), members of the RPM1-interacting protein 4 RIN4 (At3g48450), wall-associated kinase-like WAKL10 (At1g79680), and the nematode resistance protein-like HSPRO2 (At2g40000).

These findings are consistent with a role for members of the PADRE family in disease resistance.

MATERIALS AND METHODS

Pfam Domain Annotation and Enrichment Analyses

Pfam domains were annotated using hmmscan 3.1b1 with e-value threshold $1E-10$ against the Pfam-A 32.0 database. Enrichment of Pfam domains among genes induced after *S. sclerotiorum* infection was analyzed using a two-proportion Z-test in R. *Arabidopsis thaliana* gene expression from GEO accession GSE106811 (Badet et al., 2017) was used in this analysis. Briefly, total RNA was extracted from the edge of developed necrotic lesions of leaves from 4-week-old plants inoculated by *S. sclerotiorum* strain 1980, as described in Peyraud et al. (2019). Samples were collected in triplicates from three plants in independent inoculation experiments. RNA sequencing was performed on an Illumina HiSeq 2500 instrument as described in Badet et al. (2017). A composite enrichment score taking into account the significance of the Z-test and the enrichment ratio was calculated with the formula $R_Z(i) * R_r(i)$, where $R_Z(i)$ is the normalized rank of domain i for the Z-test p -value and $R_r(i)$ is the normalized rank of domain i for the enrichment ratio.

RNA Sequencing Data Analysis

Raw data for RNA sequencing experiments used in this work is available in the NCBI Gene Expression Omnibus (GEO) database with accession numbers provided in Table 1. All raw datasets were processed separately with DESeq2 to calculate normalized read counts (Basemean) and \log_2 fold change (LFC) of expression and identify genes differentially expressed between control and treated samples. Genes were considered differentially expressed for $LFC \geq 1.5$ and adjusted p -value ≤ 0.01 . In the multiple stress analysis, raw data were used for statistical analysis and LFC values were normalized for the heatmap, as follows: $(\pm)\log_2(1 + |LFC|)$.

Taxonomic Distribution of PADRE Proteins

We used MAFFT Version 7.407 (Katoh et al., 2002) to align the *Arabidopsis thaliana* DUF4228 protein sequences using default parameters. After manual curation, 24 *A. thaliana* DUF4228 proteins expressed in our RNA sequencing data (GSE106811) were kept for further analysis. This alignment (Supplementary Datasheet S1) was used in a phmmer search on the HMMER webserver² (Potter et al., 2018) against the UniProt References Proteomes in UniProtKB (The Uniprot Consortium, 2019). The search was carried out using parameters $-E\ 1e-10$ $-\text{domE}\ 1$ $-\text{incE}\ 1e-10$ $-\text{incdomE}\ 0.03$ $-\text{seqdb}\ \text{uniprotrefprot}$ identifying 3467 significant sequence hits. The 'target length' and length of the target alignment from the output of the HMM search were used to compare total protein length and DUF4228 domain length (Figure 2A). Genome sizes and total number of genes per genomes were obtained from the Phytozome 12.1 database

(Goodstein et al., 2011). Ploidy levels were obtained from the Plant DNA C-values Database on the Kew Royal Botanic Gardens website (Plant DNA C-values Database | Royal Botanic Gardens, Kew) and from the original genome papers. The timetree was generated on <http://www.timetree.org/> (Kumar et al., 2017) using species names as input. Polyploidization events described in the literature were collected from Wang et al. (2019) and Xu et al. (2019). DUF4228 proteins in complete plant genomes were identified using hmmsearch against a local instance of the Phytozome 12.1 proteome database, using the same parameters as previously.

Phylogenetic Analysis of PADRE Proteins

We extracted DUF4228 proteins from *Marchantia polymorpha*, *Physcomitrella patens*, *Sphagnum fallax*, *Selaginella moellendorffii*, *Amborella trichopoda*, *Brachypodium distachyon*, *Setaria italica*, *Aquilegia coerulea*, *Solanum lycopersicum*, *Medicago truncatula*, *Arabidopsis thaliana*, *Ricinus communis*, *Vitis vinifera*, and *Theobroma cacao* from our hmmsearch against Phytozome 12.1. Prior to alignment, we removed sequence Pp3c24_13210 for having <40 amino acids and truncated the 650, 650, and 1650 N-terminal amino acids from Solyc05g013500, Thecc1EG010515, and Medtr8g069400, respectively. A first sequence alignment was performed in ClustalO (Madeira et al., 2019), and 10 sequences were removed for being too divergent, leaving 344 sequences (Supplementary Datasheet S2). These sequences were aligned with ClustalO, and the alignment was manually edited in Jalview to keep positions with no gap in at least 172/344 sequences, yielding a final alignment of 116 amino acids long (Supplementary Datasheet S3). Phylogenetic relationships were determined by a maximum likelihood approach using PhyML (Guindon et al., 2010) with aLRT branch support in phylogeny.fr (Dereeper et al., 2008), with no alignment and no alignment curation steps (Supplementary Datasheet S4), using the LG substitution model (Le and Gascuel, 2008) and a gamma distribution with four categories. The resulting tree had a log-likelihood of -48028.9 and gamma shape parameter 1.720. The tree was rooted on *M. polymorpha* Mapoly0024s0004 and rendered with FigTree³ v1.4.3. Phylogenetic groups were defined based on the most ancestral branch with support ≥ 0.9 . The rates of birth and death of PADRE domains were calculated using BadiRate 1.35 (Librado et al., 2012) with parameters $-\text{bmodel}\ \text{FR}$ $-\text{ep}\ \text{CML}$ $-\text{family}$.

Bioinformatics Analyses of PADRE Sequence Features

Conserved motifs were identified in the 28 *A. thaliana* PADRE proteins using the alignment provided in Supplementary Datasheet S3 and rendered using WebLogo 3 (Crooks et al., 2004). ELMs were identified using the ELM webserver (Gouw et al., 2018) with *A. thaliana* as species and subcellular localization not specified. Phosphorylated peptides were identified with a 'Basic search' in 'Experiment data' in the PhosPhAt 4.0 database (Durek et al., 2009). Intrinsic disorder probability was calculated using the PrDOS webserver

²<https://www.ebi.ac.uk/Tools/hmmer/>

³<http://tree.bio.ed.ac.uk/software/figtree/>

(Ishida and Kinoshita, 2007) with a false-positive rate of 5%. Grantham residue polarity was determined using the ProtScale tool in ExPASy (Gasteiger et al., 2005) with a window size of 9.

Reconstruction of PADRE Co-expression Network

The co-expression network was built using the NetworkDrawer tool in ATTED-II version 9.2 (Obayashi et al., 2018) using the Ath-m version C7.1 platform including 14,668 microarray samples, with the Coex option “Add many genes” and PPI option “Add a few genes.” The resulting network was rendered in Cytoscape 3.6.1 (Shannon et al., 2003). Gene expression LFC upon *S. sclerotiorum* corresponds to the *A. thaliana* RNA sequencing data from Badet et al. (2017) (GSE106811), with LFC values provided as node attribute table in Cytoscape (Shannon et al., 2003). The modularity of the network was computed by the algorithm proposed by Blondel et al. (2008). Gene ontologies associated with subnetworks were determined using the GO enrichment analysis online tools⁴. Cutoff on FDR was set at 1E-2.

DATA AVAILABILITY STATEMENT

RNA sequencing read datasets are available from the NCBI Gene Expression Omnibus (GEO) database with accession numbers GSE106811, GSE138039, GSE66290, GSE83478, GSE104590, GSE70094, GSE116269, GSE72548, GSE56922, and GSE72806. The datasets generated by the analyses presented in this study are included in the article **Supplementary Material**.

AUTHOR CONTRIBUTIONS

MD, MK, and SR performed phylogenetic analyses. MD, LG, and SR performed gene expression analyses. AB and SR performed co-expression network analysis. SR conceived and designed the study. All authors contributed to writing the manuscript draft, reviewed the manuscript, and approved the final article.

FUNDING

This work was supported by a Ph.D. grant from INRAE SPE Division to MD, and the French Laboratory of Excellence Project TULIP (Grants ANR-10-LABX-41 and ANR-11-IDEX-0002-02).

⁴geneontology.org

REFERENCES

- Akiyama, K., Ueda, H., Yoda, H., and Sano, H. (2005). A novel basic pathogenesis-related protein from tobacco plants. *Plant Biotechnol.* 22, 157–159. doi: 10.5511/plantbiotechnology.22.157
- Albahlal, W. S., Obomighie, I., Blein, T., Persad, R., Chernukhin, I., Crespi, M., et al. (2018). *Arabidopsis* HEAT SHOCK TRANSCRIPTION FACTOR1b regulates multiple developmental genes under benign and stress conditions. *J. Exp. Bot.* 69, 2847–2862. doi: 10.1093/jxb/ery142
- Arai, M., Sugase, K., Dyson, H. J., and Wright, P. E. (2015). Conformational propensities of intrinsically disordered proteins influence the mechanism of

The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We thank the INRAE SPE division and Labex TULIP community for stimulating discussions and support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00491/full#supplementary-material>

FIGURE S1 | Genomic distribution and synteny analysis of DUF4228 genes in *A. thaliana*. For each gene, the phylogenetic clade and the level of expression (\log_2 fold change) upon infection by *S. sclerotiorum* is provided. Genes significantly differentially expressed upon infection are indicated with a star (*). Lines in the center of the graph show PADRE synteny blocks. The quartet AT2G23690/AT4G37240/AT5G66580/AT3G50800 is reported as two gene pairs in Yang et al. (2020), the quartet AT1G71015/AT2G01340/AT1G66480/AT5G37840 was not associated with gene duplication by Yang et al. (2020).

FIGURE S2 | Modularity of genes included in the AtPADRE co-expression network. Genes with values higher than average are labeled on the graph. Boxplots show first and third quartiles (box), median (thick line) and the most dispersed values within 1.5 times the interquartile range (whiskers).

TABLE S1 | PFAM domains enriched in genes induced upon *S. sclerotiorum* inoculation.

TABLE S2 | Expression of 24 PADRE genes from *A. thaliana* under a range of biotic and abiotic stress, with mention of their inclusion in the list of genes upregulated upon *S. sclerotiorum* inoculation, down-regulated upon *S. sclerotiorum* inoculation and top 10% most induced genes after infection by *S. sclerotiorum*.

TABLE S3 | Expression of 74 PADRE genes from *A. thaliana*, *S. lycopersicum* and *R. communis* upon *S. sclerotiorum* inoculation determined by RNA sequencing.

DATASHEET S1 | Sequence alignment of 24 *A. thaliana* DUF4228 proteins produced by MAFFT and used for phmmer search, in fasta format.

DATASHEET S2 | Full-length sequence of 344 PADRE proteins from 13 Embryophyte species in fasta format.

DATASHEET S3 | Sequence alignment of 344 PADRE proteins from 13 Embryophyte species produced by ClustalO and used for phylogenetic analyses, in fasta format.

DATASHEET S4 | Maximum likelihood phylogenetic tree of 344 PADRE proteins from 13 Embryophyte species, in Newick format.

DATASHEET S5 | Co-expression gene network shown in **Figures 6B,C**, in .xgmml format.

binding and folding. *Proc. Natl. Acad. Sci. U.S.A.* 112, 9614–9619. doi: 10.1073/pnas.1512799112

Badet, T., Léger, O., Barascud, M., Voisin, D., Sadon, P., Vincent, R., et al. (2019). Expression polymorphism at the ARPC 4 locus links the actin cytoskeleton with quantitative disease resistance to *Sclerotinia sclerotiorum* in *Arabidopsis thaliana*. *New Phytol.* 222, 480–496. doi: 10.1111/nph.15580

Badet, T., Voisin, D., Mbengue, M., Barascud, M., Sucher, J., Sadon, P., et al. (2017). Parallel evolution of the POQR prolyl oligo peptidase gene conferring plant quantitative disease resistance. *PLoS Genet.* 13:e1007143. doi: 10.1371/journal.pgen.1007143

- Bhandari, D. D., Lapin, D., Kracher, B., von Born, P., Bautor, J., Niefind, K., et al. (2019). An EDS1 heterodimer signalling surface enforces timely reprogramming of immunity genes in *Arabidopsis*. *Nat. Commun.* 10:772. doi: 10.1038/s41467-019-08783-0
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* 2008:10008. doi: 10.1088/1742-5468/2008/10/P10008
- Boisson, B., Giglione, C., and Meinel, T. (2003). Unexpected protein families including cell defense components feature in the N-myristoylome of a higher eukaryote. *J. Biol. Chem.* 278, 43418–43429. doi: 10.1074/jbc.M307321200
- Boland, G. J., and Hall, R. (1994). Index of plant hosts of *Sclerotinia sclerotiorum*. *Can. J. Plant Pathol.* 16, 93–108.
- Bozkurt, T. O., Richardson, A., Dagdas, Y. F., Mongrand, S., Kamoun, S., and Raffaele, S. (2014). The plant membrane-associated REMORIN1.3 accumulates in discrete periaustorial domains and enhances susceptibility to phytophthora infestans. *Plant Physiol.* 165, 1005–1018. doi: 10.1104/pp.114.235804
- Corwin, J. A., Copeland, D., Feusier, J., Subedy, A., Eshbaugh, R., Palmer, C., et al. (2016). The quantitative basis of the *Arabidopsis* innate immune system to endemic pathogens depends on pathogen genetics. *PLoS Genet.* 12:e1005789. doi: 10.1371/journal.pgen.1005789
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Derbyshire, M., Mbengue, M., Barascud, M., Navaud, O., and Raffaele, S. (2019). Small RNAs from the plant pathogenic fungus *Sclerotinia sclerotiorum* highlight host candidate genes associated with quantitative disease resistance. *Mol. Plant Pathol.* 20, 1279–1297. doi: 10.1111/mpp.12841
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008). Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36, W465–W469. doi: 10.1093/nar/gkn180
- Dodds, P. N., and Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.* 11, 539–548. doi: 10.1038/nrg2812
- Doerks, T., Strauss, M., Brendel, M., and Bork, P. (2000). GRAM, a novel domain in glucosyltransferases, myotubularins and other putative membrane-associated proteins. *Trends Biochem. Sci.* 25, 483–485. doi: 10.1016/s0968-0004(00)01664-9
- Dortay, H., Gruhn, N., Pfeifer, A., Schwerdtner, M., Schmülling, T., and Heyl, A. (2008). Toward an interaction map of the two-component signaling pathway of *Arabidopsis thaliana*. *J. Proteome Res.* 7, 3649–3660. doi: 10.1021/pr0703831
- Durek, P., Schmidt, R., Heazlewood, J. L., Jones, A., MacLean, D., Nagel, A., et al. (2009). PhosphoAT: The *Arabidopsis thaliana* phosphorylation site database. An update. *Nucleic Acids Res.* 38, D828–D834. doi: 10.1093/nar/gkp810
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995
- Eulgem, T. (2005). Regulation of the *Arabidopsis* defense transcriptome. *Trends Plant Sci.* 10, 71–78. doi: 10.1016/j.tplants.2004.12.006
- Fordyce, R. F., Soltis, N. E., Caseys, C., Gwinner, R., and Corwin, J. A. (2018). Digital imaging combined with genome-wide association mapping links loci to plant–pathogen interaction traits. *Plant Physiol.* 178, 1406–1422. doi: 10.1104/pp.18.00851
- Fukuoka, S., Saka, N., Koga, H., Ono, K., Shimizu, T., Ebana, K., et al. (2009). Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science* 325, 998–1001. doi: 10.1126/science.1175550
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). “Protein identification and analysis tools on the ExPASy server,” in *The Proteomics Protocols Handbook*, ed. J. M. Walker (Totowa, NJ: Humana Press), 571–607. doi: 10.1385/1-59259-584-7:531
- Gatta, A. T., Wong, L. H., Sere, Y. Y., Calderón-Noreña, D. M., Cockcroft, S., Menon, A. K., et al. (2015). A new family of StART domain proteins at membrane contact sites has a role in ER-PM sterol transport. *eLife* 4, 1–46. doi: 10.7554/eLife.07253
- Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* 43, 205–227. doi: 10.1146/annurev.phyto.43.040204.135923
- Goodacre, N. F., Gerloff, D. L., and Uetz, P. (2013). Protein domains of unknown function are essential in bacteria. *mBio* 5:e00744-13. doi: 10.1128/mBio.00744-13
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2011). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Gouw, M., Michael, S., Sámano-Sánchez, H., Kumar, M., Zeke, A., Lang, B., et al. (2018). The eukaryotic linear motif resource - 2018 update. *Nucleic Acids Res.* 46, D428–D434. doi: 10.1093/nar/gkx1077
- Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. *Science* 185, 862–864. doi: 10.1126/science.185.415.4.862
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Habermann, B. (2004). The BAR-domain family of proteins: a case of bending and binding? *EMBO Rep.* 5, 250–255. doi: 10.1038/sj.embor.7400105
- Hacquard, S., Kracher, B., Hiruma, K., Münch, P. C., Garrido-Oter, R., Thon, M. R., et al. (2016). Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nat. Commun.* 7:11362. doi: 10.1038/ncomms11362
- Hegedus, D. D., and Rimmer, S. R. (2005). *Sclerotinia sclerotiorum*: when “to be or not to be” a pathogen? *FEMS Microbiol. Lett.* 251, 177–184. doi: 10.1016/j.femsle.2005.07.040
- Ishida, T., and Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 35, W460–W464. doi: 10.1093/nar/gkm363
- Jaroszewski, L., Li, Z., Krishna, S. S., Bakolitsa, C., Wooley, J., Deacon, A. M., et al. (2009). Exploration of uncharted regions of the protein universe. *PLoS Biol.* 7:e1000205. doi: 10.1371/journal.pbio.1000205
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kersting, A. R., Bornberg-Bauer, E., Moore, A. D., and Grath, S. (2012). Dynamics and adaptive benefits of protein domain emergence and arrangements during plant genome evolution. *Genome Biol. Evol.* 4, 316–329. doi: 10.1093/gbe/evs004
- Khafif, M., Balagué, C., Huard-Chauveau, C., and Roby, D. (2017). An essential role for the VAST domain of the *Arabidopsis* VAD1 protein in the regulation of defense and cell death in response to pathogens. *PLoS One* 12:e0179782. doi: 10.1371/journal.pone.0179782
- Khafif, M., Cottret, L., Balagué, C., and Raffaele, S. (2014). Identification and phylogenetic analyses of VAST, an uncharacterized protein domain associated with lipid-binding domains in Eukaryotes. *BMC Bioinformatics* 15:222. doi: 10.1186/1471-2105-15-222
- Krattinger, S. G., Lagudah, E. S., Spielmeier, W., Singh, R. P., Huerta-Espino, J., McFadden, H., et al. (2009). A Putative ABC Transporter Confers Durable Resistance to Multiple Fungal Pathogens in Wheat. *Science* 323, 1360–1363. doi: 10.1126/science.1166453
- Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. (2017). Timetree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* 34, 1812–1819. doi: 10.1093/molbev/msx116
- Kumar Sudheeran, P., Ovadia, R., Galsarker, O., Maoz, I., Sela, N., Maurer, D., et al. (2019). Glycosylated flavonoids: fruit’s concealed antifungal arsenal. *New Phytol.* 225, 1788–1798. doi: 10.1111/nph.16251
- La Camera, S., Balagué, C., Göbel, C., Geoffroy, P., Legrand, M., Feussner, I., et al. (2009). The *Arabidopsis* patatin-like protein 2 (PLP2) plays an essential role in cell death execution and differentially affects biosynthesis of oxylipins and resistance to pathogens. *Mol. Plant Microbe Interact.* 22, 469–481. doi: 10.1094/MPMI-22-4-0469
- Le, S. Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25, 1307–1320. doi: 10.1093/molbev/msn067
- Leitch, I. J., Johnston, E., Pellicer, J., Hidalgo, O., and Bennett, M. D. (2019). *Plant DNA C-values database* (release 7.1, Apr 2019) Available online at: <https://cvalues.science.kew.org/>
- Librado, P., Vieira, F. G., and Rozas, J. (2012). BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics* 28, 279–281. doi: 10.1093/bioinformatics/btr623
- Liu, S., Kracher, B., Ziegler, J., Birkenbihl, R. P., and Somssich, I. E. (2015). Negative regulation of ABA signaling by WRKY33 is critical for *Arabidopsis*

- immunity towards *Botrytis cinerea* 2100. *eLife* 4:e07295. doi: 10.7554/eLife.07295
- Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641. doi: 10.1093/nar/gkz268
- Majeran, W., Le Caer, J. P., Ponnala, L., Meinel, T., and Giglione, C. (2018). Targeted profiling of *Arabidopsis thaliana* subproteomes illuminates co- and posttranslationally N-terminal myristoylated proteins. *Plant Cell* 30, 543–562. doi: 10.1105/tpc.17.00523
- Marin, M., and Ott, T. (2012). Phosphorylation of intrinsically disordered regions in remorin proteins. *Front. Plant Sci.* 3:86. doi: 10.3389/fpls.2012.00086
- Marin, M., Thallmair, V., Ott, T., Marin, M., Thallmair, V., and Ott, T. (2012). The intrinsically disordered N-terminal region of AtREM1.3 remorin protein mediates protein-protein interactions. *J. Biol. Chem.* 287, 39982–39991. doi: 10.1074/jbc.M112.414292
- Mbengue, M., Navaud, O., Peyraud, R., Barascud, M., Badet, T., Vincent, R., et al. (2016). Emerging trends in molecular interactions between plants and the broad host range fungal pathogens *botrytis cinerea* and *Sclerotinia sclerotiorum*. *Front. Plant Sci.* 7:422. doi: 10.3389/fpls.2016.00422
- Mine, A., Seyferth, C., Kracher, B., Berens, M. L., Becker, D., and Tsuda, K. (2018). The defense phytohormone signaling network enables rapid, high-amplitude transcriptional reprogramming during effector-triggered immunity. *Plant Cell* 30, 1199–1219. doi: 10.1105/tpc.17.00970
- Obayashi, T., Aoki, Y., Tadaka, S., Kagaya, Y., and Kinoshita, K. (2018). ATTED-II in 2018: a plant coexpression database based on investigation of the statistical property of the mutual rank index. *Plant Cell Physiol.* 59:e3. doi: 10.1093/pcp/pcx191
- O'Brien, D. P., Perez, A. C. S., Karst, J., Cannella, S. E., Enguéné, V. Y. N., Hessel, A., et al. (2018). Calcium-dependent disorder-to-order transitions are central to the secretion and folding of the CyaA toxin of *Bordetella pertussis*, the causative agent of whooping cough. *Toxicon* 149, 37–44. doi: 10.1016/j.toxicon.2018.01.007
- Pearson, W. R., and Sierk, M. L. (2005). The limits of protein sequence comparison? *Curr. Opin. Struct. Biol.* 15, 254–260. doi: 10.1016/j.sbi.2005.05.005
- Perchev, L., Balagué, C., Riou, C., Claudel-Renard, C., Rivière, N., Grezes-Besset, B., et al. (2010). Nitric oxide participates in the complex interplay of defense-related signaling pathways controlling disease resistance to *Sclerotinia sclerotiorum* in *Arabidopsis thaliana*. *Mol. Plant Microbe Interact.* 23, 846–860. doi: 10.1094/MPMI-23-7-0846
- Perraki, A., Cacas, J.-L., Crowet, J.-M., Lins, L., Castroviejo, M., German-Retana, S., et al. (2012). Plasma membrane localization of *Solanum tuberosum* Remorin from group 1, homolog 3 is mediated by conformational changes in a novel C-terminal anchor and required for the restriction of potato virus X movement. *Plant Physiol.* 160, 624–637. doi: 10.1104/pp.112.200519
- Perraki, A., Gronnier, J., Gouguet, P., Boudsocq, M., Deroubaix, A. F., Simon, V., et al. (2018). REM1.3's phospho-status defines its plasma membrane nanodomain organization and activity in restricting PVX cell-to-cell movement. *PLoS Pathog.* 14:e1007378. doi: 10.1371/journal.ppat.1007378
- Peyraud, R., Mbengue, M., Barbacci, A., and Raffaele, S. (2019). Intercellular cooperation in a fungal plant pathogen facilitates host colonization. *Proc. Natl. Acad. Sci. U.S.A.* 116, 3193–3201. doi: 10.1073/pnas.1811267116
- Poland, J. A., Balint-Kurti, P. J., Wissner, R. J., Pratt, R. C., and Nelson, R. J. (2009). Shades of gray: the world of quantitative disease resistance. *Trends Plant Sci.* 14, 21–29. doi: 10.1016/j.tplants.2008.10.006
- Popescu, S. C., Popescu, G. V., Bachan, S., Zhang, Z., Seay, M., Gerstein, M., et al. (2007). Differential binding of calmodulin-related proteins to their targets revealed through high-density *Arabidopsis* protein microarrays. *Proc. Natl. Acad. Sci. U.S.A.* 104, 4730–4735. doi: 10.1073/pnas.0611615104
- Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi: 10.1093/nar/gky448
- Raffaele, S., Bayer, E., Lafarge, D., Cluzet, S., Retana, S. G., Boubekour, T., et al. (2009). Remorin, a solanaceae protein resident in membrane rafts and plasmodesmata, impairs potato virus X movement. *Plant Cell Online* 21, 1541–1555. doi: 10.1105/tpc.108.064279
- Raffaele, S., Mongrand, S., Gamas, P., Niebel, A., and Ott, T. (2007). Genome-wide annotation of remorins, a plant-specific protein family: evolutionary and functional perspectives. *Plant Physiol.* 145, 593–600. doi: 10.1104/pp.107.108639
- Rajarammohan, S., Pradhan, A. K., Pental, D., and Kaur, J. (2018). Genome-wide association mapping in *Arabidopsis* identifies novel genes underlying quantitative disease resistance to *Alternaria brassicae*. *Mol. Plant Pathol.* 19, 1719–1732. doi: 10.1111/mpp.12654
- Rausch, S. (2016). The function of *Arabidopsis* microRNAs in defense against the necrotrophic fungal pathogen *Alternaria brassicicola*. *Biologie* 7, 1–126. doi: 10.15496/publikation-16913
- Roux, F., Voisin, D., Badet, T., Balagué, C., Barlet, X., Huard-Chauveau, C., et al. (2014). Resistance to phytopathogens e tutti quanti: placing plant quantitative disease resistance on the map. *Mol. Plant Pathol.* 15, 427–432. doi: 10.1111/mpp.12138
- Scholz, S. S., Schmidt-Heck, W., Guthke, R., Furch, A. C. U., Reichelt, M., Gershenzon, J., et al. (2018). Verticillium dahliae-*Arabidopsis* interaction causes changes in gene expression profiles and jasmonate levels on different time scales. *Front. Microbiol.* 9:217. doi: 10.3389/fmicb.2018.00217
- Shanks, C. M., Rice, J. H., Zubo, Y., Schaller, G. E., Hewezi, T., and Kieber, J. J. (2016). The Role of cytokinin during infection of *Arabidopsis thaliana* by the cyst nematode heterodera schachtii. *Mol. Plant Microbe Interact.* 29, 57–68. doi: 10.1094/MPMI-07-15-0156-R
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software Environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sun, X., Greenwood, D. R., Templeton, M. D., Libich, D. S., McGhie, T. K., Xue, B., et al. (2014). The intrinsically disordered structural platform of the plant defence hub protein RPM1-interacting protein 4 provides insights into its mode of action in the host-pathogen interface and evolution of the nitrate-induced domain protein family. *FEBS J.* 281, 3955–3979. doi: 10.1111/febs.12937
- Sun, X., Rikkerink, E. H. A., Jones, W. T., and Uversky, V. N. (2013). Multifarious roles of intrinsic disorder in proteins illustrate its broad impact on plant biology. *Plant Cell* 25, 38–55. doi: 10.1105/tpc.112.106062
- Suzuki, N., Bassil, E., Hamilton, J. S., Inupakutika, M. A., Zandalinas, S. I., Tripathy, D., et al. (2016). ABA is required for plant acclimation to a combination of salt and heat stress. *PLoS One* 11:e0147625. doi: 10.1371/journal.pone.0147625
- Tavidou, E., Pireyre, M., and Ulm, R. (2020). Degradation of the transcription factors PIF4 and PIF5 under UV-B promotes UVR8-mediated inhibition of hypocotyl growth in *Arabidopsis*. *Plant J.* 101, 507–517. doi: 10.1111/tpj.14556
- Tenhaken, R., Doerks, T., and Bork, P. (2005). DCD - A novel plant specific domain in proteins involved in development and programmed cell death. *BMC Bioinformatics* 6:169. doi: 10.1186/1471-2105-6-169
- The Uniprot Consortium. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/nar/gky1049
- Toruño, T. Y., Shen, M., Coaker, G., and Mackey, D. (2019). Regulated disorder: posttranslational modifications control the RIN4 plant immune signaling hub. *Mol. Plant Microbe Interact.* 32, 56–64. doi: 10.1094/MPMI-07-18-0212-FI
- Uversky, V. N. (2013). A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci.* 22, 693–724. doi: 10.1002/pro.2261
- Wang, J., Qin, J., Sun, P., Ma, X., Yu, J., Li, Y., et al. (2019). Polyploidy index and its implications for the evolution of polyploids. *Front. Genet.* 10:807. doi: 10.3389/fgene.2019.00807
- Weidenbach, D., Esch, L., Möller, C., Hensel, G., Kumlehn, J., Höfle, C., et al. (2016). Polarized defense against fungal pathogens is mediated by the jacalin-related lectin domain of modular poaceae-specific proteins. *Mol. Plant* 9, 514–527. doi: 10.1016/j.molp.2015.12.009
- Xu, X., Yang, Y., Liu, C., Sun, Y., Zhang, T., Hou, M., et al. (2019). The evolutionary history of the sucrose synthase gene family in higher plants. *BMC Plant Biol.* 19:566. doi: 10.1186/s12870-019-2181-4
- Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K., and Yokoyama, S. (2013). DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. *Trends Plant Sci.* 18, 267–276. doi: 10.1016/j.tplants.2012.09.001
- Yang, H., Li, Y., and Hua, J. (2006). The C2 domain protein BAP1 negatively regulates defense responses in *Arabidopsis*. *Plant J.* 48, 238–248. doi: 10.1111/j.1365-3113.2006.02869.x

- Yang, Q., Niu, X., Tian, X., Zhang, X., Cong, J., Wang, R., et al. (2020). Comprehensive genomic analysis of the DUF4228 gene family in land plants and expression profiling of ATDUF4228 under abiotic stresses. *BMC Genomics* 21:12. doi: 10.1186/s12864-019-6389-3
- Zhang, M., Mo, H., Sun, W., Guo, Y., and Li, J. (2016). Systematic isolation and characterization of cadmium tolerant genes in tobacco: a cDNA library construction and screening approach. *PLoS One* 11:e0161147. doi: 10.1371/journal.pone.0161147
- Zhang, Z., Feechan, A., Pedersen, C., Newman, M. A., Qiu, J. L., Olesen, K. L., et al. (2007). A SNARE-protein has opposing functions in penetration resistance and defence signalling pathways. *Plant J.* 49, 302–312. doi: 10.1111/j.1365-313X.2006.02961.x
- Zorzatto, C., MacHado, J. P. B., Lopes, K. V. G., Nascimento, K. J. T., Pereira, W. A., Brustolini, O. J. B., et al. (2015). NIK1-mediated translation suppression functions as a plant antiviral immunity mechanism. *Nature* 520, 679–682. doi: 10.1038/nature14171
- Zuther, E., Schaarschmidt, S., Fischer, A., Erban, A., Pagter, M., Mubeen, U., et al. (2019). Molecular signatures associated with increased freezing tolerance due to low temperature memory in *Arabidopsis*. *Plant Cell Environ.* 42, 854–873. doi: 10.1111/pce.13502

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 Didelon, Khafif, Godiard, Barbacci and Raffaele. 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.



Enzymatic Functions for Toll/Interleukin-1 Receptor Domain Proteins in the Plant Immune System

Adam M. Bayless and Marc T. Nishimura*

Department of Biology, Colorado State University, Fort Collins, CO, United States

OPEN ACCESS

Edited by:

Takaki Maekawa,
Max Planck Institute for Plant
Breeding Research, Germany

Reviewed by:

Brian Staskawicz,
University of California, Berkeley,
United States
Xin Li,
The University of British Columbia,
Canada

*Correspondence:

Marc T. Nishimura
marcusn@colostate.edu

Specialty section:

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

Received: 16 February 2020

Accepted: 04 May 2020

Published: 02 June 2020

Citation:

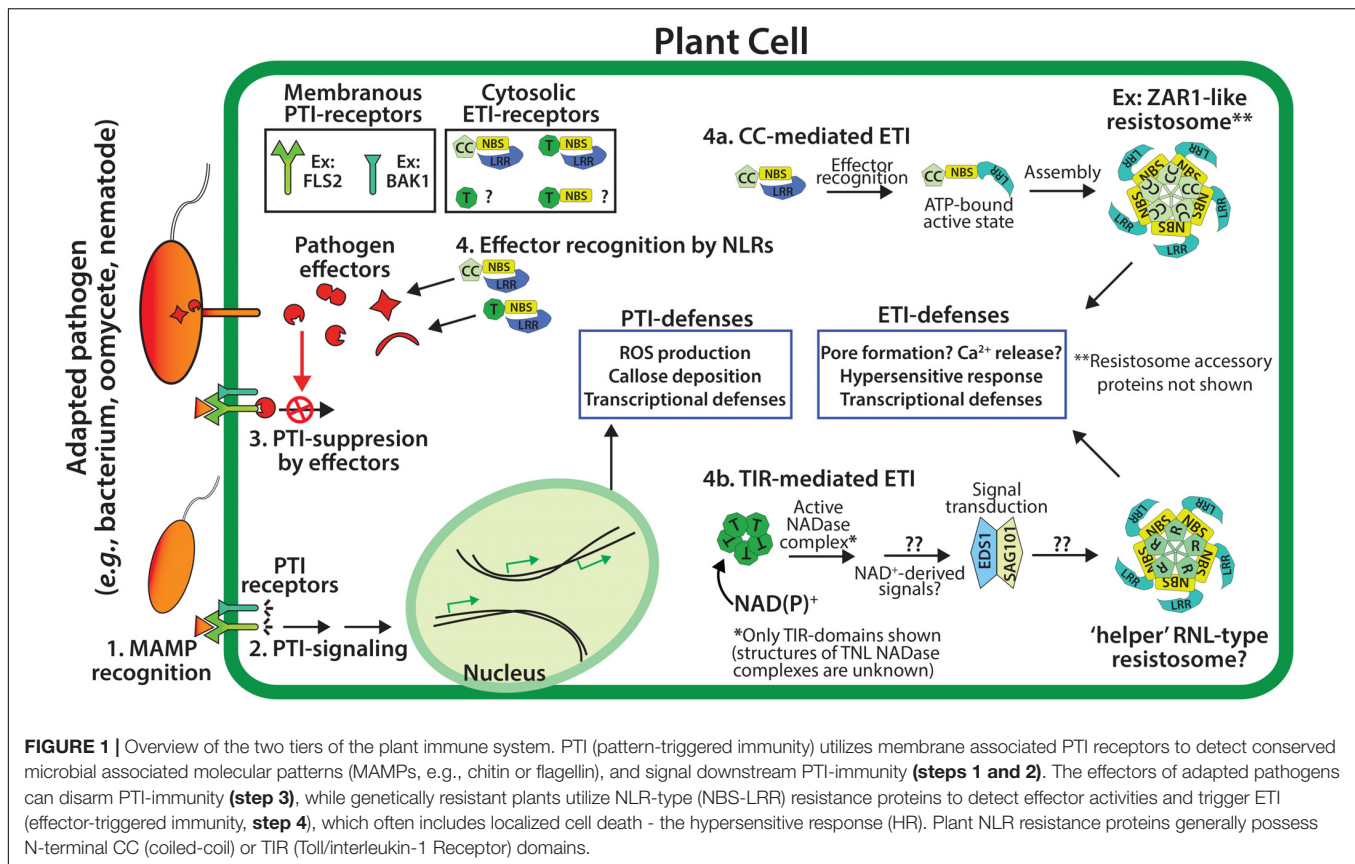
Bayless AM and Nishimura MT
(2020) Enzymatic Functions
for Toll/Interleukin-1 Receptor Domain
Proteins in the Plant Immune System.
Front. Genet. 11:539.
doi: 10.3389/fgene.2020.00539

Rationally engineered improvements to crop plants will be needed to keep pace with increasing demands placed on agricultural systems by population growth and climate change. Engineering of plant immune systems provides an opportunity to increase yields by limiting losses to pathogens. Intracellular immune receptors are commonly used as agricultural disease resistance traits. Despite their importance, how intracellular immune receptors confer disease resistance is still unknown. One major class of immune receptors in dicots contains a Toll/Interleukin-1 Receptor (TIR) domain. The mechanisms of TIR-containing proteins during plant immunity have remained elusive. The TIR domain is an ancient module found in archaeal, bacterial and eukaryotic proteins. In animals, TIR domains serve a structural role by generating innate immune signaling complexes. The unusual animal TIR-protein, SARM1, was recently discovered to function instead as an enzyme that depletes cellular NAD⁺ (nicotinamide adenine dinucleotide) to trigger axonal cell death. Two recent reports have found that plant TIR proteins also have the ability to cleave NAD⁺. This presents a new paradigm from which to consider how plant TIR immune receptors function. Here, we will review recent reports of the structure and function of TIR-domain containing proteins. Intriguingly, it appears that TIR proteins in all kingdoms may use similar enzymatic mechanisms in a variety of cell death and immune pathways. We will also discuss TIR structure–function hypotheses in light of the recent publication of the ZAR1 resistosome structure. Finally, we will explore the evolutionary context of plant TIR-containing proteins and their downstream signaling components across phylogenies and the functional implications of these findings.

Keywords: Toll/interleukin-1 receptor, TIR, NLR, NADase, innate immunity

THE PLANT IMMUNE SYSTEM

Single and multicellular organisms have evolved numerous defenses to ward off biotic challenges. The plant innate immune system consists of receptor proteins that monitor both extracellular and intracellular pathogen-related signals to activate defenses (**Figure 1**). Typically, extracellular signals are transduced across the plasma membrane by an extensive array of receptor-like kinase (RLK) and receptor-like proteins (RLPs) (Boutrot and Zipfel, 2017; Tang et al., 2017). Disease resistance conferred by the RLK/RLP pattern recognition receptor (PRR) system is triggered by a wide array of apoplastic molecules from microbes, pathogens and host damage signals. Accordingly, pathogens have evolved to extensively target PRR pathways to promote host susceptibility. A common strategy of plant pathogens is to deliver intracellular



virulence proteins (such as type III effector proteins) in order to disrupt PRR-based defense (Jones and Dangl, 2006; Dangl et al., 2013). These virulence proteins are necessary for pathogenicity, and thus serve as reliable indicators of pathogen presence. In response to pathogen immunosuppression, plants have evolved a second layer of innate immune receptors that directly or indirectly recognize the presence of pathogen virulence proteins (Jones and Dangl, 2006; Qi and Innes, 2013). As such, virulence proteins are the tools that pathogens use to suppress the host immune system, but also the signals that plants of the correct genotype (i.e., resistant plants) can recognize to reinitiate a defense response. These intracellular receptors are characterized by nucleotide-binding site (NBS) domains and a C-terminal Leucine-rich repeat (LRR). This combination of domains is present in both plant and animal NLR proteins (confusingly referring to both “NBS-LRR” and “Nod-like receptors (Nod: N-terminal oligomerization domain).” While plant and animal NLR proteins are functionally conserved in many ways, it appears that they are the product of convergent evolution (Urbach and Ausubel, 2017).

The recognition of intracellular pathogen virulence molecules promotes conformational changes in NLR proteins (Takken and Govers, 2012). The N-terminal domain of NLR proteins has signaling activities, while the C-terminal NBS-LRR domains negatively regulate signaling in the resting state. The NBS domain functions as a molecular switch depending on the bound nucleotide: ADP-bound in the resting state and ATP-bound

in the active state (Takken and Govers, 2012). Both plant and animal NLRs are auto-regulated and self-associate during signal transduction, however, the N-terminal signaling domains of plant and animal NLRs are distinct (Qi and Innes, 2013; Hu et al., 2015; Nanson et al., 2019). Generally, plant NLRs contain N-terminal TIR (Toll/Interleukin Receptor-1) or CC (coiled coil) domains, and are therefore known as TNLs or CNLs (Qi and Innes, 2013). Monocot genomes appear to lack TNL loci, however, both monocots and dicots can encode TIR-only and TIR-NBS proteins (Meyers et al., 2002; Collier et al., 2011; Nandety et al., 2013; Nishimura et al., 2017; Gao et al., 2018). TIR and CC-domains from plant NLRs are sufficient to activate immune outputs, including a localized cell death termed the hypersensitive response (HR), and transcriptional defense programs (Swiderski et al., 2009; Collier et al., 2011). The self-association and oligomerization of either TIR or CC-domains is required for plant immune signaling, however, the downstream events which follow the activation of TIR or CC resistance proteins has remained unclear (Casey et al., 2016; Wan et al., 2019; Wang et al., 2019a).

DOWNSTREAM COMPONENTS OF TIR-SIGNALING PATHWAYS IN PLANTS

Genetic screens have identified two families of proteins that appear universally required for plant TIR phenotypes

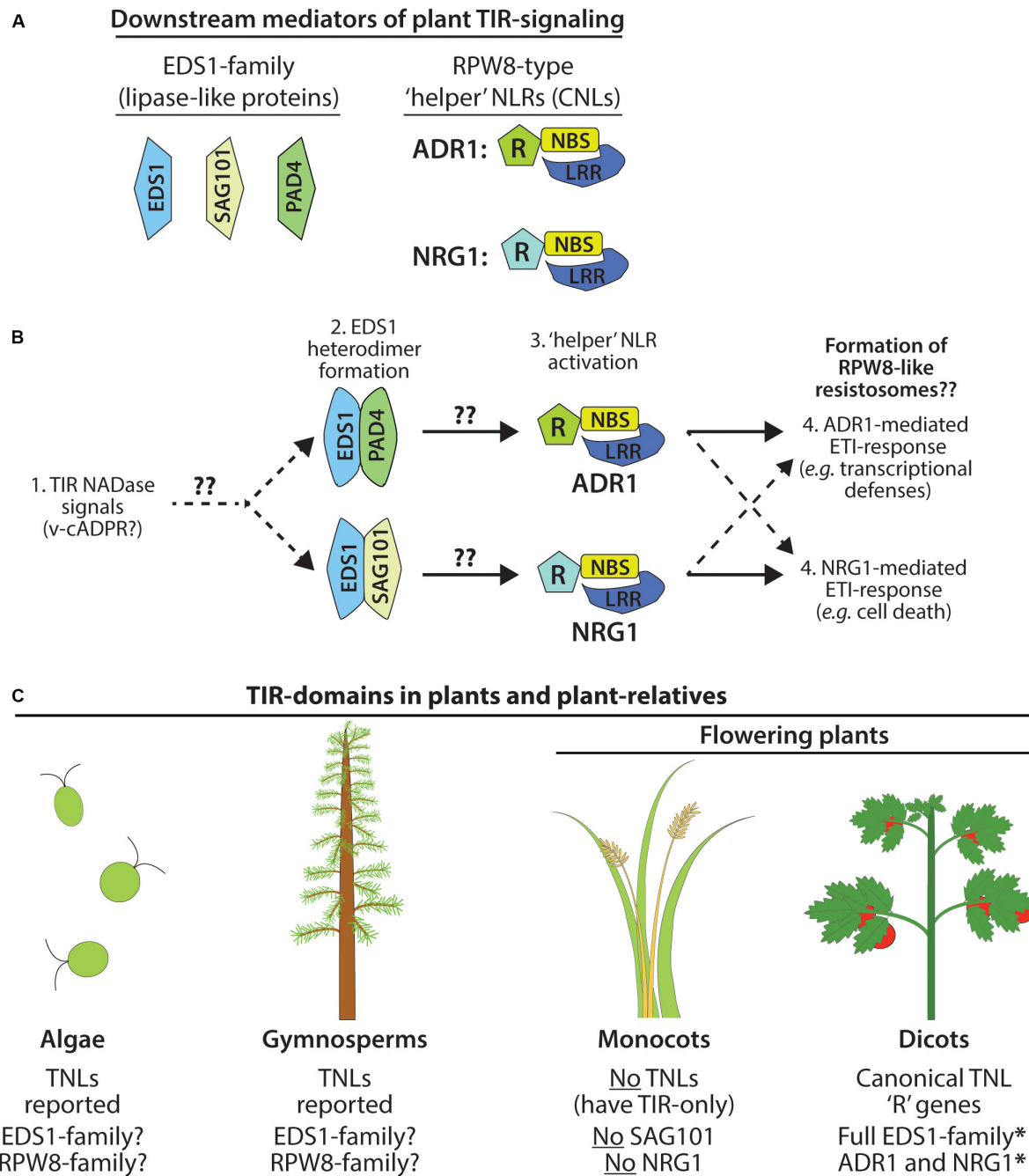


FIGURE 2 | Immune signaling by plant TIR NADases requires downstream components. **(A)** Members of the EDS1 lipase-like family: EDS1, SAG101, PAD4. Terminal mediators of TIR-signal transduction are the RPW8-type 'helper' RNLs: ADR1 and/or NRG1. **(B)** Model of ETI-pathway activation by plant TIR NADases. Perception of plant TIR signals (e.g., v-cADPR?) promotes EDS1-family heterodimerization, and subsequent activation of the 'helper' RNLs, ADR1 or NRG1. EDS1-PAD4 heterodimers may favor activation of ADR1-mediated responses (transcriptional defense programs), while EDS1-SAG101 heterodimers activate NRG1-mediated responses (cell death). Functional redundancy among NRG1 and ADR1 indicated by dashed arrows. **(C)** TIR-domain containing proteins, including TNLs, are found in the genomes of phylogenetically distant plant-lineages and in the relatives of land plants, including green algae (Sun et al., 2014; Shao et al., 2019), as well as gymnosperms (western white pine) and the moss, *Physcomitrella patens*. Monocots do not encode TNLs and lack two downstream mediators of TIR-immune signaling: SAG101 and NRG (Collier et al., 2011; Lapin et al., 2019).

(Figure 2). The first component is the EDS1 (Enhanced Disease Susceptibility 1) family of lipase-like proteins [EDS1, SAG101 (Senescence-Associated Gene 101), and PAD4 (Phytoalexin

Deficient 4)] (Feys et al., 2005; Lapin et al., 2019). The second component, the RPW8 class of 'helper' CNLs, often referred to as 'RNLs,' functions downstream of the EDS1 family (Peart et al.,

2005; Qi et al., 2018; Jubic et al., 2019; Wu et al., 2019). Helper NLRs such as the NRG1 (N-requirement Gene 1) and the ADR1 family (Activated Disease Resistance 1) are candidates for being the ultimate output of TIR pathways (Collier et al., 2011; Qi et al., 2018; Jubic et al., 2019). How these downstream components are activated by TIR oligomerization, and the organization of the overall pathway, remains a major unanswered question (Jubic et al., 2019; Lapin et al., 2019; Wan et al., 2019).

EDS1 forms exclusive heterodimers with either PAD4 or SAG101 to relay TIR-immune signals (Feys et al., 2005; Wagner et al., 2013). EDS1 and PAD4 are also reported to function in plant basal defenses and salicylic acid signaling (Cui et al., 2017). The crystal structure of the EDS1-SAG101 heterodimer suggests that binding of the N-terminal lipase-like domains establishes unique interaction interfaces at the C-terminal EP domain (Wagner et al., 2013). The C-terminal EP-domain of EDS1-members contains positively charged residues and is essential for transduction of TIR-signals (Bhandari et al., 2019; Lapin et al., 2019). The TNL RPS4 (Resistance to *Pseudomonas syringae* 4), as well as particular TIR-NBS proteins, have been reported to associate with EDS1, as has the 'helper' NLR, NRG1 (Heidrich et al., 2011; Nandety et al., 2013; Huh et al., 2017; Qi et al., 2018). The functional consequences of these physical interactions are unknown. Lapin et al. (2019) determined that the EDS1-members of *Solanaceous* species could complement a *N. benthamiana* mutant which lacks all EDS1-family members. However, the orthologous EDS1-members of *Arabidopsis* did not complement, suggesting that within species, EDS1-members may have co-evolved a high degree of specificity in the relay of TIR-signals (Lapin et al., 2019). Curiously, in the absence of downstream 'helper' NLRs, EDS1-members can still mediate limited transcriptional defense programs from an auto-active version of the TNL, Roq1 (Recognition of XopQ1) (Qi et al., 2018).

The expression of the RPW8-domains of ADR1 or NRG1 is sufficient to trigger HR, even in *eds1* null backgrounds, placing 'helper' RNLs as downstream mediators of TIR-signaling (Collier et al., 2011; Qi et al., 2018). Typically, plant genomes carry relatively few loci encoding helper RNLs, consistent with a conserved RNL function that integrates inputs channeled from upstream TNL receptors via EDS1-complexes. Additionally, functional redundancy between ADR1 and NRG1 has been reported (Castel et al., 2019; Jubic et al., 2019; Lapin et al., 2019; Wu et al., 2019). Some CNLs are also reported to signal through ADR1, suggesting that cross-talk might occur at the endpoints of certain CNL and TNL-signal pathways (Castel et al., 2019; Wu et al., 2019). The RPW8-domain of helper RNLs does share similarities with the CC-domain of CNLs; thus, the recent structure of the ZAR1 (HOPZ-ACTIVATED RESISTANCE 1, a CNL) resistosome may provide insights into the functions of the ADR1 and NRG1 helper NLRs (Wang et al., 2019a,b). The active ZAR1 complex assembles into a ring-shaped pentamer, the "resistosome," and hypothetically disrupts cell membrane integrity with a pore-forming channel (Wang et al., 2019a,b).

The mechanisms of how plant NLRs activate downstream immunity is an active area of research. While TIR-TIR interactions are well known to promote animal immune signaling

via scaffold function, a new paradigm of plant TIR function has recently emerged: signal competent plant TIR-domains are NAD⁺-(nicotinamide adenine dinucleotide)-hydrolyzing enzymes (**Figures 3A–D**) (Horsefield et al., 2019; Wan et al., 2019). Below, we review recent advances in the understanding of plant TIR-domain structure, evolution, and enzymatic (NADase) function. We also draw insights from the TIR-NADases encoded by animals and prokaryotes, and explore how the newly reported structure of the ZAR1 CNL 'resistosome' complex might inform the high order complexes of plant TIR-NADases.

TIR-DOMAINS: A CELLULAR DEFENSE MODULE FOUND IN ALL DOMAINS OF LIFE

Toll/Interleukin Receptor-1 (TIR)-domain containing proteins are found in all domains of life – Eukarya, Bacteria, and Archaea (**Figures 4A,B**) (Essuman et al., 2018). Frequently, TIR-domain containing proteins function in immunity or cell death decisions in bacteria, plants and animals, suggesting an ancient role in cellular defenses (**Figures 3A,B, 4A,B**). The core TIR-domain is typically ~120–200 residues, and is found in multi-domain and single domain proteins (Nimma et al., 2017). TIR-domains generally require TIR-TIR self-associations for function, and TIR-domains can also participate in heterotypic protein interactions. The sequence identity of TIR-domains among different species may be as low as 20–30%, however, TIR-domains share a flavodoxin-like fold, consisting of parallel beta-sheets and alpha-helices with interconnecting loops (Ve et al., 2015).

INSIGHTS TO PLANT TIR FUNCTION FROM ANIMAL SYSTEMS: SARM1 (STERILE ALPHA AND TIR MOTIF-CONTAINING 1) IS AN NADase

Typically, animal TIRs (e.g., Toll-like receptors, MyD88) couple pathogen detection to defense gene activation by nucleating the formation of large multimeric signaling complexes (**Figure 3A**) (Xu et al., 2000; O'Neill and Bowie, 2007; Kenny and O'Neill, 2008; Nimma et al., 2017). Crystal structures for numerous animal TIR-domains have acted as guides for a biochemical dissection of TIR-domain function (Xu et al., 2000; Valkov et al., 2011; Bovijn et al., 2012). The crystal structure of the TIR-domain from Toll Like Receptor 2 (TLR2) revealed residues required for TIR-TIR interactions, and the core TIR-domain structure of parallel beta-sheets and alpha-helices (Xu et al., 2000). Additional structural studies of TIR-adaptor proteins further defined TIR interfaces required for multimerization and signal complex formation (Nyman et al., 2008; Valkov et al., 2011; Bovijn et al., 2012). Animal TIR scaffolding can signal various defensive outputs, such as inflammatory responses and cytokine production (**Figure 2A**) (O'Neill and Bowie, 2007). In contrast, the unusual animal TIR protein SARM1 (*sterile alpha and TIR motif-containing 1*) was recently found to have a surprising enzymatic function (Essuman et al., 2017).

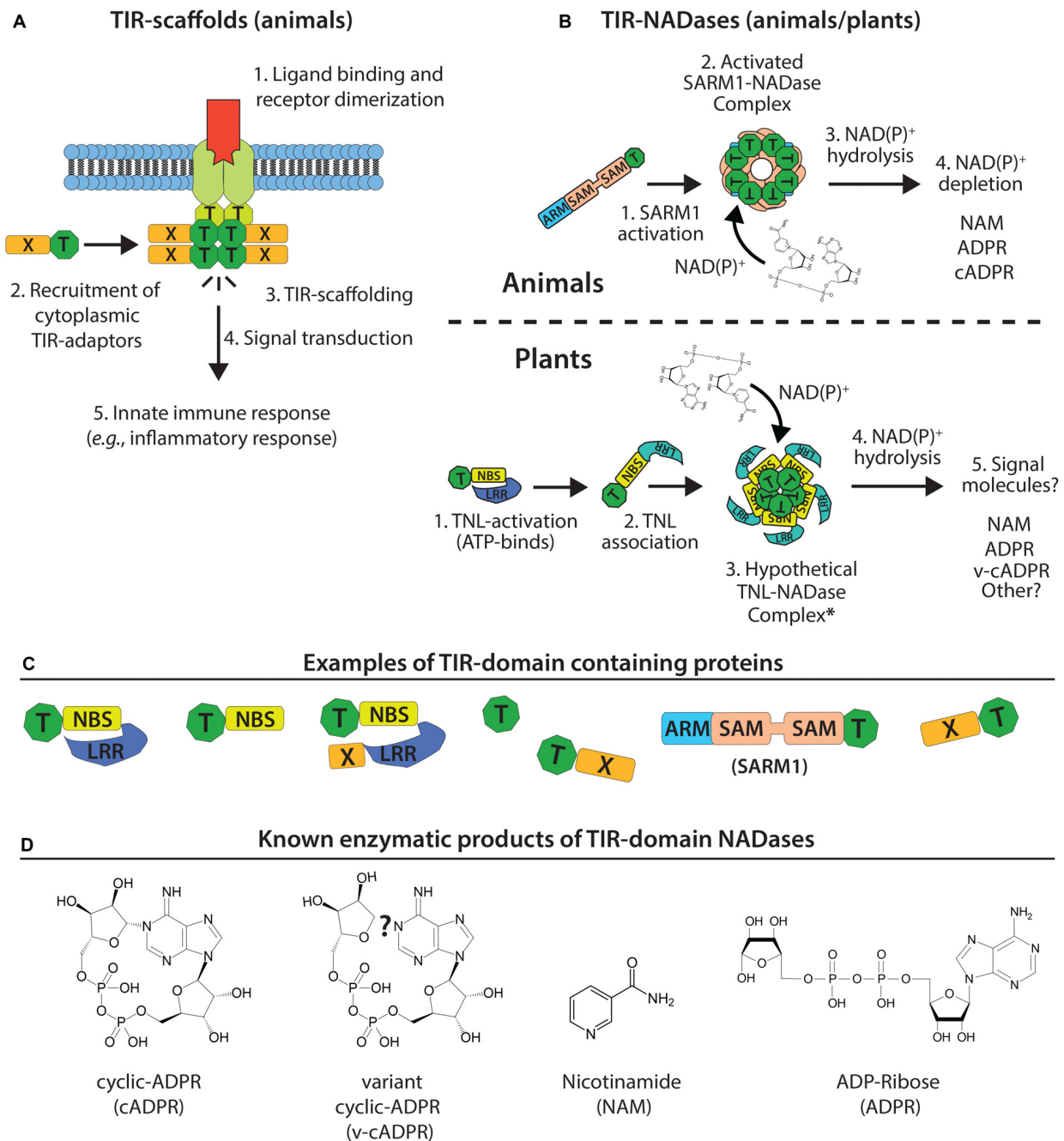


FIGURE 3 | Model of TIR-domain scaffolding (animals) and TIR-NADase activity (plants and animals). **(A)** Canonical TIR-scaffold function in animals: TIR-TIR interactions promote signal complex formation and innate immune signal transduction. **(B)** Top: animal TIR NADases (e.g., SARM1) assemble into high order complexes, and hydrolyze NAD(P)⁺ substrate and alter NAD(P)⁺ pools. Bottom: assembly of plant TIR-domains into hypothetical NADase complex (resistosome-like?) and generation of immunomodulatory signals. **(C)** Numerous TIR-domain configurations are present in animal, plant, and bacterial proteins. Plant TIR-domains are often found in modular NBS-LRRs, TIR-NBS, TIR-X or TIR-only proteins. -X corresponds to atypical or undefined domains. The animal SARM1 TIR is located at the C-terminus; the SARM1 SAM-domains promote oligomerization. **(D)** Known products of TIR NADases; plant TIRs produce variant cyclic-ADPR (v-cADPR), whose structure is currently unknown.

The animal TIR protein SARM1 functions in axon degeneration, an active process of programmed cell death in response to injury (classically known as “Wallerian degeneration”) (Gerdt et al., 2015). NAD⁺-depletion had

been associated with axon degeneration, but the SARM1-regulated NADase had remained elusive. The critical observation that the TIR domain of SARM1 is structurally similar to bacterial nucleotide-processing enzymes led to the

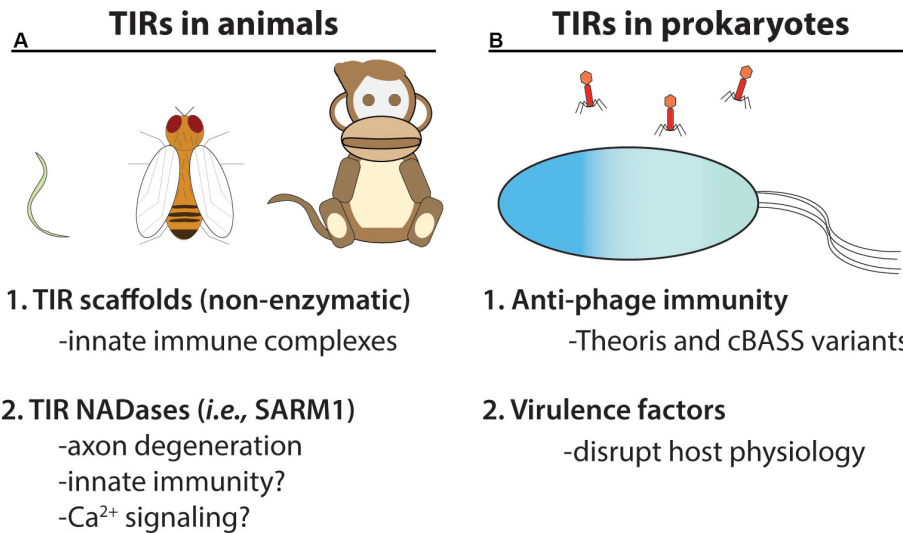


FIGURE 4 | TIR NADases were recently reported in animals and prokaryotes. **(A)** Diverse eukaryotic organisms, including invertebrates (e.g., *C. elegans*, *D. melanogaster*) and vertebrates, utilize TIR-domain containing proteins in cellular innate immunity. Non-enzymatic TIR-domain containing proteins in animals promote signal complex formation via TIR – TIR interactions. The SARM1 NADase TIR from animals functions in axon degeneration, and is reported to function in immunity in *C. elegans* (Shivers et al., 2009). **(B)** Numerous bacteria and archaeal species encode TIR-NADases. Prokaryotic TIR-domain containing proteins are reported to function in anti-phage immunity (Theoris system and variants of cBASS) (Doron et al., 2018; Cohen et al., 2019). TIR-domains from pathogenic bacteria are reported to function in virulence (Alaidarous et al., 2014).

recognition that the SARM1 TIR has an intrinsic enzymatic activity: NAD⁺-hydrolase function (**Figure 3B**) (Gerdt et al., 2015; Essuman et al., 2017). Axon degeneration requires SARM1 TIR domain NADase activities (Essuman et al., 2017). The unusual enzymatic activity of SARM1 TIR relative to other animal TIR domains is perhaps reflected in an unusual evolutionary history, as the SARM1 TIR appears to have been horizontally transferred into animals (Zhang et al., 2011). TIRs that function in canonical TLR pathways (TLR4 and MyD88) do not have NADase activity, although the family has not been exhaustively tested (Essuman et al., 2017).

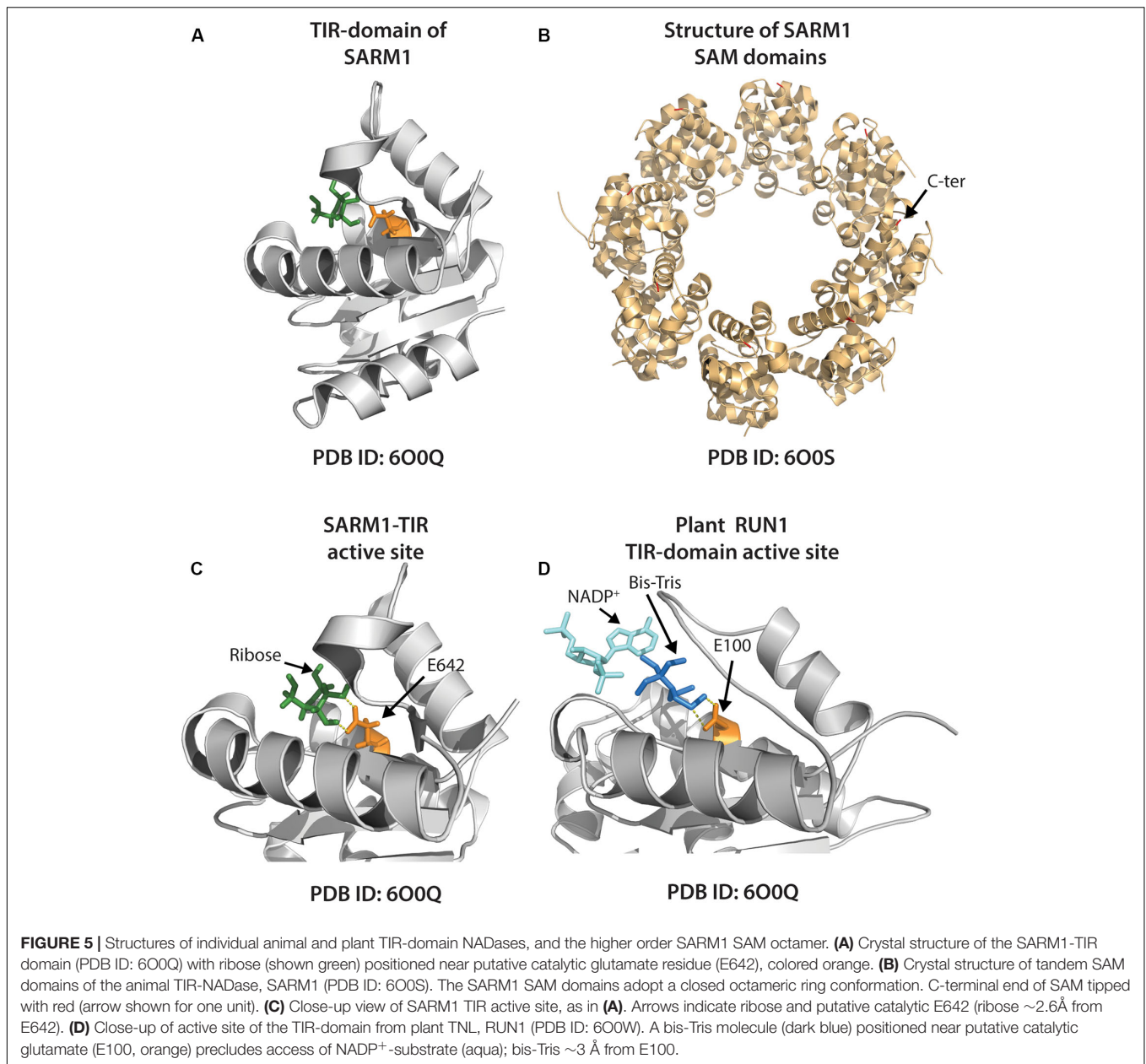
Like NLRs, SARM1 is a multidomain TIR protein that is auto-inhibited. SARM1 has two tandem sterile alpha (SAM) domains, which enable oligomerization, and an N-terminal Armadillo domain, which is required for auto-inhibition (**Figures 3B,C**) (Essuman et al., 2018). SARM1 TIR NADase function is dependent upon oligomerization and TIR-TIR associations. The mechanism of activation during axon degeneration is unclear, but NADase activity of SARM1 can be enhanced by phosphorylation or treatment with a cell-permeant mimetic of nicotinamide mononucleotide, an NAD⁺ precursor (Murata et al., 2018; Zhao et al., 2019).

NAD⁺-hydrolysis by SARM1 generates ADPR (ADP-ribose), cyclic ADPR (c-ADPR) and NAM (nicotinamide) (Essuman et al., 2017) (see **Figure 3D**). The products of SARM1-mediated NAD⁺-hydrolysis (cADPR, ADPR) are known Ca²⁺ mobilization agents and may thus effect cellular Ca²⁺ signaling (Lee, 2012; Guse, 2015; Lee and Zhao, 2019; Zhao et al., 2019). SARM1 readily hydrolyzes NADP⁺ as well as NAD-analogs with substitutions to the adenine ring, such as amino

group additions (Essuman et al., 2017). However, FAD (flavin adenine dinucleotide) and NADH or NAD-analogs lacking the amino group of the nicotinamide ring could not be hydrolyzed (Essuman et al., 2017, 2018). Depending on local cellular pH, SARM1 is also reported to generate NAAD (nicotinic acid adenine dinucleotide) (Zhao et al., 2019).

A recent crystal structure of the SARM1 TIR reveals conservation with both plant and prokaryotic TIR-domains (Horsefield et al., 2019). The active site of the SARM1 TIR-domain includes a conserved glutamic acid (E642) which is required for NAD⁺-hydrolysis (**Figure 5A**). Recent crystal and cryo-EM structures of SARM1 complexes, and of the tandem SAM-domains, indicate that the active SARM1 NADase complex forms a ring-shaped octamer (Horsefield et al., 2019; Sporny et al., 2019) (**Figure 5B**). The crystal structure of the SARM1 TIR active site revealed close proximity of ribose with the putative catalytic glutamate (E642) and may suggest potential substrate-active site interactions (**Figure 5C**) (Horsefield et al., 2019). The exact catalytic mechanism of SARM1 is unknown, but appears distinct from CD38, which also produces cADPR from NAD⁺ (Loring et al., 2020).

Strikingly, SARM1 triggers cell death when transiently expressed in the leaves of the plant, *Nicotiana benthamiana* (Horsefield et al., 2019; Wan et al., 2019). Like axon degeneration, plant cell death triggered by SARM1 requires NADase function, however, SARM1-mediated cell death occurs independently of the known plant TIR-signaling components EDS1 and NRG1 (Horsefield et al., 2019; Wan et al., 2019). Notably, supplementation of exogenous NAD⁺ reduces axon degeneration mediated by SARM1 (Gerdt et al., 2015). As such, SARM1 depletion of cellular NAD(P)⁺ is likely to underlie both animal



axon degeneration and plant cell death resulting from its ectopic expression. However, some cell lines are reported to tolerate low levels of SARM1 expression (Lee and Zhao, 2019; Zhao et al., 2019). Whether low level SARM1 activity in particular cellular contexts might generate signaling molecules vs. deplete cellular NAD⁺ stores, is not yet clear.

TIR NADases IN PROKARYOTES: PHAGE IMMUNE SYSTEMS AND VIRULENCE FACTORS

Numerous bacterial and archaeal species encode TIR-domain containing proteins, primarily of unknown function (Spear et al.,

2009; Doron et al., 2018; Essuman et al., 2018). However, some prokaryotic TIRs are reported to function in anti-phage immunity, while other TIRs may act as virulence factors which manipulate host responses (Figure 4B) (Alaidarous et al., 2014; Doron et al., 2018; Coronas-Serna et al., 2019). TIR-domains encoded by *Brucella* and *Paracoccus* are reported to mimic animal TIR-adaptors and disrupt TLR immune signaling, potentially via physical interactions with animal TIR domains (Chan et al., 2009; Alaidarous et al., 2014; Snyder et al., 2014). However, many apparently non-pathogenic bacteria encode TIR-proteins, suggesting that some TIR-domains could possess functions outside of virulence or immunity (Spear et al., 2009). NAD⁺-hydrolase activities have recently been shown for several bacterial and archaeal TIRs, and

thus, it has been suggested that ancestrally, the TIR-domain belongs to a large family of nucleotide hydrolase enzymes (Essuman et al., 2018).

Like the SARM1 TIR NADase, all examined prokaryotic TIRs also require the putative catalytic glutamate for NADase function (Essuman et al., 2018). Prokaryotic TIRs are likely to also require TIR-TIR self-associations, as local crowding (via TIR protein laden beads) enhanced NADase function (Essuman et al., 2018). Prokaryotic TIR-domains show variation in terms of NAD⁺-hydrolysis kinetics, as well as in the type and ratio of products produced from NAD⁺-hydrolysis (Essuman et al., 2017). For example, the TirS TIR domain from *Staphylococcus aureus* generated ADPR and cADPR, while the TcpO TIR domain from the archaea *Methanobrevibacter olleyae* produced a novel product initially termed metabolite X, which is likely a variant of cyclic ADPR (v-cADPR), whose structure remains unresolved (Essuman et al., 2017; Wan et al., 2019).

Recent studies from the Sorek lab may provide a glimpse into the origins of TIR-mediated immunity (**Figure 4B**) (Doron et al., 2018; Cohen et al., 2019). A survey of tens of thousands of prokaryotic genomes, coupled with functional screening, unveiled multiple new classes of anti-phage defense systems. Among these, an anti-phage system termed Thoeris, was found in ~2,000 bacterial and archaeal genomes (Doron et al., 2018). The Thoeris defense operon encodes an NAD⁺ binding protein (ThsA) and a TIR-domain protein (ThsB). Both ThsA and B are required for anti-phage immunity. Amino acid alignment of the ThsB TIR-domain with the SARM1-TIR indicated conservation of the catalytic glutamate (Doron et al., 2018). We used Phyre2 to model the *B. amyloliquefaciens* encoded ThsB (*BaThsB*), and retrieved a top-match (60% identity, 100% confidence) to the crystal structure (PDB ID: 3HYN) of a putative signal transduction factor from *Agathobacter rectalis* (**Figures 6A,B**). A comparison of the SARM1 TIR and plant RPS4 TIR structures with the *BaThsB* TIR-domain model indicates positional conservation of the putative catalytic glutamate (**Figure 6A**). The putative catalytic glutamate (E99) of ThsB was required for phage protection, suggesting that TIR domains may have an ancient enzymatic-based immune function (Doron et al., 2018). It will be interesting to assess if Thoeris functions via NAD⁺-depletion, akin to SARM1, or could generate NAD⁺-derived immunomodulatory signals.

The Sorek group further reported that some prokaryote genomes harbor an ortholog of the cGAS-STING defense system found in animals (Cohen et al., 2019). Upon detecting invading DNA, cGAS (cyclic GMP-AMP synthase) generates cyclic GMP-AMP (cGAMP) via oligonucleotide cyclase activity. The cGAMP signal then promotes host cell demise through activating a phospholipase which disrupts membrane integrity (Cohen et al., 2019). This prokaryotic system was dubbed CBASS for cyclic oligonucleotide-based anti-phage signaling system. Notably, variants of CBASS-mediated immunity can encode TIR-domains (Cohen et al., 2019). Whether the TIR-domains of particular CBASS variants require NADase function is uncertain. Nonetheless, it is becoming

clear that TIR-mediated immunity to phages is common in both bacteria and archaea. CBASS and Thoeris appear to trigger host cell death prior to the completion of viral replication, thereby restricting phage release into the bacterial population. Elucidating the molecular mechanisms of these prokaryotic TIR-based systems may provide insights into the evolution and function of both immunity and cell death in plants and animals.

TIR NADase ACTIVITY IN PLANTS

Similar to animal SARM1, plant TIRs were recently demonstrated to be NAD⁺ hydrolases, and this NADase activity is required to relay immune signals (Horsefield et al., 2019; Wan et al., 2019). Sequence analysis of the TIR-domain encoding genes from Arabidopsis, as well as ~8,000 TIR sequences found from 108 available plant genomes, indicates high conservation (~90%) of the putative catalytic glutamate required for NADase activity (Wan et al., 2019). The minority of TIR-domains that lack this conserved glutamate appear to be from 'sensor-type' TNLs which function via a signal-competent, genomically paired TNL. These sensor-type TNLs lack the ability to trigger cell death or immunity without their partner TNL (Wan et al., 2019).

Like SARM1 of animals, the NADase activity of plant TIRs was required for TIR-domain function; i.e., to relay immune signals (Horsefield et al., 2019; Wan et al., 2019). *In vitro* NADase cleavage activity was demonstrated by TIR-domains from full length TNLs, as well as TIR-only proteins from dicot plants (Horsefield et al., 2019; Wan et al., 2019). Similar to SARM1 TIR and prokaryotic TIRs, plant TIR-domains could utilize NAD⁺ and NADP⁺ as a substrate, but not the structurally related NAD⁺ precursor NAAD (nicotinic acid adenine dinucleotide) (Essuman et al., 2018; Horsefield et al., 2019; Wan et al., 2019). Intriguingly, a TIR-only protein from the monocot, *Brachypodium distachyon* (BdTIR), also displayed NAD⁺-hydrolysis, in addition to triggering an EDS1/NRG1-dependent HR, suggesting that TIR-immune signaling may be conserved among dicot and monocot plants (Wan et al., 2019). The products generated by plant TIR NADase reactions include NAM, ADPR, and v-cADPR. Unlike the SARM1 TIR, production of cyclic-ADPR by plant TIRs was not detected. v-cADPR has a near identical HPLC retention time and molecular mass to the product of an archaeal TIR, TcpO (Essuman et al., 2018; Wan et al., 2019).

A crystal structure of the plant TIR-domain, RUN1, with bound NADP⁺ substrate was determined by Horsefield et al. (2019) (**Figure 5D**). The putative catalytic glutamate of RUN1 was associated with a molecule of bis-Tris, while NADP⁺ was bound near the periphery of the proposed active site (**Figure 5D**). Accordingly, bis-Tris addition to RUN1 NADase assays inhibited activity, suggesting that bis-Tris association with active site residues may preclude NADP⁺ access and subsequent hydrolysis (Horsefield et al., 2019). How the NAD(P)⁺ substrate interacts with and positions in the active site of plant TIRs during catalysis remains to be determined.

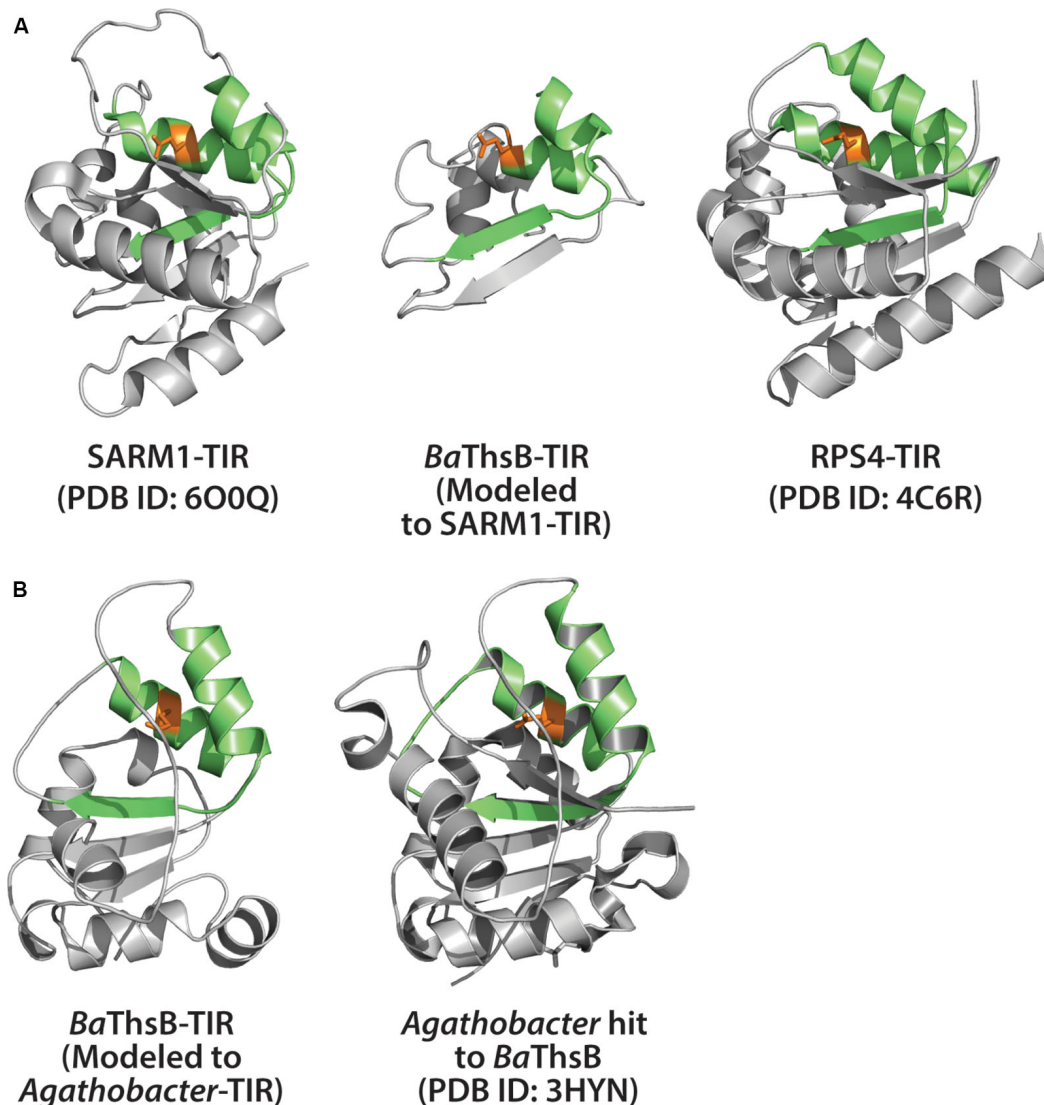
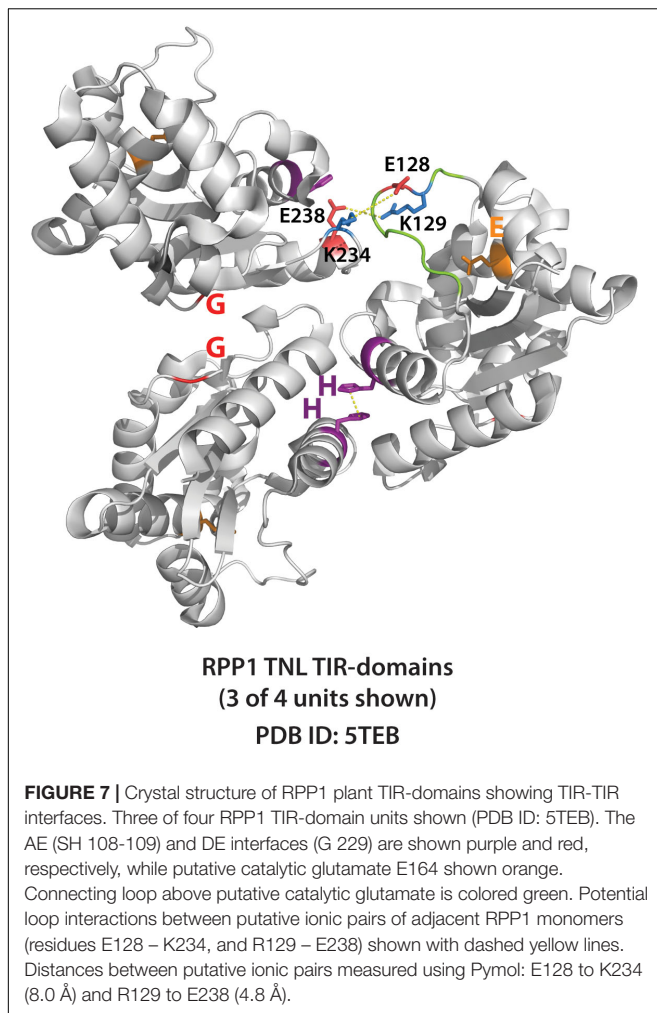


FIGURE 6 | Structural modeling of ThsB (Thoreris TIR) from *Bacillus amyloliquefaciens* (Ba). **(A)** Center: Phyre2 modeling of BaThsB TIR-domain to the SARM1-TIR structure. Phyre2 model confidence: 96.2%; BaThsB-TIR amino acid identity to SARM1-TIR: 15%. Left: Alignment of SARM1-TIR (PDB ID: 6O0Q) to BaThsB-TIR. Right: Alignment of RPS4-TIR structure (PDB ID: 4C6R) to BaThsB-TIR (and to SARM1-TIR). **(B)** Phyre2 modeling of full length BaThsB to a putative signal transduction protein from *Agathobacter rectales* (a putative Thoreris system ThsB TIR). Phyre2 model confidence: 100%; BaThsB amino acid identity to *Agathobacter* ThsB match: 59%.

PLANT TIR-DOMAIN SELF-ASSOCIATION IS NECESSARY FOR NADase ACTIVITY

Plant TIR-TIR self-association occurs through at least two known interfaces formed by pairs of alpha helices (denoted as 'α') (Bernoux et al., 2011; Williams et al., 2014, 2016). Both AE- (i.e., the αA/αE surface) and DE-type (αD/αE surface) helical interfaces are necessary for TIR-TIR self-association, and, subsequent activation of the hypersensitive response. The DE interface was first revealed by the crystal structure of the flax L6 TIR domain (Bernoux et al., 2011). The RRS1 and RPS4 TIR heterodimer crystal indicated TIR-TIR contacts at the AE

interfaces, while the RPP1 crystal revealed both AE and DE contacts (Williams et al., 2014; Zhang et al., 2017). Plant TIR-domains vary in strength of TIR-TIR self-associations and in some cases, self-association strength correlates with function (Schreiber et al., 2016; Zhang et al., 2017). The TIR-only protein, RBA1 (Response to HopBA1), self-associates using both AE and DE interfaces (Nishimura et al., 2017). RBA1 self-association is detectable via co-immunoprecipitation or yeast 2-hybrid assay, and both self-association interfaces must be intact to trigger cell death (Nishimura et al., 2017). Similarly, the isolated TIR-domain of the RPV1 TNL is sufficient to activate HR (Williams et al., 2016). However, self-association of RPV1 TIR-domains was not detectable by yeast two-hybrid analysis



or size exclusion chromatography (Williams et al., 2016), yet disruption of the AE interface did abolish RPV1-mediated HR (Williams et al., 2016). Thus, intact TIR-TIR interfaces appear necessary for TIR-immune function, and can vary in strength. Additionally, the NBS-LRR domains of modular TNLs also promote oligomerization, and whether TIR-only proteins must evolve stronger TIR-TIR interfaces due to lack of NBS-LRR mediated organization is unclear.

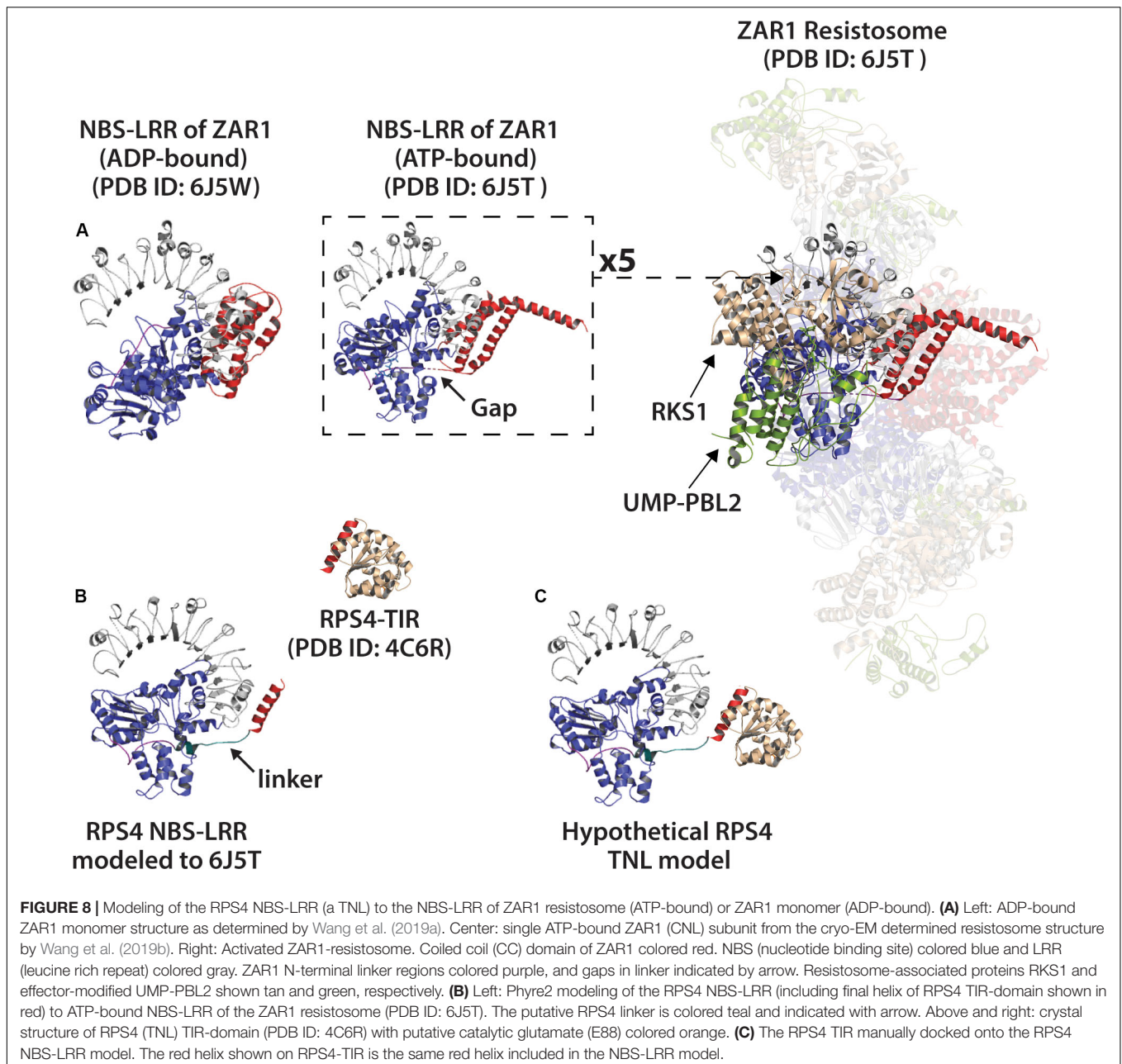
Similar to cell death and disease resistance phenotypes, the activation of plant TIR NADase function requires both AE and DE self-association interfaces (Horsefield et al., 2019; Wan et al., 2019). It seems likely that the NADase activity of plant TIRs is dependent on some higher-order oligomer that has simultaneously engaged both AE and DE interfaces. Intriguingly, the RPP1 crystal structure (Figure 7) suggests that a loop that covers the catalytic glutamate could play this role, as it is positioned near a neighboring monomer only once both interfaces are engaged. Whether or not crystal structures of isolated TIR domains reflect the orientation in the activated TNL context remains to be determined. Currently, no structure of a full length TNL is available, and thus, how TNL oligomerization mediated by the NBS domains influences TIR-TIR associations,

remains unclear. The activation of NADase activity following higher-order TIR oligomerization seems consistent with the behavior of the RBA1 E86A putative catalytic mutant (Wan et al., 2019). RBA1 E86A still self-associates (as measured by co-immunoprecipitation), suggesting that activation of NAD⁺-hydrolysis follows the self-association of TIR-domains.

OLIGOMERIC PLANT “RESISTOSOMES”

The N-terminal coiled coil (CC) domain of some CC-domain type NLRs (e.g., Sr71, NRG1, MLA) can induce HR (Collier et al., 2011; Bai et al., 2012; Casey et al., 2016). Modeling of RPW8-type CC-domains suggests that they may adopt a 4-helix bundle fold similar to that of the mixed-lineage kinase-like protein family of animals, which insert into host membranes and promote cell death (Jubic et al., 2019). Recently, cryo-EM structures for active (ATP-bound) and inactive (ADP-bound) ZAR1 ‘resistosomes’ were determined (Wang et al., 2019a,b). The ZAR1 (HOPZ-ACTIVATED RESISTANCE 1) resistosome complex forms a ring-shaped pentameric structure, and contains bound RKS1 pseudokinase, and an effector-modified kinase, PBL2. The pentameric resistosome structure is driven by the ZAR1 NBS-LRR domains, however, the presence of associated host guarder and adaptor proteins (e.g., RKS1, PBL2) will also influence overall resistosome structure (Wang et al., 2019a). The N-terminal CC-domains of ZAR1 subunits undergo a conformational change, each extending a helix to form a funnel-like structure, which is hypothesized to disrupt membrane integrity and promote cell death (Wang et al., 2019a).

Can the pentameric structure of the ZAR1 resistosome – a CC-domain type NLR – inform what higher order complexes an activated TNL might form? It is enticing to speculate that, like ZAR1 and animal NLRs, an oligomeric TNL NADase complex also forms a ring-shaped resistosome? A variety of stoichiometries are observed for the animal NLR oligomers that form the apoptosome and inflammasome rings (Zhang et al., 2015). The hypothetical TNL resistosome could be of a range of stoichiometries, and most likely forms a ring. However, given the existing structures of plant TIR domains, it seems difficult to reconcile the radial (head to tail) symmetry of a ring-shaped resistosome, no matter the stoichiometry. In these structures, the AE and DE interfaces are in a “head to head” orientation that seems at odds with a circular chain. Perhaps an increase in local concentration of TIR domains is sufficient to promote signaling. Or possibly, these interfaces will not be seen in the context of a full-length TNL oligomer structure. Fusion of the SARM1 SAM domains to either the N-terminus (Horsefield et al., 2019; Wan et al., 2019) or C-terminus (unpublished) of plant TIR-domains enables NADase activity and HR-induction. The SAM domains of SARM1 form an octameric ring (Figure 5B) (Horsefield et al., 2019; Sporny et al., 2019). Even in the context of a fusion protein with forced oligomerization, the RPS4 SAM:TIR protein still requires both AE and DE interfaces (Wan et al., 2019). These results suggest that an octameric ring structure can accommodate plant TIR function, and also that there is surprising flexibility in how functional TIR domain oligomerization can be promoted.



Using Phyre2, we modeled the NBS-LRR domains of RPS4 onto the structures of inactive and active ZAR1 NBS-LRRs (Figure 8). The NBS and N-terminal linker regions of RPS4, as compared with ZAR1, are similar in length and potentially in orientation (Figure 8). While entirely speculative, there would appear to be limits on the amount of rotational flexibility the TIR domains would have in a hypothetical resistosome to engage in simultaneous AE and DE interfaces. The oligomerization state of so-called “paired NLRs” – where individual partners typically assume a ‘sensor’ or ‘signal’ role – may be even more complex. Given that RPS4 and RRS1 appear to function in a complex (Huh et al., 2017), what would the stoichiometry and organization of a hetero-oligomeric resistosome look like? The fact that the

RRS1 TIR lacks a catalytic glutamate makes the situation even more interesting.

Plant TIR-only proteins can signal despite their lack of C-terminal NBS-LRR domains (Nishimura et al., 2017; Wan et al., 2019). In the absence of oligomerizing NBS-LRR domains, what higher order structures might naturally occurring TIR-only proteins form? The TIR-only protein, RBA1, self-associates and requires the conserved AE and DE-type interfaces. Are TIR-only oligomers different than TNL oligomers? RBA1 also requires EDS1 and NRG1, but like TNL receptors there is still no clear mechanistic link between TIR activation and downstream signal transduction (Nishimura et al., 2017; Wan et al., 2019).

HOW MIGHT PLANT TIR-NADases TRANSMIT IMMUNE SIGNALS?

NAD⁺ is a major cellular metabolite, redox carrier, and substrate for numerous processes including DNA repair, epigenetic modifications, immunity and signaling (Adams-Phillips et al., 2010; Petriacq et al., 2013; Petriacq et al., 2016). Activated plant TIR-domains are NAD⁺-hydrolases, but how might NAD⁺-consumption activate immune responses? SARM1 apparently triggers cell death by depleting NAD⁺, but plant TIRs do not cause detectable NAD⁺ reductions *in planta* (Wan et al., 2019). One possibility is that NAD⁺ consumption by plant TIRs generates signal molecules that turn on downstream immune components.

Unlike SARM1, plant TIRs did not generate c-ADPR, but instead produced v-cADPR, both *in vitro* and after transient expression in *N. benthamiana* (Wan et al., 2019). Moreover, v-cADPR was also produced by activation of RBA1 after bacterial delivery of the *Pseudomonas syringae* effector HopBA1 (Wan et al., 2019). Neither EDS1 or NRG1 – downstream TIR-signaling components – were required for v-cADPR generation by activated TIRs *in planta* (Wan et al., 2019). These results indicate that v-cADPR accumulation is upstream of both cell death and the known signaling components downstream of TIR proteins. Curiously, the *in planta* generation of v-cADPR by TIR-domains isolated from TNLs was nearly 100-fold lower than that of TIR-only proteins (Wan et al., 2019). Is this difference an artifact of truncating TNL proteins, or an intrinsic difference between TIR-only and TNL TIR-domains? Whether an auto-active variant of a full length TNL might produce comparable v-cADPR to TIR-only proteins has not been examined. It is also unclear if the context of a full length NLR could influence the ratio or type of products generated by NAD⁺-hydrolysis, apart from hydrolysis kinetics.

The v-cADPR molecule appears to uniquely identify plant TIR-driven ETI, as MLA10 expression and RPM1 activation (both CNLs) did not elevate v-cADPR (Wan et al., 2019). The chemical structure of v-cADPR is presently unknown, and could vary significantly from cyclic-ADPR. It is possible that v-cADPR shares signaling properties with other NAD⁺-derivatives such as cyclic-ADPR, ADPR, and NAAD (a product of the SARM1-TIR), which are potent Ca²⁺ channel activators (Lee, 2012; Guse, 2015). Numerous studies reveal Ca²⁺ signaling is necessary for plant immunity and HR-driven cell death (Grant et al., 2000; Ma and Berkowitz, 2007, 2011; Marcec et al., 2019). Intriguingly, cyclic-ADPR has been reported to trigger plant defense gene expression, and a calcium channel blocker, lanthanum chloride, prevents plant cell death and HR (although this is not specific to TIR phenotypes) (Durner et al., 1998; Grant et al., 2000).

At this point v-cADPR can be considered a biomarker for plant TIR activity, as its production is correlated with TIR function, however, it is not clear if it is either necessary or sufficient to trigger cell death or disease resistance. *In vitro* assays indicate that the TIR-only proteins RBA1 and BdTIR are also capable of cleaving NADP⁺ (Wan et al., 2019), and it remains to be determined what the putative v-cADPR product looks like and if it is produced *in planta*. Are

there other, as yet, unidentified products? How NADase-produced signaling products might activate immune responses is entirely speculative, but a reasonable candidate to receive a signal would be EDS1, potentially mediated by an EDS1 hetero-oligomer surface. The fact that EDS1/SAG101 and EDS1/PAD4 heterodimers can have non-redundant functions, with specificity in regards to the particular activating TIR (Cui et al., 2017; Castel et al., 2019; Lapin et al., 2019; Wu et al., 2019), complicate simple models where TIR proteins generate a generic signal.

Because NAD⁺ levels influence numerous cellular processes, the consumption of NAD⁺ by plant TIRs during immunity could impact myriad cellular responses. For instance, extracellular NAD⁺ (eNAD⁺) is a potent immunostimulatory signal and reducing NAD⁺ levels compromises disease resistance; conversely, eNAD⁺ application can bolster immunity (Zhang and Mou, 2012; Wang et al., 2016; Mou, 2017; Alferez et al., 2018). Likewise, the AvrRxo1 and RipN, virulence-promoting effectors of plant pathogens, can modulate host NAD⁺ homeostasis and defense responses (Schuebel et al., 2016; Shidore et al., 2017; Sun et al., 2019). While total NAD⁺ levels did not obviously change with TIR expression (Wan et al., 2019), it's possible that localization of NADase activity could have an impact on output.

TIR-PROTEINS ACROSS PLANT PHYLOGENIES

TIR-domain encoding genes can be found in almost all plant lineages. However, the class and abundance of encoded TIR-proteins can vary widely between species (Collier et al., 2011; Yue et al., 2012; Nandety et al., 2013; Sun et al., 2014; Gao et al., 2018). Particularly, between dicot and monocot plant species, the complement of CNL vs. TNL-type NLRs can vary greatly (Sun et al., 2014; Gao et al., 2018). Canonical TNL-type resistance genes are absent from all examined monocot genomes, as are the TIR-pathway mediators, SAG101 and NRG1 (Collier et al., 2011; Wagner et al., 2013). Remarkably, convergent loss of TNLs and downstream genes has occurred several times during plant evolution (Collier et al., 2011; Baggs et al., 2019). Monocots do, however, encode several TIR-NBS and TIR-only genes, although in low abundance relative to the high number of TNLs commonly present in dicots (Sun et al., 2014; Gao et al., 2018). Whether or not these monocot TIR proteins are functioning as immune receptors remains to be determined. However, the TIR-only protein RBA1, can trigger cell death in response a specific pathogen effector, and both TIR-NBS and TIR-X proteins from various plant species are reported to enhance immunity (Meyers et al., 2002; Staal et al., 2008; Nandety et al., 2013; Zhao et al., 2015; Nishimura et al., 2017; Chen et al., 2018; Santamaria et al., 2019). Thus, while TNLs may be absent from monocot genomes, TIR-signaling could play roles in regulating physiological responses and immunity in monocots. BdTIR, a TIR-only protein from the monocot *Brachypodium*, has many of the hallmarks of dicot TIR domains: it has the conserved putative catalytic glutamic acid, produces v-cADPR and triggers EDS1-dependent cell death in *N. benthamiana* (Wan

et al., 2019). Intriguingly, BdTIR cell death in *N. benthamiana* is also dependent on the downstream TIR signaling component NbNRG1, despite the fact that monocots have lost *NRG1* from their genomes. Therefore, it is possible that TIR-domains from distant plant phylogenies produce common signals from NAD⁺-hydrolysis, while the putative immune output depends on which downstream components (e.g., EDS1-members, NRG1) are present to enact the signal.

While TNLs are absent from monocots (and several dicot lineages), they are present broadly across the plant phylogeny, including bryophytes and conifers (see also **Figure 2C**) (Baggs et al., 2019). For instance, the moss *Physcomitrella patens* carries TNL loci, as does the western white pine, *Pinus monticola* (Liu and Ekramoddoullah, 2011; Tanigaki et al., 2014). Two pine TNL loci, TNL1 and TNL2, are correlated with blister rust resistance (Liu and Ekramoddoullah, 2011). TIR-domain-encoding genes were more recently reported in the agriculturally important red algae, *Pyropia yezoensis*, which is used for nori production (Tang et al., 2019). At least one TIR-domain encoding gene, along with several NBS genes of *Pyropia* are upregulated by challenge with the oomycete pathogen, *Pythium* (Tang et al., 2019). Genes with TIR immune receptor-like domain combinations have been found in the genomes of green algae. *Botryococcus* contains TIR-NBS encoding genes, while remarkably, *Chromochloris* has NLR-like genes that contain all three canonical NLR domains (TIR, NBS and LRR) (Shao et al., 2019). More functional evidence for algal TIRs or TNLs in immunity is needed, as well as investigation into the algal relatives of downstream TIR pathway components defined in dicots. It seems likely that TIR-domains across photosynthetic organisms harbor NADase activities, however, this has not been explored. Nor is it clear if these TIR-domains could produce similar molecules from NAD⁺-hydrolysis. An expanded collection of genomic data from algae and early plant clades will help to assess both the conservation and abundance of putative TIR-immune pathways.

(MORE) UNANSWERED QUESTIONS?

TIR-domains encoded by species from all domains of life are now known to play roles in immunity. Recent studies now suggest a new paradigm of TIR-mediated immunity in plants: the oligomerization and self-association of TIR-domains, and subsequent hydrolysis of NAD⁺ (Zhang et al., 2017; Horsefield et al., 2019; Wan et al., 2019). Many important and intriguing questions about TIR-immunity remain. For instance, the stoichiometry and confirmation of active plant TNL or TIR-immune complexes is not known. Furthermore, does the NADase activity of plant TIRs generate immunomodulatory signals? And if so, how are these signals transduced and decoded? Finally, the extent of plant TIR functional conservation is not fully known; i.e., are the TIR-domains encoded by more distantly-related photosynthetic lineages also NADases and do they function in or outside of immunity?

If plant TIRs generate immunomodulatory signals from the hydrolysis of NAD(P)⁺, then what is that signal? For

instance, might variant-cADPR *per se* be sufficient to activate transcriptional defenses, or the hypersensitivity response? Or might different TIR-derived signal molecules communicate different outputs? Additionally, plant TIR-NADases could potentially regulate NAD⁺ levels and cellular metabolism apart from immune signal generation. Do TIR-domains from all plant lineages generate the same type(s) of signals, and how has evolution shaped the components which sense and translate outputs from these signals? The subcellular localization and expression of both signal generating TIRs, and downstream signal receivers could influence potential response outcomes.

TIR-based immunity appears to have an ancient role in prokaryotes as an anti-viral defense system (Doron et al., 2018; Cohen et al., 2019). The conservation of NADase activity among animal, plant and prokaryotic TIRs suggests that an ancient enzymatic activity has been re-purposed multiple times in eukaryotic evolution to promote cell death or immune function. A particularly intriguing question is how did plant TNLs and TIRs evolve to become reliant on the downstream EDS1-family and ‘helper’ NLR partners? Presumably, these components independently provided host benefits, prior to co-evolution into overlapping networks. An in-depth analysis of genomes from early plant lineages may provide insights into how TIRs, EDS1-members and ‘helper’ NLRs co-evolved to function in a core pathway, and provide clues into the mechanisms of TIR-signaling networks of higher plants.

Combined biochemical and evolutionary approaches may provide guidance into how variation in the TIR active site or TIR association interfaces could affect immune outputs. In the future, such findings may be able to offer predictions regarding the kinetic properties of specific TIR-domains, as well as a likely profile of NAD⁺-derived products. For instance, might modulating NAD⁺-hydrolysis kinetics and/or product profile influence the type or strength of immune output? Can *in vitro* evolution enable ‘tweaking’ of TIR-active sites, or of TIR-TIR self-association interfaces, and thus alter the profile of products derived from NAD⁺-hydrolysis?

The recognition that TIR domains across the tree of life have conserved enzymatic functions has opened new avenues of investigation into the plant immune system. While much remains undiscovered, the field is poised to describe fully connected NLR signaling pathways that lead to immune outputs. This synthesis will enable rational engineering of plant immunity to help address the increasing demands on our agricultural systems.

AUTHOR CONTRIBUTIONS

AB and MN wrote the manuscript. Both authors approved the manuscript.

FUNDING

This work was supported by the National Science Foundation (IOS-1758400) and startup funds from Colorado State University to MN.

REFERENCES

- Adams-Phillips, L., Briggs, A. G., and Bent, A. F. (2010). Disruption of poly(ADP-ribose)ylation mechanisms alters responses of *Arabidopsis* to biotic stress. *Plant Physiol.* 152, 267–280. doi: 10.1104/pp.109.148049
- Alaidarous, M., Ve, T., Casey, L. W., Valkov, E., Ericsson, D. J., Ullah, M. O., et al. (2014). Mechanism of bacterial interference with TLR4 signaling by *Brucella* Toll/interleukin-1 receptor domain-containing protein TcpB. *J. Biol. Chem.* 289, 654–668. doi: 10.1074/jbc.M113.523274
- Alferez, F. M., Gerberich, K. M., Li, J. L., Zhang, Y., Graham, J. H., and Mou, Z. (2018). Exogenous nicotinamide adenine dinucleotide induces resistance to citrus canker in citrus. *Front. Plant Sci.* 9:1472. doi: 10.3389/fpls.2018.01472
- Baggs, E., Thanki, A., O'Grady, R., Schudoma, C., Haerty, W., and Krasileva, K. (2019). Convergent loss of an EDS1/PAD4 signalling pathway in several plant lineages predicts new components of plant immunity and drought response. *bioRxiv* [preprint]. doi: 10.1101/572560
- Bai, S., Liu, J., Chang, C., Zhang, L., Maekawa, T., Wang, Q., et al. (2012). Structure-function analysis of barley NLR immune receptor MLA10 reveals its cell compartment specific activity in cell death and disease resistance. *PLoS Pathog* 8:e1002752. doi: 10.1371/journal.ppat.1002752
- Bernoux, M., Ve, T., Williams, S., Warren, C., Hatters, D., Valkov, E., et al. (2011). Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. *Cell Host Microbe* 9, 200–211. doi: 10.1016/j.chom.2011.02.009
- Bhandari, D. D., Lapin, D., Kracher, B., von Born, P., Bautor, J., Niefind, K., et al. (2019). An EDS1 heterodimer signalling surface enforces timely reprogramming of immunity genes in *Arabidopsis*. *Nat. Commun.* 10, 772. doi: 10.1038/s41467-019-08783-0
- Boutrot, F., and Zipfel, C. (2017). Function, discovery, and exploitation of plant pattern recognition receptors for broad-spectrum disease resistance. *Annu. Rev. Phytopathol.* 55, 257–286. doi: 10.1146/annurev-phyto-080614-120106
- Bovijn, C., Ulrichs, P., De Smet, A. S., Catteeuw, D., Beyaert, R., Tavernier, J., et al. (2012). Identification of interaction sites for dimerization and adapter recruitment in Toll/interleukin-1 receptor (TIR) domain of Toll-like receptor 4. *J. Biol. Chem.* 287, 4088–4098. doi: 10.1074/jbc.M111.282350
- Casey, L. W., Lavrencic, P., Bentham, A. R., Cesari, S., Ericsson, D. J., Croll, T., et al. (2016). The CC domain structure from the wheat stem rust resistance protein Sr33 challenges paradigms for dimerization in plant NLR proteins. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12856–12861. doi: 10.1073/pnas.1609922113
- Castel, B., Ngou, P. M., Cevik, V., Redkar, A., Kim, D. S., Yang, Y., et al. (2019). Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. *New Phytol.* 222, 966–980. doi: 10.1111/nph.15659
- Chan, S. L., Low, L. Y., Hsu, S., Li, S., Liu, T., Santelli, E., et al. (2009). Molecular mimicry in innate immunity: crystal structure of a bacterial TIR domain. *J. Biol. Chem.* 284, 21386–21392. doi: 10.1074/jbc.C109.007591
- Chen, G., Wei, B., Li, G., Gong, C., Fan, R., and Zhang, X. (2018). TaEDS1 genes positively regulate resistance to powdery mildew in wheat. *Plant Mol. Biol.* 96, 607–625. doi: 10.1007/s11103-018-0718-9
- Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., et al. (2019). Cyclic GMP-AMP signalling protects bacteria against viral infection. *Nature* 574, 691–695. doi: 10.1038/s41586-019-1605-5
- Collier, S. M., Hamel, L. P., and Moffett, P. (2011). Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Mol. Plant Microbe Interact.* 24, 918–931. doi: 10.1094/MPMI-03-11-0050
- Coronas-Serna, J. M., Louche, A., Rodríguez-Escudero, M., Roussin, M., Imbert, P. R. C., Rodríguez-Escudero, I., et al. (2019). The TIR-domain containing effectors BtpA and BtpB from *Brucella abortus* block energy metabolism. *bioRxiv* [Preprint]. doi: 10.1101/703330
- Cui, H., Gobbato, E., Kracher, B., Qiu, J., Bautor, J., and Parker, J. E. (2017). A core function of EDS1 with PAD4 is to protect the salicylic acid defense sector in *Arabidopsis* immunity. *New Phytol.* 213, 1802–1817. doi: 10.1111/nph.14302
- Dangl, J. L., Horvath, D. M., and Staskawicz, B. J. (2013). Pivoting the plant immune system from dissection to deployment. *Science* 341, 746–751. doi: 10.1126/science.1236011
- Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., et al. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359:eaar4120. doi: 10.1126/science.aar4120
- Durner, J., Wendehenne, D., and Klessig, D. F. (1998). Defense gene induction in tobacco by nitric oxide, cyclic GMP, and cyclic ADP-ribose. *Proc. Natl. Acad. Sci. U.S.A.* 95, 10328–10333. doi: 10.1073/pnas.95.17.10328
- Essuman, K., Summers, D. W., Sasaki, Y., Mao, X., DiAntonio, A., and Milbrandt, J. (2017). The SARM1 Toll/Interleukin-1 receptor domain possesses intrinsic NAD(+) cleavage activity that promotes pathological axonal degeneration. *Neuron* 93, 1334–1343.e5. doi: 10.1016/j.neuron.2017.02.022
- Essuman, K., Summers, D. W., Sasaki, Y., Mao, X., Yim, A. K. Y., DiAntonio, A., et al. (2018). TIR domain proteins are an ancient family of NAD(+) -consuming enzymes. *Curr. Biol.* 28, 421430.e4. doi: 10.1016/j.cub.2017.12.024
- Feys, B. J., Wiermer, M., Bhat, R. A., Moisan, L. J., Medina-Escobar, N., Neu, C., et al. (2005). Arabidopsis SENESENCE-ASSOCIATED GENE101 stabilizes and signals within an ENHANCED DISEASE SUSCEPTIBILITY1 complex in plant innate immunity. *Plant Cell* 17, 2601–2613. doi: 10.1105/tpc.105.033910
- Gao, Y., Wang, W., Zhang, T., Gong, Z., Zhao, H., and Han, G. Z. (2018). Out of water: the origin and early diversification of plant R-Genes. *Plant Physiol.* 177, 82–89. doi: 10.1104/pp.18.00185
- Gerdt, J., Brace, E. J., Sasaki, Y., DiAntonio, A., and Milbrandt, J. (2015). SARM1 activation triggers axon degeneration locally via NAD(+) destruction. *Science* 348, 453–457. doi: 10.1126/science.1258366
- Grant, M., Brown, I., Adams, S., Knight, M., Ainslie, A., and Mansfield, J. (2000). The RPM1 plant disease resistance gene facilitates a rapid and sustained increase in cytosolic calcium that is necessary for the oxidative burst and hypersensitive cell death. *Plant J.* 23, 441–450. doi: 10.1046/j.1365-313x.2000.00804.x
- Guse, A. H. (2015). Calcium mobilizing second messengers derived from NAD. *Biochim. Biophys. Acta* 1854, 1132–1137. doi: 10.1016/j.bbapap.2014.12.015
- Heidrich, K., Wirthmueller, L., Tasset, C., Pouzet, C., Deslandes, L., and Parker, J. E. (2011). Arabidopsis EDS1 connects pathogen effector recognition to cell compartment-specific immune responses. *Science* 334, 1401–1404. doi: 10.1126/science.1211641
- Horsefield, S., Burdett, H., Zhang, X., Manik, M. K., Shi, Y., Chen, J., et al. (2019). NAD(+) cleavage activity by animal and plant TIR domains in cell death pathways. *Science* 365, 793–799. doi: 10.1126/science.aax1911
- Hu, Z., Zhou, Q., Zhang, C., Fan, S., Cheng, W., Zhao, Y., et al. (2015). Structural and biochemical basis for induced self-propagation of NLRC4. *Science* 350, 399–404. doi: 10.1126/science.aac5489
- Huh, S. U., Cevik, V., Ding, P., Duxbury, Z., Ma, Y., Tomlinson, L., et al. (2017). Protein-protein interactions in the RPS4/RRS1 immune receptor complex. *PLoS Pathog* 13:e1006376. doi: 10.1371/journal.ppat.1006376
- Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329.
- Jubic, L. M., Saile, S., Furzer, O. J., El Kasmi, F., and Dangl, J. L. (2019). Help wanted: helper NLRs and plant immune responses. *Curr. Opin. Plant Biol.* 50, 82–94. doi: 10.1016/j.pbi.2019.03.013
- Kenny, E. F., and O'Neill, L. A. (2008). Signalling adaptors used by Toll-like receptors: an update. *Cytokine* 43, 342–349. doi: 10.1016/j.cyt.2008.07.010
- Lapin, D., Kovacova, V., Sun, X., Dongus, J. A., Bhandari, D. D., von Born, P., et al. (2019). A coevolved EDS1-SAG101-NRG1 module mediates cell death signaling by TIR-domain immune receptors. *Plant Cell* 31, 24302455. doi: 10.1105/tpc.19.00118
- Lee, H. C. (2012). Cyclic ADP-ribose and nicotinic acid adenine dinucleotide phosphate (NAADP) as messengers for calcium mobilization. *J. Biol. Chem.* 287, 31633–31640. doi: 10.1074/jbc.R112.349464
- Lee, H. C., and Zhao, Y. J. (2019). Resolving the topological enigma in Ca(2+) signaling by cyclic ADP-ribose and NAADP. *J. Biol. Chem.* 294, 19831–19843. doi: 10.1074/jbc.REV119.009635
- Liu, J. J., and Ekramoddoullah, A. K. (2011). Genomic organization, induced expression and promoter activity of a resistance gene analog (PmTNL1) in western white pine (*Pinus monticola*). *Planta* 233, 1041–1053. doi: 10.1007/s00425-011-1353-8
- Loring, H. S., Icsó, J. D., Nemmara, V., and Thompson, P. R. (2020). Initial kinetic characterization of sterile alpha and Toll/Interleukin receptor motif-containing protein 1. *Biochemistry* 59, 933942. doi: 10.1021/acs.biochem.9b01078
- Ma, W., and Berkowitz, G. A. (2007). The grateful dead: calcium and cell death in plant innate immunity. *Cell Microbiol.* 9, 2571–2585. doi: 10.1111/j.1462-5822.2007.01031.x

- Ma, W., and Berkowitz, G. A. (2011). Ca²⁺ conduction by plant cyclic nucleotide gated channels and associated signaling components in pathogen defense signal transduction cascades. *New Phytol.* 190, 566–572. doi: 10.1111/j.1469-8137.2010.03577.x
- Marcec, M. J., Gilroy, S., Poovaiah, B. W., and Tanaka, K. (2019). Mutual interplay of Ca(2+) and ROS signaling in plant immune response. *Plant Sci.* 283, 343–354. doi: 10.1016/j.plantsci.2019.03.004
- Meyers, B. C., Morgante, M., and Michelmore, R. W. (2002). TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes. *Plant J.* 32, 77–92. doi: 10.1046/j.1365-313x.2002.01404.x
- Mou, Z. (2017). Extracellular pyridine nucleotides as immune elicitors in arabidopsis. *Plant Signal. Behav.* 12:e1388977. doi: 10.1080/15592324.2017.1388977
- Murata, H., Khine, C. C., Nishikawa, A., Yamamoto, K. I., Kinoshita, R., and Sakaguchi, M. (2018). c-Jun N-terminal kinase (JNK)-mediated phosphorylation of SAR1M regulates NAD(+) cleavage activity to inhibit mitochondrial respiration. *J. Biol. Chem.* 293, 18933–18943. doi: 10.1074/jbc.RA118.004578
- Nandety, R. S., Caplan, J. L., Cavanaugh, K., Perroud, B., Wroblewski, T., Michelmore, R. W., et al. (2013). The role of TIR-NBS and TIR-X proteins in plant basal defense responses. *Plant Physiol.* 162, 1459–1472. doi: 10.1104/pp.113.219162
- Nanson, J. D., Kobe, B., and Ve, T. (2019). Death, TIR, and RHIM: self-assembling domains involved in innate immunity and cell-death signaling. *J. Leukoc. Biol.* 105, 363–375. doi: 10.1002/JLB.MR0318-123R
- Nimma, S., Ve, T., Williams, S. J., and Kobe, B. (2017). Towards the structure of the TIR-domain signalosome. *Curr. Opin. Struct. Biol.* 43, 122–130. doi: 10.1016/j.sbi.2016.12.014
- Nishimura, M. T., Anderson, R. G., Cherkis, K. A., Law, T. F., Liu, Q. L., Machius, M., et al. (2017). TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 114, E2053–E2062. doi: 10.1073/pnas.1620973114
- Nyman, T., Stenmark, P., Flodin, S., Johansson, I., Hammarstrom, M., and Nordlund, P. (2008). The crystal structure of the human toll-like receptor 10 cytoplasmic domain reveals a putative signaling dimer. *J. Biol. Chem.* 283, 11861–11865. doi: 10.1074/jbc.C80001200
- O'Neill, L. A., and Bowie, A. G. (2007). The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat. Rev. Immunol.* 7, 353–364. doi: 10.1038/nri2079
- Pearl, J. R., Mestre, P., Lu, R., Malcuit, I., and Baulcombe, D. C. (2005). NRG1, a CC-NB-LRR protein, together with N, a TIR-NB-LRR protein, mediates resistance against tobacco mosaic virus. *Curr. Biol.* 15, 968–973. doi: 10.1016/j.cub.2005.04.053
- Petriacq, P., de Bont, L., Tcherkez, G., and Gakiere, B. (2013). NAD: not just a pawn on the board of plant-pathogen interactions. *Plant Signal. Behav.* 8:e22477. doi: 10.4161/psb.22477
- Petriacq, P., Ton, J., Patrit, O., Tcherkez, G., and Gakiere, B. (2016). NAD Acts as an Integral Regulator of Multiple Defense Layers. *Plant Physiol.* 172, 1465–1479. doi: 10.1104/pp.16.00780
- Qi, D., and Innes, R. W. (2013). Recent advances in plant NLR structure, function, localization, and signaling. *Front. Immunol.* 4:348. doi: 10.3389/fimmu.2013.00348
- Qi, T., Seong, K., Thomazella, D. P. T., Kim, J. R., Pham, J., Seo, E., et al. (2018). NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant immunity in *Nicotiana benthamiana*. *Proc. Natl. Acad. Sci. U.S.A.* 115, E10979–E10987. doi: 10.1073/pnas.1814856115
- Santamaria, M. E., Martinez, M., Arnaiz, A., Rioja, C., Burow, M., Grbic, V., et al. (2019). An *Arabidopsis* TIR-Lectin two-domain protein confers defense properties against *Tetranychus urticae*. *Plant Physiol.* 179, 1298–1314. doi: 10.1104/pp.18.00951
- Schreiber, K. J., Bentham, A., Williams, S. J., Kobe, B., and Staskawicz, B. J. (2016). Multiple domain associations within the arabidopsis immune receptor RPP1 regulate the activation of programmed cell death. *PLoS Pathog* 12:e1005769. doi: 10.1371/journal.ppat.1005769
- Schubel, F., Rocker, A., Edelmann, D., Schessner, J., Brieke, C., and Meinhardt, A. (2016). 3'-NADP and 3'-NAADP, Two Metabolites Formed by the Bacterial Type III Effector AvrRxo1. *J. Biol. Chem.* 291, 22868–22880. doi: 10.1074/jbc.M116.751297
- Shao, Z. Q., Xue, J. Y., Wang, Q., Wang, B., and Chen, J. Q. (2019). Revisiting the Origin of Plant NBS-LRR Genes. *Trends Plant Sci.* 24, 9–12. doi: 10.1016/j.tplants.2018.10.015
- Shidore, T., Broeckling, C. D., Kirkwood, J. S., Long, J. J., Miao, J., Zhao, B., et al. (2017). The effector AvrRxo1 phosphorylates NAD in planta. *PLoS Pathog* 13:e1006442. doi: 10.1371/journal.ppat.1006442
- Shivers, R. P., Kooistra, T., Chu, S. W., Pagano, D. J., and Kim, D. H. (2009). Tissue-specific activities of an immune signaling module regulate physiological responses to pathogenic and nutritional bacteria in *C. Elegans*. *Cell Host Microbe* 6, 321–330. doi: 10.1016/j.chom.2009.09.001
- Snyder, G. A., Deredge, D., Waldhuber, A., Fresquez, T., Wilkins, D. Z., Smith, P. T., et al. (2014). Crystal structures of the Toll/interleukin-1 receptor (TIR) domains from the *Brucella* protein TcpB and host adaptor TIRAP reveal mechanisms of molecular mimicry. *J. Biol. Chem.* 289, 669–679. doi: 10.1074/jbc.M113.523407
- Spear, A. M., Loman, N. J., Atkins, H. S., and Pallen, M. J. (2009). Microbial TIR domains: not necessarily agents of subversion? *Trends Microbiol.* 17, 393–398. doi: 10.1016/j.tim.2009.06.005
- Sporny, M., Guez-Haddad, J., Lebendiker, M., Ullisse, V., Volf, A., Mim, C., et al. (2019). Structural evidence for an octameric ring arrangement of SAR1M. *J. Mol. Biol.* 431, 3591–3605. doi: 10.1016/j.jmb.2019.06.030
- Staal, J., Kaliff, M., Dewaele, E., Persson, M., and Dixelius, C. (2008). RLM3, a TIR domain encoding gene involved in broad-range immunity of *Arabidopsis* to necrotrophic fungal pathogens. *Plant J.* 55, 188–200. doi: 10.1111/j.1365-313X.2008.03503.x
- Sun, X., Pang, H., Li, M., Chen, J., and Hang, Y. (2014). Tracing the origin and evolution of plant TIR-encoding genes. *Gene* 546, 408–416. doi: 10.1016/j.gene.2014.04.060
- Sun, Y., Li, P., Shen, D., Wei, Q., He, J., and Lu, Y. (2019). The *Ralstonia solanacearum* effector RipN suppresses plant PAMP-triggered immunity, localizes to the endoplasmic reticulum and nucleus, and alters the NADH/NAD(+) ratio in *Arabidopsis*. *Mol. Plant Pathol.* 20, 533–546. doi: 10.1111/mpp.12773
- Swiderski, M. R., Birker, D., and Jones, J. D. (2009). The TIR domain of TIR-NB-LRR resistance proteins is a signaling domain involved in cell death induction. *Mol. Plant Microbe Interact.* 22, 157–165. doi: 10.1094/MPMI-22-2-0157
- Takken, F. L., and Govers, A. (2012). How to build a pathogen detector: structural basis of NB-LRR function. *Curr. Opin. Plant Biol.* 15, 375–384. doi: 10.1016/j.pbi.2012.05.001
- Tang, D., Wang, G., and Zhou, J. M. (2017). Receptor kinases in plant-pathogen interactions: more than pattern recognition. *Plant Cell* 29, 618–637. doi: 10.1105/tpc.16.00891
- Tang, L., Qiu, L., Liu, C., Du, G., Mo, Z., Tang, X., et al. (2019). Transcriptomic insights into innate immunity responding to red rot disease in red alga *Pyropia yezoensis*. *Int. J. Mol. Sci.* 20:5970. doi: 10.3390/ijms20235970
- Tanigaki, Y., Ito, K., Obuchi, Y., Kosaka, A., Yamato, K. T., Okanami, M., et al. (2014). Physcomitrella patens has kinase-LRR R gene homologs and interacting proteins. *PLoS one* 9:e95118. doi: 10.1371/journal.pone.0095118
- Urbach, J. M., and Ausubel, F. M. (2017). The NBS-LRR architectures of plant R-proteins and metazoan NLRs evolved in independent events. *Proc. Natl. Acad. Sci. U.S.A.* 114, 1063–1068. doi: 10.1073/pnas.1619730114
- Valkov, E., Stamp, A., Dimaio, F., Baker, D., Verstak, B., Roversi, P., et al. (2011). Crystal structure of Toll-like receptor adaptor MAL/TIRAP reveals the molecular basis for signal transduction and disease protection. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14879–14884. doi: 10.1073/pnas.1104780108
- Ve, T., Williams, S. J., and Kobe, B. (2015). Structure and function of the Toll/interleukin-1 receptor/resistance protein (TIR) domains. *Apoptosis* 20, 250–261. doi: 10.1007/s10495-014-1064-2
- Wagner, S., Stuttmann, J., Rietz, S., Guerois, R., Brunstein, E., Bautor, J., et al. (2013). Structural basis for signaling by exclusive EDS1 heteromeric complexes with SAG101 or PAD4 in plant innate immunity. *Cell Host Microbe* 14, 619–630. doi: 10.1016/j.chom.2013.11.006
- Wan, L., Essuman, K., Anderson, R. G., Sasaki, Y., Monteiro, F., Chung, E. H., et al. (2019). TIR domains of plant immune receptors are NAD(+) -cleaving enzymes that promote cell death. *Science* 365, 799–803. doi: 10.1126/science.aax1771
- Wang, C., Zhang, X., and Mou, Z. (2016). Comparison of nicotinamide adenine dinucleotide phosphate-induced immune responses against biotrophic

- and necrotrophic pathogens in *Arabidopsis thaliana*. *Plant Signal. Behav.* 11:e1169358. doi: 10.1080/15592324.2016.1169358
- Wang, J., Hu, M., Wang, J., Qi, J., Han, Z., Wang, G., et al. (2019a). Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science* 364:eaav5870. doi: 10.1126/science.aav5870
- Wang, J., Wang, J., Hu, M., Wu, S., Qi, J., Wang, G., et al. (2019b). Ligand-triggered allosteric ADP release primes a plant NLR complex. *Science* 364:eaav5868. doi: 10.1126/science.aav5868
- Williams, S. J., Sohn, K. H., Wan, L., Bernoux, M., Sarris, P. F., Segonzac, C., et al. (2014). Structural basis for assembly and function of a heterodimeric plant immune receptor. *Science* 344, 299–303. doi: 10.1126/science.1247357
- Williams, S. J., Yin, L., Foley, G., Casey, L. W., Outram, M. A., Ericsson, D. J., et al. (2016). Structure and function of the TIR domain from the grape NLR protein RPV1. *Front. Plant. Sci.* 7:1850. doi: 10.3389/fpls.2016.01850
- Wu, Z., Li, M., Dong, O. X., Xia, S., Liang, W., Bao, Y., et al. (2019). Differential regulation of TNL-mediated immune signaling by redundant helper CNLs. *New Phytol.* 222, 938–953. doi: 10.1111/nph.15665
- Xu, Y., Tao, X., Shen, B., Horng, T., Medzhitov, R., Manley, J. L., et al. (2000). Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. *Nature* 408, 111–115. doi: 10.1038/35040600
- Yue, J. X., Meyers, B. C., Chen, J. Q., Tian, D., and Yang, S. (2012). Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. *New Phytol.* 193, 1049–1063. doi: 10.1111/j.1469-8137.2011.04006.x
- Zhang, L., Chen, S., Ruan, J., Wu, J., Tong, A. B., Yin, Q., et al. (2015). Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. *Science* 350, 404–409. doi: 10.1126/science.aac5789
- Zhang, Q., Zmasek, C. M., Cai, X., and Godzik, A. (2011). TIR domain-containing adaptor SARM is a late addition to the ongoing microbe-host dialog. *Dev. Comp. Immunol.* 35, 461–468. doi: 10.1016/j.dci.2010.11.013
- Zhang, X., Bernoux, M., Bentham, A. R., Newman, T. E., Ve, T., Casey, L. W., et al. (2017). Multiple functional self-association interfaces in plant TIR domains. *Proc. Natl. Acad. Sci. U.S.A.* 114, E2046–E2052. doi: 10.1073/pnas.1621248114
- Zhang, X., and Mou, Z. (2012). Expression of the human NAD(P)-metabolizing ectoenzyme CD38 compromises systemic acquired resistance in *Arabidopsis*. *Mol. Plant Microbe Interact.* 25, 1209–1218. doi: 10.1094/MPMI-10-11-0278
- Zhao, T., Rui, L., Li, J., Nishimura, M. T., Vogel, J. P., Liu, N., et al. (2015). A truncated NLR protein, TIR-NBS2, is required for activated defense responses in the exo70B1 mutant. *PLoS Genet.* 11:e1004945. doi: 10.1371/journal.pgen.1004945
- Zhao, Z. Y., Xie, X. J., Li, W. H., Liu, J., Chen, Z., Zhang, B., et al. (2019). A Cell-Permeant Mimetic of NMN Activates SARM1 to Produce Cyclic ADP-Ribose and Induce Non-apoptotic Cell Death. *iScience* 15, 452–466. doi: 10.1016/j.isci.2019.05.001

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 Bayless and Nishimura. 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.



Distinct Evolutionary Patterns of NBS-Encoding Genes in Three Soapberry Family (Sapindaceae) Species

Guang-Can Zhou^{1*}, Wen Li¹, Yan-Mei Zhang², Yang Liu³, Ming Zhang¹, Guo-Qing Meng¹, Min Li¹ and Yi-Lei Wang^{1*}

¹ College of Agricultural and Biological Engineering (College of Tree Peony), Heze University, Heze, China, ² Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China, ³ State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing, China

OPEN ACCESS

Edited by:

Madhav P. Nepal,
South Dakota State University,
United States

Reviewed by:

Anna Maria Mastrangelo,
Council for Agricultural
and Economics Research (CREA),
Italy

Jacqueline Batley,
The University of Western Australia,
Australia

*Correspondence:

Guang-Can Zhou
zgcan2009@163.com
Yi-Lei Wang
wangyilei001@163.com

Specialty section:

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

Received: 12 March 2020

Accepted: 19 June 2020

Published: 10 July 2020

Citation:

Zhou G-C, Li W, Zhang Y-M, Liu Y,
Zhang M, Meng G-Q, Li M and
Wang Y-L (2020) Distinct Evolutionary
Patterns of NBS-Encoding Genes
in Three Soapberry Family
(Sapindaceae) Species.
Front. Genet. 11:737.
doi: 10.3389/fgene.2020.00737

Nucleotide-binding site (NBS)-type disease resistance genes (*R* genes) play key roles in plant immune responses and have co-evolved with pathogens over the course of plant lifecycles. Comparative genomic studies tracing the dynamic evolution of NBS-encoding genes have been conducted using many important plant lineages. However, studies on Sapindaceae species have not been performed. In this study, a discrepant number of NBS-encoding genes were identified in the genomes of *Xanthoceras sorbifolium* (180), *Dinnocarpus longan* (568), and *Acer yangbiense* (252). These genes were unevenly distributed and usually clustered as tandem arrays on chromosomes, with few existed as singletons. The phylogenetic analysis revealed that NBS-encoding genes formed three monophyletic clades, RPW8-NBS-LRR (RNL), TIR-NBS-LRR (TNL), and CC-NBS-LRR (CNL), which were distinguished by amino acid motifs. The NBS-encoding genes of the *X. sorbifolium*, *D. longan*, and *A. yangbiense* genomes were derived from 181 ancestral genes (three RNL, 23 TNL, and 155 CNL), which exhibited dynamic and distinct evolutionary patterns due to independent gene duplication/loss events. Specifically, *X. sorbifolium* exhibited a “first expansion and then contraction” evolutionary pattern, while *A. yangbiense* and *D. longan* exhibited a “first expansion followed by contraction and further expansion” evolutionary pattern. However, further expansion in *D. longan* was stronger than in *A. yangbiense* after divergence, suggesting that *D. longan* gained more genes in response to various pathogens. Additionally, the ancient and recent expansion of CNL genes generated the dominance of this subclass in terms of gene numbers, while the low copy number status of RNL genes was attributed to their conserved functions.

Keywords: Sapindaceae, NBS-encoding gene, phylogeny, gene duplication/loss, evolutionary pattern

INTRODUCTION

Nucleotide-binding site (NBS)-encoding genes are the largest type (~80%) of disease resistance genes (*R* genes) found in plants and are responsible for the protection against various pathogens (Liu et al., 2007; Dangl et al., 2013). A typical NBS-encoding gene consists of a variable domain at the N-terminal, a highly conserved NBS domain in the middle, and a diverse leucine-rich repeat

(LRR) domain at the C-terminal. Therefore, according to the structure of N-terminal domains that possess a coiled-coil (CC), *Drosophila* toll and mammalian interleukin-1 receptor-like (TIR), or resistance to powdery mildew8 (RPW8) domain, NBS-encoding genes can be classified into three subclasses: CC-NBS-LRR (CNL), TIR-NBS-LRR (TNL), and RPW8-NBS-LRR (RNL) (Meyers et al., 2003; Shao et al., 2014, 2016, 2019; Zhang et al., 2016, 2020; Qian et al., 2017; Xue et al., 2020). As an example of function, the CNL gene, *RPP8*, in *Arabidopsis thaliana* provides resistance against downy mildew after *Peronospora parasitica* infection (Mcdowell et al., 1998). Another CNL gene, *Pik*, confers resistance to rice blast caused by *Magnaporthe grisea* infection (Zhai et al., 2011). The tobacco TNL gene, *N*, prevents tobacco mosaic virus invasion (Whitham et al., 1994). Functionally, CNL and TNL genes act as detectors that recognize specific pathogen effectors encoded by avirulence genes and initiate downstream hypersensitive reactions in the resistance pathway (Dangl and Jones, 2001; Mchale et al., 2006), while RNL genes appear to function downstream and transduce signals from CNL or TNL genes through interactions with corresponding partners (Bonardi et al., 2011; Collier et al., 2011; Tamborski and Krasileva, 2020). Therefore, NBS-encoding *R* genes are of great importance to plant growth and would tangibly benefit humankind if properly used in disease resistance breeding.

NBS-encoding genes constitute a large gene family found in plant genomes. With the accumulation of more plant whole-genome sequences, genome-wide evolutionary analyses of NBS-encoding genes have been performed in many plants since the first comprehensive study on NBS-encoding genes in *A. thaliana* (Meyers et al., 2003). Furthermore, comparative genomic studies on the NBS-encoding gene family have been performed among a few closely related plant species, which exhibited different evolutionary patterns. For example, frequent gene losses and limited gene duplications resulted in small number of NBS-encoding genes in three Cucurbitaceae species (Lin et al., 2013). Both Li et al. (2010) and Luo et al. (2012) investigated the NBS-encoding genes in four Poaceae species, including rice, maize, sorghum, and brachypodium, which exhibited a “contraction” evolutionary pattern and may have been caused by gene losses or frequent gene deletions and translocations. Similar analyses were performed on Fabaceae, Rosaceae, and Brassicaceae species, of which Fabaceae and Rosaceae species exhibited a “consistent expansion” evolutionary pattern (Shao et al., 2014; Jia et al., 2015), while five Brassicaceae species exhibited a “first expansion and then contraction” evolutionary pattern (Zhang et al., 2016). Moreover, although plants belonged to the same family, the evolutionary patterns of NBS-encoding genes were also diverse. For example, in three Solanaceae crop species, pepper exhibited a “contraction” pattern, tomato showed a “first expansion and then contraction” pattern, and potato presented a “consistent expansion” pattern (Qian et al., 2017). In four orchid species, *Phalaenopsis equestris* and *Dendrobium catenatum* exhibited an “early contraction to recent expansion” evolutionary pattern, while *Gastrodia elata* and *Apostasia shenzhenica* showed a “contraction” evolutionary pattern (Xue et al., 2020). Recently, a large scale analysis of the NBS-encoding genes in 22 representative angiosperms

demonstrated that CNL genes exhibited a “gradual expansion” evolutionary pattern during the first 100 million years of angiosperm evolution, then underwent intense expansion along with TNL genes, which corresponded with the explosion of fungal diversity (Shao et al., 2016).

The soapberry family (Sapindaceae) consists of 135 genera and ~1500 species, which is comprised mostly of trees or shrubs and some herbaceous climbers, which are widely distributed throughout the tropics and subtropics, especially in Southeastern Asia (Flora of China¹). Sapindaceae species possess many great economic uses. For example, the seed kernels of Yellowhorn (*Xanthoceras sorbifolium*), a major woody oil plant species, contains as much as 67% oil (Venegas-Calderón et al., 2017), and extractions from its husks improve learning and memory that could be used to treat Alzheimer’s disease (Ji et al., 2017; Zhang et al., 2018). Longan (*Dinnocarpus longan*), an important evergreen fruit tree, is mainly grown in Southern China and serves as a source of traditional medicine and timber (Chung et al., 2010; Mei et al., 2014). *Acer yangbiense* is a charismatic landscape plant that possesses a colorful foliage is a newly described critically endangered endemic maple tree found in Southwestern China (Chen et al., 2003), and it possesses many bioactive compounds (Bi et al., 2016; Yang et al., 2019).

Recently, the high-quality genome sequences of *X. sorbifolium*, *D. longan*, and *A. yangbiense* were made available and all recovered >94% complete BUSCO genes by BUSCO analysis (Lin et al., 2017; Bi et al., 2019; Yang et al., 2019). An analysis of the NBS-encoding genes in *D. longan* was performed, and researchers found high number and recent expansions/contractions of these genes that may be attributed to the genomic basis for insect, fungus, and bacteria resistance (Lin et al., 2017). However, these findings require further elucidation of this phenomenon caused by gene duplication/loss events. Additionally, the systematic evaluation and comparison of NBS-encoding genes at the genome level in more Sapindaceae species is needed to obtain a better understanding of this family’s molecular evolutionary history. In this study, *X. sorbifolium*, *D. longan*, and *A. yangbiense* genome sequence data were utilized to perform comparative genomic analyses in order to uncover the evolutionary features and patterns of NBS-encoding genes in the Sapindaceae family, as well as further investigate the mechanisms of these evolutionary changes.

MATERIALS AND METHODS

Identification and Classification of the NBS-Encoding Genes

The whole genomes of the *X. sorbifolium*, *D. longan*, and *A. yangbiense* were used in this study (Supplementary Figure S1). Genomic sequences and annotation files were obtained from the GigaScience database (*X. sorbifolium*²,

¹www.iplant.cn/foc/

²http://dx.doi.org/10.5524/100606

Bi et al., 2019; *D. longan*³, Lin et al., 2017; *A. yangbiense*⁴, Yang et al., 2019). The method for NBS-encoding gene identification was described in a previous study (Shao et al., 2016). Briefly, BLAST and hidden Markov model (HMM) searches were simultaneously conducted using the NB-ARC domain (Pfam accession No.: PF00931) as the query sequence to identify candidate NBS-encoding genes in the three genomes. The threshold expectation value was set to 1.0 for the BLAST search. For the HMM search⁵, default settings were used. Then, the identified candidate sequences were merged and the redundant hits were removed. In order to confirm the presence of the NBS domain, the remaining sequence hits were subjected to online Pfam analysis⁶ with an E-value of 10^{-4} . All of the identified NBS-encoding genes were subjected to NCBI's conserved domain database⁷ using the default settings to determine whether they encoded CC, TIR, RPW8, or LRR domains.

Chromosomal Distribution and Cluster Arrangement of Identified NBS-Encoding Genes

The chromosomal locations of all identified NBS-encoding genes in Sapindaceae genomes were determined by retrieving relevant information from the downloaded gff files. Gene cluster was determined according to the criterion used for *Medicago truncatula* (Ameline-Torregrosa et al., 2008): if two neighboring NBS-encoding genes were located within 250 kb on a chromosome, these two genes were regarded as members of the same gene cluster. Based on this criterion, the NBS-encoding genes in Sapindaceae genomes were assigned to clustered loci and singleton loci, which were mapped along the chromosomes.

Sequence Alignment and Conserved Motif Analysis of the NBS Domain

Amino acid sequences of the NBS domain were extracted from the identified NBS-encoding genes and used for multiple alignments using ClustalW integrated in MEGA 7.0 using the default settings (Kumar et al., 2016). Sequences that were too short (<190 amino acids, less than two-thirds of a regular NBS domain) or too divergent were removed to prevent interference with the alignments and subsequent phylogenetic analysis. The resulting alignments were manually corrected and improved using MEGA 7.0. The conserved protein motifs within the NBS domain of the three subclasses of NBS-encoding genes were analyzed by Multiple Expectation Maximization for Motif Elicitation (MEME) and WebLogo using the default settings (Crooks et al., 2004; Bailey et al., 2006). Additionally, structural motif annotation was performed using the Pfam analysis⁸ and SMART tools⁹.

³<http://dx.doi.org/10.5524/100276>

⁴<http://dx.doi.org/10.5524/100610>

⁵<http://hmmer.org>

⁶<http://pfam.sanger.ac.uk/>

⁷<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

⁸<http://pfam.janelia.org/>

⁹<http://smart.embl-heidelberg.de/>

Phylogenetic and Gene Loss/Duplication Analysis of the NBS-Encoding Genes

To explore the relationships of NBS-encoding genes in the three Sapindaceae genomes, a phylogenetic tree was constructed based on the amino acid sequences of the conserved NBS domain using the NBS-encoding genes of *A. thaliana* as a reference. The NBS-encoding genes of *A. thaliana* were identified using the same method, and these NBS-encoding genes were also used in a previous study (Zhang et al., 2011). Amino acid sequences were aligned as described above. The phylogenetic tree was constructed using IQ-TREE and the maximum likelihood method based on the best-fit model estimated by ModelFinder (Nguyen et al., 2015; Kalyaanamoorthy et al., 2017); branch support values were assessed using UFBoot2 tests (Minh et al., 2013). The short or divergent sequences that were removed from the phylogenetic analyses were BLASTp searched against all of the identified NBS genes to identify their potential phylogenetic positions by identifying their closest relatives. Additionally, in order to identify the gene duplication/loss events during the speciation of the three Sapindaceae species, the NBS-encoding gene phylogenetic tree was reconciled with the real species tree using Notung software (Stolzer et al., 2012). The NBS-encoding genes on the tree that formed a monophyletic branch and originated from one ancestral gene inherited from the common ancestor of the three Sapindaceae species was defined as a Sapindaceae lineage gene. One or more Sapindaceae lineage genes that formed a monophyletic branch with the *A. thaliana* NBS-encoding genes were inherited from a common ancestor of the three Sapindaceae species and *A. thaliana*, and were defined as Malvids lineage genes.

Synteny Analyses Across/Within Sapindaceae Genomes and Gene Duplication Type Determination

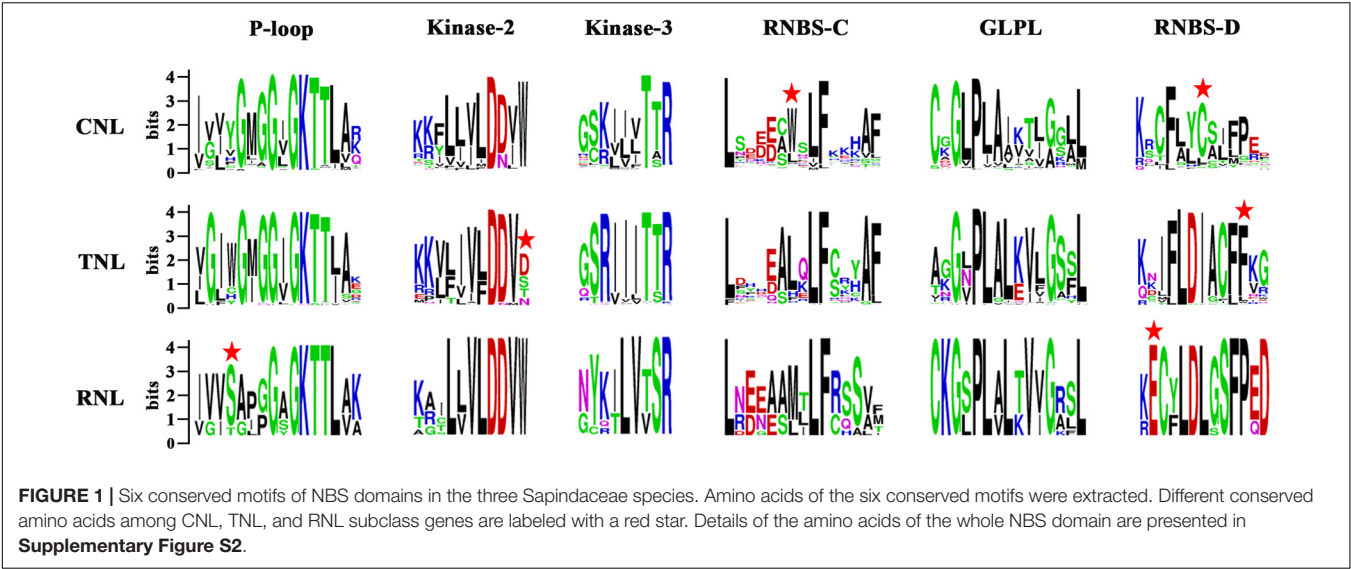
The MCScanX package (Lee et al., 2012; Wang et al., 2012) was used to identify syntenic blocks within a genome or between different genomes through pair-wise all-against-all blast of protein sequences. The purposes of synteny analysis were to explore the pattern of conservation of NBS-encoding gene loci among the Sapindaceae genomes and determine the types of NBS-encoding gene duplication. Synteny relationship of NBS-encoding genes was displayed by TBtools¹⁰.

RESULTS

Identification and Classification of the NBS-Encoding Genes

A total of 180, 252, and 568 non-redundant NBS-encoding genes were identified from the genomes of *X. sorbifolium*, *A. yangbiense*, and *D. longan*, respectively (Table 1 and Supplementary Table S1), accounting for 0.73, 0.89, and 1.83% of the 24,672, 28,320, and 31,007 annotated protein-coding

¹⁰<https://github.com/CJ-Chen/TBtools>



genes in *X. sorbifolium*, *A. yangbiense*, and *D. longan* genomes, respectively (Lin et al., 2017; Bi et al., 2019; Yang et al., 2019). The average lengths of NBS-encoding genes in *X. sorbifolium*, *D. longan*, and *A. yangbiense* were 6081, 5569, and 6199 bp, while the average lengths of the coding region of NBS-encoding genes in *X. sorbifolium*, *D. longan*, and *A. yangbiense* were 3082, 3132, and 3142 bp, respectively. *D. longan* possessed the largest number of NBS-encoding genes and was 3.16- and 2.25-times greater than *X. sorbifolium* and *A. yangbiense*, respectively. The identified NBS-encoding genes from the three Sapindaceae species were divided into the CNL, TNL, and RNL subclasses based on their domain compositions and primary phylogenies. Among the three subclasses, CNL genes overwhelmingly outnumbered TNL and RNL genes with proportions of 81.11, 89.26, and 89.68% in *X. sorbifolium*, *D. longan*, and *A. yangbiense*, respectively. The proportion of TNL genes was 17.78, 9.15, and 7.94% in *X. sorbifolium*, *D. longan*, and *A. yangbiense*, respectively, while the number of RNL genes was the smallest with only two, nine, and six genes, respectively. Not all of the identified NBS-encoding genes had intact structures possessing all three domains (CC/TIR/RNL) with many either lacking the CC/TIR/RPW8 domain at the N-terminus, the LRR domain at the C-terminus, or domains at both termini (Table 1). Additionally, some identified NBS-encoding genes were classified as “other” in TNL and CNL due to their atypical structural domain compositions (Table 1). For example, the *D. longan* genome encoded two “other” genes in TNL, including LN_{TIR}L and TLN, and 13 “other” genes in CNL, including CNLNL(2), N_{CC}LN_{CC}LN_{CC}L(1), N_{CC}N_{CC}LN_{CC}L(1), NLCNL(1), CNLCNL(4), CNLNL(1), N_{CC}N_{CC}L(2), and CNN(1).

Distribution and Organization of NBS-Encoding Genes in Sapindaceae Genomes

The NBS-encoding genes were unevenly distributed among different chromosomes. For example, Chrom (chromosome) 9 of *X. sorbifolium* genome contains the most genes (42 genes),

TABLE 1 | The number of identified NBS-encoding genes in the three Sapindaceae genomes.

Domain compositions	<i>X. sorbifolium</i>	<i>A. yangbiense</i>	<i>D. longan</i>
TNL subclass	32 (17.78%)	20 (7.94%)	52 (9.15%)
TNL (intact)	24	3	24
TN	0	0	7
NL	5	10	6
N	3	4	13
Other	0	3	2
CNL subclass	146 (81.11%)	226 (89.68%)	507 (89.26%)
CNL (intact)	55	26	218
CN	27	30	88
NL	45	95	107
N	18	72	81
Other	1	3	13
RNL subclass	2 (1.11%)	6 (2.38%)	9 (1.58%)
RNL (intact)	2	2	5
RN	0	1	0
NL	0	1	1
N	0	2	3
Other	0	0	0
Total number	180	252	568
Proportion to total protein-coding genes	0.73%	0.89%	1.83%
Average gene size (bp)	6081	6199	5569
Average length of coding sequence (bp)	3082	3142	3132

whereas the Chrom 1 and 11 contain the fewest (only two genes in each Chrom) (Supplementary Figure S3). In *A. yangbiense*, Chrom 3 contains the most genes (89 genes), whereas no NBS-encoding gene was detected on Chrom 11 (Supplementary Figure S4). Uneven distributions were also observed among the three subclasses of NBS-encoding genes (Supplementary Figures S3, S4). Chrom 9 (41 genes) of

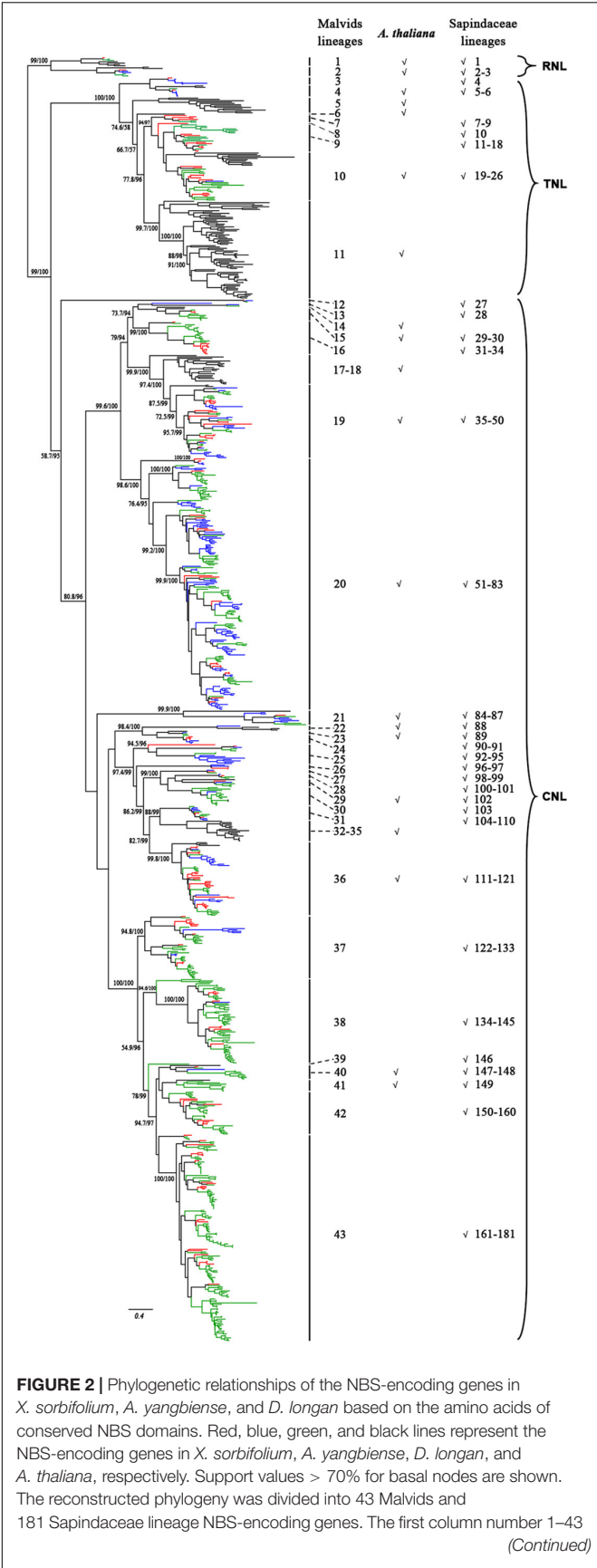


FIGURE 2 | Continued represents the Malvids lineage NBS-encoding genes, and the last column number 1–181 represents the Sapindaceae lineage NBS-encoding genes. The presence of Malvids lineage NBS-encoding genes in *A. thaliana* and the common ancestor of the three Sapindaceae species is indicated by “✓.” The scale bar indicated amino acid substitutions/sites. The detailed phylogenetic tree of the NBS-encoding genes, including gene names, evolutionary relationships among genes, and supporting values of all of the nodes, is presented in **Supplementary Figure S5**.

TABLE 2 | Organization of NBS-encoding genes in the three Sapindaceae genomes.

Loci and genes	<i>X. sorbifolium</i>	<i>A. yangbiense</i>	<i>D. longan</i>
No. of chromosome-anchored loci (and genes)	83 (175)	77 (246)	/
No. of singleton loci (no. of genes)	44 (44)	39 (39)	/
No. of clustered loci (no. of genes)	39 (131)	38 (207)	/
Clustered genes/singleton genes	2.98	5.31	/
Average no. of genes in clusters	3.36	5.45	/
No. of clusters with 10 or more genes	1	7	/
No. of genes in the largest cluster	14 (Chrom 14)	21 (Chrom 3)	/

No., number; Chrom., chromosome; /, not detected.

X. sorbifolium and Chrom 3 (86 genes) of *A. yangbiense* contained the most CNL genes, whereas Chrom 4 (20 genes) of *X. sorbifolium* and Chrom 2 (five genes) and 8 (five genes) of *A. yangbiense* contained the most TNL genes. All chromosomes of the two species contained CNL genes except Chrom 10 and 11 of *A. yangbiense*, whereas only seven chromosomes of *X. sorbifolium* and six chromosomes of *A. yangbiense* contain TNL genes. There were too few RNL genes for this analysis. The majority of NBS-encoding genes were organized into clusters rather than singletons in *X. sorbifolium* and *A. yangbiense* genomes, and their ratios were 2.98 and 5.31, respectively (Table 2). *A. yangbiense* contained more clustered genes than *X. sorbifolium* (207 genes vs 131 genes), and the largest gene cluster of *X. sorbifolium* and *A. yangbiense* were on Chrom 14 (14 genes) and Chrom 3 (21 genes), respectively. Since the NBS-encoding genes in *D. longan* genome were not anchored to chromosomes yet, the chromosomal locations and cluster assignments of these genes were not examined.

Motif Analysis of the NBS Domain

The NBS domain consists of several functional motifs that were conserved and strictly ordered among the NBS-encoding genes (Yue et al., 2012). A total of six conserved motifs were identified in each subclass of the NBS domains in the three Sapindaceae species using MEME and WebLogo (Crooks et al., 2004; Bailey et al., 2006). From the N-terminus to the C-terminus, these motifs included the P-loop, Kinase-2, Kinase-3, RNBS-C, GLPL, and

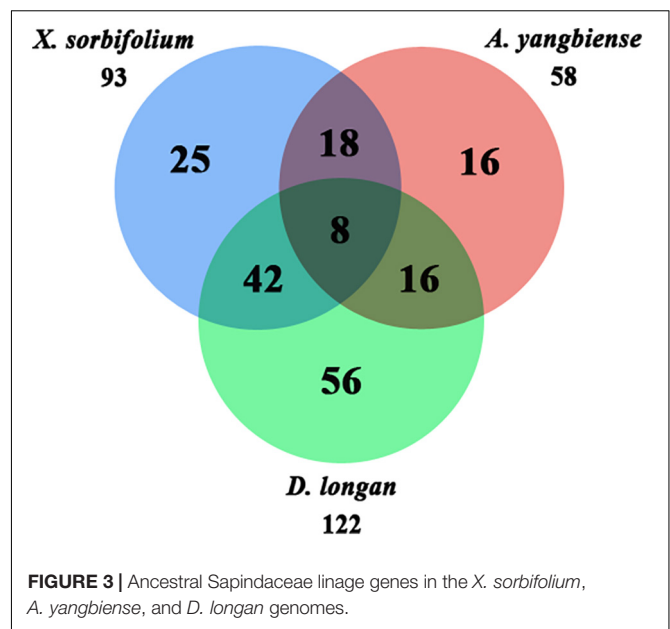
RNBS-D motifs. Comparisons of the amino acid sequences of these motifs are presented in **Figure 1**. The P-loop, Kinase-2, Kinase-3, and GLPL motifs exhibited high similarity among the three subclasses of NBS-encoding genes, suggesting that the NBS domains with critical functions regulating the immune responses were homologous. The other two motifs, especially RNBS-D, were poorly conserved among the three subclasses of NBS-encoding genes. The variation of these motifs may be responsible for further functional divergences within these subclasses. The subclass-specific signatures within the six motifs were further analyzed, which could be used as preliminary labels to identify CNL, TNL, or RNL genes. For example, Tryptophan (W) at the seventh position of the RNBS-C and Cysteine (C) at the seventh position of RNBS-D in CNL genes, phenylalanine (F) at the 11th position of the RNBS-D in TNL genes, and serine (S) at the fourth position of the P-loop and glutamic acid (E) at the second position of the RNBS-D in RNL genes. Additionally, the amino acids at the final position of Kinase-2 could also be used to distinguish TNL genes from CNL and RNL genes (**Figure 1**). Therefore, the subclass of a given NBS-encoding gene could be classified based on the amino acid signatures of the motif sequences.

Phylogenetic Analysis of the NBS-Encoding Genes

To reconstruct the phylogenetic relationship of the NBS-encoding genes in Sapindaceae, a phylogenetic tree was constructed based on the amino acid sequences of NBS domain alignments using the NBS-encoding genes of *A. thaliana* as a reference. To attain better phylogeny, too short or extremely divergent NBS domains were removed from the data matrix. A total of 957 genes (*X. sorbifolium*, 159; *D. longan*, 468; *A. yangbiense*, 173; *A. thaliana*, 157) were obtained and used to reconstruct the evolutionary history of the NBS-encoding genes. The phylogenetic tree was composed of three monophyletic clades, RNL, TNL, and CNL, with support values > 99%; many internal nodes had high (>70%) support values (**Figure 2** and **Supplementary Figure S5**). The three clades represented the divergence of RNL, TNL, and CNL genes. Compared to TNL and CNL, the branch lengths of RNL genes were short (**Figure 2**), suggesting that they had a low evolutionary rate.

To gain insight on the evolution of the NBS-encoding genes before and after divergence of *A. thaliana* and the three Sapindaceae species, Notung software was used to reconcile gene duplication/loss events of the NBS-encoding genes at each node of the phylogenetic tree (Stolzer et al., 2012). Based on the definitions (refer to section “Materials and Methods” for details), 43 Malvids lineage genes were retrospected (**Figure 2**). The ratio of RNL:TNL:CNL genes was 2:9:32. Among them, only 13 genes (two RNL, two TNL, and nine CNL) were reserved by both *A. thaliana* and the common ancestor of the three Sapindaceae species, while 19 genes (four TNL, 15 CNL) were lost by *A. thaliana* and 10 genes (three TNL, seven CNL) were lost by the common ancestor.

A total of 33 Malvids lineage genes (two RNL, six TNL, 25 CNL) were inherited by the common ancestor of the three



Sapindaceae species and these genes intensively expanded to 181 Sapindaceae lineage genes (three RNL, 23 TNL, 155 CNL) (**Figure 2**). CNL genes exhibited a more active expansion rate than RNL or TNL genes, especially the CNL Malvids lineage 19, 20, and 43 expanded 16, 33, and 21 Sapindaceae lineages, respectively (**Figure 2**). Further analysis revealed that 93, 122, and 58 of 181 Sapindaceae lineages were inherited by *X. sorbifolium*, *D. longan*, and *A. yangbiense*, respectively (**Figure 3**), while only eight lineages were reserved in all three genomes. A total of 76 Sapindaceae lineages (*X. sorbifolium* and *D. longan*, 42; *D. longan* and *A. yangbiense*, 16; *X. sorbifolium* and *A. yangbiense*, 18) were maintained in only two genomes and 97 lineages were species-specific (*X. sorbifolium*, 25; *D. longan*, 56; *A. yangbiense*, 16). The distribution patterns suggest that these ancestral gene lineages have experienced differential gene duplication/loss events when the three Sapindaceae species diverged.

Syntenic Analysis of NBS-Encoding Genes in Sapindaceae Genomes

The synteny analysis was performed between and within the three Sapindaceae genomes (the *D. longan* genome was presented as scaffold form). A total of 33 syntenic NBS-encoding genes were detected between *X. sorbifolium* and *A. yangbiense* genome (**Figure 4A** and **Supplementary Table S2**). Among them, the number of syntenic CNL and RNL subclass genes was 32 and one, respectively, whereas no syntenic TNL subclass genes were detected. The CNL, TNL, and RNL subclass syntenic genes were all detected between *X. sorbifolium* and *D. longan* genome (**Figure 4B** and **Supplementary Table S2**), and the gene number was 26, 2, and 1, respectively. However, there were only 16 CNL subclass syntenic genes were detected between *A. yangbiense* and *D. longan* (**Figure 4C** and **Supplementary Table S2**). Moreover, there were only seven syntenic NBS-encoding genes preserved by the three genomes and they all belong to the CNL subclass

TABLE 3 | Contributions of three duplication types in producing NBS-encoding genes during the evolution of Sapindaceae species.

Different types of duplication	<i>X. sorbifolium</i>	<i>A. yangbiense</i>	<i>D. longan</i>
Total no. of NBS-encoding genes	180	252	568
Local tandem duplication	108 (60%)	172 (68.2%)	/
Dispersed duplication	55 (30.6%)	61 (24.2%)	/
WGD or segmental duplication	12 (6.7%)	13 (5.2%)	/
Unanchored genes	5 (2.7%)	6 (2.4%)	568

No., number; /, not detected.

(**Supplementary Table S2**). The syntenic NBS-encoding genes revealed three different forms: singleton to singleton, singleton to cluster, and cluster to cluster. These forms were resulted from different extents of gene duplication/loss of the ancestor genes. The large number of NBS-encoding genes showing presence/absent (P/A) polymorphism and asymmetric gene numbers among co-linear loci indicate that the NBS-encoding genes have experienced dramatic gain and loss during the evolution of Sapindaceae species.

There were three types of NBS-encoding gene duplications: local tandem duplication, dispersed (or ectopic) duplication, and whole genome duplications (WGDs) or segmental duplication (Leister, 2004). MCScanX software was adopted to determine the type of gene duplications in producing NBS-encoding genes during the evolution of Sapindaceae species. As shown in **Table 3**, tandem and dispersed duplication were the main contributors to gene expansion events in *X. sorbifolium* and *A. yangbiense* genomes. However, the proportion of segmental duplicated genes might have been underestimated because of the syntenic relationship of NBS-encoding genes would be disrupted during long-term evolution. The gene duplication types of the NBS-encoding genes in *D. longan* genome were not examined, since these genes were not anchored to chromosomes yet.

Gene Duplications/Losses Resulting in Dynamic Evolutionary Patterns of the NBS-Encoding Genes

To assess the evolutionary patterns of the NBS-encoding genes during the speciation of the three Sapindaceae species, the gene duplication/loss events of the NBS-encoding genes were restored in the phylogenetic tree based on the conserved NBS domain sequences. The retrospective analysis revealed that 43 ancient NBS-encoding genes were shared by the common ancestor of *A. thaliana* and the three Sapindaceae species (**Figure 2**). Among these ancient genes, 33 were inherited and duplicated into 181 genes in the common ancestor of the three Sapindaceae species. These 181 ancestral Sapindaceae genes experienced different evolutionary patterns, which were reflected by species-specific gene duplication/loss events during the speciation of *X. sorbifolium*, *D. longan*, and *A. yangbiense*. *X. sorbifolium* duplicated 66 genes (57 CNL, nine TNL) and lost 88 genes (80 CNL, seven TNL, one RNL) during speciation (**Figures 5A,B** and **Supplementary Figure S6**), resulting in a slight decrease in the number of the NBS-encoding genes in its genome. Similarly, the common ancestor of *D. longan* and *A. yangbiense* after

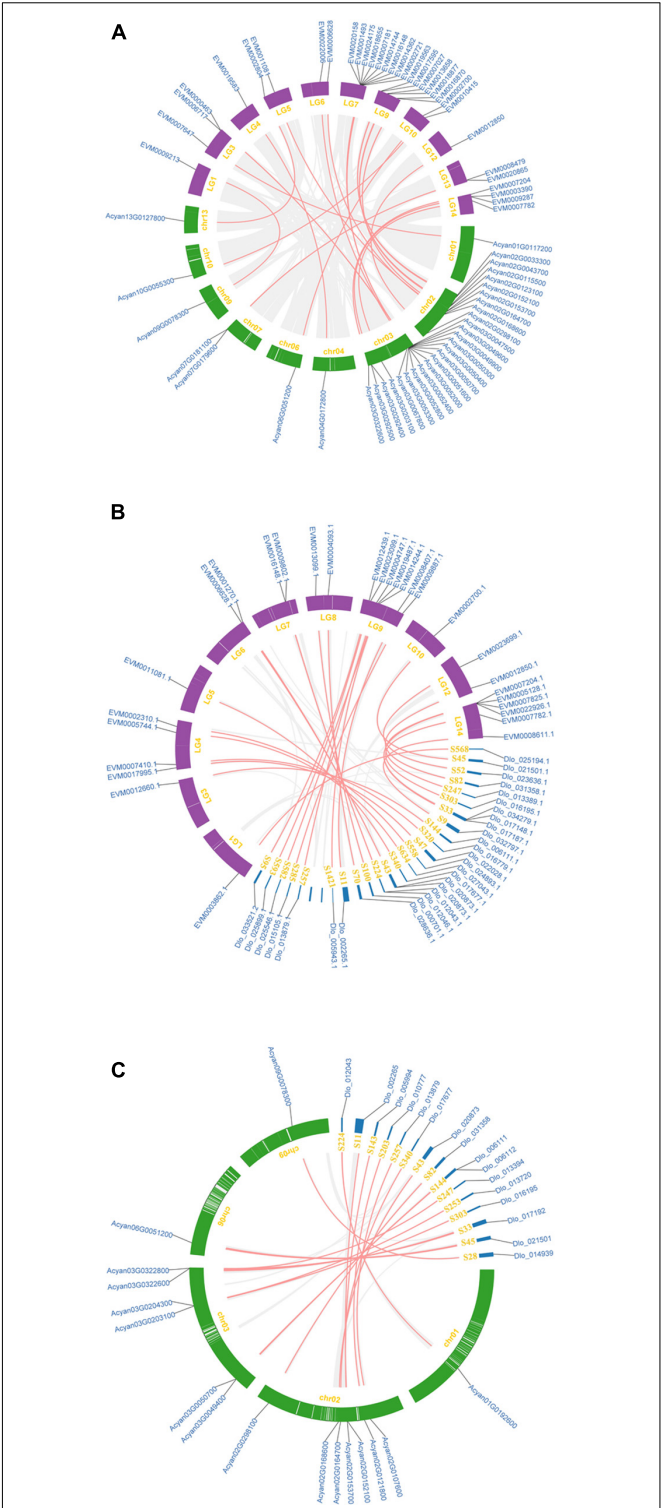


FIGURE 4 | Synteny of the three Sapindaceae species NBS-encoding genes. **(A)** Synteny of *X. sorbifolium* and *A. yangbiense* NBS-encoding genes. **(B)** Synteny of *X. sorbifolium* and *D. longan* NBS-encoding genes. **(C)** Synteny of *A. yangbiense* and *D. longan* NBS-encoding genes. LG1-15, *X. sorbifolium* chromosomes; chr01-13, *A. yangbiense* chromosomes; S, *D. longan* scaffolds. Red and gray lines represent syntenic NBS-encoding genes and non-NBS-encoding genes, respectively.

diverging from *X. sorbifolium* lost 25 genes (18 CNL, seven TNL) and duplicated 12 genes (11 CNL, one TNL), resulting in a decrease in the total number of genes (168). However, *D. longan* and *A. yangbiense* underwent considerably different evolutionary processes after divergence (Figures 5A,B and Supplementary Figure S6). Specifically, *D. longan* duplicated 340 genes (321 CNL, 16 TNL, three RNL), but lost 40 genes (37 CNL, two TNL, one RNL), thus, the gene number clearly increased in its genome. *A. yangbiense* duplicated 106 genes (97 CNL, eight TNL, one RNL) and lost 101 genes (85 CNL, 14 TNL, two RNL), resulting in a slight increase in the number of NBS-encoding genes in its genome. Therefore, the NBS-encoding genes in the three Sapindaceae species exhibited dynamic and distinct evolutionary patterns.

DISCUSSION

Copy Number Variation of the NBS-Encoding Genes Among Different Species

The NBS-encoding gene family is one of the most divergent plant genome families with high copy number discrepancies among plant lineages/species (Jacob et al., 2013). For example, several NBS-encoding gene sequences were identified in some orchid species genomes (Xue et al., 2020), which possess the smallest number of genes observed thus far. Dozens of NBS-encoding genes have been identified in papaya, melon, cucumber, and watermelon (Lin et al., 2013; Zhang et al., 2016). Hundreds of gene copies in *Arabidopsis*, rice, potato, tomato, pepper, cotton, soybean, grape, poplar, barley, and sunflower have also been identified (Yang et al., 2008; Li et al., 2010; Guo et al., 2011; Wei et al., 2013; Shao et al., 2014; Wu et al., 2014; Andersen et al., 2016; Qian et al., 2017; Neupane et al., 2018). Moreover, 1219 and 1303 copies were identified in wheat and apple genomes, respectively (Jia et al., 2013, 2015). In this study, the gene number variation among the three Sapindaceae species was also quite large, with 180, 568, and 252 genes identified in *X. sorbifolium*, *D. longan*, and *A. yangbiense*, respectively. The proportion of NBS-encoding genes to all predicted genes in the three Sapindaceae genomes (0.73–1.83%) was higher than the proportions reported for Cucurbitaceae species (0.19–0.27%), similar to Rosaceae (0.78–2.05%) and Solanaceae (0.73–1.15%) species (Jia et al., 2015; Qian et al., 2017). The discrepancy of NBS-encoding genes among the three species should not be the result of ploidy variation, because all of the three analyzed species are diploids and have similar chromosomal numbers (Lin et al., 2017; Bi et al., 2019; Yang et al., 2019). Actually, a previous study has shown that no correlation exists between the NBS-encoding genes and species phylogeny or genome size (Jacob et al., 2013). For example, in Fabaceae, soybean (465) possessed fewer NBS-encoding genes than *M. truncatula* (571), although the genomes size of the former was twice that of the latter (Shao et al., 2014). In Brassicaceae, the numbers of NBS-encoding genes were similar across some species even though these species exhibited variations in genome size (Zhang et al., 2016).

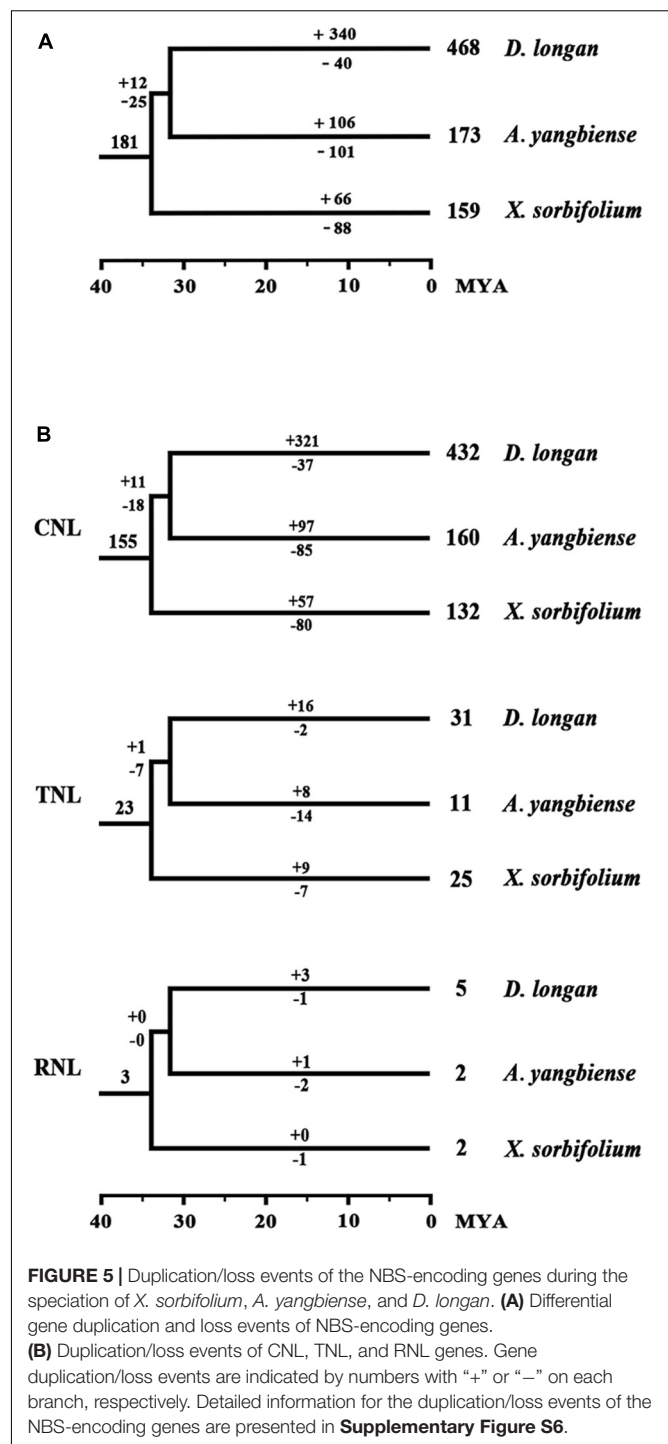
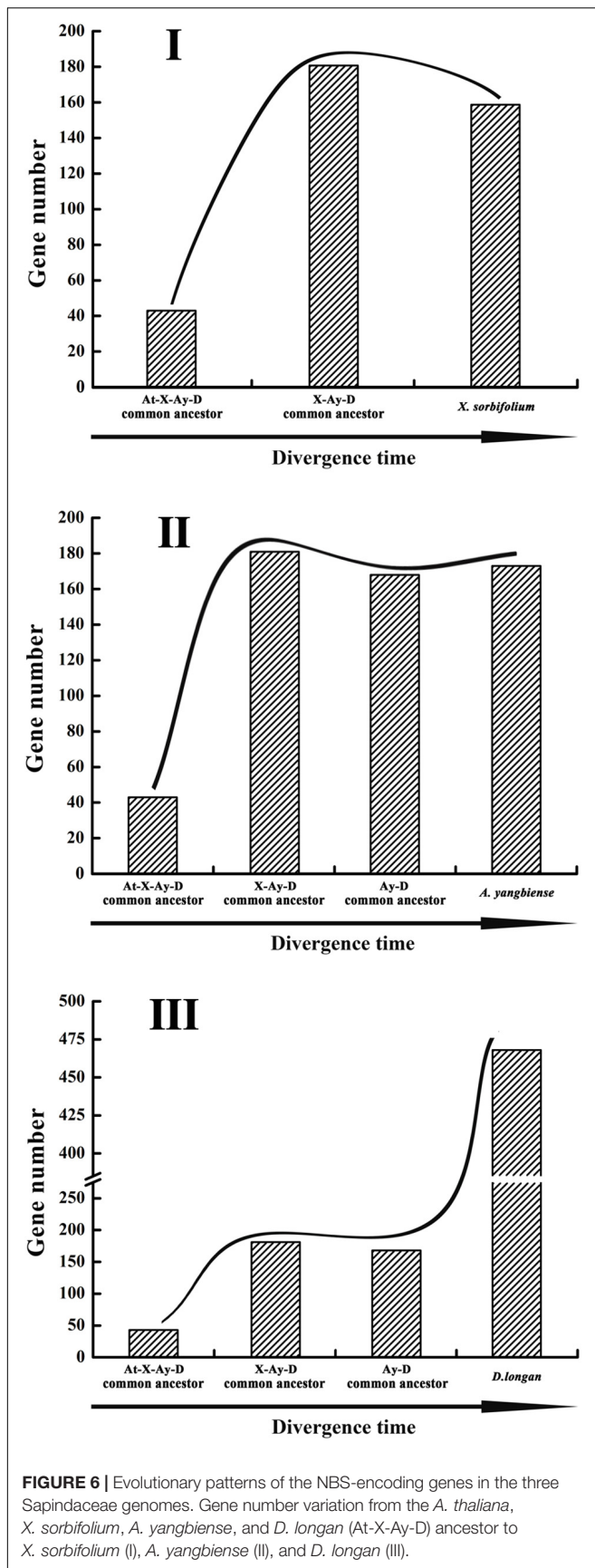


FIGURE 5 | Duplication/loss events of the NBS-encoding genes during the speciation of *X. sorbifolium*, *A. yangbiense*, and *D. longan*. **(A)** Differential gene duplication and loss events of NBS-encoding genes. **(B)** Duplication/loss events of CNL, TNL, and RNL genes. Gene duplication/loss events are indicated by numbers with “+” or “-” on each branch, respectively. Detailed information for the duplication/loss events of the NBS-encoding genes are presented in **Supplementary Figure S6**.

The *A. yangbiense* genome was sequenced using the combined Pacific Biosciences Single-molecule Real-time, Illumina HiSeq X, and Hi-C technologies, and BUSCO analysis recovered 95.5% complete BUSCO genes (Yang et al., 2019); the *X. sorbifolium* genome was sequenced using the combined Illumina HiSeq, Pacific Biosciences Sequel, and Hi-C technologies, and BUSCO analysis recovered 94.7% complete BUSCO genes (Bi et al., 2019); for *D. longan* genome sequencing: using the standard Illumina



library preparation protocols, and BUSCO analysis recovered 94% complete BUSCO genes (Lin et al., 2017). Additionally, the assembly quality of *D. longan* genome sequence was assessed by aligning the scaffolds to a *D. longan* transcriptome assembly from the NCBI Sequence Read Archive (SRA; SRA050205). Of the 96,251 *D. longan* transcriptome sequences reported previously (Lai and Lin, 2013), 97.55% were identified in the genome assembly (Lin et al., 2017). Although all of the three analyzed genomes are of high quality (all recovered > 94% complete BUSCO genes) according to the BUSCO analysis of the genome papers (Lin et al., 2017; Bi et al., 2019; Yang et al., 2019), the *D. longan* genome assembly has not reached a chromosomal level. A previous study revealed that genome sequence and annotation quality may affect the NBS-encoding gene identification (Bayer et al., 2018). This raises a possibility that the fragmented assembled genome of *D. longan* may have contributed to its high NBS-encoding gene number by splitting of one NBS-encoding gene into multiple genes during annotation. However, this possibility was largely rejected by coding sequence length analysis, which revealed that the average coding sequence length of NBS-encoding genes in *D. longan* is larger than that of *X. sorbifolium*. Furthermore, a close look at NBS-encoding gene profiles in the three species revealed that the number of intact NBS-encoding genes in *D. longan* (247) is even larger than the total NBS-encoding gene number in *X. sorbifolium* (180) and close to the total NBS-encoding gene number in *A. yangbiense*. Therefore, the relatively larger number of NBS-encoding genes in *D. longan* is more likely a consequence of evolution rather than other artificial reasons.

The rapid evolutionary feature of the NBS-encoding gene family caused by frequent gene duplications/loss events could elicit copy number discrepancies. The phylogenetic analysis and ancestral gene classification results of the NBS-encoding genes indicated that species-specific gene duplication/loss events occurred after species diverged from the common ancestor (Figures 2, 5A and Supplementary Figure S6). Lots of *D. longan*-specific lineage genes were found in *D. longan* and these specific lineage genes experienced recent sharp expansions by more frequent gene duplications and less gene loss events (+340 vs. -40) after the divergence of *D. longan* from *A. yangbiense* (Figure 4). The differences between gene duplication/loss events in *X. sorbifolium* (+66 vs. -88) or *A. yangbiense* (+106 vs. -101) were not as large (Figure 5A). Therefore, a copy number discrepancy of NBS-encoding genes was observed among the three Sapindaceae species. The NBS-encoding gene number variations caused by independent gene duplication/loss events were common in Cucurbitaceae, Fabaceae, Rosaceae, Poaceae, Brassicaceae, Solanaceae, and orchid species (Li et al., 2010; Lin et al., 2013; Shao et al., 2014; Jia et al., 2015; Zhang et al., 2016; Qian et al., 2017; Xue et al., 2020). The high copy number of NBS-encoding genes that resulted from recent expansions in *D. longan* could represent candidate resistance genes that responded to various pathogens, such as witches broom disease (Lai et al., 2000; Guo, 2002). From another perspective, high copy number of NBS-encoding genes might be a disadvantage for plants in the absence of corresponding pathogens due to the fitness

cost of resistance genes (Tian et al., 2003). There should be some balancing mechanisms between resistance and fitness cost provided by the large number of NBS-encoding genes in *D. longan* genome.

Chromosomal distribution analysis of NBS-encoding genes suggested an uneven distribution pattern among different chromosomes (**Supplementary Figure S3, S4**). The uneven distributing pattern was commonly observed in other lineages, like legume, Brassicaceae, Solanaceae species, *Vitis vinifera*, and *Populus trichocarpa* (Yang et al., 2008; Shao et al., 2014; Zhang et al., 2016; Qian et al., 2017). Tandem duplication and dispersed duplication played major roles in the NBS-encoding gene expansion in *X. sorbifolium* and *A. yangbiense* genomes (**Table 3**). Random dispersed gene duplications and gene losses likely brought out the uneven distributions of these genes on different chromosomes, and this difference was made more apparent through local tandem duplications. This strategy made the NBS-encoding gene form clusters to overcome the limitations of R gene diversity during the coevolution of plants and pathogens (Michelmore and Meyers, 1998; Le Roux et al., 2015).

The dynamic evolution of NBS-encoding genes in the three Sapindaceae species and *A. thaliana* was traced back to different time periods by reconciling ancient genes in the common ancestor of the three Sapindaceae species and the common ancestor of Sapindaceae and *A. thaliana*. In total, 43 ancient Malvids lineage genes were restored in the common ancestor of Sapindaceae and *A. thaliana* (**Figure 2**). Among these ancient genes, 10 were lost by the Sapindaceae common ancestor before the three Sapindaceae species diverged. The remaining 33 Malvids lineage genes were inherited and duplicated into 181 genes in the common ancestor of the three Sapindaceae species. These genes experienced further species-specific evolution during speciation and resulted in the currently observed NBS-encoding genes of the three genomes. Over time, the ancient NBS-encoding genes in the common ancestor of Sapindaceae and *A. thaliana* experienced different evolutionary patterns during the speciation of the three Sapindaceae species (**Figure 6**). *X. sorbifolium* exhibited a “first expansion and then contraction” evolutionary pattern (**Figure 6I**); more genes were duplicated in the common ancestor of Sapindaceae, but subsequently lost more genes during speciation. Although *A. yangbiense* and *D. longan* exhibited similar evolutionary patterns (**Figures 6II,III**), expansion followed by contraction and further expansion, subsequent expansion in *D. longan* was stronger and gained more genes than *A. yangbiense* after their divergence (**Figure 5A**). Furthermore, the evolutionary patterns of these three subclasses of NBS-encoding genes were also diverse among the three species. For example, the CNL and RNL genes exhibited similar patterns as the NBS-encoding genes of corresponding species (**Figure 5B**), while the patterns of TNL genes were altered in *A. yangbiense* (recent further contraction) and *X. sorbifolium* (recent further expansion) (**Figure 5B**). Such small changes in TNL genes had little effect on the overall trends of the NBS-encoding genes in *A. yangbiense* and *X. sorbifolium* due to the dominance of CNL genes (see discussion below). Therefore,

distinct and dynamic evolutionary patterns of the NBS-encoding genes caused by independent gene duplication/loss events appeared during the speciation of the three Sapindaceae species.

Numbers of Subclass Genes and the Causes

The number of CNL subclass genes in *X. sorbifolium*, *D. longan*, and *A. yangbiense* was considerably higher than TNL and RNL genes, accounting for 81.11, 89.26, and 89.68% of the NBS-encoding genes in each species, respectively. The greater number of CNL subclass genes is a common phenomenon observed among plant species, except for the Brassicaceae (Zhang et al., 2016). For example, CNL subclass genes are the majority in the genome of *Glycine max* (57.8%), *Phaseolus vulgaris* (66.5%), *V. vinifera* (82.8%), *Capsicum annuum* (94.1%), *Solanum lycopersicum* (87.0%), *Solanum tuberosum* (83.7%), and *Amborella trichopoda* (84.8%) (Yang et al., 2008; Shao et al., 2014, 2016; Jia et al., 2015; Qian et al., 2017). Extremely, monocots have only CNL genes due to the loss of TNL genes near dicot/monocot differentiation (Li et al., 2010; Shao et al., 2016). Previously, it was reported that CNL genes underwent twice gene expansions and expanded earlier than TNL genes during angiosperm evolution (Shao et al., 2016). A total of 155 CNL, 23 TNL, and three RNL ancestral genes were reconciled in the common ancestor of the three Sapindaceae species. Therefore, when the common ancestor of the three Sapindaceae species emerged, its genome possessed more CNL genes than TNL or RNL genes. Although these ancestral genes have experienced frequent gene duplication/loss events during the divergence of the three Sapindaceae species, gene expansion in their common ancestor maintained the large number of CNL genes. Additionally, lots of recent expansions of CNL genes in the three Sapindaceae species may have given rise to the dominance of this subclass's genes (**Figures 2, 5** and **Supplementary Figure S6**). An extremely small number of RNL genes were previously reported in angiosperms (Shao et al., 2016; Zhang et al., 2016; Qian et al., 2017; Xue et al., 2020). In this study, two, six, and nine RNL genes were identified in *X. sorbifolium*, *A. yangbiense*, and *D. longan*, respectively. It was speculated that CNL and TNL genes were involved in recognizing specific pathogens, while RNL genes helped transduce signals in the downstream pathways of plant immunity (Bonardi et al., 2011; Collier et al., 2011; Tamborski and Krasileva, 2020). Therefore, RNL genes clearly participated in basic defense responses and without the divergent selection of pathogens, these genes did not necessarily expand in large numbers. Moreover, pairwise synteny analysis detected more syntenic CNL (74 genes) subclass genes than TNL (two genes) and RNL (one gene) subclass genes among the three Sapindaceae genomes (**Figure 4** and **Supplementary Table S2**). These differences might be caused by the different evolutionary feature of the three subclass genes during angiosperm evolution: long-term contraction and fast evolution rate of TNL genes, gradual expansion of CNL genes, and conservative evolution of RNL genes (Zhang et al., 2011; Shao et al., 2016).

Notably, not all of the identified NBS-encoding genes in *X. sorbifolium*, *A. yangbiense*, and *D. longan* genomes had intact structures possessing all three domains. There were only 81 (24 TNL, 55 CNL, two RNL), 31 (three TNL, 26 CNL, two RNL), and 247 (24 TNL, 218 CNL, five RNL) NBS-encoding genes with intact structures in *X. sorbifolium*, *A. yangbiense*, and *D. longan* genomes (Table 1), respectively, accounting for 45.0, 12.3, and 43.5% of all identified NBS-encoding genes in these three species, respectively. Similarly, the relative small proportion of NBS-encoding genes with intact structures were also reported in *C. annuum* (23.2%), *S. lycopersicum* (42.7%), *S. tuberosum* (28.2%), *P. trichocarpa* (46.2%), *M. truncatula* (39.1%), *Lotus japonicus* (31.0%), and *Oryza sativa* (30.6%) genomes (Yang et al., 2008; Zhang et al., 2011; Shao et al., 2014; Qian et al., 2017). In plants, the highly conserved NBS domain has been demonstrated to be able to bind and hydrolyze ATP or GTP and act as molecular switches in immune signaling, while the LRR and N-terminal domains (TIR and CC) are typically involved in the recognition of, and the activation of, corresponding partners, respectively (Meyers et al., 1999; Dangl and Jones, 2001; Lukasik and Takken, 2009). Theoretically, in order to trigger immune responses and transfer defense signals, an NBS-encoding gene should possess an intact structure. However, transient overexpression results showed that those NBS-encoding genes without intact structures also function in plant immunity (Nandety et al., 2013; Kato et al., 2014).

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the *X. sorbifolium*: <http://dx.doi.org/10.5524/100606>, *D. longan*: <http://dx.doi.org/10.5524/100276>, and *A. yangbiense*: <http://dx.doi.org/10.5524/100610>.

AUTHOR CONTRIBUTIONS

G-CZ, WL, and Y-LW conceived and directed the project. G-CZ, WL, Y-MZ, and YL obtained and analyzed the data. Y-MZ, ML, MZ, YL, and G-QM conducted the phylogenetic analysis and constructed the discussion. G-CZ wrote the manuscript. All of the authors contributed to the discussion of the results, reviewed the manuscript, and approved of the final article.

REFERENCES

- Ameline-Torregrosa, C., Wang, B. B., Oblesness, M., Deshpande, S., Zhu, H., Roe, B. A., et al. (2008). Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol.* 146, 5–21. doi: 10.1104/pp.107.104588
- Andersen, E. J., Ali, S., Reese, R. N., Yen, Y., Neupane, S., and Nepal, M. P. (2016). Diversity and evolution of disease resistance genes in barley (*Hordeum vulgare* L.). *Evolut. Bioinform.* 12:S38085.
- Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, W369–W373.

FUNDING

This work was funded by the Doctoral Fund (No. XY19BS15) and the First Training Program of Institute of Peony of Heze University (provided to G-CZ), and Qingchuang Science and Technology Support Program of Shandong Provincial College. The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00737/full#supplementary-material>

FIGURE S1 | Phylogenetic relationship of *X. sorbifolium*, *A. yangbiense*, and *D. longan*. Time of divergence: million years ago (MYA) (Koenen et al., 2015; Liu et al., 2015; Lin et al., 2017; Bi et al., 2019; Yang et al., 2019).

FIGURE S2 | Details of the amino acid frequencies of the whole NBS domain of CNL, TNL, and RNL genes in the three Sapindaceae species (WebLogo).

FIGURE S3 | The chromosomal distribution of identified NBS-encoding genes in the *X. sorbifolium* genome.

FIGURE S4 | The chromosomal distribution of identified NBS-encoding genes in the *A. yangbiense* genome.

FIGURE S5 | The detailed ML tree of the NBS-encoding genes. Sequence names and supporting values are presented on all of the nodes. The tree was reconstructed based on the amino acids of the NBS domain of the NBS-encoding genes in *X. sorbifolium*, *A. yangbiense*, and *D. longan* using NBS-encoding genes in the *A. thaliana* genome were used as a reference.

FIGURE S6 | The reconciled NBS-encoding gene tree with real species phylogeny and restored various duplication/loss events. “n4” indicates a loss event that occurred in the common ancestor of *X. sorbifolium*, *A. yangbiense*, and *D. longan*. “n2” indicates a loss event that occurred in the common ancestor of *A. yangbiense* and *D. longan*.

TABLE S1 | A list of NBS-encoding genes identified from the three Sapindaceae genomes.

TABLE S2 | Detailed information of the synteny NBS-encoding genes in three Sapindaceae genomes.

- Bayer, P. E., Edwards, D., and Batley, J. (2018). Bias in resistance gene prediction due to repeat masking. *Nat. Plants* 4, 762–765. doi: 10.1038/s41477-018-0264-0
- Bi, Q., Zhao, Y., Du, W., Lu, Y., Gui, L., Zheng, Z., et al. (2019). Pseudomolecule-level assembly of the Chinese oil tree yellowhorn (*Xanthoceras sorbifolium*) genome. *GigaScience* 8:giz070.
- Bi, W., Gao, Y., Shen, J., He, C., Liu, H., Peng, Y., et al. (2016). Traditional uses, phytochemistry, and pharmacology of the genus *Acer* (maple): a review. *J. Ethnopharmacol.* 189, 31–60. doi: 10.1016/j.jep.2016.04.021
- Bonardi, V., Tang, S., Stallmann, A., Roberts, M., Cherkis, K., and Dangl, J. L. (2011). Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16463–16468. doi: 10.1073/pnas.1113726108

- Chen, Y., Yang, Q., and Zhu, G. (2003). *Acer yangbiense* (Aceraceae), a new species from Yunnan, China. *Novon* 13, 296–299.
- Chung, Y. C., Lin, C. C., Chou, C. C., and Hsu, C. P. (2010). The effect of Longan seed polyphenols on colorectal carcinoma cells. *Eur. J. Clin. Investigat.* 40, 713–721. doi: 10.1111/j.1365-2362.2010.02322.x
- Collier, S. M., Hamel, L.-P., and Moffett, P. (2011). Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Mol. Plant Microbe Interact.* 24, 918–931. doi: 10.1094/mpmi-03-11-0050
- Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Dangl, J. L., Horvath, D. M., and Staskawicz, B. J. (2013). Pivoting the plant immune system from dissection to deployment. *Science* 341, 746–751. doi: 10.1126/science.1236011
- Dangl, J. L., and Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411, 826–833. doi: 10.1038/35081161
- Guo, J.-H. (2002). A list of diseases on litchi and longan in China mainland. *Subtrop. Plant Sci.* 31, 48–50.
- Guo, Y.-L., Fitz, J., Schneeberger, K., Ossowski, S., Cao, J., and Weigel, D. (2011). Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in *Arabidopsis*. *Plant Physiol.* 157, 757–769. doi: 10.1104/pp.111.181990
- Jacob, F., Vernaldi, S., and Maekawa, T. (2013). Evolution and conservation of plant NLR functions. *Front. Immunol.* 4:297.
- Ji, X.-F., Chi, T.-Y., Liu, P., Li, L.-Y., Xu, J.-K., Xu, Q., et al. (2017). The total triterpenoid saponins of *Xanthoceras sorbifolia* improve learning and memory impairments through against oxidative stress and synaptic damage. *Phytomedicine* 25, 15–24. doi: 10.1016/j.phymed.2016.12.009
- Jia, J., Zhao, S., Kong, X., Li, Y., Zhao, G., He, W., et al. (2013). *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 496, 91–95.
- Jia, Y., Yuan, Y., Zhang, Y., Yang, S., and Zhang, X. (2015). Extreme expansion of NBS-encoding genes in Rosaceae. *BMC Genet.* 16:48.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14:587. doi: 10.1038/nmeth.4285
- Kato, H., Saito, T., Ito, H., Komeda, Y., and Kato, A. (2014). Overexpression of the TIR-X gene results in a dwarf phenotype and activation of defense-related gene expression in *Arabidopsis thaliana*. *J. Plant Physiol.* 171, 382–388. doi: 10.1016/j.jplph.2013.12.002
- Koenen, E. J., Clarkson, J. J., Pennington, T. D., and Chatrou, L. W. (2015). Recently evolved diversity and convergent radiations of *Rainforest mahoganies* (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytol.* 207, 327–339. doi: 10.1111/nph.13490
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Lai, Z., Chen, C., Zeng, L., and Chen, Z. (2000). “Somatic embryogenesis in longan [*Dimocarpus longan* Lour.],” in *Somatic Embryogenesis in Woody Plants. Forestry Sciences*, Vol. 67, eds S. M. Jain, P. K. Gupta, and R. J. Newton (Dordrecht: Springer), 415–431. doi: 10.1007/978-94-017-3030-3_13
- Lai, Z., and Lin, Y. (2013). Analysis of the global transcriptome of longan (*Dimocarpus longan* Lour.) embryogenic callus using Illumina paired-end sequencing. *BMC Genomics* 14:561. doi: 10.1186/1471-2164-14-561
- Le Roux, C., Huet, G., Jauneau, A., Camborde, L., Tremousaygue, D., Kraut, A., et al. (2015). A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell* 161, 1074–1088. doi: 10.1016/j.cell.2015.04.025
- Lee, T., Tang, H., Wang, X., and Paterson, A. H. (2012). PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res.* 41, 1152–1158.
- Leister, D. (2004). Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends Genet.* 20, 116–122. doi: 10.1016/j.tig.2004.01.007
- Li, J., Ding, J., Zhang, W., Zhang, Y., Tang, P., Chen, J.-Q., et al. (2010). Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. *Mol. Genet. Genom.* 283, 427–438. doi: 10.1007/s00438-010-0527-6
- Lin, X., Zhang, Y., Kuang, H., and Chen, J. (2013). Frequent loss of lineages and deficient duplications accounted for low copy number of disease resistance genes in Cucurbitaceae. *BMC Genom.* 14:335. doi: 10.1186/1471-2164-14-335
- Lin, Y., Min, J., Lai, R., Wu, Z., Chen, Y., Yu, L., et al. (2017). Genome-wide sequencing of longan (*Dimocarpus longan* Lour.) provides insights into molecular basis of its polyphenol-rich characteristics. *Gigascience* 6:gix023.
- Liu, B., Ye, J., Liu, S., Wang, Y., Yang, Y., Lai, Y., et al. (2015). Families and genera of Chinese angiosperms: a synoptic classification based on APG III. *Biodiv. Sci.* 23, 225–231. doi: 10.17520/biods.2015052
- Liu, J., Liu, X., Dai, L., and Wang, G. (2007). Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants. *J. Genet. Genom.* 34, 765–776. doi: 10.1016/s1673-8527(07)60087-3
- Lukasik, E., and Takken, F. L. W. (2009). STANDING strong, resistance proteins instigators of plant defence. *Curr. Opin. Plant Biol.* 12, 427–436. doi: 10.1016/j.pbi.2009.03.001
- Luo, S., Zhang, Y., Hu, Q., Chen, J., Li, K., Lu, C., et al. (2012). Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes in the grass family. *Plant Physiol.* 159, 197–210. doi: 10.1104/pp.111.192062
- Mcdowell, J. M., Dhondy, M., Long, T. A., Aarts, M. G., Goff, S., Holub, E. B., et al. (1998). Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of *Arabidopsis*. *Plant Cell* 10, 1861–1874. doi: 10.1105/tpc.10.11.1861
- Mchale, L., Tan, X., Koehl, P., and Michelmore, R. W. (2006). Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 7:212.
- Mei, Z., Fu, S., Yu, H., Yang, L., Duan, C., Liu, X., et al. (2014). Genetic characterization and authentication of *Dimocarpus longan* Lour. using an improved RAPD technique. *Genet. Mol. Res.* 13, 1447–1455. doi: 10.4238/2014.march.6.3
- Meyers, B. C., Dickerman, A. W., Michelmore, R. W., Sivaramakrishnan, S., Sobral, B. W. S., and Young, N. D. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* 20, 317–332. doi: 10.1046/j.1365-3113.1999.t01-1-00606.x
- 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
- Minh, B. Q., Nguyen, M. A. T., and Von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024
- Nandety, R. S., Caplan, J. L., Cavanaugh, K. A., Perroud, B., Wroblewski, T., Michelmore, R. W., et al. (2013). The role of TIR-NBS and TIR-X proteins in plant basal defense responses. *Plant Physiol.* 162, 1459–1472. doi: 10.1104/pp.113.219162
- Neupane, S., Andersen, E. J., Neupane, A., and Nepal, M. P. (2018). Genome-wide identification of NBS-encoding resistance genes in sunflower (*Helianthus annuus* L.). *Genes* 9:384. doi: 10.3390/genes9080384
- Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Qian, L.-H., Zhou, G.-C., Sun, X.-Q., Lei, Z., Zhang, Y.-M., Xue, J.-Y., et al. (2017). Distinct patterns of gene gain and loss: diverse evolutionary modes of NBS-encoding genes in three Solanaceae crop species. *G3 Genes Genomes Genetics* 7, 1577–1585. doi: 10.1534/g3.117.040485
- Shao, Z.-Q., Xue, J.-Y., Wang, Q., Wang, B., and Chen, J.-Q. (2019). Revisiting the origin of plant NBS-LRR genes. *Trends Plant Sci.* 24, 9–12. doi: 10.1016/j.tplants.2018.10.015
- 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 genes reveal three anciently diverged classes with distinct evolutionary patterns. *Plant Physiol.* 170, 2095–2109. doi: 10.1104/pp.15.01487
- Shao, Z.-Q., Zhang, Y.-M., Hang, Y.-Y., Xue, J.-Y., Zhou, G.-C., Wu, P., et al. (2014). Long-term evolution of nucleotide-binding site-leucine-rich repeat genes: understanding gained from and beyond the legume family. *Plant Physiol.* 166, 217–234. doi: 10.1104/pp.114.243626

- Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernot, B., and Durand, D. (2012). Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 28, i409–i415. doi: 10.1093/bioinformatics/bts386
- Tamborski, J., and Krasileva, K. V. (2020). Evolution of plant NLRs: from natural history to precise modifications. *Ann. Rev. Plant Biol.* 71, 355–378. doi: 10.1146/annurev-arplant-081519-035901
- Tian, D., Traw, M. B., Chen, J., Kreitman, M., and Bergelson, J. (2003). Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* 423, 74–77. doi: 10.1038/nature01588
- Venegas-Calderón, M., Ruiz-Méndez, M. V., Martínez-Force, E., Garcés, R., and Salas, J. J. (2017). Characterization of *Xanthoceras sorbifolium* Bunge seeds: lipids, proteins and saponins content. *Industr. Crops Product.* 109, 192–198. doi: 10.1016/j.indcrop.2017.08.022
- 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.
- Wei, H., Li, W., Sun, X., Zhu, S., and Zhu, J. (2013). Systematic analysis and comparison of nucleotide-binding site disease resistance genes in a diploid cotton *Gossypium raimondii*. *PLoS One* 8:e68435. doi: 10.1371/journal.pone.0068435
- Whitham, S., Dinesh-Kumar, S. P., Choi, D., Hehl, R., Corr, C., and Baker, B. (1994). The product of the tobacco mosaic virus resistance gene *N*: similarity to toll and the interleukin-1 receptor. *Cell* 78, 1101–1115. doi: 10.1016/0092-8674(94)90283-6
- Wu, P., Shao, Z.-Q., Wu, X.-Z., Wang, Q., Wang, B., Chen, J.-Q., et al. (2014). Loss/retention and evolution of NBS-encoding genes upon whole genome triplication of *Brassica rapa*. *Gene* 540, 54–61. doi: 10.1016/j.gene.2014.01.082
- Xue, J.-Y., Zhao, T., Liu, Y., Liu, Y., Zhang, Y.-X., Zhang, G.-Q., et al. (2020). Genome-wide analysis of the nucleotide binding site leucine-rich repeat genes of four orchids revealed extremely low numbers of disease resistance genes. *Front. Genet.* 10:1286.
- Yang, J., Wariss, H. M., Tao, L., Zhang, R., Yun, Q., Hollingsworth, P., et al. (2019). De novo genome assembly of the endangered *Acer yangbiense*, a plant species with extremely small populations endemic to Yunnan Province, China. *GigaScience* 8:giz085.
- Yang, S., Zhang, X., Yue, J.-X., Tian, D., and Chen, J.-Q. (2008). Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genom.* 280, 187–198. doi: 10.1007/s00438-008-0355-0
- Yue, J. X., Meyers, B. C., Chen, J. Q., Tian, D., and Yang, S. (2012). Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS–LRR) genes. *New Phytol.* 193, 1049–1063. doi: 10.1111/j.1469-8137.2011.04006.x
- Zhai, C., Lin, F., Dong, Z., He, X., Yuan, B., Zeng, X., et al. (2011). The isolation and characterization of Pik, a rice blast resistance gene which emerged after rice domestication. *New Phytol.* 189, 321–334. doi: 10.1111/j.1469-8137.2010.03462.x
- Zhang, X., Feng, Y., Cheng, H., Tian, D., Yang, S., and Chen, J. (2011). Relative evolutionary rates of NBS-encoding genes revealed by soybean segmental duplication. *Mol. Genet. Genom.* 285, 79–90. doi: 10.1007/s00438-010-0587-7
- Zhang, Y., Xiao, L., Xiao, B., Yin, M., Gu, M., Zhong, R., et al. (2018). Research progress and application prospect of *Xanthoceras sorbifolia* for treating Alzheimer's disease. *Drug Eval. Res.* 25, 912–917.
- Zhang, Y.-M., Chen, M., Wang, Y., Yin, J., Liu, J., Sun, X.-Q., et al. (2020). Genome-wide identification and evolutionary analysis of NBS-LRR genes from *Dioscorea rotundata*. *Front. Genet.* 11:484.
- Zhang, Y. M., Shao, Z. Q., Wang, Q., Hang, Y. Y., Xue, J. Y., Wang, B., et al. (2016). Uncovering the dynamic evolution of nucleotide-binding site–leucine-rich repeat (NBS–LRR) genes in Brassicaceae. *J. Integrat. Plant Biol.* 58, 165–177. doi: 10.1111/jipb.12365

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 Zhou, Li, Zhang, Liu, Zhang, Meng, Li 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.



Wheat Disease Resistance Genes and Their Diversification Through Integrated Domain Fusions

Ethan J. Andersen^{1*}, Madhav P. Nepal^{2*}, Jordan M. Purintun², Dillon Nelson³, Glykeria Mermigka⁴ and Panagiotis F. Sarris^{4,5,6}

¹ Department of Biology, Francis Marion University, Florence, SC, United States, ² Department of Biology and Microbiology, South Dakota State University, Brookings, SD, United States, ³ Department of Math, Science and Technology, Oglala Lakota College, Kyle, SD, United States, ⁴ Department of Biology, University of Crete, Crete, Greece, ⁵ Institute of Molecular Biology and Biotechnology, FORTH, Crete, Greece, ⁶ School of Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom

OPEN ACCESS

Edited by:

Genlou Sun,
Saint Mary's University, Canada

Reviewed by:

Robert A. Haney,
Ball State University, United States
Xiaoqin Sun,
Institute of Botany (CAS), China

*Correspondence:

Ethan J. Andersen
Ethan.Andersen@fmarion.edu
Madhav P. Nepal
madhav.nepal@sdstate.edu

Specialty section:

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

Received: 21 December 2019

Accepted: 20 July 2020

Published: 05 August 2020

Citation:

Andersen EJ, Nepal MP, Purintun JM, Nelson D, Mermigka G and Sarris PF (2020) Wheat Disease Resistance Genes and Their Diversification Through Integrated Domain Fusions. *Front. Genet.* 11:898. doi: 10.3389/fgene.2020.00898

Plants are in a constant evolutionary arms race with their pathogens. At the molecular level, the plant nucleotide-binding leucine-rich repeat receptors (NLRs) family has coevolved with rapidly evolving pathogen effectors. While many NLRs utilize variable leucine-rich repeats (LRRs) to detect effectors, some have gained integrated domains (IDs) that may be involved in receptor activation or downstream signaling. The major objectives of this project were to identify NLR genes in wheat (*Triticum aestivum* L.) and assess IDs associated with immune signaling (e.g., kinase and transcription factor domains). We identified 2,151 NLR-like genes in wheat, of which 1,298 formed 547 gene clusters. Among the non-toll/interleukin-1 receptor NLR (non-TNL)-like genes, 1,552 encode LRRs, 802 are coiled-coil (CC) domain-encoding (CC-NBS-LRR or CNL) genes, and three encode resistance to powdery mildew 8 (RPW8) domains (RPW8-NBS-LRR or RNL). The expansion of the NLR gene family in wheat is attributable to its origin by recent polyploidy events. Gene clusters were likely formed by tandem duplications, and wheat NLR phylogenetic relationships were similar to those in barley and *Aegilops*. We also identified wheat NLR-ID fusion proteins as candidates for NLR functional diversification, often as kinase and transcription factor domains. Comparative analyses of the IDs revealed evolutionary conservation of more than 80% amino acid sequence similarity. Homology assessment indicates that these domains originated as functional non-NLR-encoding genes that were incorporated into NLR-encoding genes through duplication events. We also found that many of the NLR-ID genes encode alternative transcripts that include or exclude IDs, a phenomenon that seems to be conserved among species. To verify this, we have analyzed the alternative transcripts that include or exclude an ID of an NLR-ID from another monocotyledon species, rice (*Oryza sativa*). This indicates that plants employ alternative splicing to regulate IDs, possibly using them as baits, decoys, and functional signaling components. Genomic and expression data support the hypothesis that wheat uses alternative splicing to include and exclude IDs from NLR proteins.

Keywords: wheat R-genes, integrated domain, pathogen resistance, alternative splicing, disease resistance

INTRODUCTION

Plant innate immune systems utilize specialized receptor proteins to detect pathogens (Jones and Dangl, 2006; Duxbury et al., 2016). Nucleotide-binding leucine-rich repeat receptor (NLR) proteins detect pathogen effectors that would otherwise inhibit host resistance responses (Jones et al., 2016). NLRs can be recognized by the domain nucleotide-binding site found in apoptotic protease activating factor 1, resistance genes, and *Caenorhabditis elegans* death-4 protein (NB-ARC) (Mermigka et al., 2020). The NB-ARC is associated with ATP/ADP binding, since the molecule uses it upon activation. In order to detect hundreds of pathogens and pests, immune receptors must be able to respond to many elicitors. To accomplish this, NLRs have radiated to form a diverse family of resistance genes in plants (Shao et al., 2016). Much of this diversity originated via gene duplication and variation in NLR leucine-rich repeats (LRRs), which allows NLRs to bind to new effectors (Deyoung and Innes, 2006). Diversification has led to the formation of networks of sensor and helper NLRs, with some NLRs dimerizing to initiate signaling (Hayashi et al., 2010; Kofoed and Vance, 2011; Bonardi et al., 2012; Bonardi and Dangl, 2012). NLRs have also diversified by gaining extra domains that may facilitate pathogen recognition or resistance signaling. These domains, called integrated domains (IDs), resulted from fusions of NLRs and other functional domains also involved in resistance, as outlined by the integrated decoy/sensor model (Cesari et al., 2014; Wu et al., 2015; Duxbury et al., 2016). Studies of ID diversity across many plant genomes have revealed a diversity of domains associated with potential roles in resistance (Sarris et al., 2016).

Wheat (*Triticum aestivum* L.) provides approximately 20% of the human population's caloric intake (FAOstat, 2013) and is afflicted by over 100 different diseases caused by various pathogen and pest species (Murray et al., 2015). Diseases that substantially reduce yield, such as biotrophic rusts and necrotrophic leaf spotting diseases, impact global markets and food supplies. Historically devastating pathogens continue to produce new strains, which overcome past sources of resistance, such as in the case of Ug99 stem rust (Saari and Prescott, 1985; Pretorius et al., 2000; Huerta-Espino et al., 2011; Singh et al., 2011). Disease resistant wheat cultivars have improved agricultural productivity and our understanding of phytopathology (Mcfadden, 1930; Borlaug, 1983). Advanced genetic and genomic technologies have been used to identify pathogen resistance genes (R-genes), such as the recently discovered Ug99 R-genes *Sr33* and *Sr35* (Periyannan et al., 2013; Saintenac et al., 2013). The hexaploid bread wheat genome (AABBDD), a draft of which has recently become available (International Wheat Genome Sequencing Consortium, 2014), formed through the hybridization of three separate species: *Triticum urartu* (A), an unknown relative of *Aegilops speltoides* (B), and *Aegilops tauschii* (D) (Jia et al., 2013; Ling et al., 2013; Marcussen et al., 2014). The polyploid origins of wheat may have resulted in novel mechanisms to regulate its multiple progenitor resistance signaling pathways. The large and redundant nature of wheat's hexaploid genome makes it a good candidate for studying R-gene evolution with respect to recent polyploidization events.

The objectives of this research were to conduct genome-wide identification of wheat R-genes encoding NB-ARC domains, identify wheat NLR-ID fusion proteins, and assess their homology in wheat relatives and other monocot species. Results showed that tandem duplications can explain many of the events that led to the diversification of these genes. We also propose a mechanism to describe how plants carry out NLR-ID regulation, which became apparent while manually assessing the variation among NLR-ID transcripts. This mechanism was investigated using transient expression of a NLR gene from rice (*Oryza sativa* Japonica) in *Nicotiana benthamiana* leaves. Identifying the evolutionary patterns of the NLR-ID fusions improves our understanding of how NLRs diversify to oppose various pathogenic molecular weapons.

MATERIALS AND METHODS

NB-ARC Identification

Triticum aestivum chromosome, gene, and protein sequences were downloaded using the Biomart application within the Ensembl Genomes (Kersey et al., 2015) and Phytozome (Goodstein et al., 2011) databases. InterProScan annotations (Jones et al., 2014) were compiled, and proteins containing NB-ARC domain (PF00931) were investigated. The locations of genes encoding NB-ARCs were analyzed for clusters as described in Jupe et al. (2012), who used the criteria that clustered genes must be within 200,000 bases of each other and must be separated by fewer than eight additional genes between them (Jupe et al., 2012). NB-ARC domain motifs were also assessed using MEME software (Bailey et al., 2009), which was used to identify those with P-loop, Kinase-2, and GLPL motifs. Clustered genes with the aforementioned motifs were aligned and manually curated using the ClustalW2 program integrated within the program Geneious (Larkin et al., 2007; Kearse et al., 2012). The program MEGA 7 was used to construct a neighbor-joining tree with 100 bootstraps (Kumar et al., 2016) to assess whether the clustered genes nested together formed by potential tandem duplication. Wheat and *Aegilops tauschii* R-gene locations from Ensembl Genomes were used to construct a genomic map using the program Circa¹. Clustered wheat R-genes were also compared using the program Circoletto (Darzentas, 2010) to visualize similarities between genes.

Integrated Domain Identification

Pfam annotations not inherently part of the NLR structure (CC/TIR, NB-ARC, and LRR) were assembled. Amino acid sequences and corresponding annotations were uploaded to the program Geneious (Kearse et al., 2012) for sequence alignment, homology assessment, and motif visualization. The IDs were manually investigated to assess protein location, potential function, homology to proteins in other species, and the presence of variant transcripts. Function was assessed partially through domain descriptions available through the Pfam database (Finn et al., 2015), allowing for inferences about

¹<http://omgenomics.com/circa>

domain activity. Genomes investigated for homology include: *Aegilops tauschii*, *Amborella trichopoda*, *Arabidopsis thaliana*, *Brachypodium distachyon*, *Hordeum vulgare*, *Musa acuminata*, *Oryza sativa*, *Setaria italica*, *Triticum urartu*, and *Zea mays*. Genomic data was not available for *Aegilops speltoides*, which is believed to be the contributor of wheat's B genome. Special attention was paid to the relationship between wheat NLRs and their homologs in the two progenitors of wheat: *Triticum urartu* and *Aegilops tauschii*. These results were visualized to show how many NLR-ID fusions are shared between wheat and its progenitors in order to assess how many fusions took place before their divergence and how many may have happened since their divergence.

Alternative transcripts, also downloaded from the databases mentioned above, were assessed for ID motifs. Among the NLR-ID accessions, genes were identified for the transcripts where IDs were present in some but absent in others. The Gene Structure Display Server 2.0 (Hu et al., 2014) was used to visualize alternative splicing of NLR-IDs. Wheat expression data was acquired from the NCBI database and Wheat Gene Expression Atlas data (Borrill et al., 2016). The alternative transcript data was mapped onto expression data, showing experimental evidence that these alternative transcripts were expressed in wheat tissue.

Plasmid Constructions

Six genomic fragments of *OsRPR1* were PCR-amplified from *Oryza sativa* Japonica genomic DNA with primers (Table 1) containing 4bp specific overhangs and *BsaI* recognition sequence. The amplified products were cloned into the pCRTM8/GW/TOPO (Invitrogen K2500-20) vector. The resulting constructs, together with a C-terminal YFP tag, were subsequently used for Golden Gate assembly in pICH86988 (a kind gift from Dr. Sylvestre Marillonnet), thus generating the pICH86988::*OsRPR1*:YFP construct. For cloning of *OsRPR1* CDS, the amplified PCR product of *OsRPR1* (see below) was cloned into pICSL01005.

Agrobacterium-Mediated Transient Expression in *Nicotiana benthamiana*

For agroinfiltration in *Nicotiana benthamiana*, *Agrobacterium tumefaciens* strain AGL1 was transformed with the binary constructs by electroporation. *Agrobacterium* strains carrying the construct were grown in 5 ml liquid LB-medium supplemented with adequate antibiotic for 24 h. Cells were harvested by centrifugation, washed twice in 10 ml of 10 mM MgCl₂ and re-suspended at OD₆₀₀ 0.5 in infiltration medium (10 mM MgCl₂, 10 mM MES pH 5.6). Agroinfiltration was performed with 1 ml needleless syringe in 4–5 week-old *N. benthamiana* leaves.

RNA Extraction and RT-PCR

Total RNA was isolated from deep-frozen plant material using the TRIzol® method (Invitrogen) according to the manufacturer's specifications. For cDNA synthesis, 1 µg of DNaseI (NEB) treated total RNA was reverse transcribed using Superscript II (Invitrogen). For the amplification of *OsRPR1* CDS, PCR was conducted with Phusion Polymerase (NEB) using the

primers *OsRPR1*-Frg1 Fw and Frg5 Rv (Table 1) for 35 cycles following manufacturer's instructions. The correct size band was harvested from the gel. For the identification of other spliceforms at the 3' end of *OsRPR1* CDS, PCR was performed using Phusion Polymerase (NEB) with the primer set *OsRPR1*-SV-Fw and *OsRPR1*-SV-Rv (Table 1) for 35 cycles following manufacturer's instructions.

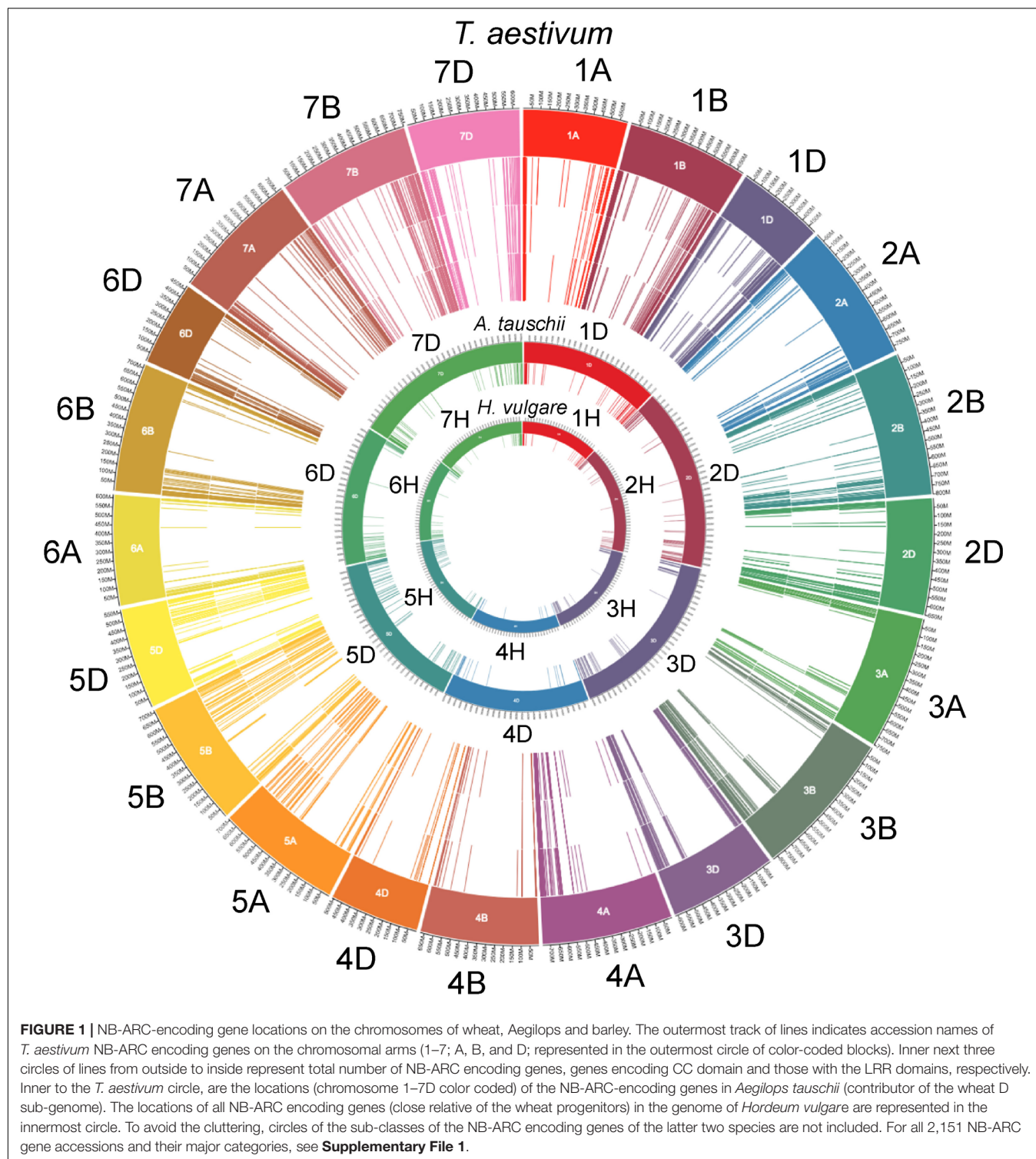
RESULTS

Wheat NB-ARC-Encoding Proteins

The wheat genome contains many genes that encode for NB-ARCs. Approximately half of wheat's 2,151 NB-ARC-encoding proteins also contained a CC domain, and approximately 75% encoded LRRs (Figure 1). Accessions and sequence data for Figure 1 are given in Supplementary File 1. Of the 2,151 NB-ARC-encoding genes, 1,505 had NB-ARCs with P-loop, kinase-2, and GLPL motifs. Among the 1,552 proteins with the LRRs, 802 contained coiled-coil (CC) domains (CNL) and three had resistance to powdery mildew 8 (RPW8) domains (RNL). Interestingly, five of the NB-ARC-encoding genes encoded a toll/interleukin-1 receptor (TIR) with no LRR (TN proteins). NB-ARC-encoding genes formed 547 gene clusters; highly similar clusters are visualized in Figure 2. Many of the clustered genes possessed greater than 75% similarity within each cluster. While chromosomes from each of the wheat sub-genomes are very similar (e.g., chromosomes 1A, 1B, and 1D), differences emerged between the gene clusters found on each chromosome. For example, in wheat's fourth homologous chromosome group (4A, 4B, and 4D), chromosome 4A contained 57 clusters (involving 119 genes), 4B contained four clusters (involving eight genes), and 4D contained three clusters (involving seven genes). It is unknown whether this diversification in 4A happened prior to the first or second hybridization event in wheat. Since *Triticum urartu* gene locations were not available, a clear understanding of the difference between

TABLE 1 | Primers used for *OsRPR1* (OsJ_34782) cloning.

Name	Sequence
OSRPR1-FRG1 FW	TTGGTCTCCAATGTTCAACCTCCCCAGGAG
OSRPR1-FRG1 RV	TTGGTCTCACTATTCTGCGATCTGGCGGCATGCTC
OSRPR1-FRG2 FW	TTGGTCTCAATAGACCTAATATTCACCACCGTAAG
OSRPR1-FRG2 RV	TTGGTCTCAGTGCTGCTAGAAATTTGCATGATG
OSRPR1-FRG3 FW	TTGGTCTCAGCACTAACAAATAATTAGAGCTTCC
OSRPR1-FRG3 RV	TTGGTCTCAAGTGAAAGTTTTCTATTCTGACAC
OSRPR1-FRG4 FW	TTGGTCTCACACTAGATGTGCACATACATCTCAAG
OSRPR1-FRG4 RV	TTGGTCTCACTCCTCATGGAGCTCCCCACTAC
OSRPR1-FRG5 FW	TGGTCTCAGGAGTGGTGAGTGCCAGGT
OSRPR1-FRG5 RV	CCGCGGTCTCACTTGAAATCAATTTTATAACAC
OSRPR1-FRG6 FW	TTGGTCTCACAAGTCTACCATGCACGCAC
OSRPR1-FRG6 RV	TGGTCTCTCGAACCACAGTAACCCAATCACAGCCC
OSRPR1-SV-FW	GGAGGATGACGGCCTCAG
OSRPR1-SV-RV	CCAGTAACCCAATCACAGCC



T. urartu chromosome 4 and wheat chromosome 4A could not be established.

Integrated Domains in Wheat

Several wheat NB-ARC-encoding genes also encode for IDs that may function as molecular baits, decoys, or signal transduction

factors. Wheat NLRs possess a diverse set of IDs, the most common of which are kinase and DNA-binding domains. **Figure 3** shows the average location of 28 different types of IDs relative to protein length, with averages calculated from every NLR-ID occurrence of that domain. Kinase domains are generally located in the *N*-terminal half of the protein, and tyrosine kinase

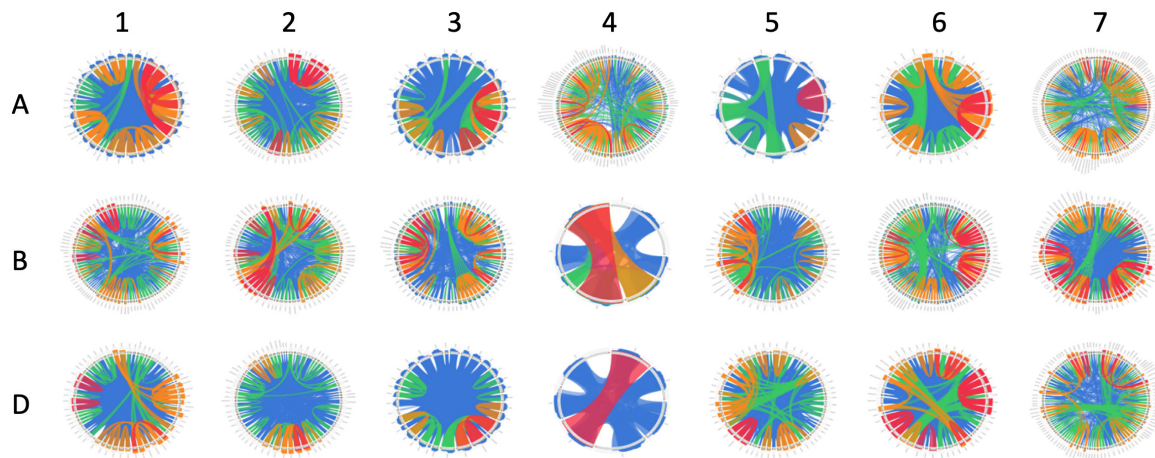


FIGURE 2 | Similarity of clustered NB-ARC-encoding genes on each chromosome. Chromosomes are shown 1–7 with sub-genomes (A–D) from superior to inferior. Clustered genes are arranged in a circle, with lines describing <50% similarity, <75% similarity, <99.999% similarity, and 100% similarity in blue, green, orange, and red, respectively.

domains are generally in the middle of the protein sequence. DNA-binding domains, which are more diverse (e.g., AP2, BED zinc finger, Myb-like, and WKRY), vary by domain type and may be present in either the N- or C-terminus. For example, Myb-like and BED zinc finger domains are generally located at the N-terminus, while B3 and WRKY domains are located at the C-terminus. Many other IDs associated with the C-terminus, including calmodulin-binding, jacalin-like lectin, thioredoxin, and ubiquitin-conjugating domains, may have roles in signaling.

Integrated Domain Homology

Many wheat IDs share homology with proteins in distantly related monocots. **Figure 4** shows the wheat accessions with high percent identity (above 70%) grouped by ID type and homolog species. The vast majority of these ID homologs in other species do not contain NB-ARC domains, as shown by the minority of homology scores surrounded by thick black lines in **Figure 4**. This lack of NB-ARCs in homologs indicates recent fusions that took place after the species diverged or loss in related species, where the few NLR-ID homologs, present mostly in *Brachypodium distachyon*, indicate an ancient fusion that took place before the species diverged. ID homologs in more distant relatives were not NLR proteins, such as the homologs in *Arabidopsis thaliana* and *Amborella trichopoda*. While other plants also possess NLR-ID fusions, many are lineage specific and are not conserved across diverse species. Barley, a close relative of wheat, possessed many of the same NLR-ID fusion proteins as wheat, with 68.5% of ID homologs in barley also possessing NLRs. The two progenitors of wheat with sequenced genomes, *T. urartu* and *A. tauschii*, also possess wheat's NLR-ID fusions. Of these progenitor homologs, 40.8% matched the expected subgenome according to the known progenitor-subgenome relationships. Genomes investigated for homology include: *Aegilops tauschii* (AT), *Amborella trichopoda* (AmT), *Arabidopsis thaliana* (ArT), *Brachypodium distachyon* (BD), *Hordeum vulgare* (HV), *Musa*

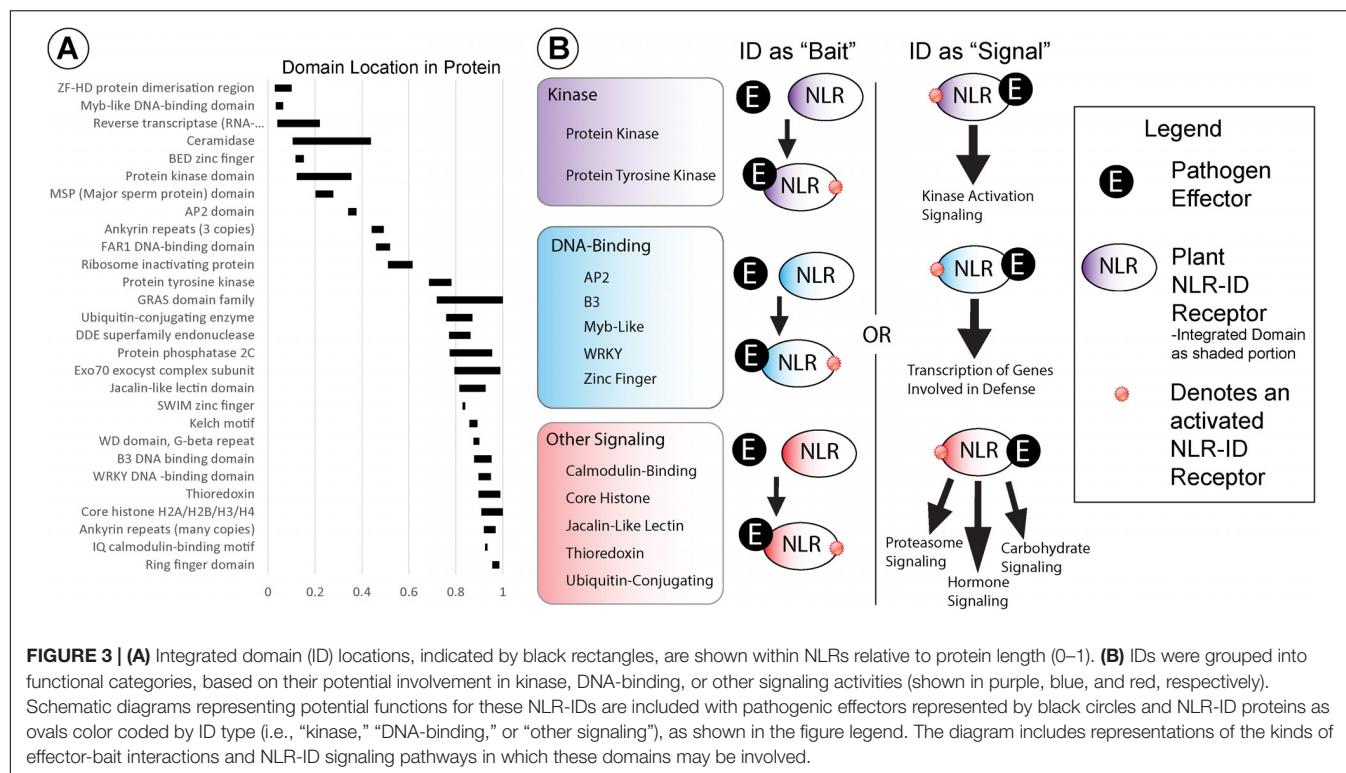
acuminata (MA), *Oryza sativa* (OS), *Setaria italica* (SI), *Triticum urartu* (TU), and *Zea mays* (ZM). Genomic data was not available for *Aegilops speltoides* (AS).

NLR-ID Regulation

Some wheat NLR-ID genes encode alternative transcripts that omit IDs or domains associated with traditionally understood NLR function. **Figure 5** illustrates a consolidation of all wheat NLR-IDs in which at least one alternative transcript of the gene excluded the ID. Alternative splicing of this kind would allow plants to regulate the use of IDs by including or excluding exons containing them. Similar characteristics were also observed in barley transcripts, indicating a conserved use of alternative splicing. Alternative transcripts may also be found in wheat progenitors, which currently lack available data. Expression data from the Wheat Gene Expression Atlas and NCBI shows differential expression between these alternative transcripts, which are also shown in **Figure 5**. This expression data verifies that these alternative transcripts are actually expressed, and that many of them are expressed at different rates depending upon experimental conditions.

Intron Retention in the Coding Sequence of OsRPR1

Alternative splicing of NLR-IDs was investigated in another monocot, *O. sativa*. We identified *in silico* an NLR-ID gene (**Figure 6**), which we termed *Resistance Paired Receptor 1* (*OsRPR1*) (Locus EEE52547.1; Os11g45750) that contains two WRKY domains at its C-terminus. The gene was cloned into a plant expression vector and overexpressed in *N. benthamiana* leaves. The coding sequence of the gene, which was amplified from cDNA generated from total RNA from the agroinfiltrated area, was cloned and sequenced. We found that exon 5 was retained in the coding sequence of *OsRPR1*. In order to investigate whether other splice variants rise from the same



C-terminal region of the gene, we performed a PCR with primers spanning this region (**Figure 6A**). Three out of the four amplified bands were sequenced (SV1, SV2, and SV4; see whole sequences in **Supplementary File 2**). As expected, the most intense band corresponded to the splice variant of *OsRPR1* which retains the 5th exon (SV1). The three other bands corresponded to splice variants containing the 4th and 5th introns (SV2), the 4th intron (SV3), or no introns (SV4) (**Figures 6B,C**). These splice variants have either longer 3' UTRs (SV1, SV2, and SV3 vs SV4) or probably code for different isoforms (SV1 and SV4 vs SV2 and SV3).

DISCUSSION

Wheat NB-ARC-Encoding Genes

Nucleotide-binding leucine-rich repeat receptor systems use groups of sensor and helper NLRs to detect and initiate defense responses when pathogenic effectors are present (Jones et al., 2016). Not all functional NLRs have all characteristic domains (TIR/CC and LRR), such as Pb1 (Hayashi et al., 2010), and some have extra domains (Bailey et al., 2018). Therefore, proteins lacking CC or LRR domains may also contribute to resistance responses, especially those with additional domains that are involved in signaling (Baggs et al., 2017). The distribution of NB-ARC-encoding genes across wheat chromosomes concurs with previous studies in barley and foxtail millet, where R-genes were also found in clusters in extra-pericentromeric regions of chromosomes (Andersen et al., 2016; Andersen and Nepal, 2017). Unequal crossing over between chromosomes as

a mechanism for duplication likely explains the formation of these clusters. Previous studies have highlighted this explanation for the locations of the quickly evolving genes (Marone et al., 2013). *A. tauschii* and *H. vulgare* share a similar pattern as wheat (**Figure 1**), with a similar number of R-genes located at the ends of chromosomes. Barley and the progenitors of wheat diverged approximately 8–9 million years ago (Middleton et al., 2014). Both barley and wheat have experienced artificial selection, as both have been grown for food production since the agricultural revolution approximately 10,000 years ago. Wheat differs from barley in that it is an allohexaploid resulting from hybridization of three species, each containing seven pairs of chromosomes, while barley remained diploid with only seven pairs of chromosomes. Wheat and *A. tauschii* differ in that *A. tauschii* did not experience selective breeding like wheat did. The wheat genome, consisting of A, B, and D subgenomes, maps to the barley genome (H), with wheat chromosomes 1A, 1B, and 1D containing much synteny to 1H of barley. Instances exist where barley possesses duplicated genes that remained unduplicated in wheat, and vice versa. The similarity between wheat and *A. tauschii* is much closer, since *A. tauschii* contributed wheat's D subgenome only a few thousand years ago. While the A subgenome progenitor, *Triticum urartu*, has limited genomic availability, future studies may be able to assess differences between NLR gene architecture in the two genomes. *A. tauschii* and barley provide excellent comparisons with wheat due to the relatively short period of time since their divergence. The similarities between R-genes in wheat relatives show that the highly diverse family of R-genes is necessary for survival, whereas the differences in number and phylogeny point to differences in selection pressure that these species each face.

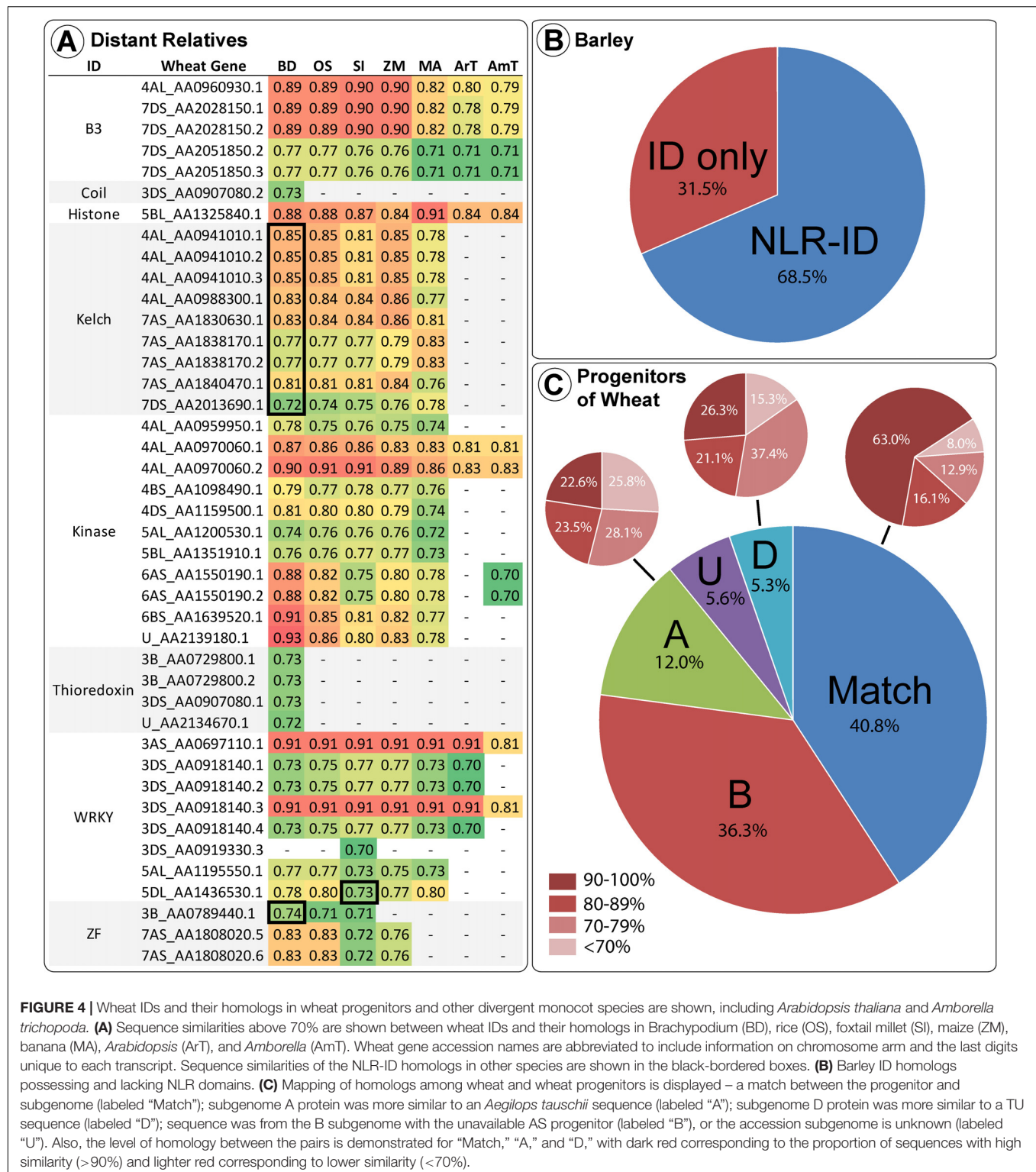


FIGURE 4 | Wheat IDs and their homologs in wheat progenitors and other divergent monocot species are shown, including *Arabidopsis thaliana* and *Amborella trichopoda*. **(A)** Sequence similarities above 70% are shown between wheat IDs and their homologs in Brachypodium (BD), rice (OS), foxtail millet (SI), maize (ZM), banana (MA), *Arabidopsis* (ArT), and *Amborella* (AmT). Wheat gene accession names are abbreviated to include information on chromosome arm and the last digits unique to each transcript. Sequence similarities of the NLR-ID homologs in other species are shown in the black-bordered boxes. **(B)** Barley ID homologs possessing and lacking NLR domains. **(C)** Mapping of homologs among wheat and wheat progenitors is displayed – a match between the progenitor and subgenome (labeled “Match”); subgenome A protein was more similar to an *Aegilops tauschii* sequence (labeled “A”); subgenome D protein was more similar to a TU sequence (labeled “D”); sequence was from the B subgenome with the unavailable AS progenitor (labeled “B”), or the accession subgenome is unknown (labeled “U”). Also, the level of homology between the pairs is demonstrated for “Match,” “A,” and “D,” with dark red corresponding to the proportion of sequences with high similarity (>90%) and lighter red corresponding to lower similarity (<70%).

Clustered genes tend to contain highly similar sequences, indicating that they result from tandem duplication. Some genes, for example, were separated by only a few hundred nucleotides and shared >90% similarity (Figure 2). Through

tandem duplication, wheat NLR genes may have diversified to respond to rapidly evolving and perhaps closely related pathogens, such as those with complex pathotype or race structures, like *Pyrenophora tritici-repentis* (Abdullah, 2017).



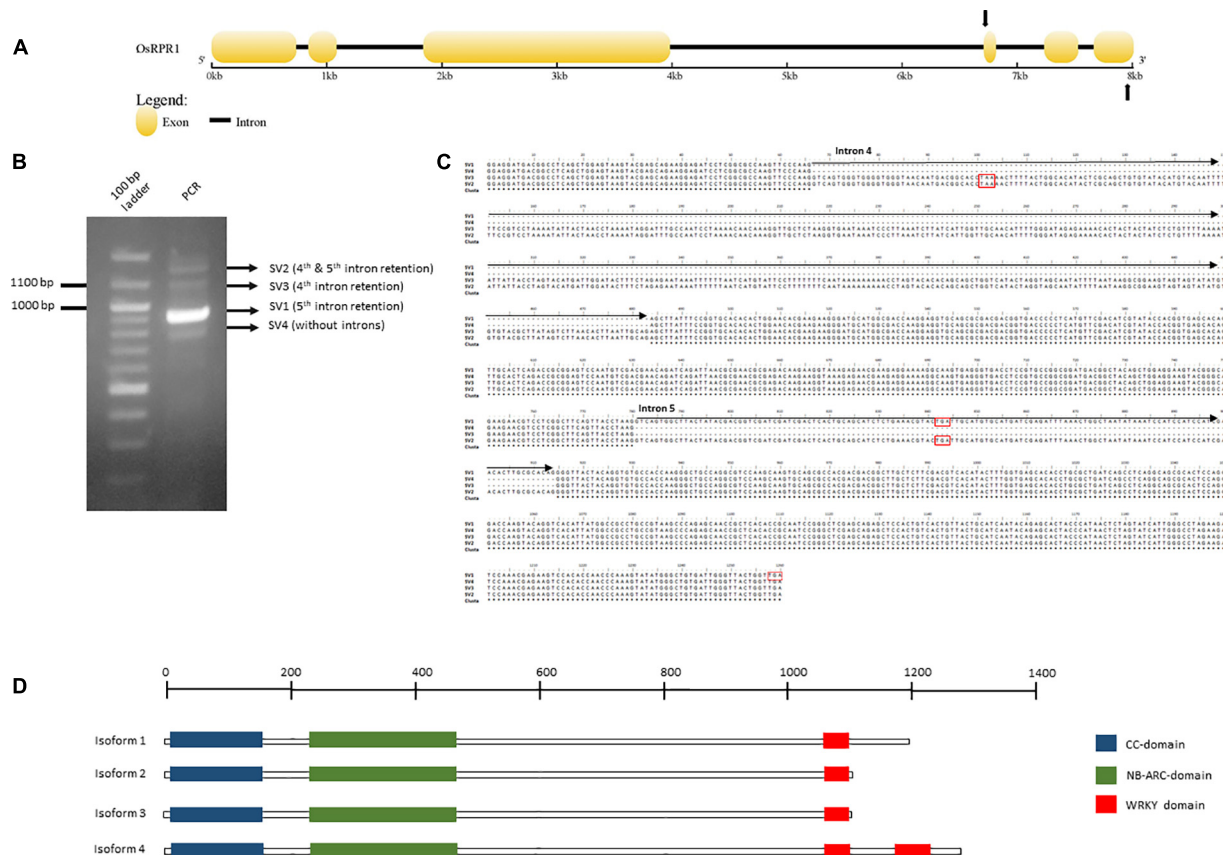


FIGURE 6 | Organization of the *OsRPR1* gene (A) and the PCR amplified splice variants produced after PCR amplification with primers spanning the fourth and sixth exon (B, C). The putative isoforms rising from this splice variants and their domains are depicted in (D). The complete DNA sequences of the putative splice variants and the corresponding isoforms are provided in **Supplementary File 2**.

Race-specific (vertical) resistance in wheat has been identified with regard to pathogens such as powdery mildew (Bourras et al., 2015). Horizontal resistance may include other signaling factors and types of receptors and may rely only partially on NLRs (Kushalappa et al., 2016). While many clustered genes were similar, several cases were identified in which genes were located close to one another but were dissimilar and did not nest together in a phylogenetic analysis. This phenomenon has two major explanations: (1) the tandem duplication took place long ago in evolutionary history and these genes had time to substantially diversify, or (2) a segmental duplication took place, causing the gene to become located next to another R-gene or R-gene cluster. R-genes are highly diversified in plants, with many species possessing hundreds of them. Ancient tandem duplications would have had time to diversify, especially if selective pressures acted upon the ancestors of modern species. However, segmental duplications cannot be discounted due to the presence of genes in some clusters that are highly similar to genes in other clusters (Leister, 2004). In these cases, transposable elements may play some role in the movement of these genes among different chromosomes or to distant locations on the same chromosome (Kim et al., 2017).

Integrated Domains May Augment NLR Function Through Signaling and Recognition

Kinase and DNA-binding IDs likely function as signaling domains that help NLRs initiate defense responses. Current models of NLR function describe a conformational shift triggered when pathogenic effectors bind to the C-terminal LRR, causing the NB-ARC to exchange ADP for ATP and opening the protein up for the N-terminus to initiate further signaling (Takken and Goverse, 2012; Michelmore et al., 2013; Cui et al., 2015). LRRs, as highly variable domains of repeating Lxx amino acid residues, allow defense receptors to bind to diverse elicitors. The NB-ARC, as a P-loop-containing nucleoside triphosphate hydrolase, functions in hydrolysis of beta-gamma phosphate bonds in ATP, binding to phosphates using the Walker A (P-loop) motif and to magnesium ions necessary for catalysis by Walker B motifs (Walker et al., 1982). This release of energy from ATP hydrolysis drives protein conformational change, allowing N-terminal domains (i.e., TIR or CC) to trigger downstream signaling. Kinase IDs found in wheat NLRs could initiate signaling through phosphorylation of transcription factors or other kinases (i.e., MAPK). Sarris et al. (2016) also found

an abundance of NLR-kinase fusions, which possibly retain their biochemical activity (Sarris et al., 2016). DNA-binding domains could move directly to the nucleus upon activation, binding to promoters of pathogenesis-related (PR) genes to recruit transcription machinery. IDs that likely bind to DNA include: AP2, B3, zinc finger, Myb, and WRKY domains, which have been shown to play roles in pathogen resistance (Gutterson and Reuber, 2004; Buscaill and Rivas, 2014). The *Arabidopsis* NLR gene AT4G12020 has been identified both as MAPKKK11 (Jonak et al., 2002) and a TNL resistance gene (Meyers et al., 2003), containing WRKY DNA-binding sites and a protein kinase domain. This gene is a homolog of SLH1, which has been associated with hypersensitive response and may function as a guard for a pathogen effector target (Noutoshi et al., 2005). Many NLR-ID fusion proteins contain transmembrane (TM) domains or nuclear localization signals (NLSs). Several proteins have multiple transmembrane domains, with proteins like 3B_AA0787000 containing seven that are characteristic of other transmembrane proteins. NLSs indicate that DNA-binding domains may functionally interact with DNA as transcription factors.

In addition to signaling, some IDs may play direct roles in effector recognition as effector-binding domains or bait domains that mimic effector targets (Figure 3). Jacalin-like lectin domains, for example, bind to carbohydrates and can recognize carbohydrates that originate directly from pathogens or from damage incurred during infection (Xiang et al., 2011; Lannoo and Van Damme, 2014; Esch and Schaffrath, 2017). Mannose-binding lectin domains were also found in NLRs, associated with disease resistance (Hwang and Hwang, 2011), along with “wall-associated receptor kinase galacturonan-binding” and “cleavage site for pathogenic type III effector avirulence factor Avr” domains. Lectin domains may distinguish proteins as helper NLRs, with carbohydrates acting as signals to initiate NLR activation. Other domains may play roles in effector recognition as bait domains that resemble effector targets. The resistance protein RRS1 becomes activated when an integrated WRKY domain interacts with *Ralstonia solanacearum* effector PopP2 and *Pseudomonas syringae* pv. *pusi* effector AvrRps4, effectors that otherwise target WRKY transcription factors (Le Roux et al., 2015; Sarris et al., 2015). Wheat NLR-WRKY fusions share homology with WRKY16, WRKY19, WRKY46, and WRKY54/70, with potential roles as targets, especially WRKY46, which is associated with bacterial resistance (Sarris et al., 2016). Variants of WRKY domains (WRKY and WSKY) were found and may provide diverse baits for effectors. Genes also encode for multiple transcription factor IDs, which may allow proteins to bind to separate promoters or act as bait for multiple effectors. Some bait proteins, such as PBS1, are kinases that pathogen effectors target for degradation, and may thus increase the utility of NLR-kinase fusions. The Rosetta stone theory posits that such associations between fused domains may indicate functional interactions (Date, 2008). Several proteins with IDs and NB-ARCs do not contain LRRs, which would not be required for activation since baits have replaced LRRs in function.

The activity of IDs as baits is further supported by ID diversity, which corresponds to the diversity of defense

regulatory components. IDs found in NLRs are also found in proteins that effectors target to interfere with defense. Several domains correspond to proteins involved in resistance signaling: calmodulin-binding (calcium signaling), Gibberellic acid insensitive (GAI) repressor of GAI and scarecrow (GRAS; gibberellin signaling), and ethylene responsive element binding (ethylene signaling). Several different domains contain IDs associated with the proteasome or ubiquitin, and these include protease subunit, proteasome component signature, cullin-repeat, RING/U-box, ubiquitin conjugating enzyme, and WD domains. Some IDs contain domains associated with regulation of DNA expression: core histone and chromatin organization modifier. Other IDs correspond to proteins involved in resistance responses: ribosome inactivating and ricin domains (disrupt ribosome activity), thioredoxin and kelch (oxidase activity in reactive oxygen species production), alpha subunit of tryptophan synthesis (synthesis of antiherbivory and antimicrobial compounds), Exo70 exocyst complex subunit (transport of antimicrobial compounds out of the cell), and DDE endonuclease (apoptosis). A few other domains are likely associated with pathogen components: major sperm protein (nematode sperm function, targeted by plant RNA interference), FNIP (found in *Dictyostelium discoideum*), and reverse transcriptase (inhibition of viral infection). Additional viral IDs include RNA-binding/recognition, retrovirus zinc finger-like domain, and integrase domains. IDs may also be associated with pathogen-derived resistance and RNAi that plants use to inhibit viruses and other pathogens. Other studies have found a similar degree of ID diversity in other plant species (Sarris et al., 2016; Baggs et al., 2017).

Our results concur with those presented in Bailey et al. (2018), where a similar set of NLR-IDs were found in the wheat genome, including kinases and transcription factors, of which AP2 was a focus of their study. Bailey et al. (2018) found that certain phylogenetic clades of NLRs contained disproportionately high contents of IDs, which were called major integration clades (MICs), which 30% of NLR-IDs belonged. Many of these were complete domains, consistent with our results. Bailey et al. (2018) proposed retrotransposition, transposition, and ectopic recombination as potential mechanisms for NLR-ID formation. A significant amount (approximately 10%) of NLRs contain exogenous domains, consistent with these results (Kroj et al., 2016; Sarris et al., 2016; Bailey et al., 2018). Our results show that IDs are diverse, but kinases and transcription factor IDs are very common and may due to the presence of MICs where several closely related genes share similar IDs.

Integrated Domains Evolve as Functional Domains Shared by Close Relatives

As a monocot, wheat shares distant relationships with other members of the family Poaceae (i.e., BD, ZM, SI, and OS). ID homologs in distant relatives generally do not contain NB-ARCs, indicating relatively recent origin of NLR-ID fusions. IDs with high percent similarity to homologs (Figure 4) may indicate functional retention. These include proteasome subunit, B3 DNA-binding, WRKY DNA-binding, core histone, protein

kinase, and kelch motif. Other domains with moderate similarity include: jacalin-like lectins, ribosome-inactivating protein, BED zinc finger, SWIM zinc finger, ZF-HD protein dimerization region, zinc knuckle, protein phosphatase 2C, tyrosine kinase, thioredoxin, major sperm protein, reverse transcriptase, and DDE endonuclease. Homologs may be obscured, since mutations accumulate in regions not essential for function or effector-bait interaction. This would cause divergence from the original sequences and would make homology difficult to assess. Some mutations may increase the functionality of NLR-IDs, since the original ID sequence was functionally optimized within a different protein. Some IDs that possess similar modification/cleavage sites may serve as baits for multiple targets (e.g., similar WRKY domains). Many IDs showed high homology in distant relatives. Kinase domains of up to 300 amino acids in length were over 80% similar to homologs. DNA-binding domains also had high homology in distant relatives. WRKY DNA binding domains present in wheat and progenitors have 90.5% similarity to several non-NLR genes in AT, BD, MA, OS, SI, and ZM. Many other IDs in wheat and its progenitors share >80% similarity with homologs in SI, ZM, BD, OS, AT, MA, and AmT. These results concur with previous investigations into IDs, where conserved IDs were identified in diverse plant species (Kroj et al., 2016; Sarris et al., 2016). Some wheat proteins are very similar to their homologs in TU and AT, whereas others provide examples of proteins in one species diversifying from the other two. The histone ID in wheat protein 5BL_AA1325840 (approximately 100 amino acids) shares strong homology (>80%) with proteins in MA, BD, OS, SI, ArT, AmT, and ZM, a recent fusion not present in wheat relatives. Greater than 90% similarity was observed between the 182 amino acid long F775_12304| EMT01588 proteasome subunit domain and proteins in BD, OS, SI, and ZM. While this indicates that these accessions are close homologs, none of the other accessions have NB-ARCs and instead include only peptidase, proteasome subunit, and nucleophile aminohydrolase domains. HV, TU, and TA homologs to this domain, while matching the sequence 100%, do not have NB-ARCs, indicating a very recent duplication and then fusion that occurred after the hybridization of hexaploid wheat. Kelch motif IDs were found in one TU, three AT, and six TA proteins. Interestingly, only one of the TA proteins is in the D sub-genome, and five were located in the A sub-genome, while sub-genome origin would suggest that the opposite pattern would be expected. Bailey et al. (2018) also found that many IDs in wheat's A and D sub-genomes correspond to IDs in TU and TA, with additional IDs that resulted after wheat split from its progenitors.

While domain homology in distantly related species indicates functional origins of IDs, homologs identified in close relatives (i.e., HV) and wheat progenitors (TU and AT) indicate recent fusions and duplications. Unlike distant relatives, barley shares many NLR-ID fusions with wheat. This indicates that many of wheat's NLR-IDs happened before the divergence of barley and wheat progenitors. Since this divergence, wheat and barley have independently gained and lost NLR-ID proteins. Many TU and AT proteins are almost identical to proteins encoded by TA genes within the A and D sub-genomes. IDs within

wheat's B sub-genome often originate from AS and do not have 100% homologs in TU and AT. In select cases, similarity was found between NLR-IDs and functional domains from non-NLR proteins, indicating potential NLR-ID fusions since the formation of wheat. Conversely, some close wheat relatives share homology with distantly related NLR-ID fusions, such as F775_00546| EMT17242 and Si008625m, with an 84.6% similarity (991 identical sites) between their whole sequences, both with NB-ARCs and kinase domains. Unexpectedly, many proteins were found in wheat but not a relative, or vice versa. These domains include histone, ribosome inactivating protein, calcium signaling, cleavage for type III effectors, RNase H type, P450, antibiotic synthesis, and ubiquitin conjugating enzyme. Other domains were found in greater or lower numbers in wheat compared to its progenitors, such as DDE endonuclease and reverse transcriptase, indicating loss or duplication in one genome.

Plants Use Alternative Splicing to Regulate NLR-IDs

Many NLR-ID protein-encoding genes possess multiple transcripts, some of which lack IDs or truncate domains within the protein (Figure 5). This indicates that plants may use alternative splicing to regulate the use of IDs within a network of NLR proteins. Previous studies have shown that resistance to some pathogens requires alternative splicing (Yang et al., 2014; Shang et al., 2017), such as *RPS4* in *Arabidopsis* (Zhang and Gassmann, 2003); and splicing is used to truncate proteins like *RCT1* in *Medicago truncatula* (Tang et al., 2013). Wheat has also shown evidence of alternative splicing of important resistance genes *Lr10* and *Sr35* (Sela et al., 2012; Saintenac et al., 2013). Splicing patterns between wheat paralogs resulting from duplication also appear to be conserved. Stop codon-containing inter-exon regions can be included in the transcript to force a truncation of the protein. The results presented in Figure 6 show that R-genes encode splice variants, and that introns were retained in the coding sequence, indicating further R gene variation is possible through alternative splicing. Truncated NB-ARCs may result in decoy proteins, where signaling function is lost but IDs 'distract' pathogenic effectors from functional target proteins. The potential involvement of alternative splicing is shown in Figure 7. In concurrence with the results, Yang et al. (2014) described NLR alternative splicing as useful for regulating NLR autoinhibition or function in signal transduction and also detailed potential transcription factor activity (Yang et al., 2014). Genes that have multiple copies of an ID can also regulate the number of copies in the protein through alternative splicing. Alternative splicing may also allow the plant to select different localization for a gene product, such as those where transmembrane helices are included in one transcript and are not included in another (Figure 5).

Wheat expression data shows that there are differences in the expression of these alternative transcripts shown in Figure 5. The expression values for the 54 transcripts present in Figure 5 were mined from datasets present in the Wheat Gene Expression Atlas and NCBI databases. Expression data from Salcedo et al. (2017) shows that alternative splicing may result from different

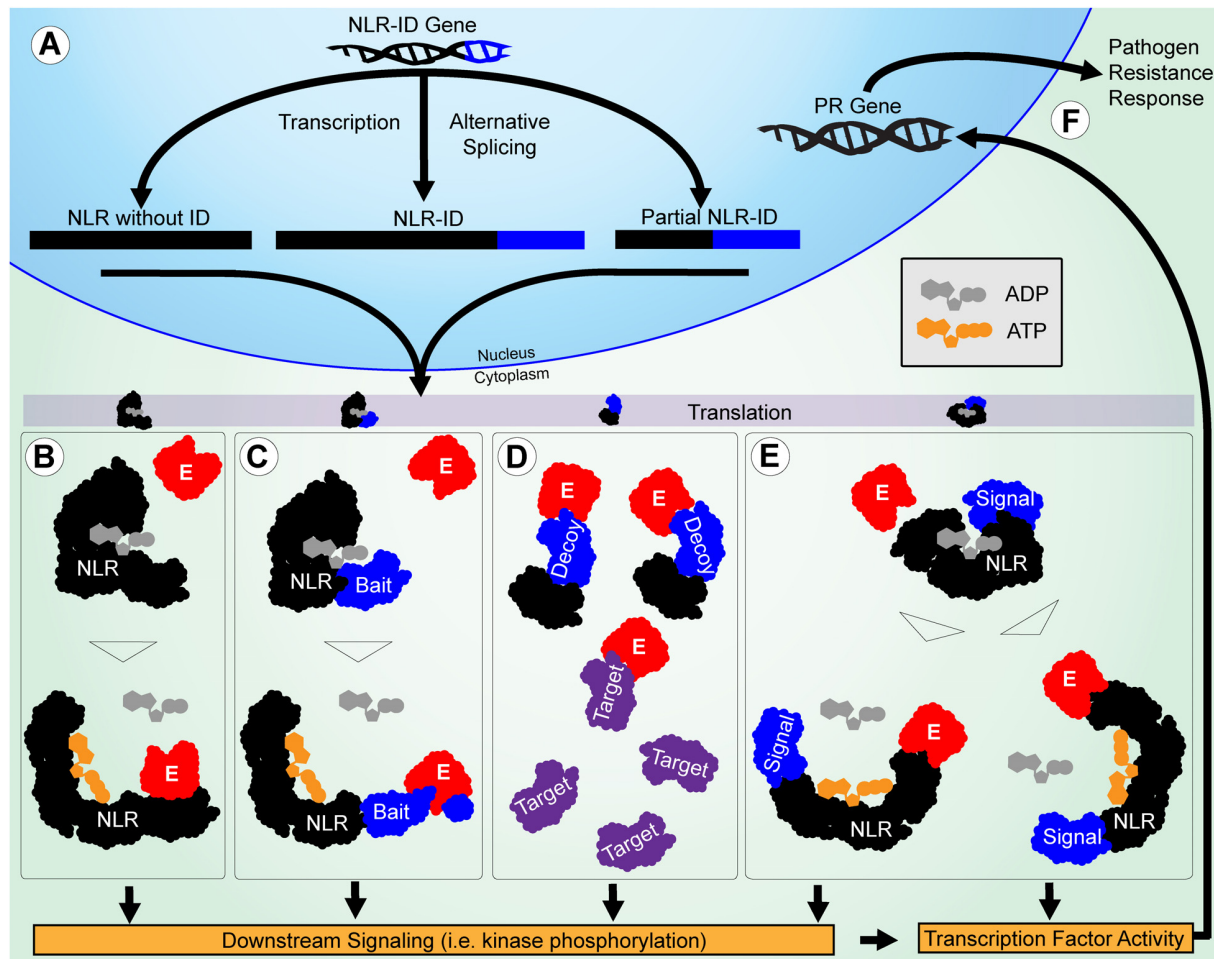


FIGURE 7 | Potential roles of IDs in functional diversification of NLRs in pathogen resistance are illustrated. **(A)** The NLR-ID gene is alternatively spliced during transcription to include or exclude IDs. The NLR and ID sequences are shown in black and blue, respectively. **(B)** The NLRs without IDs function through effector-LRR binding to activate the protein and trigger downstream signaling. Effectors are shown in red. **(C)** When IDs are used as baits, they mimic pathogen targets and cause NLR activation after they are modified. **(D)** When IDs are used as decoys, they mimic pathogen targets to reduce effector interference in resistance signaling. The targets are shown in purple. **(E)** When IDs are used in signaling, they allow NLRs to act as signal transduction factors, making them less reliant on the downstream signaling utilized by other NLRs. **(F)** Finally, transcription factor activity directly involving or triggered by NLRs causes PR genes to be expressed, leading to a resistance response.

conditions (Salcedo et al., 2017). At the very least, these data provide support for **Figure 5** accessions as confirmed alternative transcripts that differ in expression. In the Wheat Gene Expression Atlas data (Borrill et al., 2016), several transcripts with different ID contents show differential expression in wheat tissues. Several genes were more strongly expressed in the leaves, shoots, and spikes, indicating the potential tissue-specific roles that these genes play in resistance. More data is required to conclusively show differential expression based upon certain treatments and conditions. While wheat shows evidence of NLR-ID alternative splicing, barley may have evolved a more diverse set of transcripts for NLR-IDs. Several barley NLR-ID proteins have dozens of transcripts, with several of those allowing for alternative uses of IDs in NLR proteins. Many barley genes have alternative transcripts that encode NLR-ID, just NLR, just ID, or lack both. Previous studies have identified barley *Mla* genes

as utilizing alternative splicing for resistance (Halterman et al., 2003). Barley genes can also encode multiple IDs. Barley data also supports a previous prediction that alternative splicing may allow for differential cellular localization (Yang et al., 2014).

CONCLUSION

In this study, we identified 2,151 NB-ARC-encoding genes in the wheat genome, with many encoding additional domains associated with the receptors that detect pathogenic effectors. In the 21 chromosomes of wheat, 547 gene clusters were found, with many clusters containing highly similar genes. Clustering of R-genes in wheat was compared to its progenitors and barley, a close relative, and gene similarities within clusters showed that

tandem duplication explains much of the diversification among R-genes. The diversity of IDs in NLRs corresponds directly to the multiple components utilized by plant cells to initiate resistance responses. These components included kinases, transcription factors, hormone signaling receptors, and proteins involved in antimicrobial compound production. NLR-ID fusions give these immune receptors the potential to function as effector baits, decoys, and signal transduction factors. Sequence homology indicates that some IDs may retain functionality and evolve into non-NLR proteins. Genomic and gene expression data suggest that plants likely utilize alternative splicing to regulate the inclusion or exclusion of IDs in NLR proteins. We tested it in rice, another Poaceae member, where *Agrobacterium*-mediated transformation of *OsRPR1* gene exhibited alternative splicing. The ability of plants to use splicing to include or exclude IDs constitutes an important defense strategy to deal with rapidly evolving pathogen effectors. Future studies should aim to characterize the structure of NLR-ID fusion proteins, demonstrate which IDs have retained enzymatic activity, and associate the expression of alternative transcripts with specific conditions. Additional genomic data on *Aegilops speltoides*, a relative to the contributor of wheat's B sub-genome, as well as availability of data for *Triticum urartu*, contributor of wheat's A sub-genome, will allow for more thorough analyses of the evolution of disease resistance genes in wheat.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the https://figshare.com/articles/Table_S1_Expression_of_wheat_NLR-

[ID_alternative_splicing_candidate_genes_available_in_Wheat_Gene_Expression_Atlas_and_NCBI_databases_/8796449](https://figshare.com/articles/Table_S1_Expression_of_wheat_NLR-).

AUTHOR CONTRIBUTIONS

EA and GM carried out the experiment and analyzed the sequence data. MN and PS conceived the research project and supervised the experiment. JP and DN contributed drafting the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

Support for this project came from the USDA-NIFA hatch Projects to MN (SD00H469-13 and SD00H659-18), South Dakota Agriculture Experiment Station Wokini Fund, and Department of Biology and Microbiology at South Dakota State University. This manuscript has been released as a Pre-Print at biorxiv (Andersen and Nepal, 2019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00898/full#supplementary-material>

FILES 1 | Wheat accession, sequence, and location data for the genes shown in **Figure 1**. Table can be accessed at the Figshare link: <https://figshare.com/s/78d566a9c4cf4902d283>.

FILES 2 | Genomic sequences of the splice variants and the corresponding isoforms shown in **Figure 6**.

REFERENCES

- Abdullah, S. (2017). *Characterization of Pyrenophora Tritici-Repentis in Wheat and Rye to Study Tan Spot Susceptibility and Insights Into Its Relationship with Stem Rust Resistance*. Ph. D. Thesis, South Dakota State University, Brookings.
- Andersen, E. J., Ali, S., Reese, R. N., Yen, Y., Neupane, S., and Nepal, M. P. (2016). Diversity and evolution of disease resistance genes in barley (*Hordeum vulgare* L.). *Evol. Bioinform. Online* 12:99.
- Andersen, E. J., and Nepal, M. P. (2017). Genetic diversity of disease resistance genes in foxtail millet (*Setaria italica* L.). *Plant Gene* 10, 8–16. doi: 10.1016/j.plgene.2017.03.002
- Andersen, E. J., and Nepal, M. P. (2019). Diversification of disease resistance receptors by integrated domain fusions in wheat and its progenitors. *bioRxiv*[Preprint] doi: 10.1101/695148
- Baggs, E., Dagdas, G., and Krasileva, K. (2017). NLR diversity, helpers and integrated domains: making sense of the NLR IDentity. *Curr. Opin. Plant Biol.* 38, 59–67. doi: 10.1016/j.pbi.2017.04.012
- Bailey, P. C., Schudoma, C., Jackson, W., Baggs, E., Dagdas, G., Haerty, W., et al. (2018). Dominant integration locus drives continuous diversification of plant immune receptors with exogenous domain fusions. *Genome Biol.* 19:23.
- 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.
- Bonardi, V., Cherkis, K., Nishimura, M. T., and Dangl, J. L. (2012). A new eye on NLR proteins: focused on clarity or diffused by complexity? *Curr. Opin. Immunol.* 24, 41–50. doi: 10.1016/j.coi.2011.12.006
- Bonardi, V., and Dangl, J. L. (2012). How complex are intracellular immune receptor signaling complexes? *Front. Plant Sci.* 3:237. doi: 10.3389/fpls.2012.00237
- Borlaug, N. E. (1983). Contributions of conventional plant breeding to food production. *Science* 219, 689–693. doi: 10.1126/science.219.4585.689
- Borrill, P., Ramirez-Gonzalez, R., and Uauy, C. (2016). expVIP: a customisable RNA-seq data analysis and visualisation platform opens up gene expression analysis. *Plant Physiol.* 170, 2172–2186. doi: 10.1104/pp.15.01667
- Bourras, S., McNally, K. E., Ben-David, R., Parlange, F., Roffler, S., Praz, C. R., et al. (2015). Multiple avirulence loci and allele-specific effector recognition control the Pm3 race-specific resistance of wheat to powdery mildew. *Plant Cell* 27, 2991–3012.
- Buscaill, P., and Rivas, S. (2014). Transcriptional control of plant defence responses. *Curr. Opin. Plant Biol.* 20, 35–46. doi: 10.1016/j.pbi.2014.04.004
- Cesari, S., Bernoux, M., Moncuquet, P., Kroj, T., and Dodds, P. N. (2014). A novel conserved mechanism for plant NLR protein pairs: the “integrated decoy” hypothesis. *Front. Plant Sci.* 5:606. doi: 10.3389/fpls.2014.00606
- Cui, H., Tsuda, K., and Parker, J. E. (2015). Effector-triggered immunity: from pathogen perception to robust defense. *Annu. Rev. Plant Biol.* 66, 487–511. doi: 10.1146/annurev-arplant-050213-040012
- Darzentas, N. (2010). Circoletto: visualizing sequence similarity with Circos. *Bioinformatics* 26, 2620–2621. doi: 10.1093/bioinformatics/btq484
- Date, S. V. (2008). “The Rosetta stone method,” in *Bioinformatics*, ed. S. V. Date, (Berlin: Springer), 169–180. doi: 10.1007/978-1-60327-429-6_7
- Deyoung, B. J., and Innes, R. W. (2006). Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat. Immunol.* 7:1243. doi: 10.1038/ni1410

- Duxbury, Z., Ma, Y., Furzer, O. J., Huh, S. U., Cevik, V., Jones, J. D. G., et al. (2016). Pathogen perception by NLRs in plants and animals: parallel worlds. *BioEssays* 38, 769–781. doi: 10.1002/bies.201600046
- Esch, L., and Schaffrath, U. (2017). An update on jacalin-like lectins and their role in plant defense. *Int. J. Mol. Sci.* 18:1592. doi: 10.3390/ijms18071592
- FAOstat, (2013). *FAO Database* [Online]. Rome: Food and Agriculture Organization of the United Nations.
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2015). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285.
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2011). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186.
- Guttersen, N., and Reuber, T. L. (2004). Regulation of disease resistance pathways by AP2/ERF transcription factors. *Curr. Opin. Plant Biol.* 7, 465–471. doi: 10.1016/j.pbi.2004.04.007
- Halterman, D. A., Wei, F., and Wise, R. P. (2003). Powdery mildew-induced Mla mRNAs are alternatively spliced and contain multiple upstream open reading frames. *Plant Physiol.* 131, 558–567. doi: 10.1104/pp.014407
- Hayashi, N., Inoue, H., Kato, T., Funao, T., Shiota, M., Shimizu, T., et al. (2010). Durable panicle blast-resistance gene Pb1 encodes an atypical CC–NBS–LRR protein and was generated by acquiring a promoter through local genome duplication. *Plant J.* 64, 498–510. doi: 10.1111/j.1365-313x.2010.04348.x
- Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., and Gao, G. (2014). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817
- Huerta-Espino, J., Singh, R., Germán, S., McCallum, B., Park, R., Chen, W., et al. (2011). Global status of wheat leaf rust caused by *Puccinia triticina*. *Euphytica* 179, 143–160. doi: 10.1007/s10681-011-0361-x
- Hwang, I. S., and Hwang, B. K. (2011). The pepper mannose-binding lectin gene CaMBL1 is required to regulate cell death and defense responses to microbial pathogens. *Plant Physiol.* 155, 447–463. doi: 10.1104/pp.110.164848
- International Wheat Genome Sequencing Consortium, (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345:1251788.
- Jia, J., Zhao, S., Kong, X., Li, Y., Zhao, G., He, W., et al. (2013). Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 496, 91–95.
- Jonak, C., Ökrész, L., Bögre, L., and Hirt, H. (2002). Complexity, cross talk and integration of plant MAP kinase signalling. *Curr. Opin. Plant Biol.* 5, 415–424. doi: 10.1016/s1369-5266(02)00285-6
- Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329.
- Jones, J. D., Vance, R. E., and Dangl, J. L. (2016). Intracellular innate immune surveillance devices in plants and animals. *Science* 354:aaf6395. doi: 10.1126/science.aaf6395
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Jupe, F., Pritchard, L., Etherington, G. J., Mackenzie, K., Cock, P. J., Wright, F., et al. (2012). Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics* 13:75. doi: 10.1186/1471-2164-13-75
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Kersey, P. J., Allen, J. E., Armean, I., Boddu, S., Bolt, B. J., Carvalho-Silva, D., et al. (2015). Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res.* 44, D574–D580.
- Kim, S., Park, J., Yeom, S.-I., Kim, Y.-M., Seo, E., Kim, K.-T., et al. (2017). New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol.* 18:210.
- Kofoed, E. M., and Vance, R. E. (2011). Innate immune recognition of bacterial ligands by NALPs determines inflammasome specificity. *Nature* 477:592. doi: 10.1038/nature10394
- Kroj, T., Chanclud, E., Michel–Romiti, C., Grand, X., and Morel, J. B. (2016). Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytol.* 210, 618–626. doi: 10.1111/nph.13869
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Kushalappa, A. C., Yogendra, K. N., and Karre, S. (2016). Plant innate immune response: qualitative and quantitative resistance. *Crit. Rev. Plant Sci.* 35, 38–55. doi: 10.1080/07352689.2016.1148980
- Lannoo, N., and Van Damme, E. J. (2014). Lectin domains at the frontiers of plant defense. *Front. Plant Sci.* 5:397. doi: 10.3389/fpls.2014.00397
- Larkin, M. A., Blackshields, G., Brown, N., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Le Roux, C., Huet, G., Jauneau, A., Camborde, L., Trémoussaygue, D., Kraut, A., et al. (2015). A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell* 161, 1074–1088. doi: 10.1016/j.cell.2015.04.025
- Leister, D. (2004). Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends Genet.* 20, 116–122. doi: 10.1016/j.tig.2004.01.007
- Ling, H.-Q., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., et al. (2013). Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* 496, 87–90.
- Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K. S., et al. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345:1250092.
- 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
- McFadden, E. S. (1930). A successful transfer of emmer characters to vulgare wheat. *J. Am. Soc. Agron.* 22, 1020–1034. doi: 10.2134/agronj1930.00021962002200120005x
- Mermigka, G., Amprazi, M., Mentzelopoulou, A., Amartolou, A., and Sarris, P. F. (2020). Plant and animal innate immunity complexes: fighting different enemies with similar weapons. *Trends Plant Sci.* 25, 80–91. doi: 10.1016/j.tplants.2019.09.008
- 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 Online* 15, 809–834. doi: 10.1105/tpc.009308
- Michelmore, R. W., Christopoulou, M., and Caldwell, K. S. (2013). Impacts of resistance gene genetics, function, and evolution on a durable future. *Annu. Rev. Phytopathol.* 51, 291–319. doi: 10.1146/annurev-phyto-082712-102334
- Middleton, C. P., Senerchia, N., Stein, N., Akhunov, E. D., Keller, B., Wicker, T., et al. (2014). Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the *Triticeae* tribe. *PLoS One* 9:e85761. doi: 10.1371/journal.pone.0085761
- Murray, T. D., Bockus, W. W., Bowden, R. L., Hunger, R. M., Smiley, R. W., and Collators, (2015). *Diseases of Wheat (Triticum spp. L.)* [Online]. www.apsnet.org/publications/commonnames/Pages/Wheat.aspx (accessed July 28, 2017)
- Noutoshi, Y., Ito, T., Seki, M., Nakashita, H., Yoshida, S., Marco, Y., et al. (2005). A single amino acid insertion in the WRKY domain of the *Arabidopsis* TIR–NBS–LRR–WRKY–type disease resistance protein SLH1 (sensitive to low humidity 1) causes activation of defense responses and hypersensitive cell death. *Plant J.* 43, 873–888. doi: 10.1111/j.1365-313x.2005.02500.x
- Periyannan, S., Moore, J., Ayliffe, M., Bansal, U., Wang, X., Huang, L., et al. (2013). The gene Sr33, an ortholog of barley Mla genes, encodes resistance to wheat stem rust race Ug99. *Science* 341, 786–788. doi: 10.1126/science.1239028
- Pretorius, Z., Singh, R., Wagoire, W., and Payne, T. (2000). Detection of virulence to wheat stem rust resistance gene Sr31 in *Puccinia graminis* f. sp. tritici in Uganda. *Plant Dis.* 84, 203–203. doi: 10.1094/pdis.2000.84.2.203b
- Saari, E. E., and Prescott, J. (1985). World distribution in relation to economic losses. *Cereal Rusts* 2, 259–298. doi: 10.1016/b978-0-12-148402-6.5017-1
- Saintenac, C., Zhang, W., Salcedo, A., Rouse, M. N., Trick, H. N., Akhunov, E., et al. (2013). Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. *Science* 341, 783–786. doi: 10.1126/science.1239022

- Salcedo, A., Rutter, W., Wang, S., Akhunova, A., Bolus, S., Chao, S., et al. (2017). Variation in the AvrSr35 gene determines Sr35 resistance against wheat stem rust race Ug99. *Science* 358, 1604–1606. doi: 10.1126/science.aao7294
- Sarris, P. F., Cevik, V., Dagdas, G., Jones, J. D., and Krasileva, K. V. (2016). Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biol.* 14:8. doi: 10.1186/s12915-016-0228-7
- Sarris, P. F., Duxbury, Z., Huh, S. U., Ma, Y., Segonzac, C., Sklenar, J., et al. (2015). A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell* 161, 1089–1100. doi: 10.1016/j.cell.2015.04.024
- Sela, H., Spiridon, L. N., Petrescu, A. J., Akerman, M., Mandel–Gutfreund, Y., Nevo, E., et al. (2012). Ancient diversity of splicing motifs and protein surfaces in the wild emmer wheat (*Triticum dicoccoides*) LR10 coiled coil (CC) and leucine-rich repeat (LRR) domains. *Mol. Plant Pathol.* 13, 276–287. doi: 10.1111/j.1364-3703.2011.00744.x
- Shang, X., Cao, Y., and Ma, L. (2017). Alternative splicing in plant genes: a means of regulating the environmental fitness of plants. *Int. J. Mol. Sci.* 18:432. doi: 10.3390/ijms18020432
- 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 genes reveal three anciently diverged classes with distinct evolutionary patterns. *Plant Physiol.* 170, 2095–2109. doi: 10.1104/pp.15.01487
- Singh, R. P., Hodson, D. P., Huerta-Espino, J., Jin, Y., Bhavani, S., Njau, P., et al. (2011). The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. *Annu. Rev. Phytopathol.* 49, 465–481. doi: 10.1146/annurev-phyto-072910-095423
- Takken, F. L., and Goverse, A. (2012). How to build a pathogen detector: structural basis of NB-LRR function. *Curr. Opin. Plant Biol.* 15, 375–384. doi: 10.1016/j.pbi.2012.05.001
- Tang, F., Yang, S., Gao, M., and Zhu, H. (2013). Alternative splicing is required for RCT1-mediated disease resistance in *Medicago truncatula*. *Plant Mol. Biol.* 82, 367–374. doi: 10.1007/s11103-013-0068-6
- Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1, 945–951. doi: 10.1002/j.1460-2075.1982.tb01276.x
- Wu, C.-H., Krasileva, K. V., Banfield, M. J., Terauchi, R., and Kamoun, S. (2015). The “sensor domains” of plant NLR proteins: more than decoys? *Front. Plant Sci.* 6:134. doi: 10.3389/fpls.2015.00134
- Xiang, Y., Song, M., Wei, Z., Tong, J., Zhang, L., Xiao, L., et al. (2011). A jacalin-related lectin-like gene in wheat is a component of the plant defence system. *J. Exp. Bot.* 62, 5471–5483. doi: 10.1093/jxb/err226
- Yang, S., Tang, F., and Zhu, H. (2014). Alternative splicing in plant immunity. *Int. J. Mol. Sci.* 15, 10424–10445. doi: 10.3390/ijms150610424
- Zhang, X.-C., and Gassmann, W. (2003). RPS4-mediated disease resistance requires the combined presence of RPS4 transcripts with full-length and truncated open reading frames. *Plant Cell* 15, 2333–2342. doi: 10.1105/tpc.013474

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 Andersen, Nepal, Purintun, Nelson, Mermigka and Sarris. 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: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



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