INTERSPECIFIC HYBRIDIZATION IN PLANT BIOLOGY

EDITED BY: Andrew H. Paterson, Dayun Tao and Ruslan Kalendar

PUBLISHED IN: Frontiers in Plant Science

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-237-2
DOI 10.3389/978-2-83250-237-2

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: frontiersin.org/about/contact

INTERSPECIFIC HYBRIDIZATION IN PLANT BIOLOGY

Topic Editors:
Andrew H. Paterson, University of Georgia, United States
Dayun Tao, Yunnan Academy of Agricultural Sciences, China
Ruslan Kalendar, University of Helsinki, Finland

Citation: Paterson, A. H., Tao, D., Kalendar, R., eds. (2022). Interspecific Hybridization in Plant Biology. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-237-2

Table of Contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 05 September 2022
DOI 10.3389/fpls.2022.1026492

OPEN ACCESS

EDITED AND REVIEWED BY
Diego Rubiales,
Institute for Sustainable Agriculture (CSIC), Spain

*CORRESPONDENCE
Dayun Tao
taody12@aliyun.com
Ruslan Kalendar
ruslan.kalendar@helsinki.fi
Andrew H. Paterson
paterson@uga.edu

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

RECEIVED 24 August 2022
ACCEPTED 24 August 2022
PUBLISHED 05 September 2022

CITATION
Tao D, Kalendar R and Paterson AH (2022) Editorial: Interspecific hybridization in plant biology. *Front. Plant Sci.* 13:1026492. doi: 10.3389/fpls.2022.1026492

Editorial: Interspecific hybridization in plant biology

Dayun Tao [1]*, Ruslan Kalendar [2,3]* and Andrew H. Paterson [4]*

[1]Yunnan Key Laboratory for Rice Genetic Improvement, Food Crops Research Institute, Yunnan Academy of Agricultural Sciences, Kunming, China, [2]Helsinki Institute of Life Science HiLIFE, Biocenter 3, University of Helsinki, Helsinki, Finland, [3]National Laboratory Astana, Nazarbayev University, Nur-Sultan, Kazakhstan, [4]Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, United States

KEYWORDS

interspecific hybridization, crop improvement, synthetic polyploids, phenotypic variation, genomics, interspecific barriers, heterosis

Editorial on the Research Topic
Interspecific hybridization in plant biology

Many crop gene pools are derived from small numbers of founders. As a consequence of long histories of strong directional selection, crop gene pools have narrow genetic diversity available to provide inherent solutions to changing needs or challenges. Notoriously, plants can mate across taxonomically-determined species boundaries, and interspecific hybridization is widely used in plant genetics research. Interspecific hybridizations have conferred practical improvements to crops, some of which are unexpected based on the phenotypes of the parents.

Genomics has provided insights into the fundamental consequences of interspecific hybridization for plant biology. Additionally, genomics has allowed the development of molecular tools for dissecting the genetic control of phenotypic variation in interspecific hybrid populations and manipulating interspecific introgressions in crop improvement.

This Research Topic aimed to publish peer-reviewed research in plant interspecific hybridization and its consequences, both fundamental and applied. While such work is prominent in plants, consideration will also be given to salient work in other taxa. A key threshold for publication was the extent to which findings are of cross-cutting interest and importance, i.e., not only to those working on the target taxon but also to a wide range of biological scientists. As a result, 15 articles were published.

For the issue of the role of interspecific hybridization, Wong et al. used the genus *Senecio*, and revealed some of the roles interspecific hybridization could play in evolution, including transcriptome shock, genome reorganization, change in mating system and reproductive traits, and adaptive introgression. Other aspects, such as the evolution of novel compounds, gene redundancy, and the extent of adaptive allele sharing, have been explored in other *Senecio* species. For cultivated species, Zhou et al. highlighted the important role that interspecific hybridization-introgression has played in improving the genetic diversity and adaptation of *Oryza sativa*. Natural hybridization-introgression is thought to have led to

the origin of *indica*, *aus*, and *basmatic* subgroups, which adapted to changing cultivated environments, and produced feral weedy rice coexisting and competing with cultivars under production management. Artificial interspecific hybridization-introgression gained several breakthroughs in rice breeding, such as developing three-line hybrid rice, new rice for Africa (NERICA), and deploying some important pest and disease resistance genes in rice genetic improvement, contributing to the stable increase of rice production to meet the expanding human population.

For the issue of interspecific hybridization for breeding, Zhang et al. raised 6,372 agronomic trait introgression lines (ILs) from BC_2 to BC_6 based on the variations in agronomic traits by crossing 170 accessions of 7 AA genome species and 160 upland rice accessions of *O. sativa* as the donor parents, with three elite cultivars of *O. sativa*, Dianjingyou 1 (a *japonica* variety), Yundao 1 (a *japonica* variety), and RD23 (an *indica* variety) as the recurrent parents, respectively. Agronomic traits such as spreading panicle, erect panicle, dense panicle, lax panicle, awn, prostrate growth, plant height, pericarp color, kernel color, glabrous hull, grain size, 1,000-grain weight, drought resistance, and aerobic adaption, and blast resistance, were derived from more than one species. This agronomic trait introgression library from multiple species and accessions provided a powerful resource for future rice improvement and genetic dissection of agronomic traits. Tan et al. reconstructed the high-SER (stigma exsertion rate) trait based upon 18 quantitative trait loci (QTLs) for SER from *O. sativa*, *O. glaberrima*, and *O. glumaepatula* using single-segment substitution lines (SSSLs) in the genetic background of Huajingxian 74 (HJX74). A total of 29 pyramiding lines with 2-6 QTLs were developed from 10 SSSLs carrying QTLs for SER in the HJX74 genetic background. The results showed that the SER increased with increasing QTLs in the pyramiding lines. The SER in the lines with 5–6 QTLs was as high as wild rice with strong outcrossing ability. Limbalkar et al. used *Brassica carinata*-derived lines (CDLs) in *Brassica juncea* (L.) Czern. background, carrying genomic segments from *B. carinata*, to raise 105 hybrids developed from intermating 15 CDLs in half diallel fashion. The results indicated that higher productivity under drought conditions can be realized through the development of hybrids. Ullah et al. used *Trifolium occidentale*, one of the ancestral parents of *T. repens*, as a bridging species to overcome the interspecific barrier between *T. ambiguum* and *T. repens*. Recombinant chromosome segments from *T. ambiguum* were found in all five plants of *T. repens* background. Despite early chromosome imbalances, the backcross populations were fertile and produced large numbers of seeds. These hybrids represent a major new resource for the breeding of novel resilient forms of white clover.

Reproductive barriers, important in the wild to maintain species integrity, are a major obstacle to interspecific hybridization and gene introgression. Interactive RNA sequencing and proteome analysis by Du et al. revealed changes in the transcriptomic and proteomic profiles of *Fragaria viridis* styles harvested at 0 and 24 h after self-pollination. Differentially expressed genes and differentially abundant proteins associated with self-incompatible pollination were further mined, and multiple pathways were found to be involved. Moriyama et al. made a sibling cross of F_1 plants made from the cross between *Lilium* × *formolongi* cv. Hatsuki and cv. Raizan 2go, producing the pollen-sterile line, PL01. Using PL01 and its progeny, genetic analysis suggested that the male-sterile phenotype is attributed to one recessive locus *LflflTDF1*. A transcript expressed only in pollen-fertile plants was homologous to *TDF1* (*DEFECTIVE in TAPETAL DEVELOPMENT and FUNCTION1*) in *Arabidopsis*, a gene encoding a transcription factor AtMYB35 known as a key regulator of pollen development. The *LflflTDF1* allele was transferred to Easter lily to confer sterility. He et al. reported that an imbalance in parental genomes or endosperm balance number (EBN) causes hybrid seed lethality and ovary abscission in both interspecific and intraspecific-interploidy crosses in the genus *Nicotiana*. Interesting, Kopecký et al. demonstrated that in *Allium cepa* × *A. roylei* hybrids, chromosomes of *A. cepa* are frequently substituted by those of *A. roylei* and in just one generation, the genomic constitution shifts from 8 *A. cepa* + 8 *A. roylei* chromosomes in the F_1 generation to the average of 6.7 *A. cepa* + 9.3 *A. roylei* chromosomes in the F_2 generation, inferring that female meiotic drive is the key factor underlying *A. roylei* genome dominance. Salina et al. identified regions of recombination suppression in wheat chromosome 5B based on comparisons of the 5B map of a cross between the Chinese Spring (CS) variety of hexaploid wheat and CS-5Bdic with several 5B maps of tetraploid and hexaploid wheat. Pan et al. reported that wheat- *Agropyron cristatum* derivative II-11-1 was proven to contain a pair of 5P chromosomes and a pair of 2P chromosomes with 42 wheat chromosomes by analyzing the fluorescence *in situ* hybridization (FISH) and expressed sequence tag (EST) markers. Additionally, cytological identification and field investigation showed that the 5P chromosome can weaken the homologous pairing of wheat chromosomes and promote pairing between homoeologous chromosomes. This provides new materials for studying the mechanism of the alien gene affecting homologous chromosome pairing and promoting homoeologous pairing of wheat.

The major issue is genomics of interspecific hybrids and hybrid heterosis. Mo et al. made a comprehensive comparative transcriptome sequencing analysis of root samples from the hybrid G70 × GDH11 and its parental inbred lines G70 and GDH11 to elucidate the importance of the root uptake capacity of potassium (K+) in the formation of heterosis in *Nicotiana tabacum* L. The results showed that 29.53% and 60.49% of the differentially expressed genes (DEGs) exhibited dominant and over-dominant expression patterns, respectively. Li et al. applied the reciprocal interspecific hybrids and their parents (*Gossypium hirsutum* and *Gossypium*

barbadense) to elucidate the transcription regulatory mechanism of early biomass heterosis. Comparative transcriptome analysis showed that transgressive down-regulation (TDR) is the main gene expression pattern in the hybrids (*G. hirsutum* × *G. barbadense*, HB). Transgressive up-regulation (TUR) is the major primary gene expression pattern in the hybrids (*G. barbadense* × *G. hirsutum*, BH). The above results demonstrated that overdominance mediates biomass heterosis in interspecific hybrid cotton and the supervisory mechanism divergence of hybrids with different females. Cardoso-Silva et al. identified and characterized Orphan genes (OGs) in sugarcane. The results obtained suggested that sugarcane OGs are involved in several biological mechanisms, including stimulus response and defense mechanisms.

In summary, the research collected on this Research Topic facilitated understanding of issues related to interspecific hybridization in plant biology. We believe that this platform for enhancing exchange and promoting development has merit to be continued regarding some issues such as genomics of natural or synthetic polyploid formation, genomic responses to interspecific hybridization, transmission genetics across species boundaries, and genomics of interspecific hybrid heterosis.

Author contributions

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

Funding

This work was supported by National Natural Science Foundation of China (Grant No. 31991221) for DT and by the Ministry of Agriculture of the Republic of Kazakhstan in the framework of program funding for research (BR10765038) for RK.

Acknowledgments

We thank all authors and reviewers for their contributions to this Research Topic and for the support of the Editorial Office.

Conflict of interest

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

Publisher's note

frontiers in Plant Science

ORIGINAL RESEARCH
published: 25 November 2021
doi: 10.3389/fpls.2021.765645

Deployment of *Brassica carinata* A. Braun Derived *Brassica juncea* (L.) Czern. Lines for Improving Heterosis and Water Use Efficiency Under Water Deficit Stress Conditions

*Omkar Maharudra Limbalkar[1], Rajendra Singh[1], Parvesh Kumar[1], Joghee Nanjundan[2], Chiter Mal Parihar[3], Prashant Vasisth[1], Devendra Kumar Yadava[4], Viswanathan Chinnusamy[5] and Naveen Singh[1]**

[1] Division of Genetics, Indian Council of Agricultural Research-Indian Agricultural Research Institute, New Delhi, India, [2] Indian Council of Agricultural Research-Indian Agricultural Research Institute Regional Station, Wellington, India, [3] Division of Agronomy, Indian Council of Agricultural Research-Indian Agricultural Research Institute, New Delhi, India, [4] ADG (Seeds), Indian Council of Agricultural Research, New Delhi, India, [5] Division of Plant Physiology, Indian Council of Agricultural Research-Indian Agricultural Research Institute, New Delhi, India

OPEN ACCESS

Edited by:
Dayun Tao,
Yunnan Academy of Agricultural Sciences, China

Reviewed by:
Cengiz Toker,
Akdeniz University, Turkey
Devendra Ram Malaviya,
Indian Grassland and Fodder Research Institute (ICAR), India

***Correspondence:**
Naveen Singh
ns1.genet@gmail.com

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

Received: 27 August 2021
Accepted: 13 October 2021
Published: 25 November 2021

Citation:
Limbalkar OM, Singh R, Kumar P, Nanjundan J, Parihar CM, Vasisth P, Yadava DK, Chinnusamy V and Singh N (2021) Deployment of Brassica carinata A. Braun Derived Brassica juncea (L.) Czern. Lines for Improving Heterosis and Water Use Efficiency Under Water Deficit Stress Conditions. Front. Plant Sci. 12:765645. doi: 10.3389/fpls.2021.765645

Among *Brassica* species, Ethiopian mustard (*Brassica carinata* A. Braun) is known to tolerate most abiotic stresses, including drought. Drought caused by low and erratic rainfall in semi-arid regions consistently challenges rapeseed mustard productivity. Development of *B. carinata*-derived lines (CDLs) in *Brassica juncea* (L.) Czern. nuclear background, carrying genomic segments from *B. carinata*, are expected to tolerate moisture deficit stress conditions. The present study was, thus, aimed to establish the phenomenon "heterosis" for drought tolerance and water use efficiency by evaluating 105 hybrids developed from intermating 15 CDLs in half diallel fashion. Data on 17 seed yield and yield contributing traits were recorded under two different environments, *viz*., irrigated and rainfed conditions. Traits under study were found to be governed by both additive and non-additive types of gene action. Average degree of dominance was higher (>2) for yield and yield contributing traits, *viz*., secondary branches/plant, point to first siliqua on main shoot, total siliquae/plant, 1,000-seed weight, seed yield/plant, biological yield, harvest index, and seed yield/hectare under rainfed conditions, clearly indicating that higher productivity under drought conditions can be realised through the development of hybrids. Out of 15, highly significant general combining ability (GCA) effects for seven CDLs were observed under rainfed condition. Furthermore, nine and six hybrids expressed highly significant specific combining ability (SCA) effects and > 50% heterobeltiosis for yield contributing traits under rainfed and irrigated conditions, respectively. Water use efficiency (WUE) of parental CDLs and hybrids varied from 2.05 to 2.57 kg m^{-3} under rainfed, while 1.10 to 1.28 kg m^{-3} under irrigated conditions. Hybrids expressed higher WUE than parental lines under both water regimes. Furthermore, selection indices such as drought tolerance index (DTI) and mean relative performance (MRP) were identified to be efficient in the selection of productive

CDLs and hybrids under drought conditions. Nine hybrids, identified as highly productive in the present study, can further be exploited for improving the yield of Indian mustard in drought-prone areas. Usefulness of interspecific hybridisation in the development of *B. carinata*-derived *B. juncea* lines for improving heterosis and WUE is, thus, well demonstrated through the present study.

Keywords: *Brassica juncea*, interspecific hybridization, gene effects, heterosis, water use efficiency, drought tolerance

INTRODUCTION

Rapeseed-mustard group is the second most important oilseed crop in the world after soybean with a production of 69.08 million metric tons and contributed 11.89% to global oilseed production (580.51 million metric tons) in 2019–2020 (Anonymous, 2021). India is the third-largest producer of rapeseed-mustard group of crops followed by Canada and China with a total production of 7.4 million metric tons from a 7.1 mha area (Anonymous, 2021). Indian mustard, *Brassica juncea* (L.) Czern. (AABB; 2n = 36), is an amphidiploid species that originated from a spontaneous interspecific hybridisation of *Brassica rapa* L. (AA, 2n = 20) and *B. nigra* L. (BB, 2n = 16). It is a widely cultivated species among the rapeseed-mustard group of crops accounting for more than 90% of its total acreage in India. Being a pivotal crop of the Indian oilseed sector, extensive efforts have been made to improve its seed and oil yields to achieve national self-sufficiency in edible oils. Despite that, a huge amount of edible oil is being imported to meet the requirements for domestic consumption. The Solvent Extractors' Association of India reported that the import of edible oil rises to 23 million metric tons, costing more than 11 thousand million dollars in 2020 in the country, which accounts for more than 60% of the total edible oil demand.

To achieve self-sufficiency in edible oils, there is an urgent need to improve the productivity of Indian mustard. However, narrow genetic base of working gene pool is a major impediment in the improvement of this species (Chauhan et al., 2011). Most of the released Indian varieties are either derived directly from the cultivar Varuna or its derivatives (Singh and Chauhan, 2010; Chauhan et al., 2011). The yield potential of these varieties is about 4 t/ha; however, the national average yield is hovering around 1.5 t/ha. The stochastic production patterns are largely attributed to its susceptibility to different biotic and abiotic factors (Sharma et al., 2018). Indian mustard group of crops is mainly cultivated by marginal and small-farmers, wherein about 25% of the rapeseed-mustard area is rainfed, causing critical yield losses particularly in the drought-prone areas of eastern and western parts of the country (Sharma et al., 2018). Most of the cultivars of Indian mustard are susceptible to drought and encounter severe yield losses up to 94% (Sharma and Kumar, 1989). Although few varieties of Indian mustard have been released for commercial cultivation in rainfed areas of the country (Chauhan et al., 2011), these varieties failed to occupy sufficiently large area in the drought prone regions. This indicates that target environments still demand a better level of tolerance in future varieties. This also calls for the urgent development of high-yielding water use efficient genotypes and hybrids which can withstand drought conditions to minimise the yield losses, improve productivity and stabilise mustard production.

Relatives of *B. juncea* [*Sinapis alba* (Warwick, 1993), *Arabidopsis thaliana* (Kasuga et al., 1999); *B. rapa* (Seo et al., 2010), *B. napus* (Dalal et al., 2009; Fletcher et al., 2015), *B. carinata* (Malik, 1990)] harbour many useful traits, such as tolerance to drought, salinity, and cold. These could be used to incorporate abiotic stress tolerance in present-day cultivars (Warwick, 1993). Furthermore, wide hybridisation has been considered a novel method for generating selectable genetic variability in cultivated species (Prakash, 1973; Inomata, 1997).

Brassica carinata (BBCC; 2n = 34), an important oilseed crop of Ethiopian origin, is being cultivated in northeast Africa, parts of Canada, France, Spain, Australia, China, and India. Although this species has tall plant stature, poor seed characteristics, and takes longer to mature (**Supplementary Table 1**), when compared to *B. juncea*, it possesses many agronomically desirable traits including resistance or tolerance to most of the abiotic and biotic stresses like drought, heat, aphid, Sclerotinia rot, white rust, Alternaria black spot, blackleg, and powdery mildew (Raman et al., 2017; Thakur et al., 2019), which are otherwise eliminated from the predominantly cultivated *Brassica* species during domestication. Therefore, efforts toward interspecific hybridisation were made in the past to transfer desirable traits from *B. carinata* to cultivated *Brassica* species *viz*., *B. rapa* (Jiang et al., 2007; Choudhary et al., 2008), *B. napus* (Navabi et al., 2011; Sheikh et al., 2014), and *B. juncea* (Getinet et al., 1994; Sheikh et al., 2014; Singh K. H. et al., 2015). *B. carinata* is known to be the most drought tolerant among cultivated *Brassica* species as it develops sufficient biomass and seed yield even under the water deficit stress conditions. However, rainfall or irrigation leads to higher biomass accumulation, lodging of crop and, thus, reduction in seed size, oil content, and other yield related traits in this species (Thakur et al., 2019).

Keeping this in view, efforts were made at the Indian Council of Agricultural Research-Indian Agricultural Research Institute (ICAR-IARI), New Delhi to enrich the genome of *B. juncea* with *B. carinata*, through the development of *B. carinata*-derived *B. juncea* lines (CDLs), for improving their ability to withstand drought conditions. Though inter-subgenomic heterosis were reported in hybrids of *B. rapa* and *B. carinata*-derived introgression lines with *B. juncea* accessions (Wei et al., 2016), the present study aimed to use *B. carinata*-derived *B. juncea* lines for improving heterosis and water use efficiency (WUE) under water deficit stress conditions. To the best of our knowledge, this is the first case where diverse CDLs were developed and deployed for improving drought tolerance.

Hybrid breeding involving diverse parental lines is considered one of the most viable options for breaking the yield barrier (Singh N. et al., 2015). Involvement of alien introgression lines in hybrid development enables stretching of genetic diversity among parental lines required for the expression of higher heterosis, complementation of favorable novel alleles from both the introgressed parents in the F_1 generation, and, thus, improved ability of hybrids to withstand stress. Information on gene action, their relative contribution to the genetic variance, and estimation of general and specific combining ability effects (GCA and SCA) under water deficit stress conditions are pre-requisite for the identification of best parental combinations enabling the development of high yielding drought tolerant mustard hybrids (Panhwar et al., 2008; Khan et al., 2009). Several selection indices have also been suggested for the identification of water use efficient genotypes and cross combinations in crop plants, including mustard (Fischer and Maurer, 1978; Rosielle and Hamblin, 1981; Clarke et al., 1984; Fernandez, 1992; Huang, 2000). Despite this, this information was not much helpful in mitigating the ill-effect imposed by drought conditions in Indian mustard, primarily due to inherent genetic ceiling in this species.

Brassica juncea CDLs possessing useful genomic segment(s) from *B. carinata*, developed at the Indian Council of Agricultural Research – Indian Agricultural Research Institute (ICAR-IARI), New Delhi, shall be useful in the identification of high yielding drought tolerant parental lines and hybrids. Furthermore, information generated on gene action and degree of dominance for yield contributing traits and WUE under moisture deficit stress condition in *B. carinata* derived lines and hybrids shall help in devising future breeding strategies. The present study was, therefore, conducted to (i) determine the type of gene action, combining ability of parental lines and heterobeltiosis in the hybrids developed by using CDLs, (ii) assess the WUE of CDLs and their hybrids, and (iii) identify water use efficient genotypes and cross combinations.

MATERIALS AND METHODS

Selection of *Brassica carinata* Derived *Brassica juncea* Lines

A set of F_2 populations was generated from crossing *B. juncea* genotypes (DRMRIJ 31, Pusa Mustard 30 and Pusa Agrani) with *B. carinata* accessions (BC-4, BC-5, and BC-12). Within each of these F_2 populations, plants were selected on the basis of phenotype and were inter-mated in pairs – a method commonly called biparental mating. Phenotypic selection in the subsequent filial generations led to the development of *B. carinata*-derived *B. juncea* lines (CDLs). These cytologically stable homozygous CDLs possess 18 bivalents ($2n = 36$), indicating its similarity to *B. juncea* (**Supplementary Figures 1, 2**). These CDLs, in F_6 generation, were evaluated in two rows of 4-m length plots under both irrigated and rainfed field condition in the Division of Genetics, ICAR- IARI, New Delhi during 2018–2019 *rabi* season. To select CDLs differing in metric traits and response to moisture deficit stress, row-row and plant–plant spacing were kept at 30 and 15 cm, respectively (data not presented). One row was left blank before and after every plot such that plot–plot and row–row in a plot were separated by a spacing of 60 and 30 cm, respectively. The experiment was raised in augmented design both under irrigated and rainfed conditions. No irrigation was applied in the rainfed plot, while two irrigations, each of 50 mm depth, were applied in irrigated plots at 45 and 90 days after sowing (DAS). All recommended agronomic practices were adopted for raising the crop. Based on drought tolerance indices, 15 CDLs differing for various morpho-physiological traits and phenotypical and cytological resemblance to *B. juncea* (**Supplementary Table 1** and **Supplementary Figures 1, 2**) were selected for generating hybrids.

Generation of Crosses

During 2019, 15 diverse CDLs differing for morpho-physiological traits were intermated in half diallel design (excluding reciprocals) to generate seeds of 105 F_1 hybrids in ICAR-IARI, Regional Station, Wellington, Tamil Nadu, India in the off-season.

Evaluation of Hybrids and Their Parental Lines

During *rabi* 2019–2020 season, the set of 105 F_1 hybrids along with 15 parental CDLs was raised under both rainfed and irrigated conditions in the research field of Division of Genetics, ICAR-IARI, New Delhi. Two recommended irrigations of 50 mm depth each were applied at 45 and 90 DAS to the irrigated plots, while no irrigation was given to rainfed plots. Hybrids along with parents were raised in a randomised complete block design (RCBD) with three replications. Each plot consisted of a paired row of 4 m in length. Row to row and plant to plant distance were kept 30 and 15 cm, respectively. One row of 30 cm was left blank before and after every pair-row plot such that plot–plot and row–row in a plot were distanced at 60 and 30 cm, respectively. Recommended agronomical practices and plant protection measures were adopted for raising the crop.

Observations were recorded on 17 quantitative traits, *viz*., plant height (cm), point to first branch (cm), number of primary branches, number of secondary branches, main shoot length (cm), point to first siliqua on main shoot (cm), number of siliquae on main shoot, total siliquae per plant, siliqua length (cm), seeds per siliqua, days to 50% flowering, 1,000-seed weight (g), seed yield per plant (g), biological yield per plant (g), harvest index, seed yield (kg/ha), and oil content. The data were recorded on five randomly selected competitive plants from all the plots in each replication, except for seed yield per plot (g) and days to 50% flowering where observations were recorded on plot basis. Biological yield was recorded at physiological maturity after drying the harvested plants. Data on plot yield (g) were used to calculate seed yield (kg/ha).

Assessment of Drought Tolerance

Data were recorded on 17 metric traits under both irrigated and rainfed conditions. Mean values of these traits were used to analyse drought susceptibility index (DSI; Fischer and Maurer, 1978), drought tolerance index (DTI; Fernandez, 1992), tolerance

index (TOL; Rosielle and Hamblin, 1981), and mean relative performance (MRP) for the CDLs taken in this study (Raman et al., 2012).

Water-use efficiency (WUE) for parents and hybrids were calculated from the below given equation:

WUE [kg $ha^{-1}mm^{-1}$] = Seed Yield (kgha^{-1})/Water received from irrigation and rainfall (mm)

and WUE [kg m^{-3}] = 0.1 × WUE [kg $ha^{-1}mm^{-1}$].

The effective rainfall (between October 2019 and February 2020) was calculated by using Cropwat Version 8.0 by USDA SCS method. Cropwat is a computer-based decision support tool developed by the Food and Agriculture Organization (FAO) of the United Nations to estimate effective rainfall based on climate and soil data. The maximum and minimum temperatures and rainfall (mm) recorded during 2019–2020 crop seasons from ICAR-IARI meteorological observatory, located within 500 m distance from the experimental site, is presented in **Supplementary Figure 3**.

Statistical and Genetic Analyses

The mean values obtained from 17 different quantitative traits were subjected to analysis of mean and variance. Data obtained from 105 F_1 and 15 parental CDLs were subjected to Griffing's method II with a fixed model (model I) to determine the combining abilities (Griffing, 1956). The mathematical model for *ijk*th observation is as follows:

$$Yij = \mu + g_i + g_j + s_{ij} + b_k + e_{ijk}$$

where Y_{ij} is the mean of $i \times j^{th}$ genotype over k; μ is general population mean; g_i and g_j is the general combining ability effects of i^{th} and j^{th} parent, respectively; s_{ij} is the specific combining ability effects of cross involving i^{th} and j^{th} parent, respectively; and b_k is the replication effect; and e_{ijk} is the residual error.

Heterobeltiosis expressed in the hybrids was calculated by using mean values of respective F_1 and its better parent (BP) following standard formula (Coffman, 1933). The narrow-sense heritability (h^2) was calculated by using the method suggested by Mather and Jinks (1982). Analysis of variance and estimation of degree of dominance, heritability, and combining ability were done following the method suggested by Griffing (1956) using Windostat Version 9.1. The degree of dominance was estimated from variance due to dominance (σ^2_D) and variance due to additive (σ^2_A) gene action by using the formula $[\{(2 \times \sigma^2_D)/\sigma^2_A\}^{1/2}]$ given by Comstock and Robinson (1948).

RESULTS

Sufficient phenotypic variation was observed for the seed yield and its contributing traits under both rainfed and irrigated conditions. Most of the traits were normally distributed except days to 50% flowering (**Supplementary Figure 4**). There were significant differences ($p < 0.001$) among the parents and hybrids for all the recorded 17 traits, *viz*., plant height, point to the first branch, number of primary branches per plant, number of secondary branches per plant, main shoot length, point to the first siliqua on the main shoot, number of siliquae on the main shoot, total number of siliquae per plant, days to 50% flowering, 1,000-seed weight (g), seed yield per plant (g), biological yield per plant (g), harvest index, seed yield (kg/ha), and oil content (%) under irrigated (**Table 1**) and rainfed (**Table 2**) conditions. The mean squares of parents vs. hybrids under irrigated condition were found to be significant for all traits except point to first siliqua on the main shoot while under the rainfed condition, it was also significant for most of the traits except point to the first siliqua on the main shoot, seeds per siliqua, 1,000-seed weight, and oil content. Results also exhibited a significant mean sum of squares due to hybrids for all the traits under both rainfed and irrigated conditions (**Tables 1**, **2**).

The variance due to specific combining ability (σ^2sca) were higher than general combining ability variance (σ^2gca) for traits, *viz*., point to the first branch, number of primary branches, number of secondary branches, main shoot length, point to first siliqua on the main shoot, number of siliquae on the main shoot, total siliquae per plant, days to 50% flowering, 1,000-seed weight (g), seed yield per plant (g), biological yield per plant (g), harvest index, seed yield (kg/ha), and oil content (%) and, thus, the average degree of dominance for these traits were observed to be more than one (**Table 3**). For siliqua length and seeds per siliqua under both irrigated and rainfed conditions, the average degree of dominance was less than unity. For plant height, σ^2sca > σ^2gca and the degree of dominance was found to be more than one under irrigated, whereas σ^2gca > σ^2sca and degree of dominance was less than unity under rainfed condition.

Narrow-sense heritability (h^2) for different traits varied from 0.09 to 0.69 and 0.07 to 0.75 under irrigated and rainfed conditions, respectively (**Table 3**). The estimate of the degree of dominance varied from 0.74 to 4.40 and from 0.54 to 5.13 under irrigated and rainfed conditions respectively across the traits (**Table 3**). Low to moderate h^2 were observed for all the traits under irrigated and rainfed conditions, with few exceptions. High average degree of dominance along with low h^2 were observed for point to first siliqua on main shoot (cm), total number of siliquae per plant, seed yield per plant (g), biological yield per plant (g), and seed yield (kg/ha).

General Combining Ability

The general combining ability analysis revealed that each CDL possessed significant GCA effects for one or more traits under both irrigated or rainfed conditions (**Table 4**). CDL89 and CDL141 were found to be a good general combiner for point to the first siliqua on the main shoot, harvest index, seed yield per plant, seed yield (kg/ha), and oil content under rainfed conditions. CDL161, CDL112, CDL186, CDL101, and CDL25 were also found to possess high GCA for seed yield (kg/ha) and some of its contributing traits under water deficit stress conditions (**Table 4**).

Under irrigated condition, CDL161 and CDL106 expressed good general combining ability for point to the first branch, secondary branches/plant, main shoot length, total siliquae/plant, days to 50% flowering, 1,000-seed weight, seed yield/plant, biological yield/plant, and seed yield (kg/ha) (**Table 4**). CDL112, CDL186, CDL101, CDL25, and CDL89 also possessed higher

TABLE 1 | Analysis of variance for parents and hybrids generated in half diallel design for studying different yield contributing traits under irrigated conditions.

	Sources of variations							
Traits	**Replications**	**Female:male**	**Parents**	**Hybrids**	**Parents Vs. Hybrids**	**Error**	**Mean**	**CV (%)**
	d.f. = 2	119	14	104	1	238		
PH	53.50	764.78**	1047.12**	660.99**	7605.9**	140.32	230.07	5.15
	(0.38)	(5.45)	(7.46)	(4.71)	(54.2)			
PFB	12.47	717.21**	685.04**	726.95**	155.17**	5.72	32.34	7.39
	(2.18)	(125.38)	(119.76)	(127.08)	(27.12)			
PB/pt	0.64	3.66**	2.33**	3.71**	17.17**	0.92	7.60	12.65
	(0.69)	(3.96)	(2.52)	(4.01)	(18.56)			
SB/pt	3.81	51.78**	34.41**	50.68**	409.29**	1.71	17.89	7.32
	(2.23)	(30.2)	(20.07)	(29.56)	(238.68)			
MSL	4.41	185.65**	104.47**	190.97**	768.96**	37.81	79.73	7.71
	(0.12)	(4.91)	(2.76)	(5.05)	(20.33)			
FSM	1.43	40.35**	28.12**	42.37**	1.91	0.78	8.37	10.56
	(1.84)	(51.64)	(35.98)	(54.22)	(2.45)			
SMS	15.27	202.08**	121.12**	202.83**	1258.01**	38.32	53.59	11.55
	(0.4)	(5.27)	(3.16)	(5.29)	(32.82)			
TS/pt	271.49	79803.8**	29008.97**	84878.04**	263209.97**	1443.53	529.02	7.18
	(0.19)	(55.28)	(20.09)	(58.79)	(182.33)			
Siliqua length	0.127	0.517**	0.697**	0.49**	0.781**	0.062	4.15	5.99
	(2.06)	(8.37)	(11.27)	(7.93)	(12.64)			
Seeds/Siliqua	0.55	7.5**	8.33**	7.17**	30.63**	1.55	15.89	7.85
	(0.35)	(4.83)	(5.35)	(4.61)	(19.69)			
D50%F	3.13	107.05**	116.8**	98.61**	848.42**	12.69	56.73	6.28
	(0.25)	(8.43)	(9.19)	(7.76)	(66.81)			
TSW	0.17	0.87**	0.802**	0.851**	3.46**	0.079	3.96	7.1
	(2.13)	(10.94)	(10.12)	(10.73)	(43.68)			
SY/pt	2.51	302.77**	50.39**	303.21**	3790.15**	12.63	28.94	12.28
	(0.2)	(23.97)	(3.98)	(24)	(300.02)			
BY/pt	44.22	6404.17**	1012.63**	6632.03**	58188.069**	183.91	151.76	8.94
	(0.24)	(34.82)	(5.5)	(36.06)	(316.38)			
HI	0.47	33.15**	31.39**	32.4**	135.77**	1.3	19.00	6.02
	(0.36)	(25.38)	(24.03)	(24.8)	(103.94)			
SY/ha	86779.4	1071068**	1473003**	1018435**	917764.3**	39013.3	2670.5	7.39
	(2.22)	(27.45)	(37.75)	(26.10)	(23.52)			
Oil content	7.06	17.77**	27.27**	15.26**	146**	2.4	33.32	4.66
	(2.93)	(7.37)	(11.31)	(6.33)	(60.57)			

d.f., degrees of freedom; PH, plant height; PFB, point to first branch; PB/pt, primary branches/plant; SB/pt, secondary branches/plant; MSL, main shoot length; FSM, point to first siliqua on main shoot; SMS, number of siliquae on main shoot; TS/pt, total siliquae/plant; D50%F, days to 50% flowering; TSW, thousand seed weight; SY/pt, seed yield/plant; BY/pt, biological yield/plant; HI, harvest index; SY/ha, seed yield kg per hectare.
***Significant at p = 0.01 and *significant at p = 0.05.*

GCA for seed yield (kg/ha) along with some of its contributing traits under water sufficient conditions (**Table 4**).

Specific Combining Ability and Heterobeltiosis

Under the rainfed condition, a total of 36 hybrids exhibited significant SCA effects and high heterobeltiosis (>15%) for seed yield (kg/ha). Highest heterobeltiosis (99.55%) was observed for the cross CDL106 × CDL104 under limited water conditions. Out of 36, nine hybrids, *viz.*, CDL102 × CDL89, CDL102 × CDL106, CDL112 × CDL89, CDL186 × CDL121, CDL182 × CDL121, CDL 121 × CDL 103, CDL25 × CDL104, CDL89 × CDL104, and CDL104 × CDL106 exhibited highly significant SCA effects and more than 50% heterobeltiosis (**Table 5**). Similarly, under irrigated condition, 30 hybrids expressed significant SCA effects and heterobeltiosis > 15%. Highest heterobeltiosis (82.34%) for seed yield (kg/ha) was observed for the hybrid derived from parents CDL141 and CDL161 under favourable environment. Six hybrids, *viz.*, CDL161 × CDL186, CDL161 × CDL141, CDL141 × CDL182, CDL182 × CDL103, CDL182 × CDL106, and CDL104 × CDL106 recorded highly significant SCA effects and >50% heterobeltiosis for seed yield (kg/ha) under irrigated condition. It is also important to note that hybrid derived from cross CDL104 × CDL106 observed highly significant SCA effects and >50% heterobeltiosis under both rainfed and irrigated conditions (**Table 5**).

TABLE 2 | Analysis of variance for parents and hybrids generated in half diallel design for studying different yield contributing traits under rainfed conditions.

	Sources of variations							
Traits	**Replications**	**Female:male**	**Parents**	**Hybrids**	**Parents Vs. Hybrids**	**Error**	**Mean**	**CV (%)**
Df	d.f. = 2	119	14	104	1	238		
PH	39.22	729.2**	1070.57**	626.1**	6672.91**	213.8	228.26	6.4
	(0.18)	(3.41)	(5)	(2.92)	(31.21)			
PFB	11.8	566.33**	528.52**	568.31**	889.64**	22.04	48.45	10.02
	(0.54)	(25.69)	(23.97)	(25.78)	(40.35)			
PB/pt	0.194	2.83**	4.78**	2.5**	8.5**	0.318	6.90	8.16
	(0.61)	(8.88)	(15.04)	(7.88)	(26.72)			
SB/pt	0.2	26.24**	8.47**	28.68**	21.34**	1.52	12.10	10.66
	(0.13)	(17.22)	(5.56)	(18.82)	(14)			
MSL	1.24	150.71**	240.6**	137.29**	287.48**	28.37	78.12	6.82
	(0.04)	(5.31)	(8.47)	(4.83)	(10.13)			
FSM	0.15	24.57**	19.68**	25.46**	0.99	0.298	7.37	7.42
	(0.53)	(82.22)	(65.84)	(85.17)	(3.33)			
SMS	2.34	161.89**	171.46**	150.41**	1221.35**	36.25	54.84	10.98
	(0.06)	(4.47)	(4.72)	(4.14)	(33.68)			
TS/pt	1608.67	44551.2**	28892.99**	46996.99**	9402.96**	859.5	470.32	6.23
	(1.87)	(51.83)	(33.61)	(54.67)	(10.94)			
Siliqua length	0.11	0.67**	1.38**	0.57**	0.9**	0.072	4.33	6.23
	(1.52)	(9.3)	(19.02)	(7.96)	(12.48)			
Seeds/Siliqua	0.33	8.03**	10.81**	7.73**	0.38	1.47	16.06	7.59
	(0.23)	(5.44)	(7.32)	(5.23)	(0.26)			
D50F	3.811	72.66**	57.13**	74.1**	140.95**	9.79	55.61	6.71
	(0.39)	(7.42)	(5.83)	(7.56)	(14.38)			
TSW	0.58	1.14**	0.97**	1.16**	0.46	0.144	4.40	8.60
	(4.06)	(7.9)	(6.73)	(8.09)	(3.25)			
SY/pt	6.35	270.52**	44.47**	288.31**	1584.87**	9.39	31.34	9.77
	(0.68)	(28.8)	(4.73)	(30.69)	(168.73)			
BY/pt	87.68	5618.51**	757.68**	6146.71**	18737.27**	149.49	163.35	7.46
	(0.59)	(37.58)	(5.06)	(41.11)	(125.33)			
HI	0.053	81.06**	34.29**	87.13**	104.95**	0.99	19.57	5.10
	(0.05)	(81.11)	(34.31)	(87.18)	(105.01)			
SY/ha	81486	1399416**	2141042**	1310111**	304341.8**	39844.4	2819.37	7.08
	(2.04)	(35.12)	(53.73)	(32.88)	(7.63)			
Oil content	0.68	16.79**	24.12**	15.93**	3.93	2.18	34.03	4.34
	(0.31)	(7.69)	(11.04)	(7.29)	(1.801)			

d.f., degrees of freedom; PH, plant height; PFB, point to first branch; PB/pt, primary branches/plant; SB/pt, secondary branches/plant; MSL, main shoot length; FSM, point to first siliqua on main shoot; SMS, total siliqua on main shoot, TS/pt, total siliqua/plant; D50%F, days to 50% flowering; TSW, thousand seed weight; SY/pt, seed yield/plant; BY/pt, biological yield/plant; HI, harvest index; SY/ha, seed yield kg per hectare.
***Significant at p = 0.01 and *significant at p = 0.05.*

High heterobeltiosis (>15%) and higher SCA effects for seed yield were reported for hybrids derived from different combinations of parental CDLs, irrespective of their GCA effect levels (**Table 6**). Furthermore, 14 hybrids expressed significant SCA effects and high heterobeltiosis (>15%) under both rainfed and irrigated conditions (**Table 6**).

Mean Performance

The water requirement of rapeseed mustard is about 190–400 mm (Shekhawat et al., 2012; Sharma et al., 2018) and it is sensitive to water deficit stress at critical stages like pre-flowering and siliquae formation stages. In the presented experiment, the rainfed plots received 112.4 mm of water through rains up to physiological maturity stage, whereas two additional irrigations of 50 mm each were applied in the irrigated plots. Thus, the irrigated plots received a total of 212.4 mm of water throughout the crop developmental stages and met optimum water requirement. The data recorded on 17 morpho-physiological traits under rainfed and irrigated conditions were averaged separately for 15 parents and 105 hybrids. The results suggested that parental CDLs exhibited a higher mean value under irrigated than in rainfed condition for traits like plant height (cm), number of primary branches, number of secondary branches, main shoot length (cm), point to first siliqua on the

main shoot (cm), total siliquae per plant, days to 50% flowering, and seed yield (kg/ha). Whereas, under the rainfed condition, the mean value of CDLs was higher for point to first branch (cm), siliquae on the main shoot, siliqua length (cm), seeds/siliqua, 1,000-seed weight (g), seed yield/plant (g), biological yield/plant (g), harvest index, and oil content traits (**Table 7**) than in irrigated condition. Under irrigated condition, the overall mean of the hybrids was higher for plant height, primary branches per plant, secondary branches per plant, main shoot length, point to first siliqua on main shoot, total siliquae per plant and days to 50% flowering than in rainfed condition. Whereas traits like point to first branch (cm), number of siliquae on the main shoot, siliqua length (cm), seeds per siliqua, 1,000-seed weight (g), seed yield per plant (g), biological yield per plant (g), harvest index, seed yield (kg/ha), and oil content (%) observed higher mean values for hybrids when evaluated under rainfed condition (**Table 7**). Overall performance of CDLs and hybrids was found better when evaluated under rainfed condition for important yield and yield attributing traits including plant height (cm), total siliquae on main shoot (no.), siliqua length (cm), seeds/siliqua, days to 50% flowering, 1,000-seed weight (g), seed yield/plant, biological yield/plant (g), harvest index, and oil content (%) traits (**Table 7**).

Tolerance Indices and Water Use Efficiency

Drought tolerance indices, *viz*., drought susceptibility index (DSI), drought tolerance index (DTI), tolerance index (TOL), and mean relative performance (MRP) for seed yield (kg/ha) were estimated. Correlation coefficients between seed yield under

TABLE 3 | Estimates of genetic components of variance for seventeen characters in *B. carinata*-derived *B. juncea* lines (CDLs) and their hybrids evaluated under irrigated and rainfed conditions.

Components of variance		σ^2_{gca}	σ^2_{sca}	$\sigma^2_{gca}/\sigma^2_{sca}$	σ^2_A	σ^2_D	Average degree of dominance	h^2
Plant height (cm)	IR	69.75	77.81	0.90	139.50	77.81	1.06	0.52
	RF	64.27	49.03	1.31	128.53	49.03	0.87	0.52
Point to first branch (cm)	IR	46.93	162.34	0.29	93.86	162.34	1.86	0.36
	RF	36.22	123.53	0.29	72.44	123.53	1.85	0.36
Primary branches/plant	IR	0.19	0.59	0.33	0.39	0.59	1.75	0.30
	RF	0.18	0.54	0.33	0.35	0.54	1.75	0.35
Secondary branches/plant	IR	0.88	16.93	0.05	1.75	16.93	4.40	0.09
	RF	1.03	6.99	0.15	2.07	6.99	2.60	0.22
Main shoot length (cm)	IR	12.57	27.35	0.46	25.14	27.35	1.48	0.38
	RF	7.22	29.84	0.24	14.44	29.84	2.03	0.27
Point to first siliqua on main shoot (cm)	IR	1.90	10.65	0.18	3.79	10.65	2.37	0.25
	RF	0.57	7.86	0.07	1.14	7.86	3.71	0.13
Total siliqua on main shoot	IR	12.54	33.43	0.38	25.08	33.43	1.63	0.35
	RF	7.67	30.06	0.26	15.34	30.06	1.98	0.27
Total number of siliqua/plant	IR	3917.17	20723.84	0.19	7834.34	20723.84	2.30	0.27
	RF	1343.07	13461.46	0.10	2686.14	13461.46	3.17	0.16
Siliqua length (cm)	IR	0.06	0.03	1.85	0.12	0.03	0.74	0.69
	RF	0.09	0.03	2.50	0.17	0.03	0.63	0.75
Seeds/siliqua	IR	0.73	0.60	1.22	1.46	0.60	0.91	0.57
	RF	0.97	0.28	3.49	1.94	0.28	0.54	0.72
Days to 50% flowering	IR	8.98	15.29	0.59	17.96	15.29	1.30	0.48
	RF	5.08	12.25	0.41	10.15	12.25	1.55	0.40
1,000-seed weight	IR	0.05	0.19	0.26	0.10	0.19	1.95	0.32
	RF	0.03	0.30	0.11	0.07	0.30	2.96	0.17
Seed yield/plant	IR	10.58	85.62	0.12	21.16	85.62	2.84	0.19
	RF	7.27	82.16	0.09	14.54	82.16	3.36	0.15
Biological yield/plant	IR	1.88	1.92	0.98	3.76	1.92	1.01	0.16
	RF	147.00	1732.25	0.08	294.00	1732.25	3.43	0.14
Harvest index	IR	2.09	7.29	0.29	4.18	7.29	1.87	0.35
	RF	1.06	27.84	0.04	2.12	27.84	5.13	0.07
Seed yield/hectare	IR	18923.9	346993.1	0.05	37847.8	346993	4.28	0.10
	RF	47004.54	407072.46	0.12	94009	407072	2.94	0.18
Oil content	IR	0.90	3.74	0.24	1.80	3.74	2.04	0.29
	RF	1.58	1.94	0.81	3.16	1.94	1.11	0.54

IR, irrigated; RF, rainfed; h^2, narrow sense heritability.

TABLE 4 | General combining ability (GCA) effects of 15 parental lines and variance components for different traits in CDLs along with their hybrids under both rainfed and irrigated conditions.

Parents	Plant height (cm)	Point to first branch (cm)	Primary branches/ plant	Secondary branches/ plant	Main shoot length (cm)	Point to first siliqua on main shoot (cm)	Number of siliquae on main shoot	Total siliquae/ plant	Siliqua length (cm)	Seeds/ Siliqua	Day to 50% flowering	1,000-seed weight (g)	Seed yield/ plant (g)	Biological yield/ plant (g)	Harvest index	Seed yield (kg/ha)	Oil content (%)
CDL102	–14.2**	–14.63**	–0.13	2.9**	1.32	0.05	–4.69	71.12**	0.42**	2.12**	–4.3**	–0.25	–2.47	–19.53	0.63**	–222.3	–1.83
	(–12**)	(–13.2**)	(–0.87)	(–0.54)	(5.4**)	(–1.1**)	(–2.29)	(32.5**)	(0.41**)	(1.91**)	(–2.9**)	(–0.1)	(–3.04)	(5.69**)	(–2.8)	(–120.5)	(–1.8)
CDL161	–7.5**	–1.03	–0.56	–0.1	1.87*	–0.56**	0.45	–18.66	0.08	0.14	–4.12**	0.22**	3.38**	19.2**	–0.52	267.1**	–0.11
	(–3.3)	(–1.87**)	(0.27)	(0.97**)	(3.1**)	(–2.4**)	(3.67**)	(40.6**)	(–0.17)	(–0.28)	(–2.6**)	(0.43**)	(3.6**)	(21.5**)	(–0.19)	(72.5**)	(0.35)
CDL128	1.75	–1.25	–0.18	–0.48	–1.29	–0.86**	–0.43	8.36	0.05	–0.53	–2.2**	–0.26	3.5**	4.35*	1.16**	40.9	1.87**
	(1.5)	(–1.76**)	(–0.39)	(–0.05)	(–2.3)	(0.52)	(–5.83)	(–69.08)	(0.17**)	(0.54**)	(–2.4**)	(0.11)	(3.92**)	(1.7)	(2.1**)	(–142.7)	(2**)
CDL141	5.44	2.82	0.81**	–0.01	–4.68	–0.26**	0.22	–9.14	–0.2	–0.74	1.94	–0.02	0.99*	–1.77	0.47**	97.8**	1.51**
	(8.84)	(6.9)	(0.03)	(–0.89)	(–6.14)	(1.66)	(–2.97)	(–112.47)	(–0.15)	(–0.51)	(2.71)	(–0.31)	(–5.76)	(–29.34)	(–0.37)	(–265.8)	(0.5*)
CDL112	18.81	–0.6	0.39**	1.42**	3.31**	–1.33**	5.42**	51.38**	–0.35	–0.14	3.01	–0.14	4.76**	30.13**	–0.76	83.2**	1.5**
	(18.27)	(7.41)	(0.21)	(0.14)	(4.5**)	(–0.92**)	(3.27**)	(80.5**)	(–0.36)	(–0.37)	(4.73)	(–0.33)	(0.45)	(19.1**)	(–1.87)	(164**)	(–0.24)
CDL186	9.21	10.07	0.49**	–0.22	–1.25	0.23	2.33*	–32.96	–0.27	–0.13	2.07	0.23**	1.29**	2.26	0.77**	208.4**	–0.1
	(9.07)	(3.41)	(0.8**)	(1.31**)	(–3.78)	(–1.15**)	(–0.28)	(–18.8)	(0)	(0.51**)	(3.55)	(0.2**)	(–0.3)	(3.05)	(–0.97)	(54.9*)	(–0.5)
CDL182	4.58	8.6	0.34**	–0.61	–1.35	–0.01	–1.2	–71.05	0	0.30	1.25	0.07	–0.84	–7.1	0.64**	–24.18	–0.8
	(3.52)	(9.07)	(0.05)	(–1.37)	(–1.87)	(–1.35**)	(–1.78)	(–62.07)	(–0.11)	(0.45*)	(1.82)	(0.18**)	(3.04**)	(5.78**)	(1.1**)	(–66.9)	(–0.2)
CDL121	2.85	3.33	–0.12	–0.08	–0.6	0.29	–0.07	–27.22	–0.41	–1.87	1.25	0.16**	–1.92	–2.22	–0.61	–186	–0.88
	(5.61)	(9.56)	(–0.32)	(–0.73)	(1.26)	(0.94)	(1.85*)	(–49.1)	(–0.37)	(–1.77)	(2.57)	(0.2**)	(–0.71)	(–6.54)	(0.5**)	(–4.58)	(0.04)
CDL103	–6.02**	–1.03	–0.52	–1.19	0.54	0.35	–1.52	–35.57	0.16**	0.71**	–0.8**	0.05	–4.59	–14.92	–1.22	–524.9	–1.65
	(–7.2**)	(–5.11**)	(–0.35)	(–0.92)	(3.3**)	(2.16)	(–0.99)	(–48.38)	(0.12**)	(0.76**)	(–1.5**)	(0.15**)	(–1.71)	(–7.07)	(–0.12)	(–307.1)	(–0.4)
CDL101	0.12	4.81	0.53**	–0.13	–3.04	–0.11	1.75	–10.94	0.12**	0.22	1.25	–0.21	–1.04	–8.75	0.55**	164.9**	1.8**
	(–2.19)	(3.94)	(0.5**)	(–0.27)	(–2.97)	(–1.83**)	(5.58**)	(11.45*)	(0.07)	(–0.27)	(1.79)	(–0.13)	(1.4*)	(–4.6)	(1.7**)	(333**)	(1.2**)
CDL25	4.64	3.51	–0.2	0.43*	4.13**	1.07	5.65**	34.03**	–0.41	–0.75	0.74	–0.24	2.04**	3.22	1.59**	259.3**	0.4
	(4.09)	(–1.25**)	(–0.33)	(0.34)	(1.21)	(0.61)	(5.58**)	(86.6**)	(–0.22)	(–0.13)	(–0.22)	(–0.23)	(5.68**)	(15.1**)	(1.9**)	(313**)	(1.0**)
CDL89	–4.25*	–2.62**	0.15	0.07	–3.95	–0.79**	0.15	6.04	–0.12	–1.19	1.74	0.31**	0.91*	0.41	0.36*	221.4**	1.25**
	(–3.02)	(2.25)	(0.5**)	(0.38)	(–3.84)	(–0.85**)	(2.69**)	(–18.41)	(–0.12)	(–1.16)	(2.96)	(0.11*)	(–0.76)	(–12.6)	(1.4**)	(205**)	(0.6*)
CDL104	–6.5**	–2.04**	–0.22	–0.04	–1.48	–0.46**	–2.61	36.33**	0.11*	0.01	1.43	–0.07	–0.89	0.75	–0.61	40.56	–0.5
	(–8**)	(–5**)	(0.26)	(0.89**)	(–3.21)	(–1.46**)	(–3.06)	(43.8**)	(0.11**)	(–0.3)	(–1.9**)	(–0.22)	(–1.45)	(1.63)	(–1)	(35.9)	(–0.9)
CDL105	–9.02**	–6.56**	–0.57	–1.3	3.31**	0.94**	–2.69	–6.47	0.59**	1.56**	–1.61**	–0.06	–3.97	1.14	–2.47	–458.4	–0.96
	(–13**)	(–9.79**)	(–0.59)	(–1.11)	(1.18)	(0.74)	(–5.23)	(–14.96)	(0.49**)	(0.72**)	(–5.1**)	(–0.16)	(–5.46)	(–19.63)	(–1.2)	(–332.4)	(–0.79)
CDL106	0.13	–3.38**	–0.22	–0.66	3.18**	1.46	–2.75	4.74	0.21**	0.28	–1.61**	0.19**	–1.13	–7.17	0.02	32.1	–1.31
	(–2.19)	(–4.52**)	(0.18)	(1.84**)	(4.2**)	(0.6)	(–0.22)	(97.7**)	(0.11**)	(–0.09)	(–3.4**)	(0.09*)	(1.12*)	(6.29**)	(–0.23)	(60.9*)	(–0.91)
SE	1.98	0.64	0.08	0.17	0.72	0.07	0.82	3.97	0.04	0.16	0.42	0.05	0.42	1.65	0.14	27.01	0.20
	(1.6)	(0.32)	(0.13)	(0.18)	(0.83)	(0.12)	(0.84)	(5.14)	(0.03)	(0.17)	(0.48)	(0.04)	(0.48)	(1.84)	(0.16)	(26.72)	(0.21)
CD (P = 0.05)	4.24	1.36	0.16	0.36	1.55	0.16	1.75	8.51	0.08	0.35	0.91	0.11	0.89	3.55	0.29	53.2	0.43
	(3.44)	(0.69)	(0.28)	(0.38)	(1.78)	(0.26)	(1.8)	(11.02)	(0.07)	(0.36)	(1.03)	(0.08)	(1.03)	(3.94)	(0.33)	(52.63)	(0.45)
CD (P = 0.01)	5.88	1.89	0.23	0.50	2.15	0.22	2.43	11.81	0.11	0.49	1.26	0.15	1.23	4.92	0.40	70.13	0.60
	(4.77)	(0.96)	(0.39)	(0.53)	(2.48)	(0.36)	(2.49)	(15.3)	(0.10)	(0.50)	(1.44)	(0.11)	(1.43)	(5.46)	(0.46)	(69.38)	(0.63)

*Values represent GCA effects under rainfed condition; Values in parenthesis represent GCA effects under irrigated condition; **Significant at P = 0.01 and *significant at P = 0.05.*

TABLE 5 | Estimates of heterobeltiosis and specific combining ability (SCA) effects of 105 hybrids of CDLs for seed yield (kg/ha) under irrigated and rainfed conditions.

	CDL102	CDL161	CDL128	CDL141	CDL112	CDL186	CDL182	CDL121	CDL103	CDL101	CDL25	CDL89	CDL104	CDL105	CDL106
CDL102		–3.36	–27.74	–4.26	–13.98	42.82**	–36.94	–40.92	–8.82	20.32**	46.03**	51.62**	–15.33	–19.52	55.86**
		(–297.5)	(–532.5)	(–322.8)	(–235.9)	(605.6**)	(–970.1)	(–1000)	(105.6)	(626.9**)	(743.6**)	(988.7**)	(–615.4)	(–216.5)	(1093**)
CDL161	45**		34.03**	43.07**	–9.87	39.72**	12.32*	8.98	12.95*	40.04**	–3.36	29.75**	–0.35	22.99**	14.67*
	(685**)		(778.2**)	(615.7**)	(–608.6)	(416.3**)	(–78.9)	(–6)	(438.5**)	(693**)	(–779.1)	(138.3)	(–480.4)	(638.6**)	(–73)
CDL128	17.23*	44.04**		15.69**	–38.8	20.11**	32.6**	–31.93	–2.7	6.86	12.01*	–2.8	10.96*	–28.98	–27.36
	(289**)	(712.8**)		(412.9**)	(–1160.2)	(431.3**)	(1027**)	(–690.7)	(500**)	(88.7)	(144.3)	(–249.5)	(332.5**)	(–332.4)	(–775.7)
CDL141	0.5	82.34**	31.22**		19.75**	24.89**	30.16**	21.08**	12.6*	–8.68	40.25**	7.52	13.01*	–6.99	14.13*
	(8.2)	(788**)	(756**)		(402.3**)	(–31.2)	(415.3**)	(268.8**)	(398**)	(–510)	(298**)	(–438.7)	(–157.8)	(–154.4)	(–121.6)
CDL112	20.9**	0.62	–22.62	–0.86		1.76	7.33	–0.1	–2.05	–9.97	7.41	52.49**	8.5	–4.11	17.3**
	(565**)	(–178)	(–593.3)	(120.1)		(–219.3)	(171.6)	(122.3)	(406**)	(–509.2)	(–109.7)	(1209**)	(140.2)	(280.8**)	(398.7**)
CDL186	31.68**	53.41**	23.27**	17.23**	–15.56		–26.45	67.19**	–11.02	22.39**	29.24**	23.76**	49.03**	27.94**	57.43**
	(399**)	(481.7**)	(252.5*)	(46.5)	(–599.5)		(–1134.2)	(994.4**)	(–461.2)	(254.5*)	(–101)	(–141.4)	(351.1**)	(366.2**)	(552.4**)
CDL182	–25.76	27.3**	26.66**	55.83**	10.44	8.19		69.03**	–6.12	12.03*	18.91**	33.01**	–10.38	–31.04	4.92
	(–789.5)	(–396.4)	(452.3**)	(375**)	(227.9*)	(–345.7)		(1687**)	(115.9)	(195.4)	(–32.3)	(363.9**)	(–558)	(–584)	(–160.6)
CDL121	–49.24	5.92	–8.25	–12.29	22.94**	49.45**	48.49**		66.24**	–28.21	45.92**	5.86	31.69**	22.45**	69.56**
	(–1282.4)	(–103.2)	(–240.5)	(–217.9)	(504.5**)	(997.5**)	(1095**)		(394.4**)	(–776.1)	(705**)	(–196.4)	(–76.7)	(294.5**)	(163.5)
CDL103	–8.83	31.02**	–12.24	35.44**	–14.75	–17.28	60.94**	21.44**		–34.08	13.37*	–28.3	–42.11	36.64**	48.02**
	(–163.2)	(–89.5)	(–202.1)	(237.5*)	(–215.2)	(–649.9)	(505.3**)	(662.5**)		(–602.7)	(241.3*)	(–715.3)	(–1193.3)	(919.5**)	(142.9)
CDL101	1.11	25.21**	–5.43	0.8	5.82	42.82**	4.42	–5.83	18.1**		7.79	12.49*	42.31**	–14.79	28.11**
	(–89.5)	(383.6**)	(–247.8)	(47.3)	(–244.2)	(887.7**)	(–51.6)	(–397.1)	(566.6**)		(–207.5)	(–37.4)	(983.5**)	(–125.9)	(591.9**)
CDL25	37.81**	15.18**	–5.72	0.59	6.94	–6.95	–15.19	11.41	11.21	38**		27.91**	71.06**	–1.75	27.44**
	(995**)	(168.5)	(–201.4)	(98.3)	(–154.4)	(–433.6)	(–542.3)	(140)	(437**)	(546.7**)		(–88.1)	(1098**)	(–198.1)	(30.9)
CDL89	4.09	–12.61	–34.65	–43.03	31.22**	8.61	–12.29	–22.27	–13.66	–23.92	–3.97		58.19**	–16.7	33.08**
	(569**)	(–157.7)	(–646.3)	(–790.7)	(1150**)	(537.5**)	(–7.9)	(–388.9)	(188.6)	(–779.3)	(–121.8)		(891**)	(–490.7)	(269**)
CDL104	–31.15	–11.23	–1.19	–22.57	–15.56	–2.78	27.42**	8.63	–25.59	17.69**	16.57**	5.46		23.97**	99.55**
	(–774.8)	(–443.8)	(35.4)	(–404)	(–580.5)	(–204.2)	(712.1**)	(155.5)	(–441.9)	(212.3*)	(243.9*)	(455.9**)		(98.5)	(1043**)
CDL105	–45.27	–3.1	–21.64	–24.97	2.56	–19.94	–29.8	–22.11	–29.63	–27.71	–10.82	–27.14	–35.78		–17.25
	(–615.7)	(425.6**)	(98.2)	(123.8)	(499**)	(–49.7)	(–216.5)	(–53.9)	(28.7)	(–555.4)	(–40.5)	(–216.7)	(–494.4)		(–724.3)
CDL106	36.47**	26.96**	–0.28	3.82	9.42	28.72**	59.53**	–11.5	25.87**	42.33**	3.67	–8.52	71.83**	–44.51	
	(502**)	(–227.1)	(–295.1)	(–358.4)	(72.3)	(–34.7)	(573.4**)	(–525)	(130.3)	(867.9**)	(–142.2)	(–15.6)	(1752**)	(–775)	

Heterobeltiosis and SCA effects for rainfed (above diagonal) and irrigated condition (below diagonal) Values in parenthesis represent SCA effects.
***Significant at p = 0.01 and *significant p = 0.01 and *significant at P = 0.05.*

irrigated condition (Y_{IRR}), seed yield under rainfed condition (Y_{RF}), WUE_{RF}, WUE_{IRR}, DSI, DTI, TOL, and MRP were calculated. Yield under rainfed condition (Y_{RF}) were showing significant and positive association with WUE_{RF}, WUE_{IR}, DTI, MRP, and TOL. Similarly, significantly positive association of Y_{IR} with DTI and MRP was observed (**Supplementary Table 2**). WUE_{RF} exhibited a positive and significant correlation with DSI, DTI, TOL, and MRP, while WUE_{IR} showed a positive and significant correlation with DTI and MRP. DTI and MRP were identified to be significantly associated with all the traits except DSI, thus, indicating the usefulness of these two indices in indirect selection for drought tolerance.

The WUE (kg m^{-3}) was calculated for all hybrids along with parental CDLs both under rainfed and irrigated conditions. WUE among the parental lines raised under rainfed condition varied from 1.27 to 2.59 kg m^{-3} and recorded the highest and lowest values for CDL128 and CDL103, respectively (**Table 8**). Under irrigated condition it ranged from 0.73 to 1.50 kg m^{-3} with the highest and lowest values for CDL89 and CDL103, respectively. Among 105 hybrids evaluated under rainfed condition, WUE varied from 1.02 to 3.86 kg m^{-3} and the highest value was recorded for the cross CDL112 × CDL89. On the other hand, WUE varied from 0.59 to 2.13 kg m^{-3} under the irrigated condition and exhibited the highest value for CDL104 × CDL106 cross (**Table 8**). In general, the hybrids observed higher WUE than their parental CDLs. Overall, the mean of seed yield (kg/ha) and WUE of parental CDLs and hybrids was recorded higher under rainfed condition than in irrigated condition (**Figure 1**).

Further, drought tolerance index (DTI) and mean relative performance (MRP) were calculated from average values of individual parent and hybrid for all the 17 traits. All the traits, except point to first branch and point to first siliqua on the main shoot, exhibited higher DTI and MRP for hybrids than their corresponding parents (**Figure 2**). DTI for seed yield (kg/ha) varied from 0.25 to 2.55 in hybrids, whereas it varied from 0.31 to 1.12 in CDLs (**Table 8**). Some of the hybrids such as CDL112 × CDL89 (DTI: 2.55), CDL104 × CDL106 (DTI: 2.49) CDL182 × CDL121 (DTI: 2.23), CDL186 × CDL121 (DTI: 2.0), CDL101 × CDL106 (DTI: 1.99), CDL102 × CDL25 (DTI: 1.95), CDL25 × CDL104 (DTI: 1.93), CDL161 × CDL101 (DTI: 1.91),

TABLE 6 | Combining ability and heterobeltiosis of hybrids derived from CDLs with different levels of GCA effects for seed yield (kg/ha) under irrigated and rainfed conditions.

Environment	GCA effects of CDLs for Seed yield (kg/ha)		Hybrids with highly significant SCA effects and > 15% heterobeltiosis		
	Significant	**Non-significant**	**Hybrids developed from crossing CDLs with significant GCA effects**	**Hybrids developed from crossing CDLs with significant and non-significant GCA effects**	**Hybrids developed from crossing CDLs with non-significant GCA effects**
Rainfed condition (RF)	CDL161, CDL141, CDL112, CDL186, CDL101, CDL25 and CDL89	CDL102, CDL128, CDL182, CDL121, CDL103, CDL104, CDL105 and CDL106	CDL161 × CDL141, CDL161 × CDL186, CDL161 × CDL101, CDL141 × CDL112, CDL141 × CDL25, CDL112 × CDL89, CDL186 × CDL101 (7 Hybrids)	CDL161 × CDL128, CDL161 × CDL105, CDL141 × CDL128, CDL141 × CDL182, CDL141 × CDL121, CDL112 × CDL106, CDL186 × CDL102, CDL186 × CDL128, CDL186 × CDL121, CDL186 × CDL104, CDL186 × CDL105, CDL186 × CDL106, CDL101 × CDL102, CDL101 × CDL104, CDL101 × CDL106, CDL25 × CDL102, CDL25 × CDL121, CDL25 × CDL104, CDL89 × CDL102, CDL89 × CDL182, CDL89 × CDL104, CDL89 × CDL106 (22 Hybrids)	CDL102 × CDL106, CDL128 × CDL182, CDL182 × CDL121, CDL121 × CDL103, CDL121 × CDL105, CDL103 × CDL105, CDL104 × CDL106 (7 Hybrids)
Irrigated condition (IR)	CDL161, CDL112, CDL186, CDL101, CDL25, CDL89, and CDL106	CDL102, CDL128, CDL141, CDL182, CDL121, CDL103, CDL104, and CDL105	CDL161 × CDL101, CDL161 × CDL186, CDL112 × CDL89, CDL101 × CDL25, CDL186 × CDL101, CDL101 × CDL106 (6 Hybrids)	CDL161 × CDL102, CDL161 × CDL128, CDL161 × CDL141 CDL112 × CDL102, CDL112 × CDL121, CDL101 × CDL103, CDL101 × CDL104, CDL25 × CDL102, CDL25 × CDL104, CDL186 × CDL102, CDL186 × CDL128, CDL186 × CDL121, CDL106 × CDL102, CDL106 × CDL182, CDL106 × CDL104 (15 Hybrids)	CDL102 × CDL128, CDL128 × CDL141, CDL128 × CDL182, CDL141 × CDL182, CDL141 × CDL103, CDL182 × CDL121, CDL182 × CDL103, CDL182 × CDL104, CDL121 × CDL103 (9 Hybrids)
Number common in RF and IR conditions	6 CDLs	7 CDLs	4 hybrids	7 hybrids	3 hybrids

and CDL186 × CDL101 (DTI: 1.91) has a DTI value of >1, indicating their superiority under both moisture-stress and non-stress conditions. Among the CDLs highest MRP was recorded for CDL89 (2.09) and the lowest was observed for CDL103 (1.09). Among the hybrids, CDL112 × CDL89 recorded with the highest (3.11) MRP (**Table 8**). Further, nineteen hybrids, viz., CDL102 × CDL186, CDL102 × CDL25, CDL102 × CDL89, CDL102 × CDL106, CDL161 × CDL128, CDL161 × CDL141, CDL161 × CDL186, CDL161 × CDL101, CDL128 × CDL141, CDL186 × CDL121, CDL186 × CDL101, CDL182 × CDL121, CDL121 × CDL25, CDL101 × CDL104, CDL101 × CDL106, CDL25 × CDL89, CDL25 × CDL104, CDL89 × CDL104, and CDL104 × CDL106 were identified to be significantly better in productivity and possessing higher values of DTI and MRP (**Table 8**).

DISCUSSION

The present study revealed the usefulness of CDLs for improving heterosis and WUE in *B. juncea*. The existence of significant differences ($p < 0.001$) among parental CDLs and hybrids for all the 17 traits studied under irrigated and rainfed conditions indicated wide variability for the test traits in parents and their hybrids. The significance of mean squares among parents and hybrids suggested the suitability of the CDLs for combining

TABLE 7 | Mean values recorded on 15 parents and 105 hybrids for different traits under irrigated and rainfed conditions.

Traits	Parents		Hybrids	
	IR	RF	IR	RF
Plant height (cm)	217.91+	216.87	231.81*	229.88
Point to first branch (cm)	34.08	52.61+	32.09	47.86*
Primary branches/plant	7.02+	6.49	7.68*	6.96
Secondary branches/plant	15.07+	11.46	18.29*	12.19
Main shoot length (cm)	75.86+	75.76	80.28*	78.46
Point to first siliqua on main shoot (cm)	8.56+	7.51	8.34*	7.35
Number of siliqua on main shoot	48.65	49.97+	54.30	55.54*
Total siliqua/plant	457.48+	456.80	539.24*	472.25
Siliqua length (cm)	4.03	4.19+	4.17	4.35*
Seeds/siliqua	15.12	15.98+	16.00	16.08*
Days to 50% flowering	52.67+	53.96	57.31*	55.85
1,000-seed weight (g)	3.70	4.49+	4.00	4.38*
Seed yield/plant (g)	20.36	25.79+	30.17	32.14*
Biological yield/plant (g)	118.12	144.27+	156.56	166.08*
Harvest index	17.38	18.14+	19.23	19.77*
Seed yield (kg/ha)	2338+	2300	2718	2894*
Oil content (%)	36.63	38.75+	38.56	39.07*

**Traits for which higher mean value observed in hybrids.*
+Traits for which higher mean value observed in parental CDLs.
RF, rainfed; IR, irrigated.

ability and other analyses. Under both water regimes, the mean squares of parents vs. hybrids were found significant for several traits, revealed good scope for the manifestation of heterosis in hybrids generated from *B. carinata*-derived *B. juncea* lines. Further, under both irrigated and rainfed (**Tables 1**, **2**) conditions, the mean sum of squares due to hybrids was significant for all the traits, indicating the importance of both additive and non-additive variance. Earlier studies have also reported similar trends for additive and non-additive variance for the traits under study (Rao and Gulati, 2001; Yadav et al., 2005; Singh et al., 2006, 2013; Singh N. et al., 2015; Verma et al., 2011). Polygenic inheritance of drought tolerance is reported in many crop species including mustard (Fletcher et al., 2015). The quantitative traits are complex to understand hence methods like quantitative trait loci (QTL) mapping, marker-assisted breeding, and introgression from wild or related species are being used to improve drought tolerance (Mwadzingeni et al., 2016).

Combining Ability and Gene Action

Combining ability analysis partitions the genotypic variability into GCA (σ^2gca) and SCA (σ^2sca) variances representing additive and dominance effects, respectively. The higher value of σ^2sca than σ^2gca, along with an average degree of dominance more than unity observed for almost all the traits studied under irrigated and rainfed conditions, demonstrates the predominance of non-additive gene effects in the manifestation of heterosis (**Table 3**). For siliqua length and seeds per siliqua evaluated under both irrigated and rainfed condition, σ^2gca were higher than σ^2sca and average degree of dominance were less than unity demonstrating the role of additive gene effects in governing them. Similar results were also reported by Yadava et al. (2012) and Singh N. et al. (2015) in Indian mustard. Since additive and non-additive gene action were governing the yield contributing traits under irrigated and rainfed conditions, multiple or reciprocal recurrent selection could be useful for the improvement of these traits. The narrow-sense heritability (h^2) found low to moderate for almost all the traits under both the water regimes, suggesting the importance of interaction effects (Sayar et al., 2007). Since the expected genetic gain per cycle of selection will be less important, the development of hybrids involving CDLs carrying *B. carinata* genomic segments holds promise for exploitation of heterosis under rainfed conditions.

Significant GCA effects were observed for CDL161, CDL141, CDL112, CDL186, CDL101, CDL25, and CDL89 for seed yield (kg/ha) along with some of its contributing traits under rainfed condition, whereas under irrigated condition, CDL161, CDL112, CDL186, CDL101, CDL25, CDL89, and CDL106 expressed good general combining ability for seed yield (kg/ha) and its contributing traits (**Table 4**). The results suggested the importance of additive or additive × additive gene effects which are governing these traits. Similar results have also been reported earlier using a different set of Indian mustard genotypes (Mahak et al., 2008; Singh et al., 2010, 2013; Yadava et al., 2012; Singh N. et al., 2015). Furthermore, the good general combiners for seed yield traits were not showing significant GCA effects for some of its contributing traits, hence, there is a scope of improving the combining ability among parents for yield contributing traits under both irrigated and rainfed conditions. The findings of the present study suggest that CDLs, *viz.*, CDL161, CDL112, CDL186, CDL101, CDL25, and CDL89 possessing variable drought tolerance and expressing high GCA for seed yield along with other component traits under both rainfed and irrigated conditions can be incorporated in the breeding programme for accumulating desirable alleles for developing improved mustard cultivars suitable to water-scarce regions.

In the present study, significant SCA effects and high heterobeltiosis (>15%) were exhibited by 36 and 30 hybrids under rainfed and irrigated condition, respectively. Out of these, nine and six hybrids were expressing more than 50% heterobeltiosis when evaluated under rainfed and irrigated conditions, respectively (**Table 5**). This suggests that CDLs taken in this study can be exploited for developing hybrids for water-scarce regions. It is also important to note that 14 hybrids exhibited higher SCA effects and heterobeltiosis (>15%) under both water regimes (**Table 6**). These hybrids, expected to carry different introgression segments from *B. carinata*, have better resilience to changing water regimes and, thus, expressed better stability of performance. Earlier studies have recorded similar observations with a different set of conventional *B. juncea* genotypes (Singh et al., 2006; Srivastava et al., 2009). Teklewold and Becker (2005) reported significant SCA effects along with high heterobeltiosis for seed yield in *B. carinata* inbred lines, indicating substantial potential of Ethiopian mustard genotypes and their significance in the exploitation of hybrid vigour in *B. juncea* breeding programmes. Considerable heterosis was also reported in inter-subgenomic hybrids between

B. carinata-derived introgression lines and *B. juncea* accessions (Wei et al., 2016). Therefore, the development of hybrids using *B. carnata* derived lines will be an ideal approach in realizing higher yields under drought conditions.

The higher SCA effects and heterobeltiosis (>15%) for seed yield were exhibited by hybrids derived from CDLs, irrespective of their GCA effect values. Thus, hybrids generated from CDLs with significant × significant, significant × non-significant, and non-significant × non-significant GCA effects expressed high SCA effects and heterobeltiosis (**Table 6**). It demonstrates that joint effects of native and alien heterotic loci are responsible for the expression of such SCA effects and heterobeltiosis in hybrids derived from CDLs. Langham (1961) and Lal et al. (2006) have explained this anomaly where parents with other than high × high GCA effects also resulted in better heterosis and concluded that cross between parents with significantly high × low trait value resulted in superior transgressive segregants and, thus, the genes governing high trait values may express under homozygous or heterozygous conditions. Further, the genes governing lower trait values can also be expressed to their full potential in the homozygous state, but in heterozygote state, they can express very high performance due to complementation of alleles or overdominance. The existence of heterotic loci in these *B. carinata* derived lines can thus be exploited in the *B. juncea* breeding program for improving the level of heterosis and its commercial use.

Mean Performance, Tolerance Indices, and Water Use Efficiency

Brassica carinata, in general, adapted to tolerate drought conditions (Thakur et al., 2021), performed better under less availability of water and lodges under irrigated conditions due to its poor ability to withstand accumulated higher biomass. It has

TABLE 8 | Water use efficiency, drought tolerance indices, and seed yield performance of CDLs and their F_1 hybrids under rainfed (RF) and irrigated (IR) conditions.

CDLs/Cross	Seed yield (kg/ha)		WUE (kg/m^3)		DTI	MRP	CDLs/Cross	Seed yield (kg/ha)		WUE (kg/m^3)		DTI	MRP
	RF	IR	RF	IR				RF	IR	RF	IR		
CDL102	2388	2281	2.12	1.07	0.76	1.70	CDL161 × CDL186	3711	3280	3.30	1.54	1.71	2.54
CDL161	2656	1791	2.36	0.84	0.67	1.61	CDL161 × CDL182	2983	2280	2.65	1.07	0.95	1.91
CDL128	2914	2300	2.59	1.08	0.94	1.89	CDL161 × CDL121	2894	2635	2.58	1.24	1.07	2.01
CDL141	2478	1724	2.20	0.81	0.60	1.52	CDL161 × CDL103	3000	2346	2.67	1.10	0.99	1.94
CDL112	2842	2713	2.53	1.28	1.08	2.02	CDL161 × CDL101	3944	3460	3.51	1.63	1.91	2.69
CDL186	2294	2138	2.04	1.01	0.69	1.61	CDL161 × CDL25	2567	3224	2.28	1.52	1.16	2.12
CDL182	2542	1741	2.26	0.82	0.62	1.55	CDL161 × CDL89	3446	2791	3.07	1.31	1.35	2.27
CDL121	1506	2488	1.34	1.17	0.53	1.47	CDL161 × CDL104	2647	2335	2.35	1.10	0.87	1.81
CDL103	1425	1559	1.27	0.73	0.31	1.09	CDL161 × CDL105	3267	2836	2.91	1.34	1.30	2.22
CDL101	2817	2763	2.51	1.30	1.09	2.03	CDL161 × CDL106	3046	2577	2.71	1.21	1.10	2.05
CDL25	2465	2799	2.19	1.32	0.97	1.92	CDL128 × CDL141	3371	3018	3.00	1.42	1.43	2.33
CDL89	2511	3194	2.23	1.50	1.12	2.09	CDL128 × CDL112	1783	2099	1.59	0.99	0.52	1.42
CDL104	1972	2630	1.75	1.24	0.73	1.68	CDL128 × CDL186	3500	2835	3.11	1.33	1.39	2.30
CDL105	2017	2927	1.79	1.38	0.83	1.81	CDL128 × CDL182	3864	2913	3.44	1.37	1.58	2.46
CDL106	1668	2030	1.48	0.96	0.47	1.35	CDL128 × CDL121	1984	2283	1.76	1.07	0.63	1.56
CDL102 × CDL161	2567	3307	2.28	1.56	1.19	2.15	CDL128 × CDL103	2835	2019	2.52	0.95	0.80	1.76
CDL102 × CDL128	2106	2696	1.87	1.27	0.80	1.76	CDL128 × CDL101	3114	2613	2.77	1.23	1.14	2.08
CDL102 × CDL141	2372	2292	2.11	1.08	0.76	1.70	CDL128 × CDL25	3264	2639	2.90	1.24	1.21	2.15
CDL102 × CDL112	2444	3280	2.17	1.54	1.12	2.10	CDL128 × CDL89	2832	2087	2.52	0.98	0.83	1.79
CDL102 × CDL186	3411	3004	3.03	1.41	1.44	2.33	CDL128 × CDL104	3233	2599	2.88	1.22	1.18	2.12
CDL102 × CDL182	1603	1694	1.43	0.80	0.38	1.20	CDL128 × CDL105	2069	2294	1.84	1.08	0.67	1.59
CDL102 × CDL121	1411	1263	1.26	0.59	0.25	0.97	CDL128 × CDL106	2117	2294	1.88	1.08	0.68	1.61
CDL102 × CDL103	2178	2080	1.94	0.98	0.64	1.55	CDL141 × CDL112	3403	2689	3.03	1.27	1.28	2.21
CDL102 × CDL101	3389	2794	3.02	1.32	1.33	2.25	CDL141 × CDL186	3094	2506	2.75	1.18	1.09	2.04
CDL102 × CDL25	3600	3857	3.20	1.82	1.95	2.72	CDL141 × CDL182	3308	2713	2.94	1.28	1.26	2.19
CDL102 × CDL89	3807	3324	3.39	1.57	1.77	2.60	CDL141 × CDL121	3000	2182	2.67	1.03	0.92	1.88
CDL102 × CDL104	2022	1811	1.80	0.85	0.51	1.40	CDL141 × CDL103	2790	2335	2.48	1.10	0.91	1.86
CDL102 × CDL105	1922	1602	1.71	0.75	0.43	1.28	CDL141 × CDL101	2572	2785	2.29	1.31	1.00	1.96
CDL102 × CDL106	3722	3113	3.31	1.47	1.62	2.49	CDL141 × CDL25	3475	2816	3.09	1.33	1.37	2.29
CDL161 × CDL128	3906	3313	3.47	1.56	1.81	2.63	CDL141 × CDL89	2700	1819	2.40	0.86	0.69	1.64
CDL161 × CDL141	3800	3265	3.38	1.54	1.74	2.57	CDL141 × CDL104	2800	2037	2.49	0.96	0.80	1.76
CDL161 × CDL112	2561	2730	2.28	1.29	0.98	1.93	CDL141 × CDL105	2304	2196	2.05	1.03	0.71	1.64

(Continued)

TABLE 8 | (Continued)

CDLs/Cross	Seed yield (kg/ha)		WUE (kg/m^3)		DTI	MRP	CDLs/Cross	Seed yield (kg/ha)		WUE (kg/m^3)		DTI	MRP
	RF	IR	RF	IR				RF	IR	RF	IR		
CDL141 × CDL106	2828	2107	2.52	0.99	0.84	1.79	CDL121 × CDL103	2503	3021	2.23	1.42	1.06	2.02
CDL112 × CDL186	2892	2291	2.57	1.08	0.93	1.88	CDL121 × CDL101	2022	2602	1.80	1.22	0.74	1.69
CDL112 × CDL182	3050	2996	2.71	1.41	1.28	2.20	CDL121 × CDL25	3597	3119	3.20	1.47	1.57	2.44
CDL112 × CDL121	2839	3335	2.53	1.57	1.33	2.26	CDL121 × CDL89	2658	2482	2.37	1.17	0.93	1.87
CDL112 × CDL103	2783	2313	2.48	1.09	0.90	1.85	CDL121 × CDL104	2597	2857	2.31	1.35	1.04	1.99
CDL112 × CDL101	2558	2924	2.28	1.38	1.05	2.00	CDL121 × CDL105	2469	2280	2.20	1.07	0.79	1.73
CDL112 × CDL25	3052	2993	2.72	1.41	1.28	2.20	CDL121 × CDL106	2829	2202	2.52	1.04	0.87	1.83
CDL112 × CDL89	4333	4191	3.86	1.97	2.55	3.11	CDL103 × CDL101	1857	3263	1.65	1.54	0.85	1.88
CDL112 × CDL104	3083	2291	2.74	1.08	0.99	1.95	CDL103 × CDL25	2795	3113	2.49	1.47	1.22	2.16
CDL112 × CDL105	2725	3002	2.42	1.41	1.15	2.09	CDL103 × CDL89	1801	2757	1.60	1.30	0.70	1.67
CDL112 × CDL106	3333	2968	2.97	1.40	1.39	2.29	CDL103 × CDL104	1142	1957	1.02	0.92	0.31	1.14
CDL186 × CDL182	1869	2313	1.66	1.09	0.61	1.53	CDL103 × CDL105	2756	2060	2.45	0.97	0.80	1.75
CDL186 × CDL121	3836	3718	3.41	1.75	2.00	2.75	CDL103 × CDL106	2469	2555	2.20	1.20	0.88	1.83
CDL186 × CDL103	2042	1768	1.82	0.83	0.51	1.39	CDL101 × CDL25	3036	3863	2.70	1.82	1.64	2.52
CDL186 × CDL101	3447	3946	3.07	1.86	1.91	2.70	CDL101 × CDL89	3168	2430	2.82	1.14	1.08	2.03
CDL186 × CDL25	3186	2604	2.83	1.23	1.16	2.11	CDL101 × CDL104	4008	3252	3.57	1.53	1.83	2.64
CDL186 × CDL89	3108	3468	2.76	1.63	1.51	2.40	CDL101 × CDL105	2400	2116	2.14	1.00	0.71	1.64
CDL186 × CDL104	3419	2557	3.04	1.20	1.23	2.17	CDL101 × CDL106	3608	3932	3.21	1.85	1.99	2.75
CDL186 × CDL105	2936	2343	2.61	1.10	0.96	1.92	CDL25 × CDL89	3212	3067	2.86	1.44	1.38	2.29
CDL186 × CDL106	3612	2752	3.21	1.30	1.39	2.31	CDL25 × CDL104	4217	3263	3.75	1.54	1.93	2.72
CDL182 × CDL121	4296	3694	3.82	1.74	2.23	2.91	CDL25 × CDL105	2422	2610	2.16	1.23	0.89	1.84
CDL182 × CDL103	2386	2802	2.12	1.32	0.94	1.90	CDL25 × CDL106	3142	2902	2.80	1.37	1.28	2.20
CDL182 × CDL101	3156	2885	2.81	1.36	1.28	2.20	CDL89 × CDL104	3972	3368	3.53	1.59	1.88	2.67
CDL182 × CDL25	3022	2374	2.69	1.12	1.01	1.96	CDL89 × CDL105	2092	2327	1.86	1.10	0.68	1.61
CDL182 × CDL89	3381	2801	3.01	1.32	1.33	2.25	CDL89 × CDL106	3342	2921	2.97	1.38	1.37	2.28
CDL182 × CDL104	2278	3352	2.03	1.58	1.07	2.06	CDL104 × CDL105	2500	1880	2.22	0.88	0.66	1.59
CDL182 × CDL105	1753	2055	1.56	0.97	0.50	1.39	CDL104 × CDL106	3936	4520	3.50	2.13	2.49	3.09
CDL182 × CDL106	2667	3238	2.37	1.52	1.21	2.16	CDL105 × CDL106	1669	1624	1.48	0.76	0.38	1.20
							Overall mean	2819	2670	2.51	1.26	1.09	2.00
							CD	321	318				

WUE, water use efficiency; DTI, drought tolerance index; MRP, mean relative performance.

also been observed that irrigations in *B. carinata* genotypes also exhibits heavy regeneration, resulting in reduced plant height, poor seed set, and reduced seed size, which ultimately leads to reduction in seed yield per unit area. Conversely, bolder seed size is realised under rainfed than in irrigated conditions.

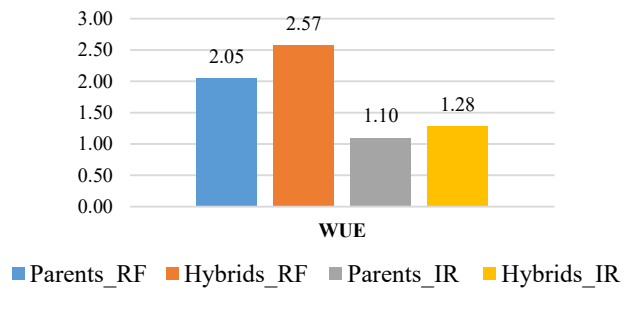


FIGURE 1 | WUE (kg m^{-3}) in *B. carinata* derived *B. juncea* lines (Parents) and their hybrids under rainfed (RF) and irrigated (IR) conditions.

In the present study, some of the *B. carinata* derived lines, along with their hybrids, observed reduced performance for seed yield under irrigated than in rainfed conditions. This could be attributed to the introduction of genomic segment(s) or substituted chromosomes from the C genome of *B. carinata* to *B. juncea*. Although, the possibility of later is remote when the biparental mating approach is being used for the development of such CDLs.

Parental CDLs and their hybrids exhibited higher mean values for seed yield (kg/ha) and other yield attributes under rainfed than in irrigated condition (**Table 7**). It is most probably due to presence of different introgression segments in CDLs which alter their performance in water deficit and surplus condition, whereas complementation of distinct genomic segments and expressed overdominance leads to heterosis in hybrids. Therefore, comparatively better performance of these hybrids under rainfed compared to irrigated conditions suggested that *B. carinata*-derived lines hold promise for the development of hybrids or cultivars for water-scarce regions.

FIGURE 2 | The drought tolerant index (DTI) and mean relative performance (MRP) observed on seventeen traits in hybrids along with parents grown in rainfed and irrigated conditions. PH, plant height; PFB, point to first branch; PB/pt, primary branches/plant, SB/pt, secondary branches/plant; MSL, main shoot length; FSM, point to first siliqua on main shoot; SMS, number of siliquae on main shoot; TS/pt, total siliquae/plant; D50%F, day to 50% flowering; TSW, thousand seed weight; SY/pt, seed yield/plant; BY/pt, biological yield/plant; HI, harvest index; SY/ha, seed yield kg per hectare.

The present study revealed that yield under rainfed condition (Y_{RF}) was showing positive correlation with WUE_{RF} ($r = 0.99$), WUE_{IR} ($r = 0.65$), and stress tolerance indices, *viz.*, DTI ($r = 0.90$), MRP ($r = 0.92$), and TOL ($r = 0.54$). Similarly, yield under irrigated conditions (Y_{IR}) was positively correlated with DTI ($r = 0.89$) and MRP ($r = 0.90$). In addition, DTI was found positively correlated with MRP ($r = 0.99$) (**Supplementary Table 2**). In the present study, a higher mean seed yield was recorded in rainfed than in irrigated conditions. Therefore, DSI would not be a reliable criterion for selection when mean values under stress condition is higher than that of normal irrigated condition. Better performance of CDLs and their hybrids under rainfed condition compared to irrigated plots attributed to excessive lodging of plants in irrigated conditions. Although the level of this lodging in CDLs and hybrids is much less than what normally happens in *B. carinata* after irrigation or rainfall. Therefore, DTI and MRP were found to be better predictors of Y_{IR} and Y_{RF} than TOL and DSI. This study is in accordance with previous reports suggesting DTI as a better parameter for identifying drought-tolerant genotypes in mungbean (Fernandez, 1992), corn (Shirinzadeh et al., 2010), wheat (Farshadfar et al., 2012), chickpea (Pour-Siahbidi and Pour-Aboughadareh, 2013), durum wheat (Mohammadi, 2016), and barley (Khalili et al., 2016). Mahajan et al. (2018) also reported that DTI is a more desirable index for stress tolerance than SSI in rice under moisture deficit stress.

The test hybrids expressed improved WUE as compared to parental CDLs. Overall, the mean of seed yield (kg/ha) and WUE of parental CDLs and hybrids was recorded higher under rainfed condition than in irrigated conditions (**Figure 1**), indicating the presence of putatively different loci/genomic segments in the CDLs for improved WUE and yield heterosis. Expression of overdominance and/or complementation of favourable alleles in hybrids appear to be the reason for improved WUE and established genetic elite in the hybrids.

Higher MRP and DTI confirmed better tolerance/adaptation of hybrids to moisture deficit stress over the parental CDLs for most of the traits under study (**Figure 2**). Hybrids expressing DTI value of >1, indicating their superiority under both moisture-stress and non-stress conditions. Lower DTI value, on the other hand, did not precisely explain susceptibility to moisture deficit stress conditions as mean yield under rainfed was higher than irrigated in present study. CDLs carry introgressions from *B. carinata* lodged after irrigation or rain and, thus, lead to a reduction in seed yield. In general, higher MRP of hybrids than CDLs indicates better performance of hybrids under rainfed and irrigated conditions. Based on higher DTI and MRP, highly productive and drought-tolerant hybrids, *viz.*, CDL102 × CDL186, CDL102 × CDL25, CDL102 × CDL89, CDL102 × CDL106, CDL161 × CDL128, CDL161 × CDL141, CDL161 × CDL186, CDL161 × CDL101, CDL128 × CDL141, CDL186 × CDL121, CDL186 × CDL101, CDL182 × CDL121, CDL121 × CDL25, CDL101 × CDL104, CDL101 × CDL106, CDL25 × CDL89, CDL25 × CDL104, CDL89 × CDL104, and CDL104 × CDL106 were identified (**Table 8**). These hybrids can be commercially exploited by converting the parental CDLs into CMS or restorer lines based on their seed production suitability. Hence, hybrid vigour expressed under moisture deficit stress

conditions is well demonstrated through the deployment of *B. carinata*-derived *B. juncea* lines.

CONCLUSION

In the present study, the availability of larger SCA variance than GCA variance for seed yield and its contributing traits under both water deficit and water sufficient conditions indicated a preponderance of dominance gene action. It suggests the manifestation of heterosis in *B. carinata*-derived *B. juncea* lines under moisture deficit stress conditions. The predominance of additive gene effects for traits like siliqua length and seeds per siliqua suggested the involvement of genetic and G × E interactions for these traits. Hybrids developed by involving CDLs with desirable SCA effects, higher heterobeltiosis, and better WUE exhibit overdominance and/or complement favourable alleles among different CDLs for expression of exploitable heterosis in hybrids. Many of the *B. carinata*-derived *B. juncea* lines that were identified in this study were heterotic and can be exploited for the development of commercial hybrids by converting them to cytoplasmic male sterile and/or restorer lines. CDLs, on the other hand, can also be used to recover desirable recombinants that would help in hastening the *Brassica* improvement programmes for the drought-prone rainfed areas. Material and information emanating from this study shall open new vistas for research involving related species for interspecific hybridization among *Brassica* species in creating novel genetic variability and deploying them for the development of high yielding hybrids with better WUE and, thus, overcoming the yield barriers in drought-prone regions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

OL conducted the field trials, generated crosses, analysed the data, and prepared the draft manuscript. RS conducted the trial, recorded observations, and maintained the CDLs. PV helped in data recording and analysis. PK and CP helped in conducting field trials, recording observations, and data analysis. JN helped in generating test hybrids. DY supervised the research. VC guided the research and edited the manuscript. NS developed the CDLs, conceptualized and supervised the research, analysed the data, and wrote the manuscript. All the authors read the manuscript and agreed with its content.

FUNDING

The research was funded by ICAR-IARI in-house project.

ACKNOWLEDGMENTS

OL is grateful to the University Grants Commission of India and Post-graduate School, ICAR-Indian Agricultural Research Institute (IARI), New Delhi for granting Senior Research Fellowship. The authors are thankful to the Division of Genetics, ICAR-IARI, New Delhi for providing facilities for hybridisation and evaluation of CDLs and hybrids. The authors are also thankful to the ICAR-IARI, Regional Station, Wellington, Tamil Nadu for extending help and resources for generating hybrid seed in the off-season.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.765645/full#supplementary-material

REFERENCES

Anonymous (2021). *Global Agricultural Information Network Circular Series, WAP 9–21*. Washington, DC: USDA Foreign Agricultural Service.

Chauhan, J. S., Singh, K. H., Singh, V. V., and Kumar, S. (2011). Hundred years of rapeseed-mustard breeding in India: accomplishments and future strategies. *Indian J. Agric. Sci.* 81, 1093–1109.

Choudhary, B. R., Joshi, P., and Ramarao, S. (2008). Interspecific hybridization between *Brassica carinata* and *Brassica rapa*. *Plant Breed.* 119, 417–420. doi: 10.1046/j.1439-0523.2000.00503.x

Clarke, J. M., Townley-Smith, F., McCaig, T. N., and Green, D. G. (1984). Growth analysis of spring wheat cultivars of varying drought resistance 1. *Crop Sci.* 24, 537–541. doi: 10.2135/cropsci1984.0011183X002400030026x

Coffman, F. A. (1933). Heterosis: Specific not general in nature. *Science* 77, 114–115. doi: 10.1126/science.77.1987.114-c

Comstock, R. E., and Robinson, H. F. (1948). The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. *Biometrics* 4, 254–266. doi: 10.2307/3001412

Dalal, M., Tayal, D., Chinnusamy, V., and Bansal, K. C. (2009). Abiotic stress and ABA-inducible Group 4 LEA from *Brassica napus* plays a key role in salt and drought tolerance. *J. Biotechnol.* 139, 137–145. doi: 10.1016/j.jbiotec.2008.09.014

Farshadfar, E., Poursiahbidi, M. M., and Abooghadareh, A. P. (2012). Repeatability of drought tolerance indices in bread wheat genotypes. *Int. J. Agric. Crop Sci.* 4, 891–903. doi: 10.1016/j.jplph.2014.04.013

Fernandez, G. C. (1992). "Effective selection criteria for assessing plant stress tolerance," in *Proceeding of the International Symposium on Adaptation of Vegetables and other Food Crops in Temperature and Water Stress*, Shanhua, 257–270.

Fischer, R. A., and Maurer, R. (1978). Drought resistance in spring wheat cultivars. I. Grain yield responses. *Austral. J. Agric. Res.* 29, 897–912. doi: 10.1371/journal.pone.0199121

Fletcher, R. S., Mullen, J. L., Heiliger, A., and McKay, J. K. (2015). QTL analysis of root morphology, flowering time, and yield reveals trade-offs in response to drought in *Brassica napus*. *J. Exp. Bot.* 66, 245–256. doi: 10.1093/jxb/eru423

Getinet, A., Rakow, G., Raney, J. P., and Downey, R. K. (1994). Development of zero erucic acid Ethiopian mustard through an interspecific cross with zero erucic acid Oriental mustard. *Can. J. Plant Sci.* 74, 793–795. doi: 10.4141/cjps94-141

Griffing, B. R. U. C. E. (1956). Concept of general and specific combining ability in relation to diallel crossing systems. *Austral. J. Biol. Sci.* 9, 463–493. doi: 10.1071/BI9560463

Huang, B. (2000). *Role of Root Morphological and Physiological Characteristics in Drought Resistance of Plants. Plant–Environment Interactions*. New York: Marcel Dekker Inc, 39–64. doi: 10.1201/9780824746568.ch2

Inomata, N. (1997). "Wide hybridization and meiotic pairing," in *Recent Advances in Oilseed Brassicas*, eds H. R. Kalia and S. K. Gupta (Ludhiana: Kalyani Publ.), 53–76.

Jiang, Y., Tian, E., Li, R., Chen, L., and Meng, J. (2007). Genetic diversity of *Brassica carinata* with emphasis on the interspecific crossability with *B. rapa*. *Plant Breed.* 126, 487–491. doi: 10.1111/j.1439-0523.2007.01393.x

Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1999). Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat. Biotechnol.* 17, 287–291. doi: 10.1038/7036

Khalili, M., Pour-Aboughadareh, A., and Naghavi, M. R. (2016). Assessment of drought tolerance in barley: integrated selection criterion and drought tolerance indices. *Environ. Exp. Biol.* 14, 33–41. doi: 10.22364/eeb.14.06

Khan, S. A., Ahmad, H., Khan, A., Saeed, M., Khan, S. M., and Ahmad, B. (2009). Using line x tester analysis for earliness and plant height traits in sunflower (*Helianthus annuus* L.). *Recent Res. Sci. Technol.* 1:316.

Lal, C., Hariprasanna, K., Rathnakumar, A. L., Gor, H. K., and Chikani, B. M. (2006). Gene action for surrogate traits of water-use efficiency and harvest index in peanut (*Arachis hypogaea*). *Ann. Appl. Biol.* 148, 165–172. doi: 10.1111/j.1744-7348.2006.00047.x

Langham, D. G. (1961). The high-low method in crop improvement. *Crop Sci.* 1, 376–378. doi: 10.2135/cropsci1961.0011183X000100050026x

Mahajan, G., Singh, K., Singh, N., Kaur, R., and Chauhan, B. S. (2018). Screening of water-efficient rice genotypes for dry direct seeding in South Asia. *Arch. Agron. Soil Sci.* 64, 103–115. doi: 10.1080/03650340.2017.1337891

Mahak, S., Bashrat, A. M., Lokendra, S., Brahm, S., and Dixit, R. K. (2008). Studies on combining ability for oil content, seed yield and its contributing characters in Indian mustard [*Brassica juncea* (L.) Czern & and Coss]. *Prog. Res.* 3, 147–150.

Malik, R. S. (1990). Prospects for *Brassica carinata* as an oilseed crop in India. *Exp. Agric.* 26, 125–129. doi: 10.1017/s0014479700015465

Mather, K., and Jinks, J. L. (1982). *Biometrical Genetics: the Study of Continuous Variation*, 3rd Edn. New York, NY: Chapman and Hall Publishers.

Mohammadi, R. (2016). Efficiency of yield-based drought tolerance indices to identify tolerant genotypes in durum wheat. *Euphytica* 211, 71–89.

Mwadzingeni, L., Shimelis, H., Dube, E., Laing, M. D., and Tsilo, T. J. (2016). Breeding wheat for drought tolerance: progress and technologies. *J. Integr. Agric.* 15, 935–943.

Navabi, Z. K., Stead, K. E., Pires, J. C., Xiong, Z., Sharpe, A. G., Parkin, I. A., et al. (2011). Analysis of B-genome chromosome introgression in interspecific hybrids of *Brassica napus* × *B. carinata*. *Genetics* 187, 659–673. doi: 10.1534/genetics.110.124925

Panhwar, S. A., Baloch, M. J., Jatoi, W. A., Veesar, N. F., and Majeedano, M. S. (2008). Combining ability estimates from line x tester mating design in upland cotton. *Proc. Pakistan Acad. Sci.* 45, 69–74.

Pour-Siahbidi, M. M., and Pour-Aboughadareh, A. (2013). Evaluation of grain yield and repeatability of drought tolerance indices for screening chickpea (*Cicer aritinum* L.) genotypes under rainfed conditions. *Iranian J. Genet. Plant Breed.* 2, 28–37.

Prakash, S. (1973). Non-homologous meiotic pairing in the A and B genomes of *Brassica*: its breeding significance in the production of variable amphidiploids. *Genet. Res. Camb.* 21, 133–137. doi: 10.1017/s0016672300013318

Raman, A., Verulkar, S., Mandal, N., Variar, M., Shukla, V., Dwivedi, J., et al. (2012). Drought yield index to select high yielding rice lines under different drought stress severities. *Rice* 5:31. doi: 10.1186/1939-8433-5-31

Raman, R., Qiu, Y., Coombes, N., Song, J., Kilian, A., and Raman, H. (2017). Molecular diversity analysis and genetic mapping of pod shatter resistance loci in *Brassica carinata* L. *Front. Plant Sci.* 8:1765. doi: 10.3389/fpls.2017.01765

Rao, N. V. P. R. G., and Gulati, S. C. (2001). Combining ability of gene action in F1 and F2 diallels of Indian mustard. *Crop Res. Hisar* 21, 72–76.

Rosielle, A. A., and Hamblin, J. (1981). Theoretical aspects of selection for yield in stress and non-stress environment 1. *Crop Sci.* 21, 943–946. doi: 10.2135/cropsci1981.0011183x002100060033x

Sayar, R., Khemira, H., and Kharrat, M. (2007). Inheritance of deeper root length and grain yield in half-diallel durum wheat (*Triticum durum*) crosses. *Ann. Appl. Biol.* 151, 213–220. doi: 10.1111/j.1744-7348.2007.00168.x

Seo, Y. J., Park, J. B., Cho, Y. J., Jung, C., Seo, H. S., Park, S. K., et al. (2010). Overexpression of the ethylene-responsive factor gene BrERF4 from *Brassica rapa* increases tolerance to salt and drought in *Arabidopsis* plants. *Mol. Cells* 30, 271–277. doi: 10.1007/s10059-010-0114-z

Sharma, D. K., and Kumar, A. (1989). Effect of water stress on plant water relations and yield of varieties of Indian mustard (*Brassica juncea*). *Indian J. Agric. Sci.* 59, 281–285.

Sharma, P., Sharma, H. O., and Rai, P. K. (2018). Strategies and technologies for enhancing rapeseed-mustard production and farmer income. *Indian Farming* 68, 44–48.

Sheikh, F. A., Bangha, S., and Banga, S. S. (2014). Broadening the genetic base of Abyssinian mustard (*Brassica carinata* A. Braun) through introgression of genes from related allotetraploid species. *Span. J. Agric. Res.* 12, 742–752. doi: 10.5424/sjar/2014123-5365

Shekhawat, K., Rathore, S. S., Premi, O. P., Kandpal, B. K., and Chauhan, J. S. (2012). Advances in agronomic management of Indian mustard (*Brassica juncea* (L.) Czernj. Cosson): an overview. *Int. J. Agron.* 2012:408284. doi: 10.1155/2012/408284

Shirinzadeh, A., Zarghami, R., Azghandi, A. V., Shiri, M. R., and Mirabdulbaghi, M. (2010). Evaluation of drought tolerance in mid and late mature corn hybrids using stress tolerance indices. *Asian J. Plant Sci.* 9, 67–73. doi: 10.3923/ajps.2010.67.73

Singh, A., Avtar, R., Singh, D., Sangwan, O., Thakral, N. K., Malik, V. S., et al. (2013). Combining ability analysis for seed yield and component traits in Indian mustard [*Brassica juncea* (L.) Czern & Coss.]. *Res. Plant Biol.* 3, 26–31.

Singh, K. H., and Chauhan, J. S. (2010). *Morphological Descriptor of Rapeseed Mustard Varieties*. Sewar: Directorate of Rapeseed Mustard Research.

Singh, K. H., Shakya, R., Singh, K. K., Thakur, A. K., Nanjundan, J., and Singh, D. (2015). "Genetic enhancement of *Brassica carinata* through interspecific hybridization and population improvement," in *Proceedings of the 14th International Rapeseed Congress*, Saskatoon.

Singh, M., Satendra, S. H., and Dixit, R. K. (2006). Combining ability of agronomic characters in Indian mustard (*Brassica juncea* (L.) Czern & and Coss). *J. Prog. Res.* 6, 69–72.

Singh, M., Singh, L., and Srivastava, S. B. L. (2010). Combining ability analysis in Indian mustard (*Brassica juncea* L. Czern & Coss). *J. Oilseed Brassica* 1, 23–27.

Singh, N., Yadava, D. K., Sujata, V., Singh, R., Giri, S. C., Dass, B., et al. (2015). Combining ability and heterobeltiosis for yield and yield contributing traits in high quality oil Indian mustard (*Brassica juncea*) genotypes. *Indian J. Agric. Sci.* 85, 498–503.

Srivastava, S. B. L., Rajshekhar, Singh, M., and Rao, M. (2009). Combining ability analysis for yield and contributing characters in Indian mustard (*Brassica juncea* (L.) Czern & and Coss.). *J. Oil Seeds Res.* 26, 58–61. doi: 10.9734/cjast/2020/v39i2030808

Teklewold, A., and Becker, H. C. (2005). Heterosis and combining ability in a diallel cross of Ethiopian mustard inbred lines. *Crop Sci.* 45, 2629–2635. doi: 10.2135/cropsci2005.0085

Thakur, A. K., Singh, K. H., Parmar, N., Sharma, D., Mishra, D. C., Singh, L., et al. (2021). Population structure and genetic diversity as revealed by SSR markers in Ethiopian mustard (*Brassica carinata* A. Braun): a potential edible and industrially important oilseed crop. *Genet. Resour. Crop Evol.* 68, 321–333. doi: 10.1007/s10722-020-00988-3

Thakur, A. K., Singh, K. H., Sharma, D., Parmar, N., and Nanjundan, J. (2019). Breeding and genomics interventions in Ethiopian mustard (*Brassica carinata* A. Braun) improvement–A mini review. *S. Afr. J. Bot.* 125, 457–465. doi: 10.1016/j.sajb.2019.08.002

Verma, O. P., Rashmi, Y., Kumar, K., Ranjeet, S., and Maurya, K. N. (2011). Combining ability and heterosis for seed yield and its components in Indian mustard (*Brassica juncea* L. Czern &,Coss.). *Plant Arch.* 11, 863–865.

Warwick, S. I. (ed.) (1993). "Wild species in the tribe *Brassicaceae* (Cruciferae) as sources of agronomic traits," in *Guide to the Wild Germplasm of Brassica and Allied Crops*, (Ottawa: Center for Land and Biological Resources Research Branch, Agriculture Canada).

Wei, Z., Wang, M., Chang, S., Wu, C., Liu, P., Meng, J., et al. (2016). Introgressing subgenome components from *Brassica rapa* and *B. carinata* to *B. juncea* for broadening its genetic base and exploring intersubgenomic heterosis. *Front. Plant Sci.* 7:1677. doi: 10.3389/fpls.2016.01677

Yadav, Y. P., Prakash, R., Singh, R., Singh, R. K., and Yadava, J. S. (2005). Genetics of yield and its component characters in Indian mustard, *Brassica juncea* (L.) Czern & Coss under rainfed conditions. *J. Oilseeds Res.* 22:255.

Yadava, D. K., Singh, N., Sujata, V., Singh, R., Singh, S., Giri, S. C., et al. (2012). Combining ability and heterobeltiosis for yield and yield attributing traits in Indian mustard (*Brassica juncea*). *Indian J. Agric. Sci.* 82, 563–570.

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 24 February 2022
doi: 10.3389/fpls.2022.844348

Identification of 5P Chromosomes in Wheat-*Agropyron cristatum* Addition Line and Analysis of Its Effect on Homologous Pairing of Wheat Chromosomes

Cuili Pan[1,2†], Qingfeng Li[1,2†], Haiming Han[1], Jinpeng Zhang[1], Shenghui Zhou[1], Xinming Yang[1], Xiuquan Li[1], Lihui Li[1] and Weihua Liu[1]**

[1] National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China, [2] School of Agriculture, Ningxia University, Yinchuan, China

OPEN ACCESS

Edited by:
Dayun Tao,
Yunnan Academy of Agricultural Sciences, China

Reviewed by:
Mu Jun Yang,
Yunnan Academy of Agricultural Sciences, China
Liqiang Song,
Agricultural University of Hebei, China

***Correspondence:**
Weihua Liu
liuweihuai@caas.cn
Lihui Li
lilihui@caas.cn

[†] These authors have contributed equally to this work

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

Received: 28 December 2021
Accepted: 27 January 2022
Published: 24 February 2022

Citation:
Pan C, Li Q, Han H, Zhang J, Zhou S, Yang X, Li X, Li L and Liu W (2022) Identification of 5P Chromosomes in Wheat-Agropyron cristatum Addition Line and Analysis of Its Effect on Homologous Pairing of Wheat Chromosomes. Front. Plant Sci. 13:844348. doi: 10.3389/fpls.2022.844348

As an important wheat wild relative, the P genome of *Agropyron cristatum* (L.) Gaertn. ($2n = 4x = 28$) is very valuable for wheat improvement. A complete set of wheat-*A. cristatum* disomic addition lines is the basis for studying the genetic behavior of alien homoeologous chromosomes and exploring and utilizing the excellent genes. In this study, a wheat-*A. cristatum* derivative II-11-1 was proven to contain a pair of 5P chromosomes and a pair of 2P chromosomes with 42 wheat chromosomes by analyzing the fluorescence *in situ* hybridization (FISH) and expressed sequence tag (EST) markers. Additionally, cytological identification and field investigation showed that the 5P chromosome can weaken the homologous pairing of wheat chromosomes and promote the pairing between homoeologous chromosomes. This provides new materials for studying the mechanism of the alien gene affecting the homologous chromosome pairing and promoting the homoeologous pairing of wheat. In addition, chromosomal structural variants have been identified in the progeny of II-11-1. Therefore, the novel 5P addition line might be used as an important genetic material to widen the genetic resources of wheat.

Keywords: ***Agropyron cristatum*, 5P chromosomes, molecular cytogenetics, homoeologous pairing, addition line**

INTRODUCTION

Wheat (*Triticum aestivum* L., $2n = 42$, AABBDD), as one of the most important food crops in the world, plays a significant role in ensuring food production and security. The homogenization of wheat varieties in major regions has narrowed their genetic background, which became the main bottleneck of breeding. Fortunately, there are plenty of beneficial genes in the wheat wild relatives that could be exploited and utilized for wheat improvement (Dong, 2000). The success of distant hybridization made it possible to transfer alien excellent genes to common wheat for creating new germplasms. Wheat disomic addition lines contain a pair of alien chromosomes in the wheat background, which is an important tool and bridge material for transferring alien excellent genes. Furthermore, they are favorable materials to identify the relationship between alien and wheat chromosomes, which could also be used for gene mapping. At present, most of the wheat wild relatives have been successfully hybridized with wheat, such as *Aegilops* (Gong et al., 2017; Du et al., 2019; Liu et al., 2019; Yi et al., 2019), *Secale cereal* L. (Li et al., 2016a, 2020; Schneider et al., 2016;

An et al., 2019), *Hordeum vulgare* L. (Szakacs and Molnal-Lang, 2010; Fang et al., 2014), *Haynaldia villosa* (Zhang et al., 2015a, 2018a,b), *Leymus racemosus* (Yang et al., 2017; Zhang et al., 2017a), and *Elytrigia repens* (Liu et al., 2017). Meanwhile, a large number of disomic addition lines and substitution lines were identified, which were used to create translocation lines and introgression lines containing desirable genes. Some introgression lines and translocation lines are widely used in wheat production. For instance, the cultivated variety Xiaoyan 6 was bred from the hybridization of wheat and *Thinopyrum* (Ma et al., 2018), and the crucial translocation lines T1RS·1BL and T6VS·6AL were from the hybridization of wheat and rye (Su et al., 2006) and *H. villosa* (Jiang et al., 2014; Gao et al., 2018), respectively.

Agropyron cristatum L. Gaertn. (2*n* = 28, PPPP) is an important wheat wild relative. It grows in arid grassland, hillside, hill, and desert and contains many desirable traits for wheat improvement, such as resistance to wheat leaf rust, powdery mildew, barley yellow dwarf, and wheat streak mosaic viruses (Dewey, 1984; Sharma et al., 1984; Ochoa et al., 2014) and tolerance to drought and low temperature (Limin and Fowler, 1987; Asay and Johnson, 1990; Dong et al., 1992), as well as with multiple spikelets and florets, small flag leaves, fertile tiller number and strong and tough stem (Dewey, 1984; Wu et al., 2006; Han et al., 2014; Jiang et al., 2018). The acquisition of wheat-*A. cristatum* disomic addition lines made it possible to utilize these desirable genes to improve wheat variety. This study has been committed to the hybridization of wheat and *A. cristatum* for a long time and has created a series of wheat-*A. cristatum* addition, translocation, and deletion lines successfully (Li and Dong, 1991, 1993; Li et al., 1995, 1997, 1998, 2016b; Luan et al., 2010; Song et al., 2013; Ye et al., 2015; Lu et al., 2016; Zhang et al., 2019). So far, wheat-*A. cristatum* 1P, 2P, 3P, 4P, 6P, and 7P disomic addition lines have been successfully created, and many excellent genes were located in specific chromosomes and transmitted into wheat (Wu et al., 2006; Han et al., 2014; Li et al., 2016a; Lu et al., 2016; Pan et al., 2017; Chen et al., 2018; Zhou et al., 2018). For instance, it has been found that *A. cristatum* 6P addition line carried gene clusters related to yield, such as multiple florets and grains per spike, and the 2P addition line possessed gene clusters related to disease resistance including powdery mildew, leaf rust, and stripe rust (Wu et al., 2006; Han et al., 2014; Li et al., 2016a). These genes were further mapped by creating translocation lines and introgression lines and developing specific markers for P chromosomes (Liu et al., 2010; Dai et al., 2012; Zhang et al., 2015b,c; Han et al., 2017, 2019; Zhou et al., 2018). However, wheat-*A. cristatum* 5P disomic addition line has not been obtained.

In this study, a wheat-*A. cristatum*-derived line II-11-1 with four *A. cristatum* chromosomes was used as a basic material for backcross and self-cross with recipient parent Fukuhokomugi (Fukuho). The purpose of this study was (1) to analyze the chromosome constitution of wheat-*A. cristatum*-derived line II-11-1, (2) to obtain wheat-*A. cristatum* 5P addition line, and (3) to analyze the effects of *A. cristatum* 5P chromosome on the chromosomes pairing. The obtained wheat-*A. cristatum* 5P addition line provided basic materials for further systematic study on genetic variation of wheat distant hybrid.

MATERIALS AND METHODS

Materials

The plant materials included *T. aestivum* cv. Fukuho (2*n* = 6*x* = 42, AABBDD), *A. cristatum* accession Z559 (2*n* = 4*x* = 28, PPPP), wheat-*A. cristatum*-derived material II-11-1 (2*n* = 46) and other wheat-*A. cristatum* homoeologous group addition lines: II-3-1a (1P) (Pan et al., 2017), II-9-3 (2P) (Li et al., 2016a), 7365 (3P) (Zhou et al., 2018), II-21-2 (4P) (Liu et al., 2010), 4844-12 (6P) (Wu et al., 2006), and II-5-1 (7P) (Lu et al., 2016). All the above materials were provided by the Center of Crop Germplasm Resources Research at the Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China.

Observation on the Mitosis of Root Tip Cells and the Meiosis of Pollen Mother Cells (PMC)

The mitotic metaphase of root tip cells and the meiotic metaphase I of PMC of Fukuho, wheat-*A. cristatum*-derived material II-11-1, and the newly obtained 5P addition line were observed. Genomic *in situ* hybridization (GISH) was performed as described by Cuadrado et al. (2000), and the meiotic metaphase I of PMC was observed following the method of Jauhar and Peterson (2006).

Fluorescence *in situ* Hybridization

The genomic DNA of *A. cristatum* accession Z559 and common wheat Fukuho were extracted using the cetyl trimethyl ammonium bromide (CTAB) method (Dellaporta et al., 1983). The P chromosome repetitive sequences, pAcTRT1 and pAcpCR2, were used as probes to identify the homologous groups of *A. cristatum* in II-11-1 and II-11-1b using the method described by Han et al. (2019). The barley clone pHvG38 contains the GAA-satellite sequence (Pedersen and Langridge, 1997), and the clone pAsl contains a 1 kb DNA repetitive sequence from *Aegilops tauschii* (Rayburn and Gill, 1986). The combination of pAs1 and pHvG38 allowed the discrimination of the three genomes in wheat. The procedure of fluorescence *in situ* hybridization (FISH) was carried out as described by Han et al. (2004) and Liu et al. (2010). Images were captured using an OLYMPUS AX80 fluorescence microscope (Olympus Corporation, Tokyo, Japan) equipped with a charge-coupled device (CCD) camera (Diagnostic Institute, Inc., Sterling Height, MI, USA) and then were processed with Photoshop CS 3.0.

Molecular Marker Analysis

A total of 236 markers were used to identify the alien P chromatin and to determine its homoeologous group (**Supplementary Table 1**), of which 160 markers were described by Zhang et al. (2017b) and 76 markers were described by Li et al. (2016b). The PCR amplification procedure was performed as described by

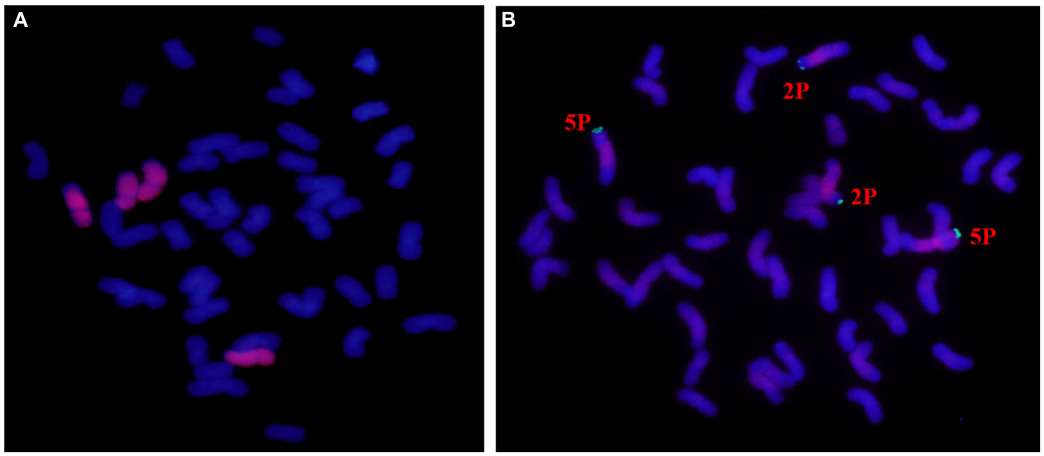


FIGURE 1 | Mitosis genomic *in situ* hybridization (GISH)/fluorescence *in situ* hybridization (FISH) identification of II-11-1. **(A)** The whole-genome DNA probe of *Agropyron cristatum* was labeled as a red signal, and wheat chromosomes were restained as blue by DAPI. **(B)** The probes pAcTRT1 and pAcpCR2 were labeled as red and green, respectively, and wheat chromosomes were restained as blue by DAPI.

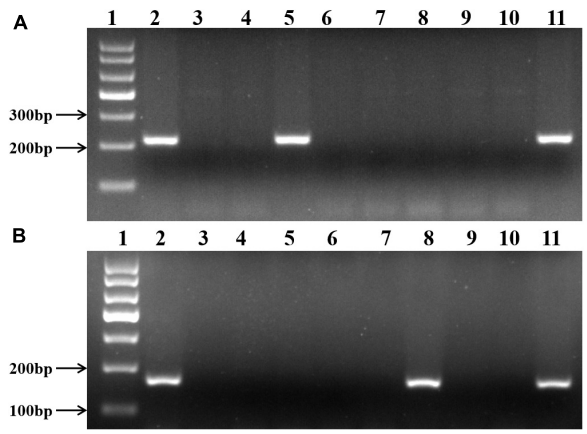


FIGURE 2 | Molecular marker identification of II-11-1. **(A,B)** 2P chromosome-specific marker, Agc3725, and 5P chromosome-specific marker, Agc737, respectively. 1: Puc19 DNA MSP/I HPA II marker; 2: *A. cristatum* Z559; 3: common wheat Fukuho; 4: Wheat-*A. cristatum* 1P addition line II-3-1a; 5: Wheat-*A. cristatum* 2P addition line II-9-3; 6: Wheat-*A. cristatum* 3P addition line 7365; 7: Wheat-*A. cristatum* 4P addition line II-21-2; 8: Wheat-*A. cristatum* 5P addition line II-11-1b; 9: Wheat-*A. cristatum* 6P addition line 4844-12; 10: Wheat-*A. cristatum* 7P addition line II-5-1; 11: Wheat-*A. cristatum* derivation line II-11-1.

Luan et al. (2010). The amplified product was verified using 6% polyacrylamide gel electrophoresis (PAGE).

Evaluation of the Agronomic Traits

All the tested materials were sown in a randomized complete block design with three replicates in the fields at Xinxiang (35°18′13.71″N, 113°55′15.05″E, Henan Province, China) during 2016–2017, 2017–2018, and 2018–2019 growing seasons. A total of 20 grains were evenly planted in 2.0 m rows spaced 0.3 m apart (Zhang et al., 2019). The agronomic traits were measured and quantified including grain number, spikelet number and kernel number per spikelet, thousand-grain weight, and effective tiller number. The Statistical Analysis System (version 9.2, SAS Institute, Cary, NC, USA) software was used for statistical analysis.

RESULTS AND ANALYSIS

The Chromosome Composition Analysis of Wheat-*A. cristatum* II-11-1

In the population composed of 50 individuals of II-11-1, 31 plants containing 42 wheat chromosomes and 4 *A. cristatum* chromosomes were identified by mitosis observation and GISH detection (**Figure 1A**). Additionally, pAcTRT1 and pAcpCR2 were used as probes to identify the additional chromosomes of *A. cristatum* in II-11-1. As shown in **Figure 1B**, the additional *A. cristatum* chromosomes were a pair of 2P and a pair of 5P according to the signal characteristics. So, the derivative II-11-1 ($2n = 46$) was preliminarily determined to be a wheat-*A. cristatum* 2P and 5P disomic addition line.

Identification of II-11-1 With Expressed Sequence Tag (EST)-STS Markers

To further confirm the identity of *A. cristatum* chromosomes in II-11-1, the EST-STS markers specific to the 2 and 5 homoeologous groups were employed to identify II-11-1 and wheat-*A. cristatum* disomic addition lines (1P, 2P, 3P 4P, 6P, and 7P). Results showed that 36 pairs of 2P chromosome-specific primers amplified specific bands for the wheat-*A. cristatum* 2P addition line and II-11-1 (**Figure 2A**) and 78 pairs of 5P chromosome-specific markers amplified specifically in II-11-1 and II-11-1b (**Figure 2B** and **Supplementary Table 2**). Therefore, it was further confirmed that II-11-1 contains 2P and 5P chromosomes, which could be used as a basic material

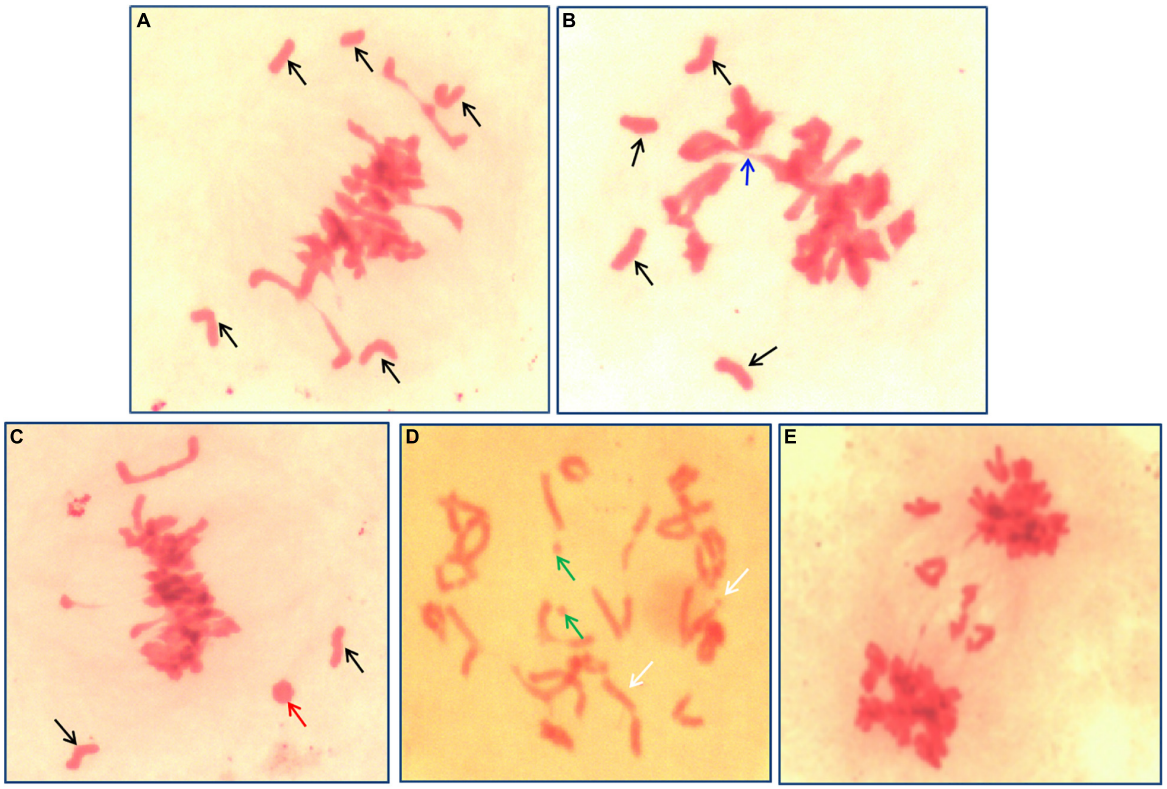


FIGURE 3 | Meiosis identification of II-11-1. **(A–D)** Meiosis metaphase I; **(E)** meiosis anaphase I, chromosome bridges, and lagging chromosomes. Black arrows refer to univalents, blue arrows to multivalents, green arrows to chromosome fragments, and red arrows to ring chromosomes.

TABLE 1 | Pollen mother cells (PMC) meiosis metaphase I chromosome configuration of wheat-*Agropyron cristatum* derivatives II-11-1 and 5P addition line II-11-1b.

Materials	2n	Number of cells	Chromosome configuration						
			Univalents	Bivalent			Trivalent	Quadrivalent	Fragment
				Rod	Ring	Total			
II-11-1	46	74	3.95 (0–8)	3.81 (1–8)	16.85 (11–19)	20.66 (18–22)	0.18 (0–1)	0.08 (0–1)	0.18 (0–2)
II-11-1b	44	71	2.08 (0–6)	2.72 (0–8)	17.56 (11–22)	20.28 (18–22)	0.23 (0–1)	0.13 (0–1)	0.14 (0–1)
Fukuho	42	50	0.06 (0–2)	1.95 (1–4)	19.32 (17–21)	20.97 (20–21)	–	–	–

for the separation and identification of wheat-*A. cristatum* 5P addition line.

Chromosome Behavior Analysis of II-11-1 During Meiosis

Meiosis pairing was observed in PMCs to analyze the chromosome behavior of II-11-1 during the generation of the gamete. Results indicated that there existed univalent and multivalent chromosomes, chromosome fragments, and ring chromosomes at metaphase (**Figures 3A–D**). Meanwhile, chromosome lagging and chromosome bridge were found at anaphase (**Figure 3E**). The chromosome configuration statistics of II-11-1 at metaphase showed that the average univalents, rod bivalents, ring bivalents, trivalents, quadrivalents and chromosome fragments were 3.95, 3.81, 16.85, 0.16, 0.08, and 0.18, respectively (**Table 1**). The above results indicated that the behaviors of chromosome pairing were abnormal during meiosis in II-11-1.

Identification of the II-11-1 Progenies by Molecular Markers

The markers specific to 2P and 5P chromosomes were used to detect the BC_1F_2 population of II-11-1 with Fukuho as a recurrent parent. According to the results, 212 BC_1F_2 individuals were divided into four types, namely, type I, type II, type III, and type IV. There were 47 plants (21.76%) with 5P chromosome only in type I, 54 plants (25.00%) with 2P chromosome only in type II, 66 plants (30.56%) with both 2P and 5P chromosomes in type III, and 45 plants (20.83%) without P chromosomes in type IV. Among them, the type I individuals with the 5P chromosome

TABLE 2 | Chromosomal constitutions of 47 plants derived from plant II-11-1b.

No. of plants	No. of chromosomes	No. of plants	Constitution	Example
15	2n = 44	12	42W + 2A	**Figure 4A**
		2	41W + $1t^{W}$ + 2A	**Figure 4B**
		1	42W + 1A + $1t^{A}$	**Figure 4C**
26	2n = 43	22	42W + 1A	**Figure 4D**
		2	41W + 2A	**Figure 4E**
		1	41W + 1W.A + 1A	**Figure 4F**
		1	41W + 1W-A + 1A	**Figure 4G**
5	2n = 42	3	40W + 2A	**Figure 4H**
		2	41W + 1A	**Figure 4I**
1	2n = 45	1	42W + 2A + $1t^{W}$	**Figure 4J**

W, wheat chromosomes; A, A. cristatum chromosomes; t, telosome; W.A, Robertsonian translocation; W-A, non-Robertsonian translocation.

only provided the candidate materials for further identifying the wheat-*A. cristatum* 5P addition line.

Identification and Molecular Cytological Detection of Wheat-*A. cristatum* 5P Addition Line

The chromosomal composition of 47 candidate individual plants in type I was further identified using GISH using the root tips (**Table 2**). Statistical results showed that there were 15 plants with 44 chromosomes, of which 12 plants were composed of 42 wheat chromosomes and 2 *A. cristatum* chromosomes (**Figure 4A**); 2 plants composed of 41 wheat chromosomes, 1 wheat telomere, and 2 *A. cristatum* chromosomes (**Figure 4B**) and 1 plant consisted of 42 wheat chromosomes, 1 *A. cristatum* telomere, and 1 *A. cristatum* chromosome (**Figure 4C**). There were 26 plants with 43 chromosomes, of which 22 plants were composed of 42 wheat chromosomes and 1 *A. cristatum* chromosome (**Figure 4D**); 2 plants contained 41 wheat chromosomes and 2 A. cristatum chromosomes (**Figure 4E**); 1 plant was composed of 41 wheat chromosomes, 1 whole arm translocation, and 1 A. cristatum chromosome (**Figure 4F**); 1 plant consisted of 41 wheat chromosomes, 1 small alien segment translocation, and 1 A. cristatum chromosome (**Figure 4G**). There were 5 plants with 42 chromosomes, of which 3 plants consisted of 40 wheat chromosomes and 2 A. cristatum chromosomes (**Figure 4H**) and 2 plants consisted of 41 wheat chromosomes and 1 A. cristatum chromosome (**Figure 4I**). One plant consists of 45 chromosomes with 42 wheat chromosomes, 1 wheat telomere, and 2 A. cristatum chromosomes (**Figure 4J**). Combined with the results of FISH identification, we found that 6D-7A translocation occurred in wheat chromosomes (**Figure 4K Supplementary Figure 1A**). Among the 47 individuals of type I, the 12 plants with 42 wheat and 2 A. cristatum chromosomes were further identified using FISH with the pAcTRT1 and pAcpCR2 probes.

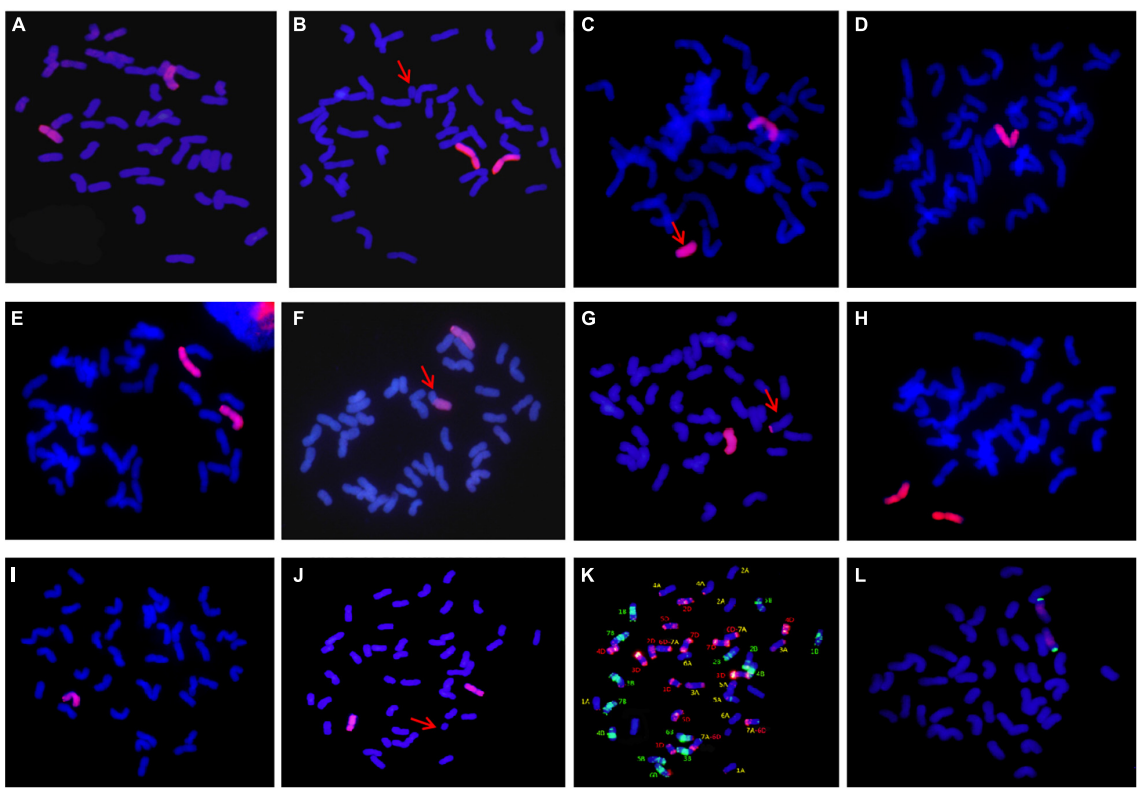


FIGURE 4 | Mitosis GISH/FISH identification of II-11-1b. **(A–J)** *A. cristatum* genome DNA probe was labeled as a red signal; **(K)** pAs1 and pHvG38 were labeled as red and green, respectively; **(L)** pAcTRT1 and pAcpCR2 were labeled as red and green, respectively, and wheat chromosomes were restained as blue by DAPI.

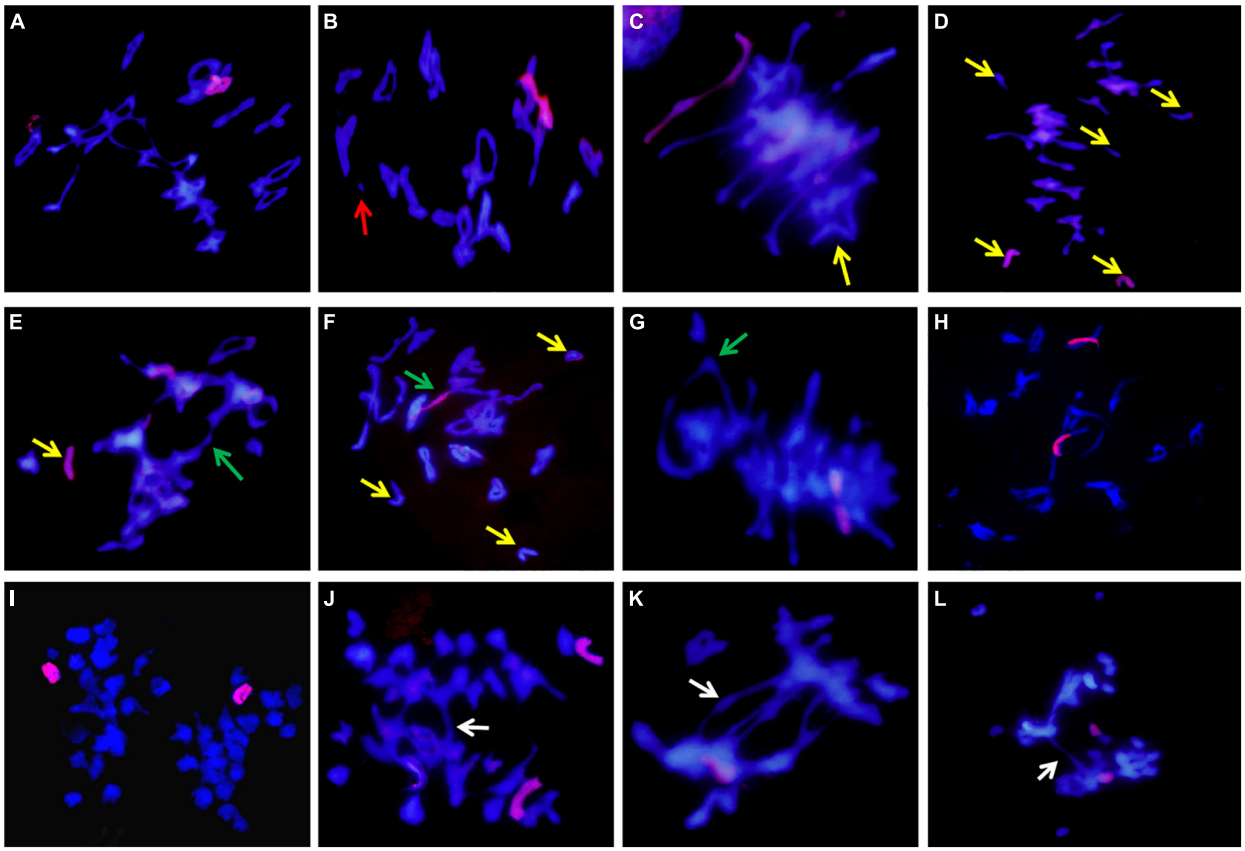


FIGURE 5 | Meiosis GISH identification of II-11-1b. **(A–L)** *A. cristatum* genome DNA probe was labeled as a red signal; wheat chromosome restained blue by DAPI. **(A–H)** Meiotic metaphase I; **(I–L)** anaphase I. The red arrows refer to chromosome fragments, the yellow arrows refer to univalents, the green arrows refer to polyvalents, and the white arrows refer to chromosome adhesion or chromosome bridge.

It was also revealed that the pair of A. cristatum chromosomes was 5P (**Figure 4L** and **Supplementary Figure 1B**). Finally, they were identified as a novel wheat-A. cristatum 5P addition line and named as II-11-1b.

Meiosis Abnormality and 5P Chromosome Functional Analysis

The meiosis metaphase I chromosome configuration of II-11-1b was observed and counted (**Table 1**). Except for the normal bivalent formation at metaphase (**Figure 5A**) and normal separation at anaphase (**Figure 5I**), chromosome behavioral abnormalities were also identified (**Figures 5B–H,J–L**). For example, there were chromosome fragments (**Figure 5B**), univalents of wheat chromosomes (**Figure 5C**) and *A. cristatum* 5P chromosomes (**Figure 5D**), multivalents, including trivalent and tetravalent, formed by wheat chromosomes only or wheat and *A. cristatum* 5P chromosomes (**Figures 5E–G**). In addition, the chromosome bridge and the division desynchrony with different numbers of lagging chromosomes were also observed (**Figures 5H–L**). Statistical analysis of the meiosis metaphase I showed that the 5P addition line II-11-1b had significantly higher frequencies of univalents and multivalents compared with the parent Fukuho (**Table 1**). The presence of multivalents indicates the existence of some homoeologous chromosome synapses or chromosomal rearrangement. At the anaphase of meiosis, the loss of univalents and the abnormal segregation of multivalents can cause changes in the chromosomal composition of their progenies. Thus, it is predicted that the chromosome 5P of *A. cristatum* has the function of inducing chromosome breakage, promoting synapses of the homoeologous chromosome.

Evaluation of Agronomic Traits

According to the results of observation, statistics, and evaluation in three growing seasons, II-11-1b progenies with 5P chromosomes showed segregation in agronomic traits (Type 1 in **Figure 6**), and progenies without 5P chromosomes gradually stabilized (Types 2-n in **Figure 6**). For example, different types of spike traits were detected in II-11-1b progenies with 5P chromosomes (**Figure 7**). It showed that 5P chromosomes resulted in a decrease in fertility, which was reflected in the significant decrease in the kernel number per spikelet and per spike and the significant increase in the number of sterile florets ($p < 0.05$) in 5P positive plants compared with negative plants (**Table 3**). Meanwhile, the variation (statistical standard deviation) in tiller number, plant height, spike length,

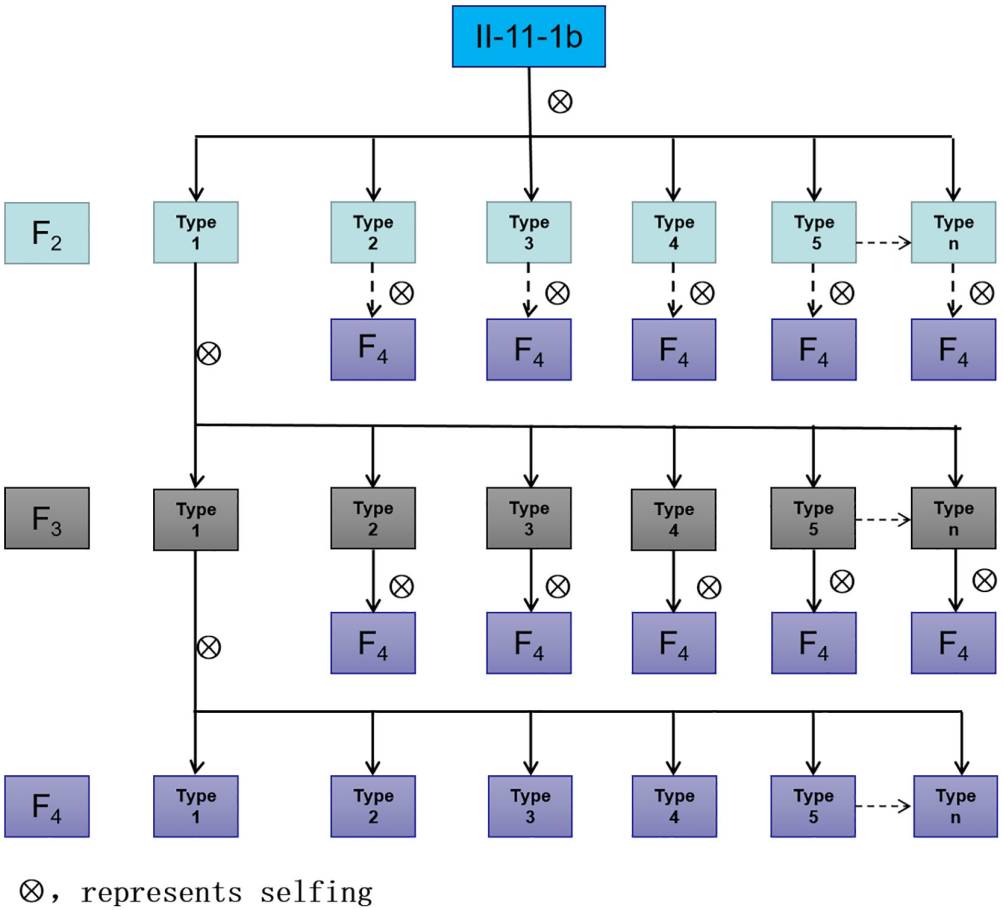


FIGURE 6 | The pattern diagram of segregation of self-crossing progenies of II-11-1b.

etc., of 5P positive plants were higher than those of negative plants (**Table 3**). Combined with the results of traits and molecular cytology, it is speculated that the presence of 5P chromosomes might influence the genetic stability by regulating the homologous and homoeologous chromosome pairing behavior during meiosis, thus leading to fertility decrease and trait separation in the progenies.

FIGURE 7 | Segregation of spike traits was detected in the F_2 population. Fukuho: parent; F_2: B83, B181, B49, B127, B173, B143, B24, B156, B132, B89, B17, B26, B25, B96, B180, B93.

DISCUSSION

The Significance of Wheat-*Agropyron cristatum* 5P Addition Line

Wheat wild relatives as gene resource pools provide abundant genetic resources for wheat improvement. Lots of alien genes of wild relatives have been introgressed into wheat by distant hybridization, which shows great potential in improving yield and quality, disease resistance, and stress resistance. Among the progenies of distant hybridization, wheat-alien disomic addition lines played a very important role in the transfer of alien excellent genes. A complete set of wheat-alien disomic addition lines is an important genetic material for studying the genetic relationship, origin and evolution of species, gene expression, and interaction of chromosomes. To fully explore the desirable genes in *A. cristatum*, it is necessary to establish a complete

TABLE 3 | Agronomic traits of II-11-1b progenies.

Year	Type	Tillering number	Plant height (cm)	Spike length (cm)	Spikelet number	Spikelet density	Under-spike internodal ratio	Kernel number per spikelet	Kernel number per spike	Number of sterile floret
2016-2017	5P +	12.90b ± 9.77 (3-37)	80.84b ± 9.64 (62.0-98.0)	8.60a ± 1.13 (4.7-9.5)	18.29a ± 2.17 (15-23)	20.19a ± 4.18 (16.09-38.29)	0.31b ± 0.08 (0.15-0.39)	3.69b ± 0.94 (1-4)	43.25b ± 12.31 (0-68)	9.53a ± 5.75 (1-24)
	5P –	17.96a ± 3.59 (9-28)	85.99a ± 4.91 (79.0-90.5)	8.92a ± 0.68 (7.1-11.0)	18.49a ± 1.81 (16-21)	19.79a ± 0.91 (17.78-24.84)	0.35a ± 0.03 (0.29-0.40)	4.49a ± 0.52 (3-5)	56.25a ± 7.61 (48-80)	4.50b ± 0.49 (0-6)
2017-2018	5P +	10.23b ± 10.02 (3-40)	86.75a ± 8.98 (59.0-98.5)	8.81a ± 0.98 (7.1-12.0)	17.61b ± 2.45 (14-23)	18.85b ± 4.73 (15.79-35.01)	0.33b ± 0.09 (0.16-0.39)	3.78b ± 0.88 (1-4)	43.15b ± 12.53 (0-69)	7.53a ± 6.31 (1-29)
	5P –	13.63a ± 3.43 (7-27)	90.35a ± 5.03 (86.0-107.0)	8.96a ± 0.57 (7.8-10.5)	18.91a ± 1.79 (15-20)	19.99a ± 0.97 (17.21-21.24)	0.37a ± 0.03 (0.29-0.40)	4.36a ± 0.49 (3-5)	58.78a ± 6.77 (44-82)	3.38b ± 0.41 (1-6)
2018-2019	5P +	12.63b ± 9.43 (4-35)	82.34a ± 9.35 (61-97.5)	8.84a ± 1.07 (6.7-10.4)	18.04a ± 2.77 (14-24)	20.07a ± 4.77 (16.53-37.65)	0.31b ± 0.07 (0.15-0.37)	3.65b ± 0.97 (1-4)	41.04b ± 12.17 (3-77)	8.33a ± 4.55 (1-24)
	5P –	16.57a ± 9.77 (6-28)	85.26a ± 5.77 (69-89.5)	9.20a ± 0.77 (6.8-11.5)	18.25a ± 1.76 (16-21)	20.22a ± 0.87 (17.83-23.37)	0.33a ± 0.03 (0.28-0.38)	4.17a ± 9.77 (3-5)	53.01a ± 5.67 (42-77)	3.58b ± 0.37 (0-6)

5P+ represents plants with 5P chromosomes and 5P- represents plants without 5P chromosomes. Means followed by different letters ("a" and "b") indicate the significant differences at 0.05 level.

wheat-*A. cristatum* addition line. In the previous studies, all the wheat-*A. cristatum* addition lines except 5P have been obtained. In this study, we identified the presence of 5P chromosomes in II-11-1. By further backcrossing and selfing, the wheat-*A. cristatum* 5P addition line was identified. So far, we have obtained a complete set of wheat-*A. cristatum* addition lines (1P–7P), which provided materials for the systematic study of the excellent exogenous genes from *A. cristatum*. However, we found that as long as there are 5P chromosomes, there will always be meiosis abnormality and trait separation in the progenies (**Table 3**). This increases the difficulty in the identification and preservation of the wheat-*A. cristatum* 5P addition line and also explains its unavailability for a long time. In addition, the 5P addition line is an excellent material for studying the abnormal behavior of chromosomes during meiosis.

The Role and Value of 5P Chromosome: Induced Homoeologous Recombination

To date, some plentiful wild relatives have been successfully hybridized with wheat. However, due to the presence of genes that control homologous chromosome pairing (such as the *Ph1* gene on the 5B chromosome), chromosome pairing is difficult to occur between wheat and its wild relatives. This mechanism not only ensures the genetic stability of wheat but also restricts the chromosome recombination and the exogenous gene transfer between wheat and its wild relatives, thus hindering the application of excellent genes in breeding. The discovery of genes that inhibit homologous chromosome pairing has laid a foundation for the transfer of exogenous excellent genes and the creation of new germplasm resources (Ceoloni and Donini, 1993). At present, the *Ph* suppressor gene has been found in several wild relatives of wheat. For example, the 5Mg chromosome of *Aegilops geniculata* contained genes that promote synapsis and crossing in prophase I of meiosis in wheat (Tiwari et al., 2015; Koo et al., 2017). *Aegilops speltoides* 5S chromosome contained a QTL (*QPh.ucd-5S*) that could increase homeologous chromosome pairing and regulate recombination between homologous chromosomes in *T. aestivum* × *Ae. speltoides* hybrids (Dvorak et al., 2006). Genes that affect chromosome pairing during meiosis were also found in the 5U chromosome of *Aegilops umbrella* (Riley et al., 1973), $4M^g$ chromosome of *A. geniculata* (Kynast et al., 2000), and 3S chromosome of *A. speltoides Tausch* (Li et al., 2019).

The offspring of distant hybridization might exhibit abnormal chromosome pairing behavior during meiosis. Observing the chromosomal configurations during meiosis of the alien addition line is a crucial way to analyze its stability. Based on the previous studies, it is speculated that there may also exist genes affecting homologous pairing in *A. cristatum*. For instance, by evaluating the *Ph*-suppressing effect of P chromosomes (1P–6P) and deletion *ph1b* of the *Ph1* gene, it showed that they all displayed a significantly higher level of homoeologous pairing than the control except for 2PL and 2PS, but allosyndetic associations between P and ABD genomes were very rare, which had no prospect in the transfer of alien genes (Jubault et al., 2006). The results of the 5P addition line were inconsistent with this study, which may be due to the different sources of *A. cristatum*. In this study, the theoretical chromosomal configurations of II-11-1 and II-11-1b should be 21 wheat chromosomes bivalents plus 2 *A. cristatum* bivalents and 21 wheat chromosomes bivalents plus 1 *A. cristatum* bivalent, respectively. However, the statistical results showed that the ratio of univalents in II-11-1 and II-11-1b was 9.85 and 2.08, respectively, trivalents was 0.18 and 0.23, respectively, and quadrivalents was 0.08 and 0.13 respectively (**Table 1**). Meanwhile, the wheat Fukuho had a significantly low univalents ratio of 0.06, and there is no trivalent or quadrivalent. In the anaphase II and telophase II of meiosis, there also existed an abnormal phenomenon including fragments, bridges, lagging chromosomes, and the chromosome adhesions in II-11-1 and II-11-1b. In the progenies of II-11-1b, wheat-*A. cristatum* 5P translocation lines were identified (**Figures 4F–G**). Thus, it is supposed that the 5P chromosome might play a comprehensive and complex role in regulating chromosomal behavior during meiosis, including the inhibition of the *Ph* gene to promote the synapses of the homoeologous chromosome, and the function similar to the gametocidal chromosome, which

induces chromosome breakage and recombination and promotes the formation of chromosomal translocation in the progenies.

Meiotic homoeologous recombination could facilitate gene introgression to diversify the wheat genome for germplasm development. Therefore, the wheat-*A. cristatum* 5P addition line II-11-1b is a potential and valuable material for gene introgression and gene mapping based on recombination between homoeologous chromosomes in wheat. The studies of the 5P chromosome further enhance our understanding of the wheat genome and its homoeologous counterparts *A. cristatum* and expand the genetic variability of the wheat genome.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

WL and LL conceived the research. CP and QL performed the research and wrote the manuscript. CP, QL, HH, JZ, SZ, XY, and XL participated in the preparation of both the reagents and materials. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the National Key Research and Development Program of China (No. 2016YFD0100102), the National Natural Science Foundation of China (No. 31801349), and the Natural Science Foundation of Ningxia Province (No. 2018AAC03036).

ACKNOWLEDGMENTS

We are grateful to Hongjie Li (The Institute of Crop Sciences, Chinese Academy of Agricultural Sciences) and Ainong Gao (The Institute of Crop Sciences, Chinese Academy of Agricultural Sciences) for their useful advice and English language editing of the manuscript. We also thank AJE (https://www.aje.cn/#) for its linguistic assistance and scientific consultation 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/fpls.2022.844348/full#supplementary-material

REFERENCES

An, D. G., Ma, P. T., Zheng, Q., Fu, S. L., Li, L. H., Han, F. P., et al. (2019). Development and molecular cytogenetic identification of a new wheat-rye 4R chromosome disomic addition line with resistances to powdery mildew, stripe rust and sharp eyespot. *Theor. Appl. Genet.* 132, 257–272. doi: 10.1007/s00122-018-3214-3

Asay, K. H., and Johnson, D. A. (1990). Genetic variances for forage yield in crested wheatgrass at six levels of irrigation. *Crop Sci.* 30, 79–82. doi: 10.2135/cropsci1990.0011183x003000010018x

Ceoloni, C., and Donini, P. (1993). Combining mutations for the two homoeologous pairing suppressor genes *Ph1* and *Ph2* in common wheat and in hybrids with alien Triticeae. *Genome* 36, 377–386. doi: 10.1139/g93-052

Chen, H. X., Han, H. M., Li, Q. F., Zhang, J. P., Lu, Y. Q., Yang, X. M., et al. (2018). Identification and genetic analysis of multiple P chromosomes of *Agropyron cristatum* in the background of common wheat. *J. Integr. Agr.* 17, 1697–1705. doi: 10.1016/s2095-3119(17)61861-6

Cuadrado, A., Schwarzacher, T., and Jouve, N. (2000). Identification of different chromatin classes in wheat using *in situ* hybridization with simple sequence repeat oligonucleotides. *Theor. Appl. Genet.* 101, 711–717. doi: 10.1007/s001220051535

Dai, C., Zhang, J. P., Wu, X. Y., Yang, X. M., Li, X. Q., Liu, W. H., et al. (2012). Development of EST markers specific to *Agropyron cristatum* chromosome 6P in common wheat background. *Acta Agronomica Sinica* 38, 1791–1801.

Dellaporta, S. L., Wood, J., and Hicks, J. B. (1983). A plant DNA minipreparation: version II. *Plant Mol. Biol. Rep.* 1, 19–21.

Dewey, D. R. (1984). *The Genomic System of Classification as a Guide to Intergeneric Hybridization with the Perennial Triticeae//Gene Manipulation in Plant Improvement*. Berlin: Springer, 209–279.

Dong, Y. C. (2000). Gene pools of common wheat. *J. Triticeae Crops* 20, 78–81.

Dong, Y. C., Zhou, R. H., Xu, S. J., Li, L. H., Cauderon, Y., and Wang, R. R. C. (1992). Desirable characteristics in perennial Triticeae collected in China for wheat improvement. *Hereditas* 116, 175–178. doi: 10.1111/j.1601-5223.1992.tb00224.x

Du, X. Y., Jia, Z. Z., Yu, Y., Wang, S., Che, B. J., Ni, F., et al. (2019). A wheat-*Aegilops umbellulata* addition line improves wheat agronomic traits and processing quality. *Breed Sci.* 69, 503–507. doi: 10.1270/jsbbs.18200

Dvorak, J., Deal, K. R., and Luo, M. C. (2006). Discovery and mapping of wheat *Ph1* suppressors. *Genetics* 174, 17–27. doi: 10.1534/genetics.106.058115

Fang, Y. H., Yuan, J. Y., Wang, Z. J., Wang, H. Y., Xiao, J., Yang, Z. X., et al. (2014). Development of *T. aestivum* L.-*H. californicum* alien chromosome lines and assignment of homoeologous groups of *Hordeum californicum* chromosomes. *J. Genet. Genomics* 41, 439–447. doi: 10.1016/j.jgg.2014.06.004

Gao, X. H., Ren, C. C., Song, J., Li, X. J., and Ru, Z. G. (2018). Diversity and detection of 1BL/1RS and *Pm21* in wheat cultivars (lines) from regional trial in southern of Yellow and Huai River facultative winter wheat region. *Acta Agr. Boreali-occidentalis Sinica* 27, 779–785.

Gong, W. P., Han, R. P., Li, H. S., Song, J. M., Yan, H. F., Li, G. Y., et al. (2017). Agronomic traits and molecular marker identification of wheat-*Aegilops caudata* addition lines. *Front. Plant Sci.* 8:1743. doi: 10.3389/fpls.2017.01743

Han, F. P., Liu, B., Fedak, G., and Liu, Z. H. (2004). Genomic constitution and variation in five partial amphiploids of wheat-*Thinopyrum intermedium* as revealed by GISH, multicolor GISH and seed storage protein analysis. *Theor. Appl. Genet.* 109, 1070–1076. doi: 10.1007/s00122-004-1720-y

Han, H. M., Bai, L., Su, J. J., Zhang, J. P., Song, L. Q., Gao, A. N., et al. (2014). Genetic rearrangements of six wheat-*Agropyron cristatum* 6P addition lines revealed by molecular markers. *PLoS One* 9:e9106. doi: 10.1371/journal.pone.0091066

Han, H. M., Liu, W. H., Lu, Y. Q., Zhang, J. P., Yang, X. M., Li, X. Q., et al. (2017). Isolation and application of P genome-specific DNA sequences of *Agropyron* Gaertn. in Triticeae. *Planta* 245, 425–437. doi: 10.1007/s00425-016-2616-1

Han, H. M., Liu, W. H., Zhang, J. P., Zhou, S. H., Yang, X. M., Li, X. Q., et al. (2019). Identification of P genome chromosomes in *Agropyron cristatum* and wheat-*A. cristatum* derivative lines by FISH. *Sci. Rep.* 9:9712. doi: 10.1038/s41598-019-46197-6

Jauhar, P. P., and Peterson, T. S. (2006). Cytological analyses of hybrids and derivatives of hybrids between durum wheat and *Thinopyrum bessarabicum*, using multicolour fluorescent GISH. *Plant Breed.* 125, 19–26. doi: 10.1111/j.1439-0523.2006.01176.x

Jiang, B., Liu, T. G., Li, H. H., Han, H. M., Li, L. H., Zhang, J. P., et al. (2018). Physical mapping of a novel locus conferring leaf rust resistance on the long arm of *Agropyron cristatum* chromosome 2P. *Front. Plant Sci.* 9:817. doi: 10.3389/fpls.2018.00817

Jiang, Z., Wang, Q. L., Wu, J. H., Xue, W. B., Zeng, Q. D., Huang, L. L., et al. (2014). Distribution of powdery mildew resistance gene *Pm21* in Chinese winter wheat cultivars and breeding lines based on gene-specific marker. *Sci. Agr. Sin.* 47, 2078–2087. doi: 10.3864/j.issn.0578-1752.2014.11.002

Jubault, M., Tanguy, A. M., Abélard, P., Coriton, O., Dusautoir, J. C., and Jahier, J. (2006). Attempts to induce homoeologous pairing between wheat and *Agropyron cristatum* genomes. *Genome* 49, 190–193. doi: 10.1139/g05-074

Koo, D. H., Liu, W. X., Friebe, B., and Gill, B. S. (2017). Homoeologous recombination in the presence of Ph1 gene in wheat. *Chromosoma* 126, 531–540. doi: 10.1007/s00412-016-0622-5

Kynast, R. G., Friebe, B., and Gill, B. S. (2000). Fate of multicentric and ring chromosomes induced by a new gametocidal factor located on chromosome 4Mg of *Aegilops geniculata*. *Chromosome Res.* 8, 133–139. doi: 10.1023/a:1009294519798

Li, H., Wang, L., Luo, M. C., Nie, F., Zhou, Y., McGuire, P. E., et al. (2019). Recombination between homoeologous chromosomes induced in durum wheat by the *Aegilops speltoides Su1-Ph1* suppressor. *Theor. Appl. Genet.* 132, 3265–3276.

Li, J. B., Dundas, I., Dong, C. M., Li, G. R., Trethowan, R., Yang, Z. J., et al. (2020). Identification and characterization of a new stripe rust resistance gene *Yr83* on rye chromosome 6R in wheat. *Theor. Appl. Genet.* 133, 1095–1107. doi: 10.1007/s00122-020-03534-y

Li, L. H., and Dong, Y. C. (1991). Hybridization between *Triticum aestivum* L. and *Agropyron michnoi* Roshev. *Theor. Appl. Genet.* 81, 312–316.

Li, L. H., and Dong, Y. C. (1993). Progress in studies of *Agropyron* Gaertner. *Acta Genet. Sin.* 15, 45–48.

Li, L. H., Dong, Y. C., Zhou, R. H., Li, X. Q., and Li, P. (1995). Cytogenetics and self-fertility of hybrids between *Triticum aestivum* L. and *Agropyron cristatum* (L.) Gaertner. *Chin. J. Genet.* 22, 105–112.

Li, L. H., Li, X., Li, P., Dong, Y. C., and Zhao, G. (1997). Establishment of wheat-*Agropyron cristatum* alien addition lines. I. Cytology of F_3, F_2BC_1, BC_4, and BC_3F_1 progenies. *Acta Genet. Sin.* 24, 154–159.

Li, L. H., Yang, X. M., Zhou, R. H., Li, X. Q., and Dong, Y. C. (1998). Establishment of wheat-*Agropyron cristatum* alien addition lines II. Identification of alien chromosomes and analysis of development approaches. *Acta Genet. Sin.* 25, 538–544.

Li, Q. F., Lu, Y. Q., Pan, C. L., Zhang, J. P., Liu, W. H., Yang, X. M., et al. (2016a). Characterization of a novel wheat-*Agropyron cristatum* 2P disomic addition line with powdery mildew resistance. *Crop Sci.* 56, 2390–2400. doi: 10.2135/cropsci2015.10.0638

Li, H. H., Lv, M. J., Song, L. Q., Zhang, J. P., Gao, A. N., Li, L. H., et al. (2016b). Production and identification of wheat-*Agropyron cristatum* 2P translocation lines. *PLoS One* 11:e0145928. doi: 10.1371/journal.pone.0145928

Limin, A. E., and Fowler, D. B. (1987). Cold hardiness of forage grasses grown on the Canadian prairies. *Can. J. Plant Sci.* 67, 1111–1115. doi: 10.4141/cjps87-150

Liu, C., Gong, W. P., Han, R., Guo, J., Li, G. R., Li, H. S., et al. (2019). Characterization, identification and evaluation of a set of wheat-*Aegilops comosa* chromosome lines. *Sci. Rep.* 9:4773. doi: 10.1038/s41598-019-41219-9

Liu, H. P., Dai, Y., Chi, D., Huang, S., Li, H. F., Duan, Y. M., et al. (2017). Production and molecular cytogenetic characterization of a durum wheat-*Thinopyrum elongatum* 7E disomic addition line with resistance to *Fusarium* Head Blight. *Cytogenet. Genome Res.* 153, 165–173. doi: 10.1159/000486382

Liu, W. H., Luan, Y., Wang, J. C., Wang, X. G., Su, J. J., Zhang, J. P., et al. (2010). Production and identification of wheat-*Agropyron cristatum* (1.4P) alien translocation lines. *Genome* 53, 472–481. doi: 10.1139/g10-023

Lu, M. J., Lu, Y. Q., Li, H. H., Pan, C. L., Guo, Y., Zhang, J. P., et al. (2016). Transferring desirable genes from *Agropyron cristatum* 7P chromosome into common wheat. *PLoS One* 11:e0159577. doi: 10.1371/journal.pone.0159577

Luan, Y., Wang, X. G., Liu, W. H., Li, C. Y., Zhang, J. P., Gao, A. N., et al. (2010). Production and identification of wheat-*Agropyron cristatum* 6P translocation lines. *Planta* 232, 501–510. doi: 10.1007/s00425-010-1187-9

Ma, F. F., Xu, Y. F., Ma, Z. Q., Li, L. H., and An, D. G. (2018). Genome-wide association and validation of key loci for yield-related traits in wheat founder parent Xiaoyan 6. *Mol. Breed.* 38:91.

Ochoa, V., Madrid, E., Said, M., Rubiales, D., and Cabrera, A. (2014). Molecular and cytogenetic characterization of a common wheat *Agropyron cristatum* chromosome translocation conferring resistance to leaf rust. *Euphytica* 201, 89–95.

Pan, C. L., Li, Q. F., Lu, Y. Q., Zhang, J. P., Yang, X. M., Li, X. Q., et al. (2017). Chromosomal localization of genes conferring desirable agronomic traits from *Agropyron cristatum* chromosome 1P. *PLoS One* 12:e0175265. doi: 10.1371/journal.pone.0175265

Pedersen, C., and Langridge, P. (1997). Identification of the entire chromosome complement of bread wheat by two-colour FISH. *Genome* 40, 589–593. doi: 10.1139/g97-077

Rayburn, A. L., and Gill, B. S. (1986). Isolation of a D-genome specific repeated DNA sequence from *Aegilops squarrosa*. *Plant Mol. Biol. Rep.* 4, 102–109.

Riley, R., Chapman, V., and Miller, T. E. (1973). "The determination of meiotic chromosome pairing," in *Proceeding 4th International Wheat Genetics Symposium*, (Columbia, MO: University of Missouri), 713–738.

Schneider, A., Rakszegi, M., Molnár, L. M., and Szakács, É (2016). Production and cytomolecular identification of new wheat-perennial rye (*Secale cereanum*) disomic addition lines with yellow rust resistance (6R) and increased arabinoxylan and protein content (1R, 4R, 6R). *Theor. Appl. Genet.* 129, 1045–1059. doi: 10.1007/s00122-016-2682-6

Sharma, H. C., Gill, B. S., and Uyemoto, J. K. (1984). High levels of resistance in *Agropyron* species to barley yellow dwarf and wheat streak mosaic viruses. *J. Phytopathol.* 110, 143–147.

Song, L. Q., Jiang, L. L., Han, H. M., Gao, A. N., Yang, X. M., Li, L. H., et al. (2013). Efficient induction of wheat-*Agropyron cristatum* 6P translocation lines and GISH detection. *PLoS One* 8:e69501. doi: 10.1371/journal.pone.0069501

Su, Y. R., Wang, Z. C., Zhang, D. L., Gao, H. Y., and Li, S. P. (2006). Genetic diversity of Yellow-Huai River Zone wheat varieties which originated from 1B/1R by SSR. *J. Plant Genet. Res.* 7, 321–326.

Szakacs, E., and Molnal-Lang, M. (2010). Identification of new winter wheat-winter barley addition lines (6HS and 7H) using fluorescence *in situ* hybridization and the stability of the whole 'Martonvasari 9 kr1'- 'Igri' addition set. *Genome* 53, 35–44. doi: 10.1139/g09-085

Tiwari, V. K., Wang, S. C., Danilova, T., Koo, D. H., Vrána, J., Kubaláková, M., et al. (2015). Exploring the tertiary gene pool of bread wheat: sequence assembly and analysis of chromosome 5M(g) of *Aegilops geniculata*. *Plant J.* 84, 733–746. doi: 10.1111/tpj.13036

Wu, J., Yang, X. M., Wang, H., Li, H. J., Li, L. H., Li, X. Q., et al. (2006). The introgression of chromosome 6P specifying for increased numbers of florets and kernels from *Agropyron cristatum* into wheat. *Theor. Appl. Genet.* 114, 13–20. doi: 10.1007/s00122-006-0405-0

Yang, X. F., Li, X., Wang, C. Y., Chen, C. H., Tian, Z. R., and Ji, W. Q. (2017). Isolation and molecular cytogenetic characterization of a wheat-*Leymus mollis* double monosomic addition line and its progenies with resistance to stripe rust. *Genome* 60, 1029–1036. doi: 10.1139/gen-2017-0078

Ye, X. L., Lu, Y. Q., Liu, W. H., Chen, G. Y., Han, H. M., Zhang, J. P., et al. (2015). The effects of chromosome 6P on fertile tiller number of wheat as revealed in wheat-*Agropyron cristatum* chromosome 5A/6P translocation lines. *Theor. Appl. Genet.* 128, 797–811. doi: 10.1007/s00122-015-2466-4

Yi, Y. J., Zheng, K., Ning, S. Z., Zhao, L. B., Xu, K., Hao, M., et al. (2019). The karyotype of *Aegilops geniculata* and its use to identify both addition and substitution lines of wheat. *Mol. Cytogenet.* 12:15. doi: 10.1186/s13039-019-0428-2

Zhang, A. C., Li, W. Y., Wang, C. Y., Yang, X. F., Chen, C. H., Zhu, C., et al. (2017a). Molecular cytogenetics identification of a wheat -*Leymus mollis* double disomic addition line with stripe rust resistance. *Genome* 60, 375–383. doi: 10.1139/gen-2016-0151

Zhang, J. P., Liu, W. H., Lu, Y. Q., Liu, Q. X., Yang, X. M., Li, X. Q., et al. (2017b). A resource of large-scale molecular markers for monitoring *Agropyron cristatum* chromatin introgression in wheat background based on transcriptome sequences. *Sci. Rep.* 7:11942. doi: 10.1038/s41598-017-12219-4

Zhang, J., Jiang, Y., Wang, Y., Guo, Y. L., Long, H., Deng, G. B., et al. (2018a). Molecular markers and cytogenetics to characterize a wheat-*Dasypyrum villosum* 3V (3D) substitution line conferring resistance to stripe rust. *PLoS One* 13:e0202033. doi: 10.1371/journal.pone.0202033

Zhang, R. Q., Fan, Y. L., Kong, L. N., Wang, Z. J., Wu, J. Z., Xing, L. P., et al. (2018b). *Pm62*, an adult-plant powdery mildew resistance gene introgressed

from *Dasypyrum villosum* chromosome arm 2VL into wheat. *Theor. Appl. Genet.* 131, 2613–2620. doi: 10.1007/s00122-018-3176-5

Zhang, R. Q., Hou, F., Feng, Y. G., Zhang, W., Zhang, M. Y., and Chen, P. D. (2015a). Characterization of a *Triticum aestivum-Dasypyrum villosum* T2VS·2DL translocation line expressing a longer spike and more kernels traits. *Theor. Appl. Genet.* 128, 2415–2425.

Zhang, J. P., Liu, W. H., Han, H. M., Song, L. Q., Bai, L. L., Gao, Z. H., et al. (2015b). De novo transcriptome sequencing of *Agropyron cristatum* to identify available gene resources for the enhancement of wheat. *Genomics* 106, 129–136. doi: 10.1016/j.ygeno.2015.04.003

Zhang, Y., Zhang, J. P., Huang, L., Gao, A. N., Zhang, J., Yang, X. M., et al. (2015c). A high-density genetic map for P genome of *Agropyron* Gaertn. based on specific-locus amplified fragment sequencing (SLAF-seq). *Planta* 242, 1335–1347. doi: 10.1007/s00425-015-2372-7

Zhang, Z., Han, H. M., Liu, W. H., Song, L. Q., Zhang, J. P., Zhou, S. H., et al. (2019). Deletion mapping and verification of an enhanced-grain number per spike locus from the 6PL chromosome arm of *Agropyron cristatum* in common wheat. *Theor. Appl. Genet.* 132, 2815–2827. doi: 10.1007/s00122-019-03390-5

Zhou, S. H., Zhang, J. P., Che, Y. H., Liu, W. H., Lu, Y. Q., Yang, X. M., et al. (2018). Construction of *Agropyron* Gaertn. genetic linkage maps using a wheat 660K SNP array reveals a homoeologous relationship with the wheat genome. *Plant Biotechnol. J.* 16, 818–882. doi: 10.1111/pbi.12831

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 March 2022
doi: 10.3389/fpls.2022.854127

Genome Dominance in *Allium* Hybrids (*A. cepa* × *A. roylei*)

David Kopecký[1]†, Olga Scholten[2†], Joanna Majka[1,3], Karin Burger-Meijer[2], Martin Duchoslav[4] and Jan Bartoš[1]*

[1] Institute of Experimental Botany, Czech Academy of Sciences, Center of the Region Hana for Biotechnological and Agricultural Research, Olomouc, Czechia, [2] Plant Breeding, Wageningen University and Research, Wageningen, Netherlands, [3] Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland, [4] Department of Botany, Palacký University, Olomouc, Czechia

Genome dominance is a phenomenon in wide hybrids when one of the parental genomes becomes "dominant," while the other genome turns to be "submissive." This dominance may express itself in several ways including homoeologous gene expression bias and modified epigenetic regulation. Moreover, some wide hybrids display unequal retention of parental chromosomes in successive generations. This may hamper employment of wide hybridization in practical breeding due to the potential elimination of introgressed segments from progeny. In onion breeding, *Allium roylei* (*A. roylei*) Stearn has been frequently used as a source of resistance to downy mildew for cultivars of bulb onion, *Allium cepa* (*A. cepa*) L. This study demonstrates that in *A. cepa* × *A. roylei* hybrids, chromosomes of *A. cepa* are frequently substituted by those of *A. roylei* and in just one generation, the genomic constitution shifts from 8 *A. cepa* + 8 *A. roylei* chromosomes in the F1 generation to the average of 6.7 *A. cepa* + 9.3 *A. roylei* chromosomes in the F2 generation. Screening of the backcross generation *A. cepa* × (*A. cepa* × *A. roylei*) revealed that this shift does not appear during male meiosis, which is perfectly regular and results with balanced segregation of parental chromosomes, which are equally transmitted to the next generation. This indicates that female meiotic drive is the key factor underlying *A. roylei* genome dominance. Single nucleotide polymorphism (SNP) genotyping further suggested that the drive has different strength across the genome, with some chromosome segments displaying Mendelian segregation, while others exhibiting statistically significant deviation from it.

Keywords: onion, meiotic drive, interspecific hybridization, homoploid, female meiosis, genome stability, homoeologous recombination

OPEN ACCESS

Edited by:
Dayun Tao,
Yunnan Academy of Agricultural Sciences, China

Reviewed by:
Chris Cramer,
New Mexico State University, United States
Ludmila Khrustaleva,
Moscow Timiryazev Agricultural Academy, Russia
Michael Havey,
University of Wisconsin-Madison, United States

****Correspondence:***
David Kopecký
kopecky@ueb.cas.cz

† These authors have contributed equally to this work

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

Received: *13 January 2022*
Accepted: *09 February 2022*
Published: *10 March 2022*

Citation:
Kopecký D, Scholten O, Majka J, Burger-Meijer K, Duchoslav M and Bartoš J (2022) Genome Dominance in Allium Hybrids (A. cepa × A. roylei). Front. Plant Sci. 13:854127. doi: 10.3389/fpls.2022.854127

INTRODUCTION

Introgression breeding is the way to efficiently transfer agronomically beneficial alleles from wild relatives to crops. This involves interspecific mating followed by one or more rounds of backcrossing to the recipient parent. Many traits have been improved *via* introgression breeding, including resistance to pests and diseases, tolerance to abiotic stresses, and root-related traits (Anamthawat-Jónsson, 2001; Scholten et al., 2007; Placido et al., 2013; Molnár-Láng, 2015). However, introgression lines frequently suffer from the instability of the introgressed segment(s) in the successive generations (Kopecký et al., 2019; Pernickova et al., 2019). Combining two genomes in a single nucleus, it opens a way for genome dominance, a phenomenon, when one parental

genome becomes dominant, while the other tends to be submissive in a hybrid progeny. Such dominance can manifest itself in several ways, including altered gene expression and epigenetic regulation (Glombik et al., 2020). Most, if not all, allopolyploids retain the expression level of one (dominant) parent (so-called expression level dominance) and/or display a preferential expression from the alleles of the dominant genome (so-called homoeolog expression bias). Such dominance does not involve all the expressed genes, as some genes can be overexpressed from the submissive genome or display the overall expression at the level of the submissive parent (Edger et al., 2017; Bird et al., 2018; Glombik et al., 2021).

Another expression of genome dominance is elimination of chromosomes of the submissive genome or their replacement by those of the dominant genome (Glombik et al., 2020; Majka et al., 2020). Chromosome elimination usually occurs in hybrids where there is no pairing of homoeologous chromosomes (i.e., chromosomes from two more or less distinct parental species). Restriction of homoeologous pairing can be a consequence of either DNA sequence dissimilarity, which precludes homoeologous recognition and initiation of pairing during prophase I of meiosis or the action of a molecular mechanism preventing dissimilar DNA sequences from forming crossovers. The textbook example of such a mechanism is *pairing homoeologous 1* (*Ph1*) in wheat (Sears and Okamoto, 1958), which is also capable of modifying chromosome pairing of other species when transferred from wheat (Lukaszewski and Kopecky, 2010). In wheat-rye hybrids, rye chromosomes are more prone to elimination during meiosis, despite strict homoeologous chromosome pairing (Tsunewaki, 1964; Lukaszewski et al., 1987; Pernickova et al., 2019).

While considered rare in the past, wide hybrids with (extensive) homoeologous chromosome pairing readily develop in nature and can also be created artificially, for example, as part of breeding programs. Chromosomes of ryegrass (*Lolium* spp.) pair and recombine freely with those of fescue (*Festuca* spp.) in xFestulolium hybrids (Kopeckyì et al., 2008; Zwierzykowski et al., 2008). Similarly, various hybrids of ornamental plants, such as lily hybrids, *Alstroemeria aurea* × *Alstroemeria inodora* and *Gasteria lutzii* × *Aloe aristata*, show homoeologous chromosome pairing (Takahashi et al., 1997; Kamstra et al., 1999; Karlov et al., 1999; Khan et al., 2009). The ability of homoeologous chromosomes to pair in meiosis opens the way for competition between parental chromosomes. While male meiosis is symmetrical with all the four products producing gametes that can contribute to the successive generation, female meiosis is asymmetrical where only one product generates a gamete (the egg cell), while the genetic information in the remaining cells is not passed onto the next generation. This aspect of female meiosis creates an opportunity for chromosome competition and this phenomenon is called "meiotic drive" (Sandler and Novitski, 1957). In hybrid (homoploid) mice with regular (homoeologous) chromosome pairing, Akera et al. (2017) observed biased orientation of parental chromosomes on the karyokinetic spindle during female meiosis. Chromosomes from the dominant genome tended to orient toward the pole eventually producing the egg cell more frequently than those from the submissive genome and their frequency among progeny was higher than expected from random segregation.

As mentioned above, interspecific hybridization is often used in breeding programs to introgress one or more desired traits in crops, usually from wild relatives. Using this approach, lines of cultivated bulb onion *Allium cepa* (*A. cepa*) L. with chromosome segments carrying downy mildew [*Peronospora destructor* (Berk.) Casp.] resistance gene(s) introgressed from its wild relative *Allium roylei* (*A. roylei*) Stearn. were developed by research programs of Wageningen University and Research Center (Netherlands) and Russian State Agrarian University (Russia) (van der Meer and de Vries, 1990; Khrustaleva and Kik, 1998; Khrustaleva et al., 2019). A combination of phenotyping, genotyping with DNA markers, and cytogenetic analyses of advanced backcross (BC1) generations allocated the putative downy mildew resistance locus *Pd1* to the region spanning the most distal ∼18% of the long arm of chromosome 3 (van Heusden et al., 2000). By controlled intercrosses and BC1, homozygous introgression lines were obtained that were resistant to downy mildew. Reduction of the introgressed segment length was an important step further, also because of a recessive lethal factor located in a close vicinity of the *Pd1* locus and probably expressed only in the *A. cepa* background (Scholten et al., 2007; Kim et al., 2016; Khrustaleva et al., 2019). Separating that lethal factor from *Pd1* gene by crossing over, it was observed in a single plant out of 215 plants screened, suggesting a tight linkage (Scholten et al., 2007). *A. roylei* has also been proposed to use as a bridge for introgression of traits from the Welsh onion [*Allium fistulosum* (*A. fistulosum*) L.] into cultivated bulb onion, as direct introgression is difficult because of a very low fertility of the hybrids between bulb onion and Welsh onion (Khrustaleva and Kik, 1998, 2000; Stevenson et al., 1998; Budylin et al., 2016).

The main aim of this study was to evaluate the stability of the newly established hybrid genome in hybrids of *A. cepa* × *A. roylei*, to shed light on the mechanisms underlying possible genome dominance, and to estimate the retention rate of individual chromosomal segments of *A. roylei* and *A. cepa* in successive hybrid generation(s).

MATERIALS AND METHODS

Plant Material

A total of 104 F2 plants from a cross *A. cepa* ♀ × *A. roylei* ♂ were analyzed in this study. The F2 population was generated by selfing of one plant of the interspecific F1 (CxR) (PRI 93103), a hybrid genotype between *A. cepa* and *A. roylei*. Of those 104 plants, 75 plants were analyzed by genomic *in situ* hybridization (GISH) (to determine their genomic constitution). All the plants were genotyped by single nucleotide polymorphism (SNP) markers, of which 80 plants were included earlier in the production of a linkage map (Scholten et al., 2016). Part of the F2 family was the original population used in the study of van Heusden et al. (2000) and all the plants shared the same parental lineages. In addition, GISH was used to analyze the genome composition of 21 BC1 plants produced by a BC1 of the F1 hybrid used as a pollinator with a male-sterile *A. cepa*.

Chromosome Preparations and Genomic *in situ* Hybridization

Roots of individual plants were collected and their cell cycle was synchronized using iced distilled water for 28 h following fixation in Carnoy's solution (absolute ethanol/glacial acetic acid, 3:1 v/v). For meiotic analyses, flower buds were fixed in Carnoy's solution at 37°C for 7 days. Individual anthers that were confirmed to be in the proper meiotic stage were squashed in a drop of acetic acid and used for GISH. Chromosome preparations were made according to a study by Masoudi-Nejad et al. (2002). GISH analyses were done on the mitotic and meiotic chromosome spreads according to a study by Ferreira et al. (2021). Total genomic DNA (gDNA) of *A. cepa* was used as blocking DNA and total gDNA of *A. roylei* was labeled with digoxigenin (DIG) using the DIG-Nick Translation Kit (Roche Applied Science, United States) according to the manufacturer's instructions and used as a probe. The probe/blocking DNA ratio was ∼1:150. Signal detection was made with anti-DIG-fluorescein isothiocyanate (FITC) conjugate (Roche Applied Science). Chromosomes were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) in Vectashield (Vector Laboratories, Oberkochen, Germany). Chromosome analyses were done using an Olympus AX70 microscope equipped with epifluorescence and a SensiCam B/W camera. Images were captured with Microimage software and processed with Adobe Photoshop version 6 software (Adobe Systems Corporation, San Jose, CA, United States). The proportions of *A. roylei* and *A. cepa* chromosomes in hybrids were tested against the assumption of a 1:1 ratio representing Mendelian inheritance. The number of *A. roylei* chromosomes was expressed as the proportion p of the total number of chromosomes within the cell and H_0:$p = 0.50$ was tested by the one-sample t-test in R (R Core Team, 2021).

To determine the positions of the crossovers along chromosomes in F2 and BC1 hybrids, we measured the lengths of introgressed segments and the lengths of both the arms of recombined chromosomes using the Scion Image software (Scion Corporation, Frederick, MD, United States). The difference in the distribution of crossovers between male meiosis and both the meioses was evaluated by comparing their distributions along chromosome arms divided into 10 segments (bins) of 10% of their length. Two empirical distributions were compared using the function ks.boot in the Matching library in R. The function uses a bootstrap version of the Kolmogorov–Smirnov test, providing accurate coverage even when the distributions being compared are not entirely continuous and ties occur in the dataset (Sekhon, 2011).

A total of 50 pollen mother cells (PMCs) were evaluated in each of the meiotic stages (prophase I, metaphase I, anaphase I, and telophase I) in a single F1 *A. cepa* × *A. roylei* plant (PRI 93103).

Single Nucleotide Polymorphism Genotyping

The 104 F2 hybrids were genotyped using SNPs markers (all the 75 plants used for GISH karyotyping and 29 others from the same cross), as described previously (Scholten et al., 2016). Only markers enabling unambiguous discrimination of *rr* (homozygote for *A. roylei* allele), *rc* (heterozygote), and *cc* (homozygote for *A. cepa* allele) genotypes were selected. The statistically significant deviation of frequencies of three genotype classes in F2 hybrids from the theoretical 1 *rr*:2 *rc*:1 *cc* of Mendelian inheritance (H_0) was assessed using multinomial test in R, separately for each SNP marker.

RESULTS

Chromosomes of *A. cepa* Are Replaced by Those of *A. roylei* in F2 Hybrids

Among 75 individuals of the F2 generation, we detected a significant shift in genome composition from eight chromosomes of *A. cepa* plus eight chromosomes of *A. roylei* in the F1 genotype toward the *A. roylei* genome. However, homoeologous crossover events occurred complicating classification of parental chromosomes/genomes. Thus, we consider the origin of the chromosome based on the fluorescence signal spanning its centromeric region (**Figure 1A**). On average, there were 9.33 chromosomes of *A. roylei*—4.07 complete and 5.26 recombined (homoeologous recombination) and 6.67 chromosomes of *A. cepa*—1.69 complete and 4.97 recombined (**Figure 2**). This proportion of *A. roylei* and *A. cepa* chromosomes significantly deviated from the 1:1 ratio (two-sided one sample t-test, mean proportion p of *A. roylei* chromosomes ± SD: 0.58 ± 0.12, $t = 6.11$, $df = 74$, $P < 0.001$) and roughly accounts for the ratio of 1.4:1 of *A. roylei* vs. *A. cepa* centromeres (**Supplementary Table 1**).

Male vs. Female Meiosis

We analyzed the consequence of male meiosis by screening the BC1 generation (male sterile *A. cepa* ♀ × F_1 hybrid ♂). Female meiosis could not be assessed in the same fashion, as all the attempts to produce seed from the reciprocal BC1 (F_1 hybrid ♀ × *A. cepa* ♂) failed. Therefore, the contribution of female meiosis can only be assessed by subtraction of the detected effects of male meiosis from the combined contribution of both the sexes to the F2 generation. The number of crossovers per bivalent calculated from the frequency of recombined chromosomes among progeny was about the same: 1.81 ± 0.48 (mean ± SD) in male meiosis and 1.65 ± 0.38 in both the meioses (two-sided equal-variance t-test, $t = -1.51$, $df = 91$, $P = 0.133$). Similarly, the difference in numbers of crossovers per recombined chromosome was also non-significant: 1.35 ± 0.24 in male meiosis vs. 1.29 ± 0.16 in both the meioses (two-sided Aspin–Welch unequal-variance t-test, $t = -1.08$, $df = 20.74$, $P = 0.292$). It, therefore, appears that the recombination rate in male and female meiosis was the same (not different).

Based on the lengths of chromosome segments in recombined chromosomes and the positions of the crossover points, we were able to estimate the distribution of recombination events along the chromosomes. We did not include double crossovers (two crossovers in a single arm), as these are subjected to crossover interference (Ferreira et al., 2021) and may bias the

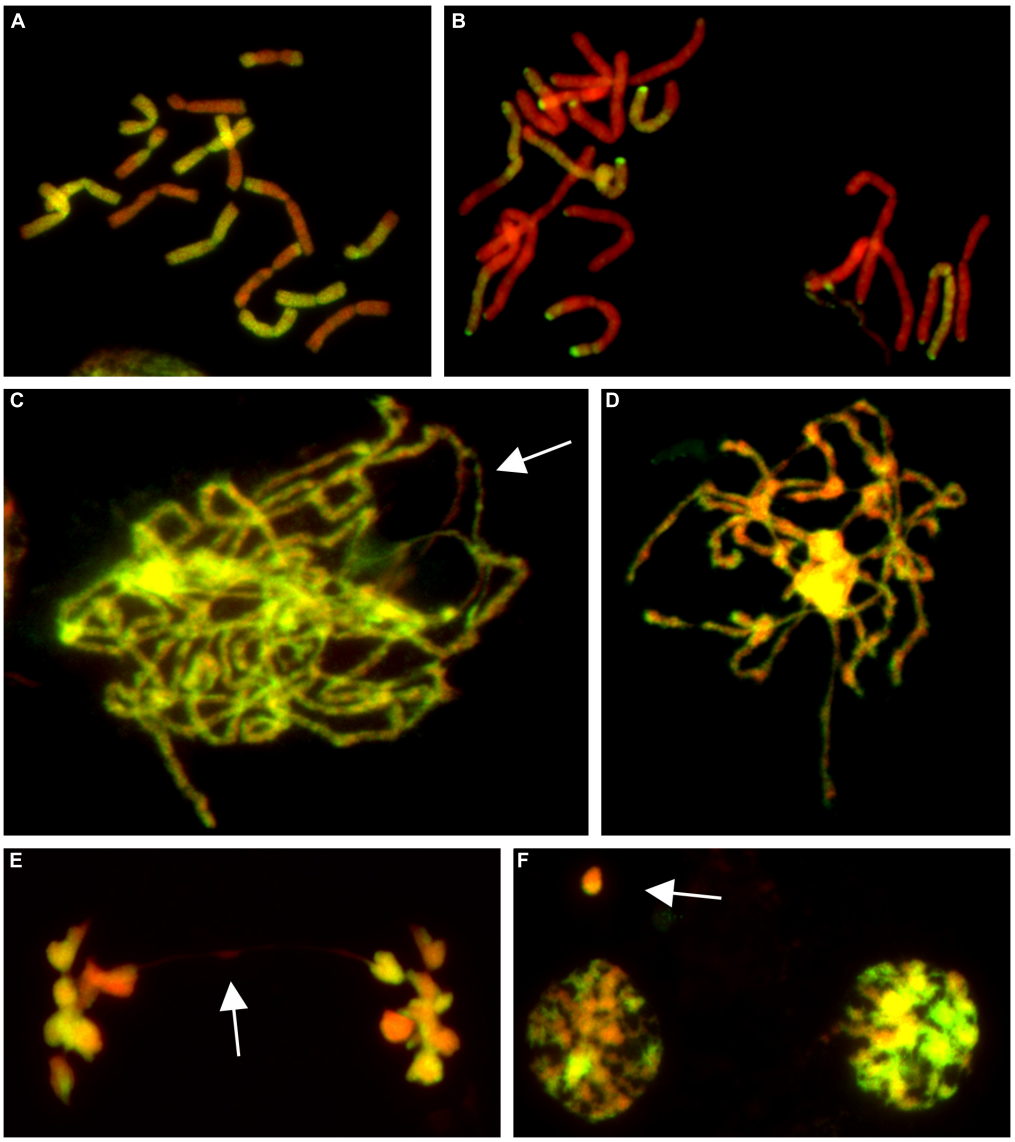


FIGURE 1 | Molecular cytogenetic analysis of *Allium cepa* (*A. cepa*) × *Allium roylei* (*A. roylei*) hybrids. Mitotic cells of F2 **(A)** and backcross (BC1) **(B)** plants and meiotic cells of F1 hybrid **(C–F)** after genomic *in situ* hybridization (GISH). During meiosis, homoeologous chromosomes initiate pairing in zygotene [**(C)**; so far unpaired segments indicated by arrow] with complete pairing in pachytene **(D)**. During anaphase I, chromosomes segregate to opposite poles **(E)** with only rare bridges (arrow) forming diads in TI **(F)** with rare micronuclei (arrow). Total guide DNA (gDNA) of *A. roylei* was labeled with digoxigenin (green/yellow color) and sheared DNA of *A. cepa* was used as blocking DNA (red pseudocolor).

overall results. However, two crossovers per chromosome, one in each arm, were included, as the centromere does act as a barrier to crossover interference. In this material, arms of a chromosome appear to be independent units in the process of crossing over (Ferreira et al., 2021). The plants of the F2 generation (contribution of both the meioses) show a pattern of homoeologous recombination distributed unevenly along chromosomes, with a higher frequency in distal regions, except for the terminal bin and highly reduced frequency in proximal regions around centromeres. A reduction in homoeologous recombination in (sub)telomeric and (peri)centromeric regions was also observed in BC1 plants (male meiosis); the highest frequency was found in interstitial regions of the chromosome arms (**Figure 3**). The difference between male and both the meioses in the distribution of homoeologous recombination was statistically significant (two-sided two-sample Kolmogorov–Smirnov test, $D = 0.7$, bootstrap $P = 0.007$).

Male Meiosis Does Not Introduce the Bias in Genome Composition

Male meiosis of F1 seems to be regular with the formation of bivalents consisting of homoeologous chromosomes during prophase I (**Figures 1C,D**) and metaphase I. Ring bivalents

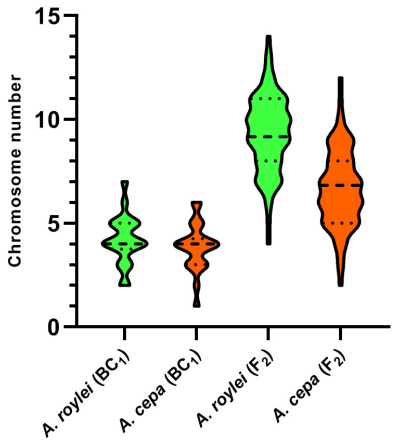


FIGURE 2 | Genome composition of F2 hybrids of *A. cepa* × *A. roylei* and genome composition of the male gametes (pollen grains) calculated from the genome composition of BC1 progeny of *A. cepa* × (*A. cepa* × *A. roylei*). Number of parental chromosomes is based on the origin of the centromere region of a particular chromosome. A dashed line in a violin plot represents median and dotted lines represent quartiles.

predominated over rod bivalents (5.06 vs. 2.94 per cell), suggesting 1.63 crossovers per bivalent. In anaphase I, segregation of homoeologous chromosomes toward opposite poles without lagging chromosomes was observed and we observed a chromosome bridge in only one out of 50 cells (**Figure 1E**). Similarly, micronuclei were observed in only two cells out of 51 cells in telophase I, one being of *A. cepa* origin and the other seemed to be composed by chromatin of both the progenitors, probably as a result of homoeologous recombination (**Figure 1F**). These results demonstrate that male meiosis is fairly regular (**Supplementary Figure 1**).

The analysis of the BC1 (*A. cepa* ♀ × F_1 hybrid ♂) provided us with an estimate of the effect of male meiosis on the genome composition of the progeny in *A. cepa* × *A. roylei* hybrids (**Figure 1B**). As the gamete (egg) of *A. cepa* had eight chromosomes of *A. cepa*, remaining chromosomes of the progeny had to come from male gamete (pollen grain) of the F1 hybrid. Once the F1 hybrids have eight chromosomes of *A. cepa* and eight chromosomes of *A. roylei*, a theoretical average constitution of a pollen grain is four chromosomes of *A. cepa* + four chromosomes of *A. roylei*. This is close to what we found among 21 plants of the BC1 generation: 11.86 chromosomes of *A. cepa* (eight of them being from the egg of *A. cepa*) and 4.19 chromosomes of *A. roylei* (**Figure 2**). With the assumption that female gamete contributed to the embryo with eight *A. cepa* chromosomes, the average genome composition of pollen grain is to be 3.86 *A. cepa* and 4.19 *A. roylei* chromosomes. The difference between the observed ratio and theoretical 1:1 ratios of Mendelian inheritance was non-significant (two-sided one-sample *t*-test, mean proportion *p* of *A. roylei* chromosomes ± SD: 0.52 ± 0.14, $t = 0.63$, $df = 20$, $P = 0.537$). While 20 plants were euploid with 16 chromosomes, one plant was aneuploid with 17 chromosomes. These results indicate that male meiosis produces viable gametes with almost equal proportions of parental chromosomes and does not significantly contribute to *A. roylei* genome dominance. By subtraction, it appears that female meiosis is likely the driving force of this phenomenon. Considering 3.86 chromosomes of *A. cepa* and 4.19 chromosomes of *A. roylei* transmitted by the pollen grain, the average egg cell must have contributed 2.81 *A. cepa* and 5.14 *A. roylei* chromosomes to achieve genome composition observed in the F2 hybrids (6.67 chromosomes of *A. cepa* + 9.33 chromosomes of *A. roylei*).

Genome Dominance Seems to Be Chromosome Specific

Single nucleotide polymorphism genotyping of 104 progeny provided another measure of the genome composition of the F2 hybrids along all the eight linkage groups representing individual chromosomes [based on genetic map of this population previously published as supplemental data by Scholten et al. (2016)]. We selected 119 SNP markers, which clearly distinguished all the three genotype classes: *rr* (homozygote for *A. roylei* allele), *rc* (heterozygote), and *cc* (homozygote for *A. cepa* allele) distributed over all the eight chromosomes (ranging from 8 to 24 per chromosome).

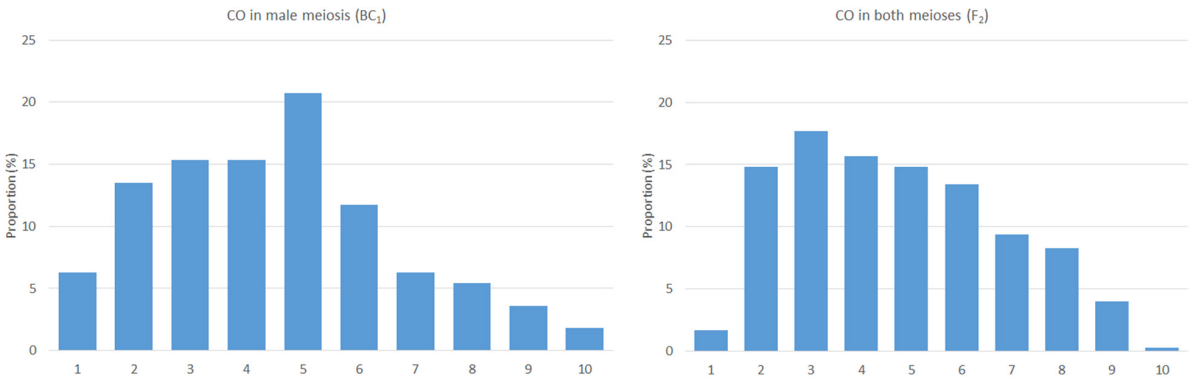


FIGURE 3 | The frequency and distribution of crossovers in male (left) and both the meioses (right). The x-axis represents a chromosome arm (from the telomere on the left to the centromere on the right) divided into bins of 10% of relative arm length.

Unfortunately, positions of centromeres are not known on the genetic maps, which hamper direct comparison of GISH analysis and SNP genotyping. However, it is evident that results from SNP genotyping are in line with those from GISH: we observed 3,647 *rr*, 6,749 *rc*, and 1,691 *cc* genotypes. This can be translated into the ratio (3,647 × 2) + 6,749 *r* allele vs. (1,691 × 2) + 6,749 *c* allele or a ratio of 1.39 *r*:1 *c*. This is almost identical to the ratio obtained from the GISH results of centromeres (1.4:1), suggesting high fidelity of our results. However, the variability in the genome composition revealed by SNP genotyping was large between individual plants and ranged from 78.2 *r*:21.8 *c* (roughly 3.6:1) to 32.8 *r*:67.2 *c* (roughly 1:2).

When focusing on individual chromosome regions along the entire genomes, we found high variability in the genome composition and frequent statistically significant deviations from the Mendelian 1:2:1 ratio of *rr*:*rc*:*cc* genotypes as tested by multinomial test, separately for each SNP marker (**Figure 4**).

LG1: Entire chromosome displayed genome dominance of *A. roylei*. Distal parts of the chromosome showed non-significant deviation, while the segment between 18 and 81 cM was deviated significantly from the theoretical 1:2:1 in favor of the *rr* and *rc* genotypes ($P < 0.01$).

LG2: All but one marker showed a statistically significant deviation ($P < 0.01$) from the theoretical ratio of 1:2:1 toward the *rr* and *rc* genotypes.

LG3: There was a strong deviation from the 1:2:1 toward the *rr* and *rc* genotypes ($P < 0.01$ for eight out of nine markers).

LG4: There was a non-significant difference from 1:2:1 in the distal part of one arm (from 0 to 8 cM), while all the remaining segments of the chromosome displayed a shift toward *rc* heterozygote constitution ($P < 0.001$).

LG5: Approximately, one-half of the linkage group showed a non-significant deviation, while the other half displayed a significant deviation ($P < 0.01$) in favor of *rr* and *rc* genotypes. Interestingly, the distal part of the chromosome represented by three markers (from 0 to 20 cM) was one out of two regions of the genome displaying higher number of *cc* genotypes than *rr* genotypes (but the deviation is non-significant).

LG6: Approximately, one-half of the linkage group (from 0 to 64 cM) showed a non-significant deviation from the theoretical, while other half significantly deviated ($P < 0.05$) toward the *rr* and *rc* genotypes. A strong deviation ($P < 0.001$) was found at the distal part of the chromosome (from 95 to 120 cM).

LG7: Distal regions showed no significant deviation, while the interstitial part (from 10 to 50 cM) deviated from 1:2:1 toward the *rr* and *rc* genotypes ($P < 0.05$).

LG8: Three out of eight markers in the distal region of one arm show an excess of the *cc* genotypes (non-significant deviation from 1:2:1), while other parts deviated significantly (three markers) or non-significantly (two markers) from theoretical 1:2:1 toward the *rr* and/or *rc* genotypes.

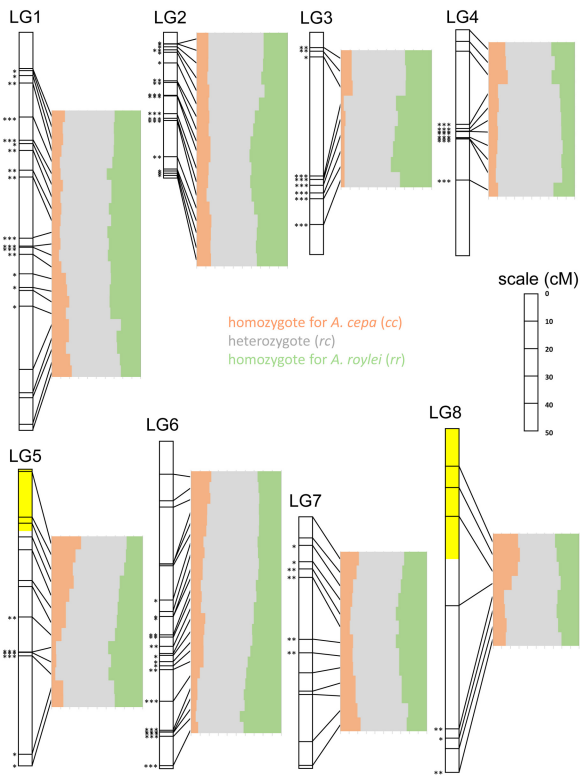


FIGURE 4 | Segregation of *A. cepa* and *A. roylei* alleles in F2 generation. Application of single nucleotide polymorphism (SNP) markers enabled visualization of segregation distortion of the regions along individual chromosomes [linkage groups based on genetic map of Scholten et al. (2016)]. Statistically significant deviation from the Mendelian 1:2:1 ratio of *rr*:*rc*:*cc* genotypes was tested by multinomial test, separately for each SNP marker (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). Two regions showing (statistically non-significant) distortion toward *A. cepa* alleles are highlighted with yellow color.

DISCUSSION

Cultivated crops are usually limited in diversity, by the domestication bottleneck and long-lasting selection. Thus, introgression of genetic diversity from alien sources, in general, or of alleles of agronomically beneficial loci never present in a crop or lost during the evolution and/or selection is a step toward the development of superior cultivars. In bulb onion (*Allium cepa*) breeding, alleles for downy mildew [caused by *Peronospora destructor* (Berk.) Casp.], leaf blight (caused by *Botrytis squamosa* Walker), and anthracnose [caused by *Colletotrichum gloeosporioides* (Penz.) Penz. and Sacc.] resistance can be introgressed from a wild relative, *A. roylei* (Kofoet et al., 1990; de Vries et al., 1992; Galvan et al., 1997; Scholten et al., 2016). However, alien introgressions are not always stable in the host genome of a crop and may be lost over generations. Various studies indicate that merging two genomes from different species results in massive changes at different levels, including modifications of gene expression and epigenetic regulation, genome down- or upsizing, and chromosome reshuffling (Wendel, 2015; Van de Peer et al., 2021). In many hybrids, one of the parental genome becomes "dominant," whereas the other turns to be "submissive." Such genome dominance can be expressed at various levels including

elimination of chromosomes from the submissive genome or replacement of such chromosomes by those from the dominant genome (Glombik et al., 2020). In triticale, a hybrid of wheat and rye, chromosomes of rye are more prone to elimination than their wheat counterparts, which may lead to the reversion to pure wheat forms in successive generations (Orellana et al., 1984). In xFestulolium, a hybrid of ryegrass (*Lolium* spp.) and fescue (*Festuca* spp.), chromosomes of *Festuca* are gradually replaced by those of *Lolium* (Kopeckyì et al., 2006; Zwierzykowski et al., 2006). In introgression cultivars, complete elimination of *Festuca* segments is expected to happen within 3–4 generations of multiplication (Kopecký et al., 2019). Therefore, studies on the genome stability of the hybrid genomes and the transmission of the introgressed segment(s) to successive generations may offer some guidance in assessing the potential of the introgression breeding.

In hybrid onion *A. cepa* × *A. roylei*, homoeologous chromosomes pair and recombine, but there is only scarce information on the genome dominance. When compiling genetic maps of the parental species *via* genotyping of F2 hybrids (from the same cross as those in this study) with amplified fragment length polymorphism (AFLP) markers, van Heusden et al. (2000) mentioned that the *cepa*-specific markers were not amplified in 28% of the F2 plants compared to 16% of the *roylei*-specific markers. From that, the authors estimated that the contribution of the *cepa*-specific and *roylei*-specific alleles in the F2 generation was about 44 and 56%, respectively. This is in line with the results obtained in this study, where results from SNP genotyping indicate the proportion of 42–58%, respectively, while GISH karyotyping revealed 42% of centromeres being of *A. cepa* origin and 58% of centromeres being of *A. roylei* origin. These results indicate violation of the Mendelian law of random segregation. However, the *roylei* genome dominance was not observed in all the plants and much variation was observed between individual plants, with the percentage of *cepa*-specific alleles ranging from 15 to 68% (van Heusden et al., 2000). Similar variation was found in this study: the frequency of *c* alleles ranged from 22 to 67% among individual plants. The correspondence of our results from GISH and SNP genotyping with AFLP markers of a study by van Heusden et al. (2000) indicates that the genome dominance is consistent at the level of about 42:58 in *A. cepa* × *A. roylei* hybrids. However, studies on different wide hybrids provided different results, ranging from synthetic *Brassica napus* (*B. rapa* × *B. oleracea*) with proportion of parental genomes 43:57, lily hybrid (*Lilium longiflorum* × Asiatic) with the ratio of 54:46 to synthetic *Tragopogon miscellus* (*T. pratensis* × *T. dubius*) with the ratio of 49:51 (Karlov et al., 1999; Xiong et al., 2011; Chester et al., 2012). Thus, it is evident that the strength of the genome dominance depends on a cross combination and the divergence of parental genomes. Interestingly, genome dominance was also observed in another *Allium* hybrid developed from a cross between *A. fistulosum* L. and *A. cepa*. Despite previously reported problems with fertility, a number of F2 and BC1 plants (with bulb onion as pollinator) were obtained and showed the dominance toward *A. fistulosum* for three out of four isozyme markers that were tested and which showed a statistically significant violation of the theoretical 1:2:1 ratio in the F2 population (Ulloa et al., 1995). The authors hypothesized that the genome dominance might be at least partly caused by the cytoplasmic effect: their hybrids displaying *A. fistulosum* dominance were developed by pollinating *A. fistulosum* flowers with pollen from *A. cepa* and thus, the hybrids likely possessed the cytoplasm from *A. fistulosum*. However, the role of the cytoplasm could probably be ruled out in our hybrids, as the dominant genome (*A. roylei*) was used as pollinator for hybrid development. Moreover, the *A. roylei* dominance was evidenced only in F2, but not in BC1 plants.

Based on our results and other reports, it appears that genome dominance is chromosome specific. In our previous study, we found that chromosome 5 of *Festuca pratensis* (*F. pratensis*) is more prone to be replaced by its *Lolium multiflorum* (*L. multiflorum*) homoeolog in *L. multiflorum* × *F. pratensis* allotetraploids than any other chromosome (Kopecký et al., 2019). Similarly, preferential elimination of some chromosomes and a higher transmission of others were also observed in *Gossypium hirsutum* × *G. sturtianum* and *G. hirsutum* × *G. australe* hybrids (Lopez-Lavalle and Brubaker, 2007). In this study, we also found that some chromosomes or even chromosome segments of *A. cepa* are less likely to be transmitted than others (**Figure 4**). Specifically, chromosome 8 showed an almost equal (random) segregation, while chromosomes 2 and 3 of *A. roylei* were transmitted much more frequently to successive generation than their counterparts from *A. cepa*. Similarly, the segregation distortion was localized on all the chromosomes, except chromosomes 7 and 8 based on the results of SNP markers applied to F2 *A. cepa* × *A. roylei* (Scholten et al., 2016). This would indicate that not all the chromosomes are the subjects of the *roylei* genome dominance.

Despite the majority of the *A. cepa* genome being prone for replacement by *A. roylei*, there are two regions in *A. cepa* genome, which are transmitted at the frequency exceeding 50% in *A. cepa* × *A. roylei* hybrids. One of the regions is represented by three markers at positions 0.84, 17.49, and 19.68 cM on chromosome 5. The other segment showing > 50% transmission of *A. cepa* allele is represented by three markers at the distal part of chromosome 8 (at positions 13.65, 21.46, and 32.01 cM). One would expect a potential link between these two regions, such as *trans*-acting regulation of one by the other; however, comparison of the frequencies of the *rr*, *rc*, and *cc* classes suggests that these two regions segregate independently with no linkage. Interestingly, one would expect a segregation distortion for the segment carrying a lethal factor, previously identified at the distal end of chromosome 3 (van Heusden et al., 2000). This locus is in a close vicinity of *Pd1*, the downy mildew resistance locus. While *Pd1* is of great interest to breeders, the lethal factor, once introduced to *A. cepa* in a double dose (i.e., *rr* homozygote for this chromosome segment), is assumed to cause lethality (Scholten et al., 2007). However, we identified 8 out of 104 plants as being *rr* homozygous for all the nine markers distributed from 5.6 to 70 cM of the genetic map of chromosome 3. This might be potentially caused by double crossover of one homoeolog in the large region, where markers are absent (between 9 and 52 cM).

Recent studies have indicated that genome dominance at the chromosome level is caused by meiotic drive, a phenomenon of non-Mendelian transmission of chromosomes to the next

generation. It is worth to mention differences between male and female meiosis. Male meiosis is symmetric and all the four products participate equally to successive generation. However, meiotic drivers increase the chance of the sperm cells carrying them to fertilize the eggs and, thus, violate the random transmission of sperm cells with or without the driver (Kruger and Mueller, 2021). Male meiotic drivers seem to benefit themselves and confer negative effects on the counterpart (from the other parental genome in the case of interspecific hybrids) such as reduced motility of sperms, differences in the pollinating rate of pollen grains, and failure to develop to maturity. On the contrary, female meiosis is asymmetric when one homoeolog is transmitted to the egg cell, while the other homoeolog is transferred to polar bodies, which are not participating in the next generation. This opens a way for competition between homoeologous chromosomes, where meiotic drivers act to alter the orientation of particular chromosomes in bivalents toward the developing egg cell and the polar body; in other words, chromosomes of the dominant genome are transmitted more often to egg cell and chromosomes of the submissive genome remain more frequently in polar bodies.

The first insight on the regularity of male meiosis in onion hybrids (*A. cepa* × *A. roylei*) has been provided by de Vries et al. (1992). They observed only a limited number of univalents in metaphase I and no abnormalities as bridges, fragments, or micronuclei in the later stages. Our observations fully confirm this, as bivalents were formed regularly, with both the arms paired (ring bivalents) more frequently than one (rod bivalents). The regularity of male meiosis and regular (random) transmission of chromosomes to progeny are supported by the genome composition of the BC1 generation. There were no significant differences between the numbers of parental chromosomes transmitted through pollen. Similarly, reciprocal BC1 progeny (BC1F1) of *F. pratensis* × *L. multiflorum*, where the effects of male vs. female meiosis could be studied, also showed a large difference in the genome composition. While male meiosis of hybrid produced gametes with almost equal contribution of parental chromosomes, female meiosis dramatically shifted the composition toward *L. multiflorum* (Kopecký et al., unpublished results).

Overall, our results indicate that in these hybrids, male meiosis does not contribute significantly, or not at all, to genome dominance and that it has to be female meiosis responsible for the shift toward *roylei* genome. This is in line with Courret et al. (2019) and Kruger and Mueller (2021), who hypothesized that meiotic drivers function exclusively in the male or female germline, but not both. However, we cannot completely rule out other mechanisms including differences in the gamete viability and preferential fertilization.

The mechanisms underlying the female meiotic drive have been intensively studied in hybrid mice [reviewed in Clark and Akera (2021)]. Chromosomes from the dominant genome were oriented toward the egg cell more frequently than those from the submissive genome. Molecular mechanism is so far unclear; however, the candidate is CDC42, which is signaling unequal regulation of microtubule tyrosination. This unequal tyrosination is probably caused by the difference in the copy number of kinetochore proteins between the two genomes. The abundance of the kinetochore proteins is presumably affected by the centromeric minor satellite repeats that are twice as high, while the major pericentromeric satellite repeats were almost undetectable on the chromosomes of the dominant genome (Akera et al., 2017). Hence, the dominant genome (having more copies of minor centromeric repeats and, consequently, more copies of kinetochore proteins) is preferentially transmitted to the egg cell, while the submissive genome is directed into the polar bodies. Thus, it is evident that the sequences of centromere and pericentromeric region are likely the key component of the drive (Chmatal et al., 2014). Centromeric drive has also been observed in *Drosophila* and monkey flower (*Mimulus* spp.) hybrids (Fishman and Saunders, 2008; Wei et al., 2017). Apart from the centromere, other chromosome regions in several organism were occasionally found to violate random segregation including the well-known knob-mediated meiotic drive in maize (Dawe et al., 1999). Thus, we cannot rule out other sequences outside of the native centromere and pericentromeric region to be incharge of female meiotic drive present in onion hybrids.

CONCLUSION

This study reveals dominance of the *A. roylei* genome in hybrids with *A. cepa*. This dominance appears to be caused by the female meiotic drive; male meiosis seemed to be regular and produced gametes with equal proportion of chromosome from parental genomes and these chromosomes were randomly transmitted to progeny. SNP genotyping revealed that the drive had different strength across the genome, with some chromosome segments showing Mendelian segregation, while others showed statistically significant deviation from it. Meiotic drive may hamper introgression breeding in development of elite onion cultivars and cause instability of introgressed segment(s) in successive generations.

Further investigation of the centromeres/kinetochores using immunoGISH of the centromeric variant of histone H3 (CenH3) and kinetochore proteins and/or allele-specific expression profiling of genes involved in the establishment of the apparatus of the meiotic spindle may shed light on this phenomenon.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

DK, OS, and JB conceived and designed the study. DK and JM performed GISH analyses of mitotic and meiotic chromosomes. OS made SNP genotyping. KB-M developed all the hybrids. MD provided statistical treatments. DK drafted the manuscript with contribution from OS, JM, and JB. All authors contributed to the final version of the manuscript.

FUNDING

This study was funded by the Czech Science Foundation (grant award 20-10019S) and by the European Regional Development Fund OPVVV project "Plants as a tool for sustainable development" number CZ.02.1.01/0.0/0.0/16_019/0000827 supporting Excellent Research at CRH.

ACKNOWLEDGMENTS

We would like to thank Adam J. Lukaszewski (University of California, Riverside) for his critical reading and valuable comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.854127/full#supplementary-material

Supplementary Figure 1 | Meoitic configurations of *A. cepa* × *A. roylei* hybrids. GISH analysis showed regular homoeologous chromosome pairing in pachytene of prophase I **(A)** and metaphase I **(B)** and segregation of chromosomes to opposite poles during anaphase I **(C)** and reaching the poles in telophase I **(D)**. Total guide DNA (gDNA) of *A. roylei* was labeled with digoxigenin (green color) and sheared DNA of *A. cepa* was used as blocking DNA (blue color).

Supplementary Table 1 | Results of genomic *in situ* hybridization (GISH) and single nucleotide polymorphism (SNP) analyses in *Allium cepa* (*A. cepa*) × *Allium roylei* (*A. roylei*) hybrids.

REFERENCES

Akera, T., Chmatal, L., Trimm, E., Yang, K., Aonbangkhen, C., Chenoweth, D. M., et al. (2017). Spindle asymmetry drives non-Mendelian chromosome segregation. *Science* 358, 668–672. doi: 10.1126/science.aan0092

Anamthawat-Jónsson, K. (2001). Molecular cytogenetics of introgressive hybridization in plants. *Methods Cell Sci.* 23, 139–148. doi: 10.1007/978-94-010-0330-8_14

Bird, K. A., VanBuren, R., Puzey, J. R., and Edger, P. P. (2018). The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* 220, 87–93. doi: 10.1111/nph.15256

Budylin, M. V., Kan, L. Y., Romanov, V. S., and Khrustaleva, L. I. (2016). GISH study of advanced generation of the interspecific hybrids between *Allium cepa* L. and *Allium fistulosum* L. with relative resistance to downy mildew. *Russian J. Genet.* 50, 387–394. doi: 10.1134/S1022795414040036

Chester, M., Gallagher, J. P., Symonds, V. V., da Silva, A. V. C., Mavrodiev, E. V., Leitch, A. R., et al. (2012). Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus (Asteraceae)*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1176–1181. doi: 10.1073/pnas.1112041109

Chmatal, L., Gabriel, S. I., Mitsainas, G. P., Martinez-Vargas, J., Ventura, J., Searle, J. B., et al. (2014). Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. *Curr. Biol.* 24, 2295–2300. doi: 10.1016/j.cub.2014.08.017

Clark, F. E., and Akera, T. (2021). Unravelling the mystery of female meiotic drive: where we are. *Open Biol.* 11:210074. doi: 10.1098/rsob.210074

Courret, C., Gerard, P. R., Ogereau, D., Falque, M., Moreau, L., and Montchamp-Moreau, C. (2019). X-chromosome meiotic drive in *Drosophila simulans*: a QTL approach reveals the complex polygenic determinism of Paris drive suppression. *Heredity* 122, 906–915. doi: 10.1038/s41437-018-0163-1

Dawe, R. K., Reed, L. M., Yu, H. G., Muszynski, M. G., and Hiatt, E. N. (1999). A maize homolog of mammalian CENPC is a constitutive component of the inner kinetochore. *Plant Cell* 11, 1227–1238. doi: 10.1105/tpc.11.7.1227

de Vries, J. N., Wietsma, W. A., and de Vries, T. (1992). Introgression of leaf-blight resistance from *Allium roylei* Stearn. into onion (*Allium cepa* L). *Euphytica* 62, 127–133. doi: 10.1007/BF00037938

Edger, P. P., Smith, R., McKain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y. W., et al. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* 29, 2150–2167. doi: 10.1105/tpc.17.00010

Ferreira, M. T. M., Glombik, M., Pernickova, K., Duchoslav, M., Scholten, O., Karafiatova, M., et al. (2021). Direct evidence for crossover and chromatid interference in meiosis of two plant hybrids (*Lolium multiflorum* x *Festuca pratensis* and *Allium cepa* x *A. roylei*). *J. Exp. Bot.* 72, 254–267. doi: 10.1093/jxb/eraa455

Fishman, L., and Saunders, A. (2008). Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. *Science* 322, 1559–1562. doi: 10.1126/science.1161406

Galvan, G. A., Wietsma, W. A., Putrasemedja, S., Permadi, A. H., and Kik, C. (1997). Screening for resistance to anthracnose (*Colletotrichum gloeosporioides* Penz) in *Allium cepa* and its wild relatives. *Euphytica* 95, 173–178. doi: 10.1023/A:1002914225154

Glombik, M., Bačovský, V., Hobza, R., and Kopecký, D. (2020). Competition of parental genomes in plant hybrids. *Front. Plant Sci.* 11:200. doi: 10.3389/fpls.2020.00200

Glombik, M., Copetti, D., Bartos, J., Stoces, S., Zwierzykowski, Z., Ruttink, T., et al. (2021). Reciprocal allopolyploid grasses (*Festuca* x *Lolium*) display stable patterns of genome dominance. *Plant J.* 107, 1166–1182. doi: 10.1111/tpj.15375

Kamstra, S. A., Kuipers, A. G. J., De Jeu, M. J., Ramanna, M. S., and Jacobsen, E. (1999). The extent and position of homoeologous recombination in a distant hybrid of *Alstroemeria*: a molecular cytogenetic assessment of first generation backcross progenies. *Chromosoma* 108, 52–63. doi: 10.1007/s004120050351

Karlov, G. I., Khrustaleva, L. I., Lim, K. B., and van Tuyl, J. M. (1999). Homoeologous recombination in 2n-gametes producing interspecific hybrids of *Lilium* (*Liliaceae*) studied by genomic *in situ* hybridization (GISH). *Genome* 42, 681–686. doi: 10.1139/g98-167

Khan, N., Barba-Gonzalez, R., Ramanna, M. S., Visser, R. G. F., and Van Tuyl, J. M. (2009). Construction of chromosomal recombination maps of three genomes of lilies (*Lilium*) based on GISH analysis. *Genome* 52, 238–251. doi: 10.1139/G08-122

Khrustaleva, L. I., and Kik, C. (1998). Cytogenetical studies in the bridge cross *Allium cepa* x (*A-fistulosum* x *A-roylei*). *Theor. Appl. Genet.* 96, 8–14. doi: 10.1007/s001220050702

Khrustaleva, L. I., and Kik, C. (2000). Introgression of *Allium fistulosum* into *A-cepa* mediated by *A-roylei*. *Theor. Appl. Genet.* 100, 17–26. doi: 10.1007/s001220050003

Khrustaleva, L., Mardini, M., Kudryavtseva, N., Alizhanova, R., Romanov, D., Sokolov, P., et al. (2019). The power of genomic *in situ* hybridization (GISH) in onterspecific breeding of bulb onion (*Allium cepa* L.) resistant to downy mildew (*Peronospora destructor* [Berk.] Casp.). *Plants-Basel* 8:36. doi: 10.3390/plants8020036

Kim, S., Kim, C. W., Choi, M. S., and Kim, S. (2016). Development of a simple PCR marker tagging the Allium roylei fragment harboring resistance to downy mildew (*Peronospora destructor*) in onion (*Allium cepa* L.). *Euphytica* 208, 561–569. doi: 10.1007/s10681-015-1601-2

Kofoet, A., Kik, C., Wietsma, W. A., and de Vries, J. N. (1990). Inheritance of resistance to downy mildew (*Peronospora destructor* [Berk] Casp) from *Allium roylei* Stearn in the backcross *Allium cepa* L by (*A roylei* by *A cepa*). *Plant Breed.* 105, 144–149. doi: 10.1111/j.1439-0523.1990.tb00467.x

Kopecký, D., Horáková, L., Duchoslav, M., and Dolezel, J. (2019). Selective elimination of parental chromatin from introgression cultivars of xFestulolium (*Festuca* x *Lolium*). *Sustainability* 11:3153. doi: 10.3390/su11113153

Kopeckyì, D., Loureiro, J., Zwierzykowski, Z., Ghesquière, M., and Doležel, J. (2006). Genome constitution and evolution in *Lolium* × *Festuca* hybrid cultivars (*Festulolium*). *Theor. Appl. Genet* 113, 731–742. doi: 10.1007/s00122-006-0341-z

Kopeckyì, D., Lukaszewski, A. J., and Dolezžel, J. (2008). Meiotic behaviour of individual chromosomes of *Festuca pratensis* in tetraploid *Lolium multiflorum*. *Chromosome Res.* 16:987. doi: 10.1007/s10577-008-1256-0

Kruger, A. N., and Mueller, J. L. (2021). Mechanisms of meiotic drive in symmetric and asymmetric meiosis. *Cell. Mol. Life Sci*. 78, 3205–3218. doi: 10.1007/s00018-020-03735-0

Lopez-Lavalle, L. A. B., and Brubaker, C. L. (2007). Frequency and fidelity of alien chromosome transmission in *Gossypium* hexaploid bridging populations. *Genome* 50, 479–491. doi: 10.1139/G07-030

Lukaszewski, A. J., and Kopecky, D. (2010). The Ph1 locus from wheat controls meiotic chromosome pairing in autotetraploid rye (*Secale cereale* L.). *Cytogenet. Genome Res.* 129, 117–123. doi: 10.1159/000314279

Lukaszewski, A. J., Apolinarska, B., Gustafson, J. P., and Krolow, K. D. (1987). Chromosome pairing and aneuploidy in tetraploid triticale. *I. Stabilized karyotypes*. *Genome* 29, 554–561. doi: 10.1139/g87-093

Majka, J., Majka, M., Kopecky, D., and Dolezel, J. (2020). Cytogenetic insights into Festulolium. *Biol. Plant.* 64, 598–603. doi: 10.32615/bp.2020.095

Masoudi-Nejad, A., Nasuda, S., Mcintosh, R. A., and Endo, T. R. (2002). Transfer of rye chromosome segments to wheat by a gametocidal system. *Chromosome Res.* 10, 349–357. doi: 10.1023/a:1016845200960

Molnár-Láng, M. (2015). "The crossability of wheat with rye and other related species. In: Alien Introgression," in *Wheat: Cytogenetics, Molecular Biology, and Genomics*, eds M. Molnár-Láng, C. Ceoloni, and J. Doležel (Chem: Springer International Publishing), 103–120. doi: 10.1007/978-3-319-23494-6_4

Orellana, J., Cermeno, M. C., and Lacadena, J. R. (1984). Meiotic pairing in wheat rye addition and substitution lines. *Can. J. Genet. Cytol.* 26, 25–33. doi: 10.1139/g84-005

Pernickova, K., Linc, G., Gaal, E., Kopecky, D., Samajova, O., and Lukaszewski, A. J. (2019). Out-of-position telomeres in meiotic leptotene appear responsible for chiasmate pairing in an inversion heterozygote in wheat (*Triticum aestivum* L.). *Chromosoma* 128, 31–39. doi: 10.1007/s00412-018-0686-5

Placido, D. F., Campbell, M. T., Folsom, J. J., Cui, X. C., Kruger, G. R., Baenziger, P. S., et al. (2013). Introgression of novel traits from a wild wheat relative improves drought adaptation in wheat. *Plant Physiol* 161, 1806–1819. doi: 10.1104/pp.113.214262

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Sandler, L., and Novitski, E. (1957). Meiotic drive as an evolutionary force. *Am. Nat*. 91:105. doi: 10.1086/281969

Scholten, O. E., van Heusden, A. W., Khrustaleva, L. I., Burger-Meijer, K., Mank, R. A., Antonise, R. G. C., et al. (2007). The long and winding road leading to the successful introgression of downy mildew resistance into onion. *Euphytica* 156, 345–353. doi: 10.1007/s10681-007-9383-9

Scholten, O. E., van Kaauwen, M. P. W., Shahin, A., Hendrickx, P. M., Keizer, L. C. P., Burger, K., et al. (2016). SNP-markers in *Allium* species to facilitate introgression breeding in onion. *BMC Plant Biol.* 16:187. doi: 10.1186/s12870-016-0879-0

Sears, E. R., and Okamoto, M. (1958). Intergenomic chromosome relationships in hexaploid wheat. *Proc. Int. Congr. Genet.*, Montreal 2, 258–259.

Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *J. Stat. Soft.* 42, 1–52. doi: 10.18637/jss.v042.i07

Stevenson, M., Armstrong, S. J., Ford-Lloyd, B. V., and Jones, G. H. (1998). Comparative analysis of crossover exchanges and chiasmata in *Allium cepa* x *fistulosum* after genomic *in situ* hybridization (GISH). *Chromosome Res.* 6, 567–574. doi: 10.1023/A:1009296826942

Takahashi, C., Leitch, I. J., Ryan, A., Bennett, M. D., and Brandham, P. E. (1997). The use of genomic *in situ* hybridization (GISH) to show transmission of recombinant chromosomes by a partially fertile bigeneric hybrid. *Gasteria lutzii x Aloe aristata (Aloaceae)*, to its progeny. *Chromosoma* 105, 342–348. doi: 10.1007/s004120050193

Tsunewaki, K. (1964). Genetic studies of 6X-derivative from an 8x *Triticale*. *Can. J. Genet. Cytol.* 6, 1–11. doi: 10.1139/g64-001

Ulloa, M., Corgan, J. N., and Dunford, M. (1995). Evidence for nuclear-cytoplasmic incompatibility between *Allium fistulosum* and *Allium cepa*. *Theor. Appl. Genet*. 90, 746–754. doi: 10.1007/BF00222143

Van de Peer, Y., Ashman, T. L., Soltis, P. S., and Soltis, D. E. (2021). Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* 33, 11–26. doi: 10.1093/plcell/koaa015

van der Meer, Q. P., and de Vries, J. N. (1990). An interspecific cross between *Allium roylei* Stearn and *Allium cepa* L and its backcross to *Allium cepa*. *Euphytica* 47, 29–31. doi: 10.1007/BF00040359

van Heusden, A. W., van Ooijen, J. W., Vrielink-van Ginkel, R., Verbeek, W. H. J., Wietsma, W. A., and Kik, C. (2000). A genetic map of an interspecific cross in *Allium* based on amplified fragment length polymorphism (AFLP (TM)) markers. *Theor. Appl. Genet.* 100, 118–126. doi: 10.1007/s001220050017

Wei, K. H. C., Reddy, H. M., Rathnam, C., Lee, J., Lin, D. A. N., Ji, S. Q., et al. (2017). A pooled sequencing approach identifies a candidate meiotic driver in *Drosophila*. *Genetics* 206, 451–465. doi: 10.1534/genetics.116.197335

Wendel, J. F. (2015). The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102, 1753–1756. doi: 10.3732/ajb.1500320

Xiong, Z. Y., Gaeta, R. T., and Pires, J. C. (2011). Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7908–7913. doi: 10.1073/pnas.1014138108

Zwierzykowski, Z., Kosmala, A., Zwierzykowska, E., Jones, N., Joks, W., and Bocianowski, J. (2006). Genome balance in six successive generations of the allotetraploid *Festuca pratensis* x *Lolium perenne*. *Theor. Appl. Genet.* 113, 539–547. doi: 10.1007/s00122-006-0322-2

Zwierzykowski, Z., Zwierzykowska, E., Taciak, M., Jones, N., Kosmala, A., and Krajewski, P. (2008). Chromosome pairing in allotetraploid hybrids of *Festuca pratensis* x *Lolium perenne* revealed by genomic *in situ* hybridization (GISH). *Chromosome Res.* 16, 575–585. doi: 10.1007/s10577-008-1198-6

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 18 March 2022
doi: 10.3389/fpls.2022.858714

Introgression of *Trifolium ambiguum* Into Allotetraploid White Clover (*Trifolium repens*) Using the Ancestral Parent *Trifolium occidentale* as a Bridging Species

*Ihsan Ullah[1,2†‡], Helal A. Ansari[1‡], Isabelle M. Verry[1‡], Syed Wajid Hussain[1], Nick W. Ellison[1], Michael T. McManus[2] and Warren M. Williams[1,2]**

[1]AgResearch (Grasslands Research Centre), Palmerston North, New Zealand, [2]College of Sciences, Massey University, Palmerston North, New Zealand

White clover (*Trifolium repens*) is an allotetraploid pasture legume widely used in moist temperate climates, but its vulnerability to drought, grazing pressure and pests has restricted its wider use. A related species, Caucasian clover (*Trifolium ambiguum*), is a potential source of resistances to drought, cold, grazing pressure and pests that could potentially be transferred to white clover by interspecific hybridization. Although direct hybridization has been achieved with difficulty, the hybrids have not been easy to backcross for introgression breeding and no interspecific chromosome recombination has been demonstrated. The present work shows that interspecific recombination can be achieved by using *Trifolium occidentale*, one of the ancestral parents of *T. repens*, as a bridging species and that large white clover breeding populations carrying recombinant chromosomes can be generated. A 4*x* hybrid between *T. ambiguum* and *T. occidentale* was crossed with *T. repens* and then backcrossed for two generations. Five backcross hybrid plants with phenotypes appearing to combine traits from the parent species were selected for FISH-GISH analyses. Recombinant chromosome segments from *T. ambiguum* were found in all five plants, suggesting that recombination frequencies were significant and sufficient for introgression breeding. Despite early chromosome imbalances, the backcross populations were fertile and produced large numbers of seeds. These hybrids represent a major new resource for the breeding of novel resilient forms of white clover.

Keywords: interspecific hybridization, backcross breeding, genomic *in situ* hybridization, fluorescence *in situ* hybridization, bridge cross

OPEN ACCESS

Edited by:
Ruslan Kalendar, University of Helsinki, Finland

Reviewed by:
Devendra Ram Malaviya, Indian Grassland and Fodder Research Institute (ICAR), India
Rainer Hofmann, Lincoln University, New Zealand
Baoliang Zhou, Nanjing Agricultural University, China

***Correspondence:**
Warren M. Williams
warren.williams@agresearch.co.nz

***Present address:**
Ihsan Ullah, Pakistan Agricultural Research Council, Islamabad, Pakistan

These authors have contributed equally to this work

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

Received: *20 January 2022*
Accepted: *16 February 2022*
Published: *18 March 2022*

Citation:
Ullah I, Ansari HA, Verry IM, Hussain SW, Ellison NW, McManus MT and Williams WM (2022) Introgression of Trifolium ambiguum Into Allotetraploid White Clover (Trifolium repens) Using the Ancestral Parent Trifolium occidentale as a Bridging Species. Front. Plant Sci. 13:858714. doi: 10.3389/fpls.2022.858714

INTRODUCTION

In this study, full sub-genomes ($x = 8$) are coded as follows: A from 4*x Trifolium ambiguum*, O from *Trifolium occidentale*, P^r, O^r from *Trifolium repens*, representing the ancestral parent genomes from *Trifolium pallescens* and *T. occidentale,* respectively (Williams et al., 2012). R also from *T. repens* in cases where sub-genome origin is unclear. Partial sub-genomes are coded

as the expected number of chromosomes (assuming normal segregation), e.g., O_4 (four *T. occidentale* chromosomes).

Interspecific hybridization is a useful tool for the agronomic improvement of crops by enabling genetic diversity to be incorporated from wild and non-agronomic species. White clover, *Trifolium repens* L. ($P^rP^rO^rO^r$, $2n = 4x = 32$), although an important pasture legume in moist temperate climates throughout the world, is vulnerable to stresses, including drought (Barbour et al., 1996; Brink and Pederson, 1998), grazing pressure (Forde et al., 1989), and pests (Alconero et al., 1986; Gaynor and Skipp, 1987; Latch and Skipp, 1987; Pederson and McLaughlin, 1989; Pederson and Windham, 1989). Breeding efforts for stress tolerances using the limited genetic variation within the primary gene pool have met with limited success (Williams et al., 2007).

Trifolium ambiguum M. Bieb. (Caucasian or Kura clover) is a rhizomatous species with resistance to drought, cold and pest attacks but also with several limitations that make it unsuitable for intensive grazing systems (Bryant, 1974; Spencer et al., 1975; Dear and Zorin, 1985; Mercer, 1988; Pederson and McLaughlin, 1989; Pederson and Windham, 1989; Sheaffer et al., 1992). However, it would be desirable if the resistance traits could be transferred to white clover by interspecific hybridization. Natural hybridization with white clover does not occur due to strong post hybridization barriers and, even with embryo rescue or ovule culture methods, very few fertile 4*x* hybrids have been produced by crossing 4*x* *T. ambiguum* (AAAA, $2n = 4x = 32$) with *T. repens* (Williams and Verry, 1981; Meredith et al., 1995; Williams, 2014). Furthermore, when these AAP^rO^r 4*x* hybrids were backcrossed to *T. repens*, most of the progeny were 6*x*, ($AAP^rP^rO^rO^r$) as the result of functional 2*n* gametes from the F_1 hybrid (Anderson et al., 1991; Meredith et al., 1995), although rare 4*x* and aneuploid forms were also reported (Williams et al., 1982; Anderson et al., 1991). A second backcross produced 5*x* BC_2 hybrids ($AP^rP^rO^rO^r$) with four sub-genomes ($4x = 32$) of *T. repens* and one sub-genome ($x = 8$) of *T. ambiguum* (Meredith et al., 1995). Interspecific chromosome pairing and recombination in these hybrids could potentially result in introgression and the transfer of *T. ambiguum* traits into *T. repens* by further backcrossing (Williams, 1987a; Meredith et al., 1995). However, such chromosome pairing occurred only at very low frequency (Williams et al., 1982; Anderson et al., 1991; Meredith et al., 1995) and to-date there has been no report of this strategy leading to introgression through interspecific meiotic recombination. Nevertheless, a breeding program using this backcrossing strategy was carried out in the United Kingdom, leading ultimately to cv. 'Aberlasting' (Abberton, 2007; Marshall et al., 2015).

Tetraploid (4*x*) *T. ambiguum* can also be crossed with a close relative and progenitor of white clover, *T. occidentale*, using embryo rescue. When colchicine-doubled 4*x* *T. occidentale* (OOOO) was used, the 4*x* (AAOO, *T. ambiguum* × *T. occidentale*) F_1 hybrids were fertile and showed evidence of frequent interspecific meiotic chromosome pairing (Williams et al., 2019a). In addition, the hybrids were able to be crossed to white clover to generate tri-species (P^rO^rAO) hybrids (Williams et al., 2019a). The knowledge that white clover is an allotetraploid and that one of its ancestral sub-genomes was from *T. occidentale* (Ellison et al., 2006; Williams et al., 2012), led to the suggestion of using *T. occidentale* as a species bridge to achieve introgression of *T. ambiguum* into white clover. The rationale is that interspecies recombination in *T. ambiguum* × *T. occidentale* hybrids would create recombinant chromosomes having *T. occidentale* centromeres with *T. ambiguum* genomic segments on the arms, and vice versa. Then crossing of these hybrids, having recombinant chromosomes with *T. repens*, followed by further backcrossing to *T. repens* might lead to the introgression of *T. ambiguum* genome segments into *T. repens* genomes as a result of the substitution of, or meiotic exchange between, the *T. occidentale* chromosomes from both parents (Williams, 2014; Williams et al., 2019a).

The proposed introgression breeding strategy (**Figure 1**) would start with 4*x* tri-species *T. repens* × (*T. ambiguum* × *T. occidentale*) hybrids, derived from crosses between white clover plants and 4*x* *T. ambiguum* × *T. occidentale* hybrids. These would be backcrossed to white clover and the progenies screened for desired *T. ambiguum* traits. Selected plants would then be further backcrossed to white clover and the process completed until new forms of white clover carrying introgressed *T. ambiguum* characteristics were produced.

Here, we test the feasibility of this strategy by carrying out a hybridization and backcrossing program starting with a tri-species hybrid and backcrossing it to white clover. Advanced families were screened for *T. ambiguum* traits and for chromosome pairing and genomic exchange using both conventional and molecular (FISH-GISH) cytogenetics. The frequent incorporation of chromosome segments from 4*x* *T. ambiguum* into hybrids with *T. repens* by meiotic recombination is reported for the first time.

MATERIALS AND METHODS

The starting materials for the study (**Figure 1**) were first crosses (F_1) between a *T. repens* plant (Red-1) and a 4*x* *T. ambiguum* × 4*x* *T. occidentale* hybrid (BL-1), with the expected genomic composition P^rO^rAO, as described by Williams et al. (2019a) and listed in **Table 1**.

Plant Hybridization and Selection Procedures

Seeds were scarified with sandpaper and germinated on moist filter paper in Petri dishes at room temperature. The seedlings were planted in plastic pots containing potting mix (peat and sand in equal ratio) in a greenhouse under natural daylight without heating. Plants were grown in an insect-free unheated glasshouse at the AgResearch Grasslands Research Centre, Palmerston North, New Zealand. The plants were watered as needed and fed once fortnightly in solution form with a commercially available complete nutrient solution (Yates Thrive®).

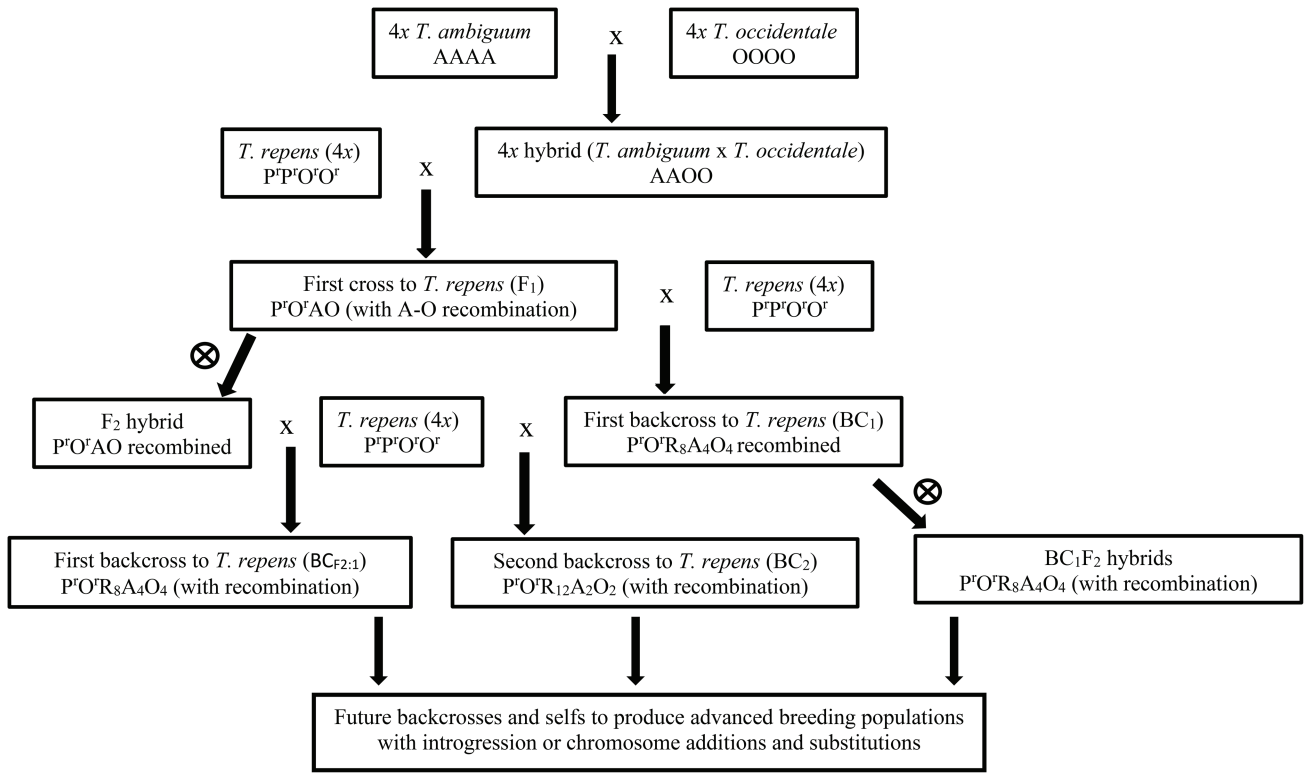


FIGURE 1 | Flow chart of the hybridization and breeding program. A, O, P^r, O^r represent haploid sub-genomes ($x = 8$) of *Trifolium ambiguum*, *Trifolium occidentale*, ancestral *Trifolium pallescens* from *Trifolium repens*, ancestral *T. occidentale* from *T. repens*, respectively. R represents chromosomes from *T. repens* where the sub-genomes could not be defined. Partial sub-genomes are designated by subscripts, e.g., O_4 represents four *T. occidentale* chromosomes. Numbers are those expected assuming random assortment. Female parents are generally on the left side, with one exception (designated ♂). BC_1 families were produced using crosses in either direction (designated ♂♀). ⊗ indicates self-pollination.

TABLE 1 | Flow cytometric estimates of ploidies of 4*x* F_1 (Red-1 × BL-1) hybrids [*Trifolium repens* × (4*x Trifolium ambiguum* × 4*x Trifolium occidentale*)], with the expected genomic composition P^{r}O^{r}AO.

Hybrid	Pedigree	Estimated ploidy (*x*)
F_1-119	(Red-1 × BL-1)-1	3.9
F_1-123	(Red-1 × BL-1)-6	3.8
F_1-124	(Red-1 × BL-1)-7	3.6
F_1-125	(Red-1 × BL-1)-8	3.9
F_1-136	(Red-1 × BL-1)-19	3.9

Crosses were made after forceps emasculation (Williams, 1954) of the female parent by brushing anthers of the male parent over the exposed stigmas. This was usually repeated daily for the subsequent 2–3 days. Self-pollination was done by gently rolling inflorescences between thumb and fingers and repeating over 2–3 days. Dried inflorescences (heads) were harvested and threshed between two pieces of corrugated rubber. Self-compatibility of all plants used as female parents was determined by using the simple techniques suggested by Williams (1987b). To estimate male fertility, anthers were squashed on a glass slide and the pollen stained in 2% (w/v) aceto-carmine. The percentage of full stained pollen grains was recorded among a minimum of 300 counted at 200x magnification.

Selection of plants for evaluation and crossing was based on vigour, flow cytometric determinations of ploidy, somatic chromosome counts and above-ground appearance combining characteristics from the parent species. Ploidy analysis was based on DNA content using flow cytometry using the procedure described by Williams et al. (2011). For the above-ground appearance, emphasis was placed on *T. repens*-like plants with the additional appearance of one or more *T. ambiguum* characteristics – sometimes coarser or more elongated leaflets, determinate stems with weak nodal rooting and thick root bases visible at the soil surface.

The F_1 plants were backcrossed to white clover plants to produce BC_1 families and were also inter-crossed or self-pollinated to produce F_2 families (**Figure 1**; **Table 1**). One hundred seeds from the crosses to white clover were germinated and the seedlings grown in pots in the greenhouse. Among the most robust survivors, 19 were subjected to further fertility and ploidy analyses. From these, six plants (five BC_1 and one F_2, **Table 2**) were selected for detailed characterization and backcrossing based on their vigour, fertility, estimated chromosome number and above-ground appearance combining characteristics from the parent species. Subsequently, the BC_1 plants were backcrossed to *T. repens* to produce BC_2 families (**Figure 1**) and were also self-pollinated to give BC_1F_2 families

(**Figure 1**). The F_2 plants were crossed with *T. repens* to produce BC_1 families, here designated $BC_{F2:1}$ to indicate derivation from F_2 (**Figure 1**). Approximately 10 plants of each of these advanced families were grown in pots in the glasshouse and plants (**Table 3**) selected as before underwent further analysis.

Most of the white clover plants used as male parents carried co-dominant purple anthocyanin leaf color markers, whereas those used as females were usually unmarked (i.e., green). C numbers relate to accessions from the Margot Forde Germplasm Centre, Palmerston North.

DNA Analysis, Cytology and Molecular Cytogenetics

The procedures were carried out in two separate experiments. In the first, root tip cells and meiotic PMCs from BC_2-126 were subjected to FISH-GISH using 5S rDNA and total genomic DNA from *T. ambiguum*. The results showed evidence of recombination and so four more plants were sampled. These included two BC_1 plants (BC_1-120, BC_1-132) and two BC_2 plants (Kopu II × BC_1-132, BC_2-133) which underwent FISH using 5S and 18S rDNA followed by sequential FISH-GISH with the centromeric probe TrR350 (Ansari et al., 2004) and *T. ambiguum* total genomic DNA. The methods for somatic chromosome preparations, Giemsa staining, DNA preparation and FISH using 5S and 18S rDNA probes were as described by Ansari et al. (1999). DNA preparation and FISH using TrR350 were as described by Ansari et al. (2004). The GISH procedures were described in Ansari et al. (2008). Total genomic DNA of *T. ambiguum* was extracted from 2*x* cv. Summit. Before meiotic GISH preparations were implemented, an additional step of pepsin treatment was included to remove the proteinaceous background caused by the dense cytoplasm of pollen mother cells (PMCs). Two plants (BC_1-120, BC_2-126) were analyzed for meiotic metaphase I configurations using conventional cytology as described by Williams et al. (2011).

Hybrid Plant Phenotypes

After their selection based on the characteristics described above, the plants listed in **Tables 2**, **3** were clonally propagated as cuttings and grown in randomized field trials and measured for a range of phenotypic traits. The methods and results are provided as **Supplementary Material**. Because they were done later, the descriptions did not influence the selection process. However, they did give retrospective indications of the variability of the hybrids in vigour, fertility and the extent to which they had combined traits from the parent species.

RESULTS

Flow cytometry estimates indicated that the five starting P^rO^rAO parent plants were all near-4*x* (**Table 1**). Two plants, F_1-125 and (Red-1 × BL-1)-9 (the latter not used in crosses) were checked for chromosome numbers and were both $2n = 32$ with the expected three satellited chromosomes (one from each of *T. ambiguum*, *T. occidentale,* and *T. repens*).

Following the crosses shown in **Table 1**, five BC_1 plants were selected for further analysis (**Table 2**). The expected genomic composition of these plants was $P^rO^rR(A_4O_4)$ (32 chromosomes), including partial genomes from *T. ambiguum* and *T. occidentale*. One of the *T. repens*-derived genomes (designated R) was expected to be an unspecified mixture of chromosomes derived from both sub-genomes of white clover. The Giemsa-stained somatic chromosome counts in these hybrids ranged from 31 to 34. Also selected for analysis was F_2-133, which had 35 chromosomes, and was derived from a self-pollination. Pollen staining results (30–67%) indicated low-medium male fertilities. These plants were self-pollinated and, as they had leaf color markers, they were also used as male parents to pollinate a totally green *T. repens* plant (Kopu-II, a commercial white clover variety). All these selected plants produced seed when crossed with white clover (**Table 2**). All except F_2-133 also proved to be self-compatible, setting large quantities of seed following self-pollination.

The resulting BC_2 and BC_1F_2 progeny of 63 plants from a mixture of backcrosses and selfs were grown in the greenhouse and, using plant phenotypes, including leaf color markers, seven plants were selected for further analysis (**Table 3**). The expected genomic formula in the three backcrosses, BC_2-126, −129 and-131, was $P^rO^r(R_{12}A_2O_2)$ and chromosome counts of 33, 32 and 32, respectively, all were within the range of expectations. $BC_{F2:1}$−136 resulted from the backcross of F_2-133 (35 chromosomes) with white clover. The female gamete from white clover in this case was fertilized by a male gamete from F_2-133 having 19 chromosomes. BC_1F_2−137, −138 and − 140 were the respective self-progeny of BC_1-128, −131 and − 132 and had somatic chromosome numbers of 32, 34 and 32, within the

TABLE 2 | BC_1 and F_2 plants selected for detailed analysis, including actual chromosome numbers, pollen fertility, seeds obtained per inflorescence (approx. 50 florets) when crossed with *T. repens*, and whether self-compatible (SC) or self-incompatible (SI).

Plant	Parentage (♀ × ♂)	Actual 2*n* number	Pollen staining (%)	Seeds/infl.	SC/SI
BC_1-120	(PB-5 × F_1-123)-1	33	43	7	SC
BC_1-128	(F_1-119 × PB-5)-1	31	30	69	SC
BC_1-130	(F_1-119 × PB-5)-2	31	67	66	SC
BC_1-131	(F_1-119 × C2418-2)-1	34	58	23	SC
BC_1-132	(F_1-136 × PB-17)-1	33	66	29	SC
F_2-133	F_1-119 self-5	35	43	3	SI

PB white clover plants were purple leaved.

TABLE 3 | Selected BC_2, $BC_{F2:1}$ and BC_1F_2 progeny with expected and actual chromosome numbers and pollen fertilities.

Hybrid	Pedigree	Expected 2*n* number	Actual 2*n* number	Pollen staining (%)
BC_2-126	(Kopu II-906 × BC_1-120)-1	32–33	33	51
BC_2-129	(Kopu II-901 × BC_1-128)-10	31–32	32	94
BC_2-131	(Kopu II-910 × BC_1-130)-4	31–32	32	70
$BC_{F2:1}$–136	(Kopu II-918 × F_2-133)-5	32–35	35	32
BC_1F_2–137	BC_1-128-Self-1	30–32	32	62
BC_1F_2–138	BC_1-131-Self-8	34	34	61
BC_1F_2–140	BC_1-132-Self-3	32–34	32	52

expected ranges (**Table 3**). Male fertility in all these hybrids was above 50%, except for $BC_{F2:1}$–136 (32%).

Molecular Cytogenetic Analysis of BC_2-126 Somatic Cells

A BC_2 hybrid, BC_2-126 with 33 chromosomes, was derived from a backcross of BC_1-120 to white clover. DAPI staining of somatic cells of BC_2-126 revealed the presence of one very large chromosome and two chromosomes with decondensed nucleolus organizer regions (NOR; **Figure 2A**). GISH using labeled genomic DNA from *T. ambiguum* (green) painted the very large chromosome, indicating that it was *T. ambiguum*-derived (**Figures 2A,B**). GISH consistently also revealed recombination between *T. ambiguum* chromosomes and two *T. repens* or *T. occidentale*-derived chromosomes. This manifested as one large and one smaller green signal on two chromosomes (**Figure 2B**). This hybrid gave six signals when probed with 5S rDNA. Two 5S signals were on a pair of *T. repens* or *T. occidentale* chromosomes bearing de-condensed NOR sequences and three were on non-NOR *T. repens* or *T. occidentale* chromosomes. The *T. ambiguum*-derived chromosome also carried a 5S rDNA signal (**Figures 2A,B**).

Chromosome Pairing Analysis in BC_2-126

In conventionally stained metaphase I cells of BC_2-126 ($2n = 33$) chromosome pairing was highly variable from cell to cell (**Table 4**). On average, approximately 73% of the chromosomes paired as bivalents with the remainder as univalents and multivalents. Sequential FISH-GISH analysis was carried out on meiotic chromosome preparations of BC_2-126 using labeled genomic DNA from *T. ambiguum* (green) and 5S rRNA (red). Among 21 PMCs analyzed at metaphase I, 9 (43%), had the *T. ambiguum*-derived chromosome unpaired (**Figures 2C,D**) while in the remaining 12 cells this chromosome paired with chromosomes from either white clover or *T. occidentale*. In six of these the *T. ambiguum* chromosome associated as a bivalent and in six it paired to form multivalents. The recombinant chromosomes also paired in bivalents or multivalents with *T. repens* or *T. occidentale* chromosomes (**Figure 2E**).

At anaphase I, univalents frequently showed precocious separation of sister chromatids, followed by movement to opposite poles. This is illustrated for a *T. ambiguum* chromosome in **Figures 2F,G** where chromatid separation was clear because this chromosome also had a 5S red signal and, after splitting, the two 5S signals could be seen moving to opposite poles (**Figure 2G**), or sometimes lagging behind. A laggard chromosome is shown in **Figures 2F,G**. In many cases at telophase II, the lagging chromatids did not become part of the tetrads but formed micronuclei (**Figures 2H,I**), leading to meiotic elimination of chromosomes including, in this case, the *T. ambiguum*-derived chromosome.

Molecular Cytogenetic Analysis of BC_1-120 Somatic Cells

Somatic cells of BC_1-120 showed 33 chromosomes when DAPI-stained, including two very large chromosomes (**Figure 3A**). These proved to be *T. ambiguum*–derived as shown by GISH using genomic DNA from *T. ambiguum* (**Figure 3C**). GISH also identified two recombinant chromosomes: one with centromeric TrR350 and with most of one arm consisting of *T. ambiguum* DNA and the other lacking TrR350 and having *T. ambiguum* DNA spanning the centromeric region (**Figure 3C**). FISH using the TrR350 probe also gave pericentromeric signals on all except three *T. repens* and *T. occidentale* chromosomes and the *T. ambiguum* chromosomes. FISH using the 18S rDNA probe labeled the 18-26S rDNA regions of the NORs which were highly decondensed and spanning chromosome segments that otherwise appear to be unattached (**Figures 3A,B**). FISH using the 5S rDNA probe produced signals on six chromosomes. One was on the long arm of one of the *T. ambiguum* chromosomes (**Figures 3B,C**) and two were on two *T. repens* or *T. occidentale* NOR chromosomes with 18-26S rDNA on the opposite arms. The remainder were also on *T. repens*- or *T. occidentale*-derived chromosomes (**Figures 3B,C**).

Chromosome Pairing Analysis in BC_1-120

Conventional cytogenetic analysis of chromosome pairing in BC_1-120 revealed that an average of just over half the chromosomes paired as bivalents and there were significant numbers of univalents and multivalents including a very low frequency of apparent pentavalents (Vs; **Table 4**).

Molecular Cytogenetic Analysis of Somatic Cells of BC_1-132 and a Derived BC_2 Plant

Hybrid BC_1-132, with different parents from BC_1-120 (**Table 2**), was also subjected to FISH and sequential FISH-GISH treatments. This hybrid had 33 chromosomes. GISH using *T. ambiguum* genomic DNA revealed the presence of two large *T. ambiguum* chromosomes and two recombinant chromosomes, each apparently with one *T. ambiguum* arm (**Figures 3D–F**). FISH with TrR350 showed that one of the recombined chromosomes was clearly labeled in the pericentromeric region while the other showed only a very

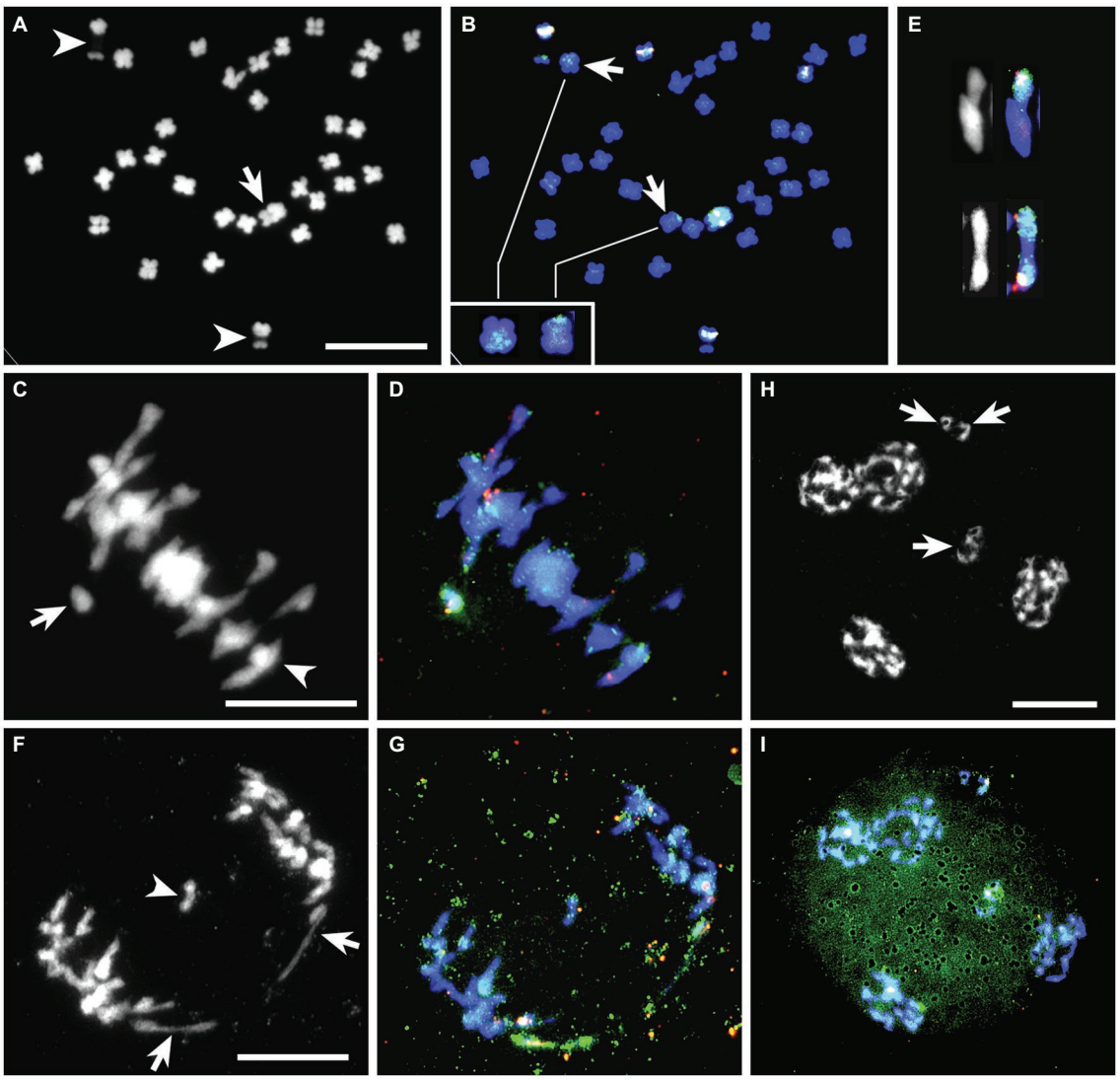


FIGURE 2 | BC_2-126 sequential FISH-GISH results with 5S rDNA (pink, white) and *T. ambiguum* genomic DNA (green). **(A,B)**, A somatic cell **(A)** DAPI stained (gray scale) and **(B)** the same cell after FISH-GISH. In **(A)**, the *arrow* depicts a single *T. ambiguum* chromosome, and *arrowheads* show NOR-chromosomes from *T. repens* or *T. occidentale*. In **(B)**, the *arrows* show recombinant chromosomes, which are magnified in the inset. **(C,D)**, A meiotic metaphase I (MI) cell with the same probes as in **(A,B)**. In **(C)**, the arrow shows the unpaired *T. ambiguum* chromosome as a univalent and the arrowhead marks a trivalent involving two recombinant chromosomes and a 5S rDNA-bearing chromosome from *T. repens* or *T. occidentale*. **(E)**, Cut-outs from other MI cells of a trivalent (top) and a bivalent (bottom) involving *T. ambiguum* or recombinant chromosomes. **(F,G)**, Anaphase I cell with the same probes as in **(A,B)**. In **(F)**, the *arrows* show the two halves of the *T. ambiguum* chromosome which has undergone precocious separation of the sister chromatids (PSSC) and the *arrowhead* marks a laggard chromosome. **(H,I)**, A telophase II with the same probes as in **(A,B)**. In **(H)**, the *arrows* mark micronuclei with chromosomes likely to be eliminated from the gametes. In **(I)**, the elimination of the lone *T. ambiguum* chromosome is shown. Bars 10 μm.

minor signal (**Figure 3F**). FISH using the 5S and 18S rDNA probes identified two *T. repens* or *T. occidentale* chromosomes with decondensed NORs (**Figures 3D,E**). One of the *T. ambiguum* chromosomes carried an 18-26S rDNA (NOR) locus that was condensed in all observed cells (**Figures 3E,F**). There were also two *T. repens* or *T. occidentale* chromosomes with 5S signals (**Figures 3E,F**).

BC_1-132 was backcrossed as male to a white clover plant Kopu II-905 and one BC_2 progeny plant (Kopu II-905 × BC_1-132)-6 was subjected to FISH and sequential FISH-GISH procedures. This plant was highly fertile (80% pollen staining) and had 32 chromosomes. GISH showed that there were no complete *T. ambiguum* chromosomes, but *T. ambiguum* DNA occurred as one full arm of a single recombinant chromosome. This chromosome had TrR350 in the centromeric region (**Figure 3I**). FISH with the rDNA probes revealed, as in BC_1-132, two *T. repens* or *T. occidentale* chromosomes with decondensed NORs and two further *T. repens* or *T. occidentale* chromosomes with 5S signals (**Figures 3G–I**).

TABLE 4 | Meiotic chromosome associations at diakinesis/metaphase I in PMCs of BC_1-120 ($2n = 33$) and BC_2-126 ($2n = 33$).

Plant	No PMCs	Mean frequency (range) of meiotic configurations					
		I	II	III	IV	V	Pollen stain (%)
BC_1-120	35	2.5 (0–7)	9.2 (5–15)	2.0 (0–5)	1.3 (0–4)	0.1 (0–1)	43
BC_2-126	47	3.5 (0–9)	12.0 (6–15)	1.1 (0–4)	0.6 (0–2)	0	51

Molecular Cytogenetic Analysis of BC_2-133 Somatic Cells

The analysis of a third BC_2 hybrid plant (BC_2-133) is presented in **Figures 3J–L**. Among the 32 chromosomes there was one *T. ambiguum* chromosome with a 5S signal (**Figures 3J,K**), and two NOR chromosomes and three 5S chromosomes from *T. repens* or *T. occidentale* (**Figures 3J,K**). Two recombinant chromosomes are shown, both apparently with nearly whole *T. ambiguum* arms attached to TrR350 labeled centromeres (**Figure 3L**).

Hybrid Plant Phenotypes

Plant morphology traits were measured in separate outdoor experiments for a small sample of plants in each of the BC_1 and the BC_2 and BC_1F_2 generations (**Supplementary Material**). In general, the BC_2 plants were more white clover-like than the BC_1 plants, and the hybrid populations were highly variable in phenotype and contained plants with good vigor and fertility while combining traits from the parent species. No rhizomes were observed as these take up to 18 months to develop and the experiments were of short duration.

Figure 4 compares the root systems of hybrid BC_2-126 with two white clover plants. The hybrid (**Figure 4A**) showed sparse thick nodal roots while the white clovers both had a concentration of fine nodal roots (**Figures 4B,C**).

DISCUSSION

In the present study we tested the hypothesis that recombination between *T. ambiguum* and *T. occidentale* chromosomes can occur and the recombinant chromosomes can be transmitted by hybridization to *T. repens*. White clover-like plants with some *T. ambiguum* traits were produced and, in all five backcross hybrids that were studied in detail at the chromosome level, there was molecular evidence of *T. ambiguum* introgression through chromosomal recombination. This is the first report of chromosomal recombination involving these species, and the results experimentally confirm the hypothesis. This opens the way for the breeding of new forms of white clover incorporating parts of the of *T. ambiguum* genome by using *T. occidentale* as a bridging species.

To-date, only first and second backcross generations have been analyzed and these are still genomically unrefined. All but one of the analyzed plants still carried whole *T. ambiguum* chromosomes and, in addition to recombinant chromosomes, there were unresolved chromosome substitutions, deletions and additions. These can be expected to be resolved by further backcross breeding and selection. Despite the genomic imbalances, the hybrids from F_1 onward were fertile and able to produce both viable pollen and seeds in sufficient numbers to enable large numbers of progeny to be produced by normal pollination.

The F_1 parent plants listed in **Table 1** (putatively P^rO^rAO) were, with one exception, likely to have been true to expectation. However, the AAOO parent of these plants (BL-1) was shown to have 15–17 disjunction at meiosis I in about 10% of PMCs (Williams et al., 2019a), and so some aneuploidy was probable. One F_1 parent with a flow cytometric estimate of 3.9 was confirmed as $2n = 32$. The remainder, with estimates of 3.8–3.9 were also probably euploid. However, F_1-124, had an estimate of 3.6 and may have been aneuploid.

Molecular Cytogenetic Analysis of Somatic Chromosomes of the Hybrids

In conventional cytological preparations, the somatic chromosomes of *T. ambiguum*, *T. repens*, and *T. occidentale* were small and metacentric to sub-metacentric (Ansari et al., 1999), and so, apart from the NOR-chromosomes, the individual chromosome pairs were indistinguishable from one another within genomes. The chromosomes of *T. ambiguum* were notably larger than those of the other species and could often be recognized by their size in conventional somatic cell preparations of hybrids (e.g., **Figures 3A,J**).

The use of 5S and 18S rDNA probes together enabled identification of two chromosome pairs in each of *T. repens*, *T. occidentale*, and *T. ambiguum* (Ansari et al., 1999). The NOR-chromosomes of *T. repens* and *T. occidentale* both have a 5S signal on the long arm opposite to the NOR on the short arm (**Figures 3B,E,H,K**) and could not be distinguished from each other. However, the *T. ambiguum* NOR-chromosome could be identified as it lacked a 5S signal (**Figures 3E,F**). The 5S rDNA probe could detect one further chromosome pair from *T. repens* (a single large signal on the long arm) and one from *T. occidentale* (a minor signal on the short arm) while the remaining six chromosome pairs of both species remain unmarked and indistinguishable (Ansari et al., 1999). In the present study, resolution of these markers was variable, and no attempt was made to distinguish between the 5S signals from *T. repens* and *T. occidentale* in the hybrid genomes. *T. ambiguum* also had a distinctive chromosome pair with a 5S locus (**Figures 3B,C,K,L**). Use of the FISH probe TrR350 (Ansari et al., 2004) could usefully distinguish most *T. ambiguum* chromosomes (one pair with TrR350) from *T. repens* (all chromosomes) and *T. occidentale* (all or most chromosomes).

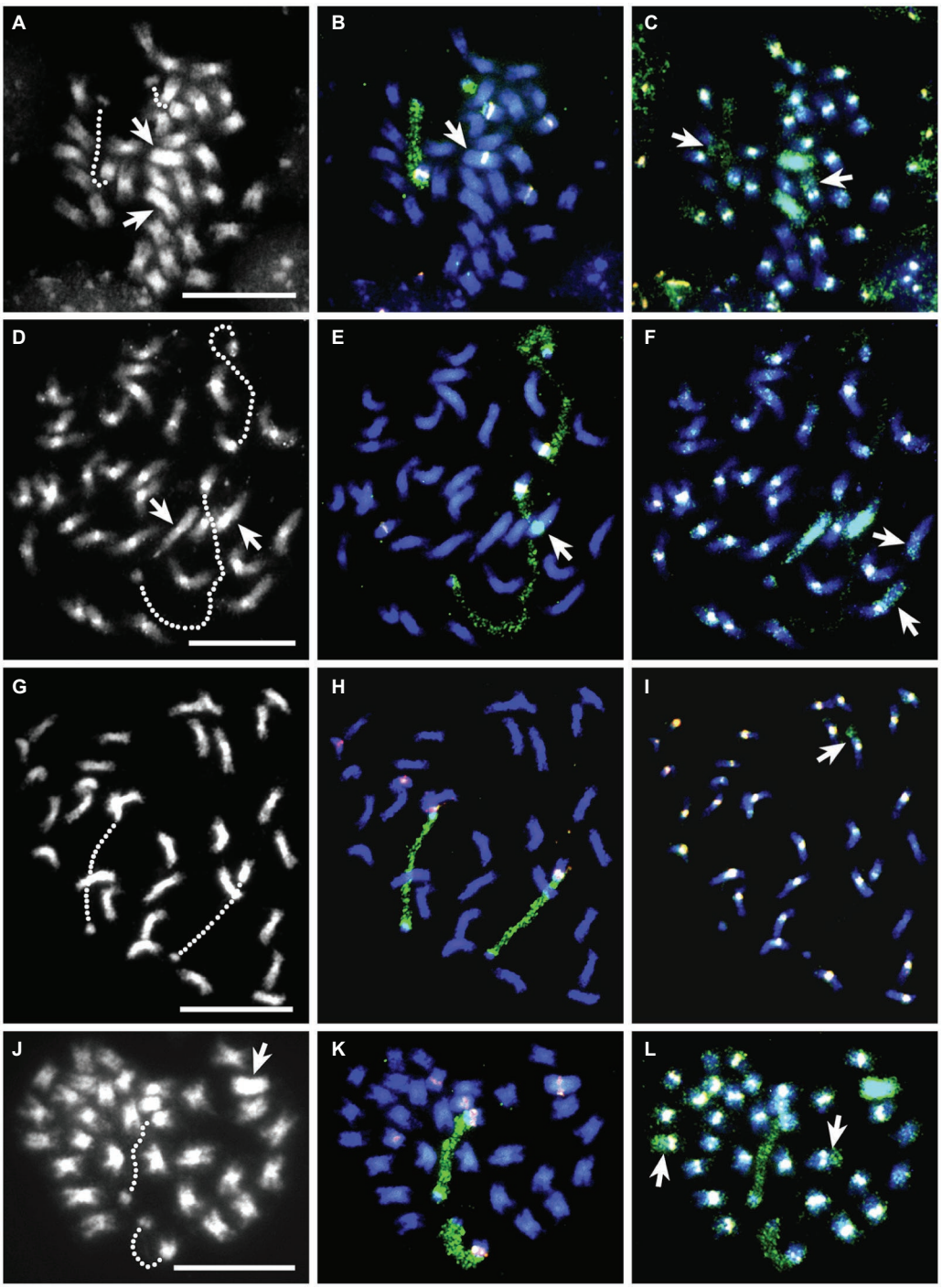


FIGURE 3 | FISH-GISH results for somatic cells of four interspecific hybrids. The left column is DAPI stained (gray scale), the center column the same cell with 5S rDNA (pink) and 18S rDNA (green) FISH probes and the right column the same cell with sequential FISH-GISH using the centromeric probe TrR350 (white) and *T. ambiguum* genomic DNA (green). Hatched lines in the left column show the decondensed NOR-regions of each pair of NOR-chromosomes from *T. repens* or *T. occidentale* that show as green 18S rDNA signals in the middle column. **(A–C)**, BC_1-120; in **(A)**, the arrows show two large *T. ambiguum* chromosomes that were apparent following GISH in **(C)**. In **(B)**, the arrow indicates that one of the *T. ambiguum* chromosomes carried a 5S rDNA signal. In **(C)**, the arrows mark recombinant chromosomes, one of which lacks a TrR350 centromeric signal. **(D–F)**, BC_1-132; in **(D)**, the arrows show two large *T. ambiguum* chromosomes that were apparent following GISH in **(F)**. In **(E)**, the arrow shows that one of the *T. ambiguum* chromosomes carried a condensed NOR region (detected by the 18S rDNA probe). In **(F)**, the arrows mark recombinant chromosomes, one of which has a very weak TrR350 centromeric signal. **(G–I)**, Kopu II × BC_1-132; in **(I)**, the arrow shows a recombinant chromosome. **(J–L)**, BC_2-133; in **(J)**, the arrow shows a large *T. ambiguum* chromosome which is visible in **(L)** and carries a 5S rDNA signal **(K)**. In **(L)**, the arrows show two recombinant chromosomes, both with strong centromeric TrR350 signals. Bars 10 μm.

Application of the GISH technique to the hybrid genomes using a *T. ambiguum* genomic DNA probe was potentially able to detect the presence of any part of the *T. ambiguum* (A) genome, including intact A chromosomes and chromosomes involving A recombinations.

Aneuploidy and Chromosome Imbalances

All the analyzed BC_1 and F_2 progeny derived from the F_1 parents were aneuploid (**Table 2**). Three BC_1 and one F_2 plant derived from a single F_1 (F_1-119) varied both up and down from the expected $2n = 32$ (**Table 2**). Thus, whatever the parental chromosome number, all the gametes involved in producing the analyzed BC_1s from this plant varied from the expected $n = 16$ (The other parent, *T. repens*, reliably produces euploid $n = 16$ gametes). Meiotic chromosome pairing was not studied in the P^rO^rAO hybrids, but genomically balanced gametes could not occur, and it is highly likely that several chromosomes (especially A chromosomes) would have been unpaired at meiosis, leading to aneuploid gametes. Consequently, progeny plants would have had unpredictable numbers of chromosomes from each of the four contributing sub-genomes. Thus, while the average expected gametic constitution would have been $P^r_4O^r_4A_4O_4$, it was apparent that none of the tested BC_1 plants were derived from such gametes, and that chromosome additions, deletions and substitutions were inevitable in the following generations. This was supported by meiotic pairing analysis of one of these plants (BC_1-120, **Table 4**), where univalent and multivalent meiotic chromosome configurations were frequent. In the next generation, the moderately fertile BC_2 plant with two introgressed segments from *T. ambiguum* (BC_2-126, $2n = 33$) had one full *T. ambiguum* chromosome and 32 combined *T. occidentale* and *T. repens* chromosomes. However, these 32 were not a balanced set as evidenced by the presence at meiosis in this plant of more than one univalent, on average, as well as trivalents (**Table 4**). All three BC_2 plants had 32 or 33 chromosomes and moderate to high fertilities (**Table 3**), suggesting that any sub-genomic imbalances did not drastically reduce fertility. A high frequency of asynapsis (i.e., univalents) at meiosis can be associated with the generation of $2n$ gametes in interspecific hybrids (Ansari et al., 2022). However, in the materials analyzed, no products of $2n$ gametes were observed.

FISH-GISH analyses of BC_1-120 and BC_1-132 showed that, in both cases, there were 33 chromosomes, but with only two rather than the expected four A chromosomes (**Table 5**). The A chromosomes had apparently been lost at a greater rate than expected, probably due to lack of pairing partners at meiosis I. The presence of 33 chromosomes indicated that chromosomes derived from *T. repens* and *T. occidentale* had more than substituted for the lost *T. ambiguum* chromosomes.

FIGURE 4 | Root morphology of BC_2-126 **(A)** as compared with control white clover genotypes colored white clover, PxB **(B)** and Kopu II **(C)**. The root system in BC_2-126 was thicker and deeper compared to the white clover parents.

TABLE 5 | Summary of the results of FISH-GISH analyses of five interspecific hybrid plants using 18S rDNA (NOR), 5S rDNA (5S), *T. ambiguum* genomic DNA (A), and TrR350 probes.

	Number of chromosomes					
	Total 2*n*	R or O NOR	R or O 5S non-NOR	A	Recombinant chromosomes	
Hybrid plant					Total	Without TrR350
BC_1-120	33	2	2	2	2	1
BC_1-132	33	2	2	2	2	0
BC_2-126	33	2	3	1	2	NA
Kopu II × BC_1-132	32	2	2	0	1	0
BC_1-133	32	2	3	1	2	0

NA, not analyzed.

Consistent with the apparent rapid loss of A chromosomes, the analyzed BC_2 plants had one or no A chromosomes instead of the expected two (**Table 5**). Meiotic analysis of BC_2-126 showed that, nearly 50% of the time, the A chromosome was unpaired at meiosis and so was likely to be lost by unequal disjunction from half of the gametes. In addition, it appeared that A univalents sometimes underwent premature disjunction (separation of sister chromatids) at anaphase I, followed by incomplete movement to the poles (lagging; **Figures 2F–I**). It is probable that this further exacerbated the loss of the A chromosome.

Non-NOR chromosomes with 5S loci also revealed chromosome imbalances in both BC_1 and BC_2. On average, one such pair was expected, as in *T. repens* and *T. occidentale*, but with potential variation from 1 to 3, assuming random assortment in the previous generation. The BC_1 and BC_2 plants had two or three such chromosomes (**Table 5**).

Meiotic Analyses and Recombinant Chromosomes

In this study, GISH signals that were consistently visible in multiple cells were regarded as definite signals. There were several cases in which small recombinant segments were apparent in a few cells but were not definite enough to confirm. Consequently, the most consistently visible sectors have been recorded, but this a minimum estimate and there may have been more recombination than identified here. Nevertheless, it was significant that recombinant chromosomes were found in all five analyzed BC plants and that four plants had at least two recombinant chromosomes. Therefore, the inter-genomic recombination frequency was significant and sufficient to encourage further development of an introgression breeding program.

Despite aneuploidy, the BC_1 plants had pollen staining of 30–67% indicating reasonable fertility and production of functional gametes. Leading up to the generation of the BC_1 hybrids, there had been two opportunities for pairing and recombination between A and O sub-genomes (**Figure 1**). First, in the AAOO hybrid used to generate the P^rO^rAO plant there could have been A-O recombination leading to gametes with *T. occidentale* chromosomes carrying *T. ambiguum* segments and vice versa (Williams et al., 2019a). Second, in the P^rO^rAO hybrid there was further opportunity for interchanges among A and O and perhaps A and O^r chromosomes and, also, P^r and O^r and O chromosomes, with potentially some of those chromosomes carrying A segments.

In the BC_1 plants there would also have been further opportunities for inter-species pairing and recombination. Analysis of BC_2-126 showed that after two backcrosses the genomes had not regained balance and departed significantly from a stable pattern of 16 bivalents and one univalent produced by the lone A chromosome. Instead, univalent frequency was 3.5 (range 0–9), indicating that R and O chromosomes were also unbalanced, probably lacking homologs or competing for homologs, and not regularly pairing. The presence of multivalent configurations was indicative of homoeologous pairing. The absence of univalents in some cells (**Table 4**) was indicative that the *T. ambiguum* chromosome sometimes paired with *T. repens* or *T. occidentale* chromosomes, and this was confirmed by FISH-GISH of meiosis which showed it to be paired in over 50% of the cells studied. The recombinant chromosomes were also shown to pair with *T. repens* or *T. occidentale* chromosomes (**Figure 2E**). Thus, further backcross progeny from this and the other BC_2 plants could potentially have led to further recombination and integration of *T. ambiguum* into a predominantly white clover genetic background. This is supported by the finding that, despite the presence of recombinant chromosomes and the potential genomic imbalances, BC_2-126 was moderately fertile and able to produce large numbers of progeny for further breeding (**Supplementary Table 2B**).

The R sub-genomes (P^r and O^r) never pair at meiosis in white clover, probably because of a genetic system preventing homoeologous pairing (Williams, 2014). However, they do pair in interspecific hybrids, as shown in 4*x T. repens* × *T. uniflorum* and 4*x T. repens* × *T. occidentale* hybrids (Hussain et al., 2016, 2017; Hussain and Williams, 2016). Probably they could also pair with each other in the P^rO^rAO hybrids, leading to novel P^r-O^r recombination. A and O chromosomes also show pairing affinities (Williams et al., 2011, 2019a), enabling A-O recombination. P^r-O and O^r-O pairing is also likely (Hussain and Williams, 2016), enabling introgression of *T. occidentale* into a white clover background. On the other hand, as already noted, in *T. ambiguum* × *T. repens* hybrids the A and R sub-genomes have shown very poor pairing affinity, suggesting that P^r-A and O^r-A pairings occur infrequently in those hybrids. The reason why

O^r-A pairing is apparently much less frequent than O-A pairing has not been investigated but could be related to the restriction of homoeologous pairing in *T. repens* (Williams, 2014).

Of the recombinant chromosomes identified in BC_1 and BC_2 plants, all except one involved a *T. ambiguum* segment attached to a *T. occidentale* chromosome, as identified by the centromeric probe, TrR350 (**Table 5**). The one exception occurred in hybrid BC_1-120, where one of the recombinant chromosomes had a *T. occidentale* segment attached to a *T. ambiguum* centromere. Here, the centromere lacked TrR350 and the *T. ambiguum* DNA signal spanned the centromeric region (**Figure 3C**). This difference in the frequencies of the recombinant chromosome classes might reflect the numerical presence of two *T. occidentale* (O and O^r) genomes and one only *T. ambiguum* genome in the F_1 hybrids. Alternatively, or in addition, there was lower than expected transmission of *T. ambiguum* chromosomes, as already noted.

The BC_2 plant, Kopu II×BC_1-132, although probably still unbalanced, came closest among the analyzed plants to achieving the aims of the backcross program. This plant had 32 chromosomes, including one with a large translocated segment from *T. ambiguum*. Otherwise, it had no additional A chromosomes and showed the desired numbers of NOR chromosomes and 5S carrying chromosomes from *T. repens* or *T. occidentale*. Further backcrossing would be expected to produce families with large numbers of plants of this type, with near-balanced karyotypes and potential for achieving the introgression-breeding aims.

Epigenetic Interactions

BC_1 plants consistently had a pair of NOR chromosomes each with a 5S locus, as in *T. repens* and *T. occidentale*. In all cases these NORs were active (decondensed). Hybrid BC_1-132 had an extra A chromosome carrying an NOR region (**Figures 3E,F**) which, by contrast, was inactive (condensed) in all observed cells. This nucleolar dominance (Chen and Pikaard, 1997) reflects epigenetic suppression of the *T. ambiguum* derived NOR locus. Similar suppression of the *T. ambiguum* NOR was observed in diploid *T. ambiguum*×*T. occidentale* hybrids (Williams et al., 2011) and some other *Trifolium* interspecific hybrids (Williams et al., 2019b). However, other hybrids, e.g., BL-1, the 4*x* hybrid parent of the present study, showed activity (decondensation) of both *T. ambiguum* and both *T. occidentale* NORs (Williams et al., 2019a). A concern for breeding by incorporation of genes from one species into another is whether the expression of donated genes will be epigenetically suppressed by the host genome (Adams et al., 2003). In the present case, the expression of *T. ambiguum* traits in the backcross hybrids suggested that this might not be a problem with these hybrids. However, in all but one BC hybrid there were remnant A chromosomes and it remains to be tested whether introgressed *T. ambiguum* traits will be fully expressed in advanced generations once whole A chromosomes have been eliminated.

The Consequences of Substitution of O^r by O-A Recombinant Chromosomes

The original proposition for this work invoked a two-stage transfer of *T. ambiguum* DNA to *T. repens* chromosomes *via* a *T. occidentale* chromosome bridge. However, recent DNA sequence analyses of *T. repens* and its two progenitors, *T. occidentale* and *T. pallescens* have revealed that the ancestral *T. occidentale* (O^r) sub-genomes currently in *T. repens* are largely unchanged from the genomes of extant *T. occidentale* (Griffiths et al., 2019). This suggests that substitution of O^r sub-genome chromosomes by O chromosomes from *T. occidentale* might have very little disruptive effect on genome functions. If so, then attachment of *T. ambiguum* segments to a predominantly *T. occidentale* chromosome and the substitution of this hybrid chromosome for its homolog in white clover by backcrossing could achieve introgression without the need for the second crossover into the white clover homolog. This requires only A-O crossover events followed by chromosome substitution during backcrossing and removes the need for a final R-A/O recombination event.

The Significance for Clover Breeding

The hybrids represent a potentially powerful breeding resource for extending the white clover gene-pool. First, the chromosome pairing patterns in these hybrids are of special significance because of their potential for novel genomic recombinations leading to introgression. Second, as already noted, they provide potential for chromosome substitutions, additions and subtractions. A breeding program involving large populations of interspecific hybrids has been established for the purpose of incorporating *T. ambiguum* traits into white clover.

The selection strategy used in the present study was based on above-ground traits of BC hybrid plants for expression of *T. ambiguum* traits along with those of the recurrent parent, *T. repens*. Even though this would have favored selection of plants with remnant A chromosomes, all the tested plants carried recombinant chromosomes, and one carried no A chromosomes. Although only a small sample, this would tend to suggest that recombination frequencies were high, and enough to provide confidence that this introgression strategy is workable. To-date, there has been no testing for resilience traits such as drought tolerance (which requires complex time-consuming experiments), or rhizome development (which take up to 18 months to develop). These tests will be needed in later generations now that the introgression strategy is viable.

As noted earlier, *T. ambiguum* exists in 2*x*, 4*x* and 6*x* forms, each with distinctive but overlapping ecological adaptations. The breeding scheme described (**Figure 1**), and tested here, used a 4*x* form of *T. ambiguum*. However, we have previously shown that a colchicine-doubled form derived from 2*x T. ambiguum* showed similar behavior in hybrids with *T. occidentale* (Williams et al., 2019a). It is probable that the scheme can be used to incorporate genetics from both 2*x* and 4*x T. ambiguum*, thus widening the pool of adaptations available for incorporation into *T. repens*. The 6*x* form has also been incorporated into three-way hybrids with *T. occidentale* and *T. repens* (Williams et al., 2019b) but to-date, these have not been developed. The early generation hybrids were near-6*x* due to reliance on unreduced gametes ($2n=4x$) and several generations of further breeding may

be necessary to produce white clover plants carrying 6*x* *T. ambiguum* genes (Williams et al., 2019b).

The progressive and relatively rapid loss of A chromosomes was desirable for an introgression breeding strategy. While some A chromosomes were present for a few generations, facilitating recombination, their relatively rapid disappearance should contribute to a timely return to a *T. repens* genomic background and the production of plants that are essentially *T. repens* integrated with short genomic segments from *T. ambiguum*. It was also notable that different A chromosomes were retained in some of the analyzed BC hybrids, indicating no evidence to-date of any undesirable retention of particular A chromosomes.

Three of the plants reported in **Table 3** were the progeny from self-pollination of BC_1 hybrids. These were expected to retain some *T. ambiguum* chromosomes and *T. occidentale* chromosomes in a near-4*x* *T. repens* genetic background. The rationale for producing these selfs was to move toward the production of *T. ambiguum* substitution and/or addition lines in which pairs of *T. ambiguum* chromosomes stably replace or supplement *T. repens* chromosomes. Such stable lines could be expected to further enhance clover breeding by providing an alternative strategy to transfer the desirable resilience traits from *T. ambiguum* into *T. repens*. These hybrids produced progeny for further study (**Supplementary Table 2B**).

These new results contrast with earlier attempts to combine the favorable traits of 4*x* *T. ambiguum* into white clover which have not revealed introgression by recombination. A strategy based on backcrossing 4*x* *T. ambiguum* × *T. repens* F_1 hybrids to *T. repens* was developed by Meredith et al. (1995). As outlined earlier, this strategy required transitions from 4*x* (F_1) to 6*x* (BC_1) and then back to 5*x* (BC_2) and on through presumed aneuploid generations (BC_3 onwards) and, finally, 4*x* in later generations. Expression of rhizome formation (Abberton et al., 1998, 2003) and drought tolerance (Marshall et al., 2001, 2015) were reported up to the BC_3 generation when aneuploidy, potentially involving the addition or substitution of several whole *T. ambiguum* chromosomes into the *T. repens* genome, was expected (Abberton et al., 2003). Reports of the phenotypes of selected populations beyond the BC_2 generation have not included any chromosome numbers or evidence of introgression *via* chromosome recombination. There is very limited pairing between *T. ambiguum* and *T. repens* chromosomes (Williams et al., 1982; Anderson et al., 1991; Meredith et al., 1995) and to-date there is no evidence of meiotic recombination leading directly to introgression of *T. ambiguum* segments into *T. repens* chromosomes. Nevertheless, as noted, chromosome additions or substitutions could occur instead. Although the details are unclear, the strategy produced the commercially sold interspecific hybrid (cv. 'Aberlasting') with some *T. ambiguum* traits. By contrast, the breeding strategy developed here achieved introgression by recombination and avoided higher ploidy (6*x* and 5*x*) generations.

To achieve the breeding objective of attaining white clover populations incorporating genes for greater resilience from *T. ambiguum*, further steps are needed. First, because the BC_2 plants had slightly unbalanced genomes, further backcrosses (with selection) are needed to obtain stable families with the desired phenotypes. This needs to be applied to large numbers of BC plants from diverse families to build a gene-pool of introgressed white clover families each carrying different arrays of introgressed material. Then new cultivars will be developed by selection for the desired resilient phenotypes using the full arsenal of available breeding methods. The molecular cytogenetic methods used here will probably play a useful but minor role in large scale breeding. Other genomic methods will have a major role for all traits, and especially for those that are difficult to measure, e.g., rhizomes, deep roots, drought tolerance; as applied for example by Abberton et al. (2003) to selection for rhizomes. Knowledge of the genomes of *T. repens*, *T. pallescens* and *T. occidentale* (Griffiths et al., 2019) will facilitate the unlocking of the *T. ambiguum* genome and enable the use of genomic methods to further track *T. ambiguum* introgressions against the *T. repens* background. Determinations of the breeding values of each introgression through genotyping and phenotypic characterization should provide rapid genetic gains in future clover breeding.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

IU carried out the experiments, analyzed the results and wrote part of the manuscript as part of a PhD degree program. HA performed and interpreted the molecular cytogenetic work. IV conceived the breeding strategy, carried out the interspecific hybridization and provided advice throughout the project. SH provided guidance for the conventional cytogenetics. NE carried out the molecular biology work and supplied the probes used for molecular cytogenetics. MM co-supervised the PhD component. WW led the project and co-wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The research was funded primarily by the NZ Ministry of Business, Innovation and Employment Grant C10X0711 with support by PGG-Wrightson Seeds. IU gratefully acknowledges financial assistance from the Higher Education Commission of Pakistan and overseas study leave from the Pakistan Agricultural Research Council. AgResearch and Massey University funded the publication costs.

ACKNOWLEDGMENTS

The authors dedicate this work to the memory of Professor Michael McManus. The authors thank statisticians John

Koolaard and Dongwen Luo for their valuable advice and Michelle Williamson and Helen van den Ende for technical assistance. Climate data were obtained from https://niwa.co.nz, Part of this research was included in the PhD thesis of IU (Ullah, 2013).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.858714/full#supplementary-material

REFERENCES

Abberton, M. T. (2007). Interspecific hybridization in the genus *Trifolium*. *Plant Breed.* 126, 337–342. doi: 10.1111/j.1439-0523.2007.01374.x

Abberton, M. T., Michaelson-Yeates, T. P. T., Bowen, C., Marshall, A. H., Prewer, W., and Carlile, E. (2003). Bulked segregant AFLP analysis to identify markers for the introduction of the rhizomatous habit from *Trifolium ambiguum* into *T. repens* (white clover). *Euphytica* 134, 217–222. doi: 10.1023/B:EUPH.0000003912.58022.e4

Abberton, M. T., Michaelson-Yeates, T. P. T., Marshall, A. H., Holdbrook-Smith, K., and Rhodes, I. (1998). Morphological characteristics of hybrids between white clover, *Trifolium repens* L., and Caucasian clover *Trifolium ambiguum* M. Bieb. *Plant Breed.* 117, 494–496. doi: 10.1111/j.1439-0523.1998.tb01981.x

Adams, K. L., Cronn, R., Percifield, R., and Wendel, J. F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4649–4654. doi: 10.1073/pnas.0630618100

Alconero, R., Fiori, B., and Sherring, W. (1986). Relationships of virus infections to field performance of six clover species. *Plant Dis.* 70, 119–121. doi: 10.1094/PD-70-119

Anderson, J. A., Taylor, N. L., and Williams, E. G. (1991). Cytology and fertility of the interspecific hybrid *Trifolium ambiguum* × *T. repens* and backcross populations. *Crop Sci.* 31, 683–687. doi: 10.2135/cropsci1991.0011183X003100030027x

Ansari, H. A., Ellison, N. W., Griffiths, A. G., and Williams, W. M. (2004). A lineage-specific centromeric satellite sequence in the genus *Trifolium*. *Chromosom. Res.* 12, 357–367. doi: 10.1023/B:CHRO.0000034099.19570.b7

Ansari, H. A., Ellison, N. W., Reader, S. M., Badaeva, E. D., Friebe, B., Miller, T. E., et al. (1999). Molecular cytogenetic organisation of 5S and 18S-26S rDNA loci in white clover (*Trifolium repens* L.) and related species. *Ann. Bot.* 83, 199–206. doi: 10.1006/anbo.1998.0806

Ansari, H. A., Ellison, N. W., Verry, I. M., and Williams, W. M. (2022). Asynapsis and unreduced gamete formation in a *Trifolium* interspecific hybrid. *BMC Plant Biol.* 22:14. doi: 10.1186/s12870-021-03403-w

Ansari, H. A., Ellison, N. W., and Williams, W. M. (2008). Molecular and cytogenetic evidence for an allotetraploid origin of *Trifolium dubium* (Leguminosae). *Chromosoma* 117, 159–167. doi: 10.1007/s00412-007-0134-4

Barbour, M., Caradus, J. R., Woodfield, D. R., and Silvester, W. B. (1996). "Water stress and water use efficiency of ten white clover cultivars," in *White Clover: New Zealand's Competitive Edge. Grassland Research and Practice Series No. 6*. ed. D. R. Woodfield (Palmerston North: New Zealand Grassland Association), 159–162.

Brink, G. E., and Pederson, G. A. (1998). White clover response to a water application gradient. *Crop Sci.* 38, 771–775. doi: 10.2135/cropsci1998.0011183X003800030025x

Bryant, W. G. (1974). Caucasian clover (*Trifolium ambiguum* M. Bieb.): a review. *J. Aust. Inst. Agric. Sci.* 40, 11–19.

Chen, Z. J., and Pikaard, C. S. (1997). Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. *Genes Dev.* 11, 2124–2136. doi: 10.1101/gad.11.16.2124

Dear, B. S., and Zorin, M. (1985). Persistence and productivity of *Trifolium ambiguum* M. Bieb. (Caucasian clover) in a high altitude region of South-Eastern Australia. *Aust. J. Exp. Agric.* 25, 124–132. doi: 10.1071/EA9850124

Ellison, N. W., Liston, A., Steiner, J. J., Williams, W. M., and Taylor, N. L. (2006). Molecular phylogenetics of the clover genus (*Trifolium* – Leguminosae). *Mol. Phylogenet. Evol.* 39, 688–705. doi: 10.1016/j.ympev.2006.01.004

Forde, M. B., Hay, M. J. B., and Brock, J. L. (1989). Development and growth characteristics of temperate perennial legumes. in: *Persistence of forage legumes. Proc. Australian/New Zealand/United States Workshop, Honolulu, HI*. eds. G. C. Marten et al. 18–22 July 1988. ASA, CSSA, and SSSA, Madison, WI, 91–109.

Gaynor, D. L., and Skipp, R. A. (1987). "Pests," in *White Clover*. eds. M. J. Baker and W. M. Williams (Wallingford, UK: CAB International), 461–492.

Griffiths, A. G., Moraga, R., Tausen, M., Gupta, V., Bilton, T. P., Campbell, M. A., et al. (2019). Breaking free: the genomics of allopolyploidy-facilitated niche expansion in white clover. *Plant Cell* 31, 1466–1487. doi: 10.1105/tpc.18.00606

Hussain, S. W., Verry, I. M., Jahufer, Z. Z., and Williams, W. M. (2017). Cytological and morphological evaluation of interspecific hybrids between *Trifolium repens* L. and *T. uniflorum* L. *Crop Sci.* 57, 2617–2625. doi: 10.2135/cropsci2017.05.0314

Hussain, S. W., Verry, I. M., and Williams, W. M. (2016). Development of breeding populations from interspecific hybrids between *Trifolium repens* L. and *T. occidentale* Coombe. *Plant Breed.* 135, 118–123. doi: 10.1111/pbr.12326

Hussain, S. W., and Williams, W. M. (2016). Chromosome pairing and fertility of interspecific hybrids between *Trifolium repens* L. and *T. occidentale* Coombe. *Plant Breed.* 135, 239–245. doi: 10.1111/pbr.12344

Latch, G. C. M., and Skipp, R. A. (1987). "Diseases," in *White clover*. eds. M. J. Baker and W. M. Williams (Wallingford, UK: CABI), 421–460.

Marshall, A. H., Lowe, M., and Collins, R. P. (2015). Variation in response to moisture stress of young plants of interspecific hybrids between white clover (*T. repens* L.) and Caucasian clover (*T. ambiguum* M. Bieb.). *Agriculture* 5, 353–366. doi: 10.3390/agriculture5020353

Marshall, A. H., Rascle, C., Abberton, M. T., Michaelson-Yeates, T. P. T., and Rhodes, I. (2001). Introgression as a route to improved drought tolerance in white clover (*Trifolium repens* L.). *Agron. Crop Sci.* 187, 11–18. doi: 10.1046/j.1439-037X.2001.00495.x

Mercer, C. F. (1988). "Reaction of some species of Trifolium to *Meloidgyne hapla* and *Heterodera trifolii*." in: *Proc. 5th Aust. Conf. Grasslands invertebrate Ecol.* ed. P. P. Stahle. August 15-19, 1988; Melbourne, Australia, 275–280.

Meredith, M. R., Michaelson-Yeates, T. P. T., Ougham, H., and Thomas, H. (1995). *Trifolium ambiguum* as a source of variation in the breeding of white clover. *Euphytica* 82, 185–191. doi: 10.1007/BF00027065

Pederson, G. A., and McLaughlin, M. R. (1989). Resistance to viruses in *Trifolium* inter-specific hybrids related to white clover. *Plant Dis.* 73, 997–999. doi: 10.1094/PD-73-0997

Pederson, G. A., and Windham, G. L. (1989). Resistance to *Meloidogyne incognita* in *Trifolium* inter-specific hybrids and species related to white clover. *Plant Dis.* 70, 119–121.

Sheaffer, C. C., Marten, G. C., Jordan, R. M., and Ristau, E. A. (1992). Forage potential of Kura clover and birdsfoot trefoil when grazed by sheep. *Agron. J.* 84, 176–180. doi: 10.2134/agronj1992.00021962008400020010x

Spencer, K., Hely, F. W., Govars, A. G., Zorin, M., and Hamilton, L. J. (1975). Adaptability of *Trifolium ambiguum* Bieb. to a Victorian montane environment. *Aust. J. Agric. Sci.* 41, 268–270.

Ullah, I. (2013). "Investigation of the possibility of introgression from *Trifolium ambiguum* M. Bieb. into *T. repens* L. A thesis submitted in the partial fulfilment of the requirements for the degree of Doctor of Philosophy in Plant Breeding and Genetics, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand". PhD diss., Massey University, 2013.

Williams, W. (1954). An emasculation technique for certain species of *Trifolium*. *Agron. J.* 46, 182–184. doi: 10.2134/agronj1954.00021962004600040013x

Williams, W. M. (1987a). "Taxonomy and biosystematics of *Trifolium repens*," in *White Clover*. eds. M. J. Baker and W. M. Williams (Wallingford, UK: CAB International), 323–342.

Williams, W. M. (1987b). "Genetics and breeding," in *White Clover*. eds. M. J. Baker and W. M. Williams (Wallingford, UK: CAB International), 343–419.

Williams, W. M. (2014). *Trifolium* interspecific hybridisation: widening the white clover gene-pool. *Crop Pasture Sci.* 65, 1091–1106. doi: 10.1071/CP13294

Williams, W. M., Easton, H. S., and Jones, C. S. (2007). Future options and targets for pasture plant breeding in New Zealand. *N. Z. J. Agric. Res.* 50, 223–248. doi: 10.1080/00288230709510292

Williams, W. M., Ellison, N. W., Ansari, H. A., Verry, I. M., and Hussain, S. W. (2012). Experimental evidence for the ancestry of allotetraploid *Trifolium repens* and creation of synthetic forms with value for plant breeding. *BMC Plant Biol.* 12:55. doi: 10.1186/1471-2229-12-55

Williams, E. G., Plummer, J., and Phung, M. (1982). Cytology and fertility of *Trifolium repens*, *T. ambiguum*, *T. hybridum*, and interspecific hybrids. *N. Z. J. Bot.* 20, 115–120. doi: 10.1080/0028825X.1982.10426411

Williams, E. G., and Verry, I. M. (1981). A partially fertile hybrid between *Trifolium repens* and *T. ambiguum*. *N. Z. J. Bot.* 19, 1–7. doi: 10.1080/0028825X.1981.10425182

Williams, W. M., Verry, I. M., Ansari, H. A., Hussain, S. W., Ullah, I., and Ellison, N. W. (2019a). 4*x Trifolium ambiguum* and 2*x T. occidentale* hybridise despite wide geographic separation and polyploidisation: implications for clover breeding. *Theor. Appl. Genet.* 132, 2899–2912. doi: 10.1007/s00122-019-03395-0

Williams, W. M., Verry, I. M., Ansari, H. A., Hussain, S. W., Ullah, I., and Ellison, N. W. (2019b). A Eurasia-wide polyploid species complex involving 6*x Trifolium ambiguum*, 2*x T. occidentale* and 4*x T. repens* produces interspecific hybrids with significance for clover breeding. *BMC Plant Biol.* 19:438. doi: 10.1186/s12870-019-2030-5

Williams, W. M., Verry, I. M., Ansari, H. A., Hussain, S. W., Ullah, I., Williamson, M. L., et al. (2011). Eco-geographically divergent diploids, Caucasian clover (*Trifolium ambiguum*) and western clover (*T. occidentale*) retain most requirements for hybridisation. *Ann. Bot.* 108, 1269–1277. doi: 10.1093/aob/mcr226

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 24 March 2022
doi: 10.3389/fpls.2022.856514

A Genetic Resource for Rice Improvement: Introgression Library of Agronomic Traits for All AA Genome *Oryza* Species

*Yu Zhang†, Jiawu Zhou†, Peng Xu, Jing Li, Xianneng Deng, Wei Deng, Ying Yang, Yanqiong Yu, Qiuhong Pu and Dayun Tao**

Yunnan Key Laboratory for Rice Genetic Improvement, Food Crops Research Institute, Yunnan Academy of Agricultural Sciences, Kunming, China

Rice improvement depends on the availability of genetic variation, and AA genome *Oryza* species are the natural reservoir of favorable alleles that are useful for rice breeding. To systematically evaluate and utilize potentially valuable traits of new QTLs or genes for the Asian cultivated rice improvement from all AA genome *Oryza* species, 6,372 agronomic trait introgression lines (ILs) from BC_2 to BC_6 were screened and raised based on the variations in agronomic traits by crossing 170 accessions of 7 AA genome species and 160 upland rice accessions of *O. sativa* as the donor parents, with three elite cultivars of *O. sativa*, Dianjingyou 1 (a *japonica* variety), Yundao 1 (a *japonica* variety), and RD23 (an *indica* variety) as the recurrent parents, respectively. The agronomic traits, such as spreading panicle, erect panicle, dense panicle, lax panicle, awn, prostrate growth, plant height, pericarp color, kernel color, glabrous hull, grain size, 1,000-grain weight, drought resistance and aerobic adaption, and blast resistance, were derived from more than one species. Further, 1,401 agronomic trait ILs in the Dianjingyou 1 background were genotyped using 168 SSR markers distributed on the whole genome. A total of twenty-two novel allelic variations were identified to be highly related to the traits of grain length (GL) and grain width (GW), respectively. In addition, allelic variations for the same locus were detected from the different donor species, which suggest that these QTLs or genes were conserved and the different haplotypes of a QTL (gene) were valuable resources for broadening the genetic basis in Asian cultivated rice. Thus, this agronomic trait introgression library from multiple species and accessions provided a powerful resource for future rice improvement and genetic dissection of agronomic traits.

Keywords: rice, AA genome, introgression line, grain size, allelic variation

OPEN ACCESS

Edited by:
Mallikarjuna Swamy, International Rice Research Institute, Philippines

Reviewed by:
Hu Zhao, Huazhong Agricultural University, China
Sung-Ryul Kim, International Rice Research Institute, Philippines

****Correspondence:***
Dayun Tao
taody12@aliyun.com

† These authors have contributed equally to this work

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

Received: *17 January 2022*
Accepted: *07 February 2022*
Published: *24 March 2022*

Citation:
Zhang Y, Zhou J, Xu P, Li J, Deng X, Deng W, Yang Y, Yu Y, Pu Q and Tao D (2022) A Genetic Resource for Rice Improvement: Introgression Library of Agronomic Traits for All AA Genome Oryza Species. Front. Plant Sci. 13:856514. doi: 10.3389/fpls.2022.856514

INTRODUCTION

Rice is one of the most important staple crops for almost half of the world's population. The Food and Agriculture Organization of the United Nations predicts that rice yield will have to be increased 50 to 70% by 2050 to meet human's demands, which increases that rice yield is still central for maintaining global food security (Ray et al., 2013). Whereas rice yield potential has been stagnant since the introduction of semidwarf gene into cultivated rice and the utilization of heterosis (Virmani et al., 1982; Monna et al., 2002; Sasaki et al., 2002), the narrow genetic basis that

results from the overuse of few parental materials and the lack of favorable variations led to yield bottleneck in rice breeding (Tanksley and Mccouch, 1997).

Genus *Oryza* contains twenty-two wild species and two cultivated rice species that represent 11 genomes: AA, BB, CC, BBCC, CCDD, EE, FF, GG, HHJJ, HHKK, and KKLL (Khush, 1997). Among these, six wild rice species (*O. nivara*, *O. rufipogon*, *O. barthii*, *O. glumaepatula*, *O. longistaminata*, and *O. meridionalis*) and two cultivated species (*O. sativa* and *O. glaberrima*) were classified into the AA genome. Asian cultivated rice (*O. sativa* L.) was domesticated from wild species *O. rufipogon* thousands of years ago (Khush, 1997; Huang et al., 2012). Previous reports indicated that 40% of the alleles of *O. rufipogon* was lost during the domestication from common wild rice to the cultivated rice (Sun et al., 2002), and only 10–20% of the genetic diversity in *O. rufipogon* and *O. nivara* was retained in two subspecies of the cultivated rice (Zhu et al., 2007). Since sharing the same AA genome, *O. glaberrima* and the six wild rice species are the most accessible gene pool for rice improvement (Ren et al., 2003). Thus, the exploitation and utilization of the useful alleles of AA genome species may overcome yield plateaus of *O. sativa* (Xiao et al., 1998). However, it is difficult to utilize the natural genetic diversity because of reproductive isolation, linkage drag, and background noise. Moreover, many important agronomic traits that include yield are controlled by quantitative trait loci (QTL) with smaller effect, which can be influenced by the environment. It is difficult to understand the QTL-controlled agronomic traits because of their complex inheritance and the genetic background noise.

Introgression lines are genetic resource in which the whole genome of a donor genotype (DG) is represented by the different segments in the genetic background of elite varieties. Genetic background noise of ILs can be eliminated significantly, which can be evaluated for any traits' improvement over the recurrent parents for rice breeding, also for QTL mapping and gene discovering as a single Mendelian factor; in addition, potential favorable genes hidden in the background of related species could be expressed in the genetic background of cultivated rice (Ballini et al., 2007; Eizenga et al., 2009; Bian et al., 2010; Rama et al., 2015; Jin et al., 2016; Yang et al., 2016; Bhatia et al., 2017; Bhatia et al., 2018; Yamagata et al., 2019). Thus, ILs that eliminate hybrid sterility, linkage drags, and background noise are one of the most important genetic resources for QTL mapping, gene identification, and discovery and rapid utilization for commercial breeding. Though a series of introgression lines developed by the genome-wide marker selections were obtained from the intersubspecific crosses between *indica* and *japonica* varieties and from the interspecific crosses between Asian cultivated rice and wild relatives of *Oryza sativa* (Bhatia et al., 2017; Divya et al., 2019), some lines showed the remarkable phenotype, whereas others did not exhibit obvious agronomic traits, which were difficult to be used for QTL identification, gene cloning, and breeding improvement. Establishing the introgression lines based on agronomic trait selection might be time-consuming and less laborious strategy.

AA genome species distributed in the natural and wild environment, which contains amount of useful allelic genes for improving rice yield and resistance to biotic and abiotic stresses (Khush, 1997). Different AA genome wild species and native varieties with unique characteristics and ecological adaptability represented the independent center of genetic diversity in rice. Comprehensively and systemically developing the ILs from all the AA genome species in different elite cultivar varieties background will help us to achieve sustainable yield improvement, diverse requirements for quality, and broad-spectrum resistance so as to meet the demand of the modern breeding program.

In this study, to explore and utilize wild relatives in rice improvement, we systematically introduced foreign segments from eight different AA genome species (*O. longistaminata*, *O. barthii*, *O. glumaepatula*, *O. meridinalis*, *O. nivara*, *O. rufipogon*, *O. glaberrima,* and upland rice of *O. sativa*) into three elite, highly productive *O. sativa* varieties (Dianjingyou 1, Yundao 1 and RD23). A total of six thousand three hundred and seventy-two agronomic ILs in three different backgrounds were screened and developed based on the repeated evaluation and selection of agronomic traits. One thousand four hundred and one of 6,372 agronomic ILs in the Dianjingyou 1 background were used to analyze genotype and discover novel alleles for grain size. Thus, this agronomic introgression library provided a powerful resource for future rice improvement and genetic dissection of agronomic traits.

MATERIALS AND METHODS

Plant Materials

The plant materials included 1 accession of *O. longistaminata*, 13 accessions of *O. barthii*, 6 accessions of *O. glumaepatula*, 8 accessions of *O. meridionalis*, 19 accessions of *O. rufipogon*, 20 accessions of *O. nivara*, 103 accessions of *O. glaberrima,* and 160 upland rice varieties of *O. sativa* (**Supplementary Table 1**). Three elite varieties, Dianjingyou 1 (a *japonica* variety), Yundao 1 (a *japonica* variety), and RD23 (an *indica* variety), were used as the recurrent parents.

A total of three hundred and twenty-nine accessions of AA genome species as the donor parents except for *O. longistaminata* were crossed with Dianjingyou 1 as the recurrent parent. A total of two hundred and twenty-six accessions as the donor parents, except for *O. longistaminata* and *O. glaberrima*, were used to cross with the recurrent parent Yundao 1. All the F_1 plants were used as female parents to backcross to their respective recurrent parents to produce BC_1F_1 generation. More than 200 BC_1F_1 seeds were generated for each of the combinations. The moderate heading date of individuals was selected to backcross with the recurrent parents, and about 200 BC_2F_1 seeds were obtained. From each of the BC_2F_1 progeny, individuals that showed a significant agronomic difference from the recurrent parents were selected for further backcrossing or selfing. After 2–6 times backcrossing and 2–7 times selfing, the progeny with stable and different target traits from their recurrent parents was developed as agronomic ILs.

The F_1 plants were obtained by embryo rescue technique from the cross between 1 accession of *O. longistaminata* as the donor parent and an *indica* variety RD23 as the recurrent parent, and crossing and selfing from BC_1F_1 generation were performed as above mentioned procedure.

All materials were grown at the Sanya Breeding Station, Sanya (18.24° N, 109.50° E), Hainan province, China. Ten individuals per row were planted at a spacing of 20 cm × 25 cm. All materials were grown and managed according to the local protocol.

Agronomic Trait Evaluation

A randomized complete block design was carried out with three replications for agronomic trait evaluation under two different environments (E1: winter and dry season, December to April 2008–2009; E2: summer and rainy season, July to November 2009), respectively. Each line was planted in three rows with 10 individuals per row. The five plants in the middle of each row were used for scoring traits. The recurrent parents, Dianjingyou 1, Yundao 1, and RD23, were used as controls in the experiment, respectively.

Prostrate growth habit was observed for the tiller angle in three main stages, which includes booting stage, heading stage, and grain filling stage. That tiller angle in ILs was larger than that in recurrent parent, which was regarded as the prostrate growth.

Primary branches at the base of panicle of the lines extend outward were regarded as the spreading panicle. Erect panicle or drooping panicle was evaluated according to the angle between the lines that connecting panicle pedestal with panicle tip and the elongation line of stem; spikelet numbers were measured as the total number of spikelets of the whole plant divided by its total number of panicles. Dense panicle was scored by the ratio of spikelet numbers to panicle length.

Tiller number was recorded from five random plants; plant height was measured from the ground level to the tip of the tallest panicle.

To measure the grain size, grains were selected from primary panicle and stored at room temperature for at least 3 months before testing. Twenty grains were used to measure grain length (GL), grain width (GW), and the ratio of grain length to grain width (RLW) from each plant. Photographs of grains per individual were taken using stereomicroscope, and then, grain size was measured by software Image J. The average value of 20 grains was used as phenotypic data. The weight of one thousand grains was measured by weighting fertile, fully mature grains from five panicles.

Aerobic adaptation was evaluated by biomass, yield, harvest index, heading date, and plant height in both aerobic and irrigated environments. Drought tolerance was assessed by the same traits as the aerobic adaptation in both upland and irrigated environments. For the aerobic and upland treatments, we used direct sowing with 4 seeds per hole and retained one seedling at the three-leaf stage. There is some difference in water management, and the rainfall provided the essential water for plant growth without the extra irrigation under aerobic treatment, whereas mobile sprinkler irrigation facilities were used to maintain a humid soil environment at the sow, tiller, and heading stage under upland treatment. For the irrigating treatment, sowing and transplanting single seedlings were done, and the field was managed according to the local standard practices.

To evaluate blast resistance, introgression lines were inoculated with *Magnaporthe oryzae* for 3 weeks after sowing by spraying with conidial suspension. After 7 days, lesion types on rice leaves were observed and scored according to a standard reference scale based on a dominant lesion type (Xu P. et al., 2015).

For a simply inherited trait, awn, pericarp color, and kernel color were observed directly in the field.

DNA Extraction and PCR Protocol

The experimental procedure for DNA extraction was performed as previously described (Edwards et al., 1991); A total of 168 SSR markers were selected from the Gramene database[1] or previously published polymorphic SSR markers within the *Oryza* AA genome species (McCouch et al., 2002; Orjuela et al., 2010). PCR was performed as follows: a total volume of 10 μl containing 10 ng template DNA, 1 × buffer, 0.2 μM of each primer, 50 μM of dNTPs, and 0.5 unit of Taq polymerase (Tiangen Company, Beijing, China). The reaction mixture was incubated at 94°C for an initial 4 min, followed by 30 cycles of 94°C 30 s, 55°C 30 s, and 72°C 30 s, and a final extension step of 5 min at 72°C. PCR products were separated on 8% non-denaturing polyacrylamide gel and detected using the silver staining method.

Determination of the Length of a Substituted Segment in Introgression Lines

The substituted segment was counted based on the SSR markers distributed on twelve chromosomes (Paterson et al., 1988; Young and Tanksley, 1989; McCouch et al., 2002). Intervals between two markers homozygous for the DG were regarded as 100% introgression segment, and a chromosome segment flanked by one marker of the DG and one marker of the recurrent type (DR) was considered as 50% introgression segment, whereas intervals between two markers homozygous for the recurrent genotype (RG) represented the background genotype. Thus, the physical distance of both DD and half of DR was used to estimate the length of introgression segments. The expected introgression length of the genome is divided by the total genome size to yield the expected proportion of introgression.

Exploring Loci for Grain Size Based on the Introgression Lines

Statistical analyses were performed on the SAS software package. The linkage between loci and grain size was scored by binomial distribution based on the genotype and phenotype between ILs and the recurrent parent. The genotypes of ILs that showed a significant difference from the recurrent parent in grain size were used to perform QTL analysis, if the rate of DG at some loci was significantly higher than that of the theoretical prediction, this

[1] http://www.gramene.org

locus could be linked with the grain size. Significant level was determined by the comparison between the ILs and the recurrent parents. Dunnett's *t*-test at $p < 0.0001$ was set to decrease the false probability (Eshed and Zamir, 1995; Xu et al., 2014).

RESULTS

Agronomic Trait Diversity in Introgression Library From AA Genome Donors

To systematically explore potentially valuable genes hidden in the AA genome wild relatives and two cultivated species, 67 accessions of AA genome wild rice, 103 accessions of *O. glaberrima,* and 160 upland rice varieties of *O. sativa* as the donors were used to raise the agronomic introgression line library. Of these accessions, all accessions except for *O. longistaminata* were used for generating ILs in the Dianjingyou 1 background, 233 donors except for the accessions of *O. glaberrima* and *O. longistaminata* were used for developing ILs in the background of Yundao 1, and 1 accession of *O. longistaminata* was used for raising ILs in the RD23 background. A total of 6,372 introgression lines with multiple donors showed a remarkable difference in the agronomic traits, which includes spreading panicle, erect panicle, dense panicle, lax panicle, awn, prostrate growth, plant height, pericarp color, kernel color, glabrous hull, grain size, 1,000-grain weight, drought resistance, and aerobic adaption, compared with their recurrent parents (**Figures 1A–D**). Among these, 74, 61, 179, 824, 135, 251, and 1561 ILs that show distinguished traits in the Dianjingyou 1 background were selected from the donors of *O. barthii*, *O. glumaepatula*, *O. meridionalis*, *O. rufipogon*, *O. nivara*, *O. glaberrima,* and upland rice of *O. sativa*, respectively (**Figures 1A,C** and **Supplementary Tables 2**, **3**). Additionally, 244, 85, 547, 714, 858, and 825 ILs that exhibit different agronomic traits in the Yundao 1 background were developed from the donors of *O. barthii*, *O. glumaepatula*, *O. meridionalis*, *O. rufipogon*, *O. nivara*, and upland rice of *O. sativa*, respectively (**Figure 1A** and **Supplementary Table 4**). A total of two hundred and sixty-five ILs were derived from the cross between 1 accession of *O. longistaminata* as the donor and an *indica* variety RD23 as the recurrent parent (**Figure 1D** and **Supplementary Table 5**). Thus, the agronomic introgression library derived from the multiple donors of AA genome species in the different backgrounds showed the abundant genetic variations in the agronomic traits.

For the same donor parent, phenotype variations for the agronomic traits varied with the genetic background. The numbers of ILs that show erect panicle, dense panicle, lax panicle, awn, plant height, pericarp color, 1,000-grain weight, drought-resistance, and aerobic adaption in Yundao 1 background was more than those of Dianjingyou 1 background, whereas the number of ILs that exhibit spreading panicle, prostrate growth, kernel color, glabrous hull, GL, and GW in Yundao 1 background was less than those of Dianjingyou 1 background (**Figure 1B**). It suggested that target trait expression was depended on the background of recurrent parent difference to a certain degree. Developing an introgression library in the different backgrounds will be beneficial to express hidden genes in the donor and discover more genetic variations for further study.

Characteristics of Chromosome Substituted Segments in the Introgression Library

A total of 168 SSR markers distributed on 12 chromosomes were selected to genotype agronomic introgression library in Dianjingyou 1 background (**Supplementary Figure 1**). The length of the interval between two markers ranged from 0.2 to 5.5 Mb, with an average of 2.22 Mb on the rice physical map (**Table 1** and **Supplementary Figure 1**). The polymorphism rate displayed from 82.74 to 98.43% between seven AA genome species and Dianjingyou 1 (**Table 1**).

A total of one thousand four hundred and one IL in the Dianjingyou 1 background were used to detect the characteristics of chromosome segments from seven AA genome species, which include 29 ILs from *O. barthii*, 30 ILs from *O. glumaepatula*, 76 ILs from *O. meridionalis*, 380 ILs from *O. nivara*, 74 ILs from *O. rufipogon*, 81 ILs from *O. glaberrima*, and 731 ILs from upland rice of *O. sativa* (**Supplementary Table 6**). In the 29 ILs from the donor of *O. barthii*, the length of introgression segments ranged from 2.66 to 28.98 Mb, averaging 6.99 Mb (**Table 2**). Different coverage rate was observed in a different chromosome. Chromosomes 3, 8, and 9 had 100% coverage rate, whereas chromosome 11 only had 39.14% coverage rate (**Supplementary Figure 2** and **Supplementary Table 7**).

A total of thirty ILs were obtained from an interspecific backcross between the cultivated rice *O. sativa* Dianjingyou 1 and the wild relative *O. glumaepatula*, and the length of introgression segments ranged from 660 kb to 25.6 Mb, averaging 5.35 Mb (**Table 2**). The ILs covered 73.11% of *O. glumaepatula* genome in the Dianjingyou 1 background (**Supplementary Figure 3** and **Supplementary Table 8**).

In the Dianjingyou 1/*O. meridionalis* introgression library, the length of introgression segments was detected from 0.27 to 23.77 Mb, averaging 5.83 Mb (**Table 2**). The donor introgressions covered 89.17% *O. meridionalis* genome. Chromosomes 3, 6, and 8 exhibited complete coverage, whereas chromosome 11 showed the least coverage rate of 60.71% (**Supplementary Figure 4** and **Supplementary Table 9**).

A total of three hundred and eight ILs were developed with the donor of *O. nivara*, and the average length of introgression segments was 5.84 Mb (**Table 2**). Average coverage rate per chromosome was about 97.17% (**Supplementary Figure 5** and **Supplementary Table 10**).

In the introgression library with the donor of *O. rufipogon*, the length of introgression segments varied from 280 kb to 23.46 Mb, averaging 6.37 Mb, and the average number of segments per chromosome was 52.08 (**Table 2**). Chromosomes 3, 8, and 9

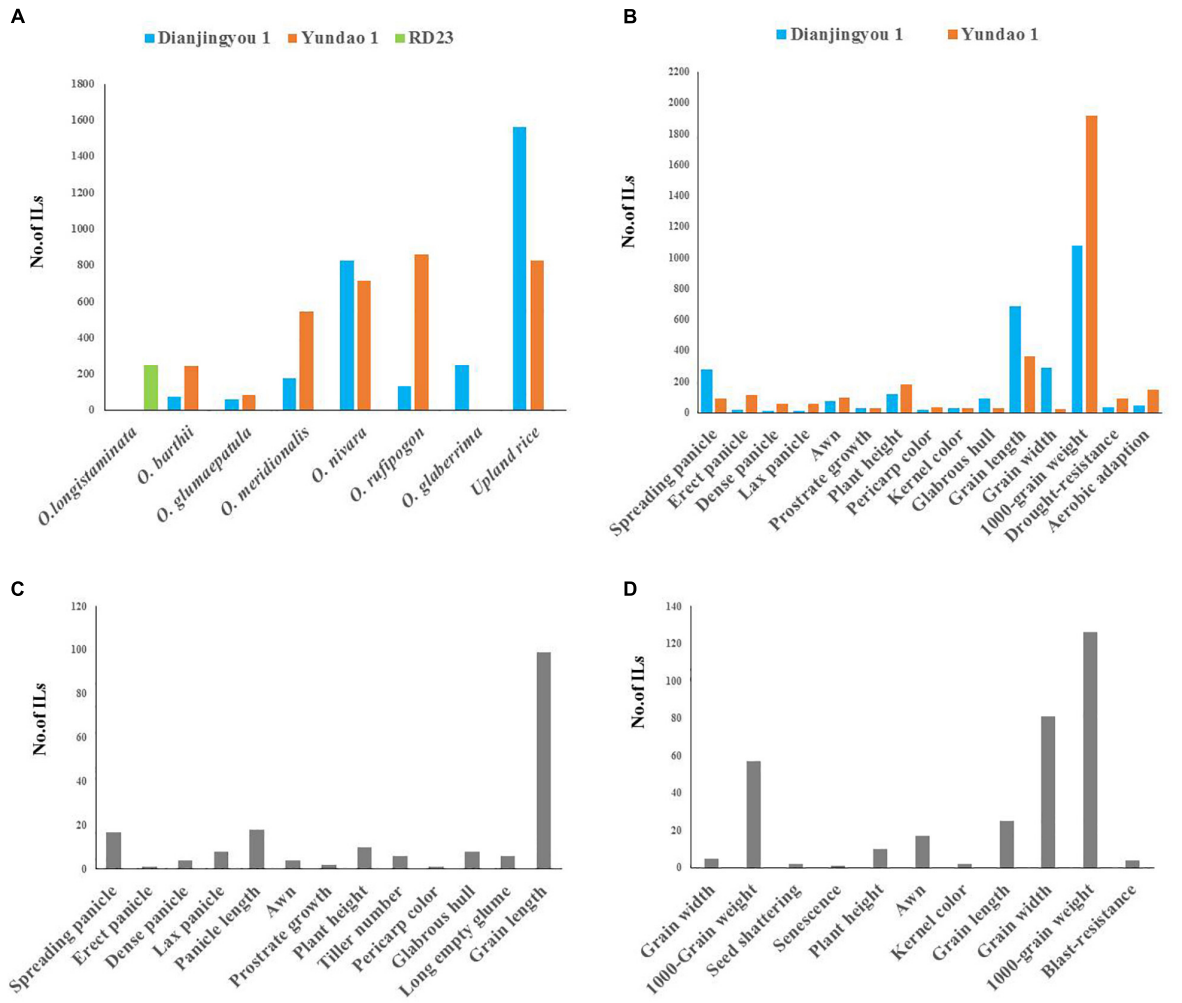


FIGURE 1 | The summary of introgression libraries with the donor of 8 AA genome species in three different genetic backgrounds. **(A)** A number of ILs from different donors in the Dianjingyou 1, Yundao 1, and RD23 background. **(B)** A number of ILs derived from *O. barthii*, *O. glumaepatula*, *O. meridionalis*, *O. nivara*, *O. rufipogon,* and upland rice, respectively, showed agronomic traits distinguishing from their recurrent parents Dianjingyou 1 and Yundao 1. **(C)** Agronomic traits of ILs derived from crosses between *O. glaberrima* and Dianjingyou 1. **(D)** Agronomic traits of ILs derived from cross between *O. longistaminata* and RD23.

exhibited 100% coverage rate, whereas chromosome 2 showed the least coverage rate of 74.43% (**Supplementary Figure 6** and **Supplementary Table 11**).

We developed 81 ILs with the *O. glaberrima* species as the donor, and the length of introgression segments varied from 270 kb to 24.17 Mb, averaging 5.80 Mb (**Table 2**). Chromosomes 3, 6, 9, and 11 exhibited complete coverage, whereas chromosome 10 showed the least coverage rate of 78.65% (**Supplementary Figure 7** and **Supplementary Table 12**).

In the introgression library derived from the donor of upland rice in *O. sativa*, the average length of introgression segments was 6.67 Mb (**Table 2**). All the chromosomes except for chromosome 10 showed 100% upland rice genome coverage (**Supplementary Figure 8** and **Supplementary Table 13**).

Taken together, ILs covered 81.19%, 73.11%, 89.17%, 97.17%, 89.19%, 90.70%, and 99.10% of *O. barthii*, *O. glumaepatula*, *O. meridionalis*, *O. nivara*, *O. rufipogon*, *O. glaberrima,* and upland rice genome information, respectively, which suggests that systematic and comprehensive agronomic IL library with the donor of AA genome species was developed by agronomic trait selection.

Detection of Potential Allelic Variations for Grain Size in the Introgression Library

Seed size plays an important role in rice yield (Xing and Zhang, 2010). Grain size not only determines rice appearance, but also affects milling, cooking, and eating quality of rice (Fan et al., 2006). Significant variations were observed for GL, GW, and the RLW in the introgression library with multiple donors in the background of Dianjingyou 1. Some ILs for GL, GW, and RLW were found to be significantly superior to the recurrent parent Dianjingyou 1. For GL, 133 and 125 ILs were found to be significantly longer than the

TABLE 1 | The description of markers used for genotyping introgression library.

Chr	Number of Markers	Density (Mb)	The marker polymorphism rate (%)						
			O. barthii	*O. glumaepatula*	*O. meridionalis*	*O. nivara*	*O. rufipogon*	*O. glaberrima*	Upland rice
1	17	2.55	94.12	94.12	100.00	100.00	100.00	100.00	100
2	18	2.00	77.78	77.78	88.89	83.33	88.89	94.44	77.78
3	16	2.28	100.00	100.00	100.00	100.00	100.00	100.00	100.00
4	15	2.37	80.00	100.00	93.33	100.00	100.00	100.00	93.33
5	15	2.00	93.33	60.00	86.67	93.33	93.33	86.67	80.00
6	13	2.40	92.31	84.62	100.00	100.00	100.00	100.00	100.00
7	14	2.12	92.86	78.57	92.86	92.86	92.86	100.00	85.72
8	14	2.03	100.00	57.14	100.00	100.00	100.00	100.00	78.58
9	10	2.30	100.00	90.00	100.00	100.00	100.00	100.00	83.34
10	12	1.93	91.67	100.00	100.00	100.00	91.67	100.00	100.00
11	12	2.42	66.67	66.67	83.33	100.00	83.33	100.00	100.00
12	12	2.29	83.33	83.33	100.00	100.00	100.00	100.00	100.00
Mean (%)		2.22	89.29	82.74	95.24	97.02	95.83	98.43	91.56

TABLE 2 | The characterization of introgression lines.

Donors	Number of introgression lines	Number of introgression Segments	Average number of segments per chromosome	Range of Segment length (Mb)	Average length of segments (Mb)	Average background recovery rate(%)
O. barthii	29	349	29.08	2.66–28.98	6.99	88.92
O. glumaepatula	30	210	17.50	0.66–25.60	5.35	91.94
O. meridionalis	76	434	36.17	0.27–23.77	5.83	93.21
O. nivara	380	3332	277.67	0.30–27.43	5.84	92.74
O. rufipogon	74	625	52.08	0.28–23.46	6.37	89.92
O. glaberrima	81	1138	94.83	0.27–24.17	5.80	85.14
Upland rice	731	9198	766.50	0.19–43.27	6.67	84.03

Dianjingyou 1 in two seasons, respectively. For GW, 412 and 508 ILs were observed to be significantly wider than the recurrent parent in two environments, respectively. For the RLW, 277 and 178 ILs were found to be higher than the Dianjingyou 1 in different seasons, respectively (**Figure 2**). In addition, the same traits in the different environments showed a highly significant correlation (**Supplementary Table 14**). These results suggested that abundant genetic variations for grain size existed in the wild and cultivated accessions of rice.

To explore favorable allelic variation for grain size, QTLs were detected based on the genotype and phenotype data. A total of forty-one loci linking with GL, forty-four loci linking with GW, and thirty-two loci linking with RLW were identified. It indicated that abundant gene pool for grain size existed in the AA genome species (**Figures 3–5**). Among these, 26 loci for GL were detected from multiple donors, and 12, 11, 2, and 1 loci were found from the donors of two species, three species, four species, and six species, respectively. It suggested that the same locus that contributes to GL is a potential allelic variation from different donors. Moreover, 4 loci from the different donors were only responsible for long grain, 12 loci derived from the multiple donors only contributed to short grain, and 10 loci from the different species controlled both long grain and short grain. Moreover, 22 novel allelic variations from multiple donors contributed to GL (**Figure 3**).

A total of 27 loci for GW were examined from the multiple donors, which include 13 loci from the only two species, 9 loci from three species, 3 loci from four species, and 1 locus from five species (**Figure 4**). Moreover, 12 loci from the different donors were only responsible for wide grain, 4 loci from multiple species only led to thin grain, and 10 loci from the different species controlled both wide and thin grains. In addition, 22 novel allelic variations for GW were found in agronomic IL library (**Figure 4**).

Nineteen loci for RLW from multiple donors were explored on 12 chromosomes, which include 12, 6, and 1 locus detected simultaneously in two, three, and four donor species, respectively (**Figure 5**).

These results indicated that detection of favorable genes using multiple donors could help us to find the novel allelic variations. The allelic genes were detected in the different donors, which suggest that some loci for grain size were conserved in Genus *Oryza*. Some loci controlled the opposite phenotype, long grain vs. short grain, wide grain vs. thin grain, validating these loci's functions in forward and reverse direction and also suggesting that the loci functioned divergence in the different donors. Taken together, these results would provide the information that the loci for grain size from the different donors were the same or different

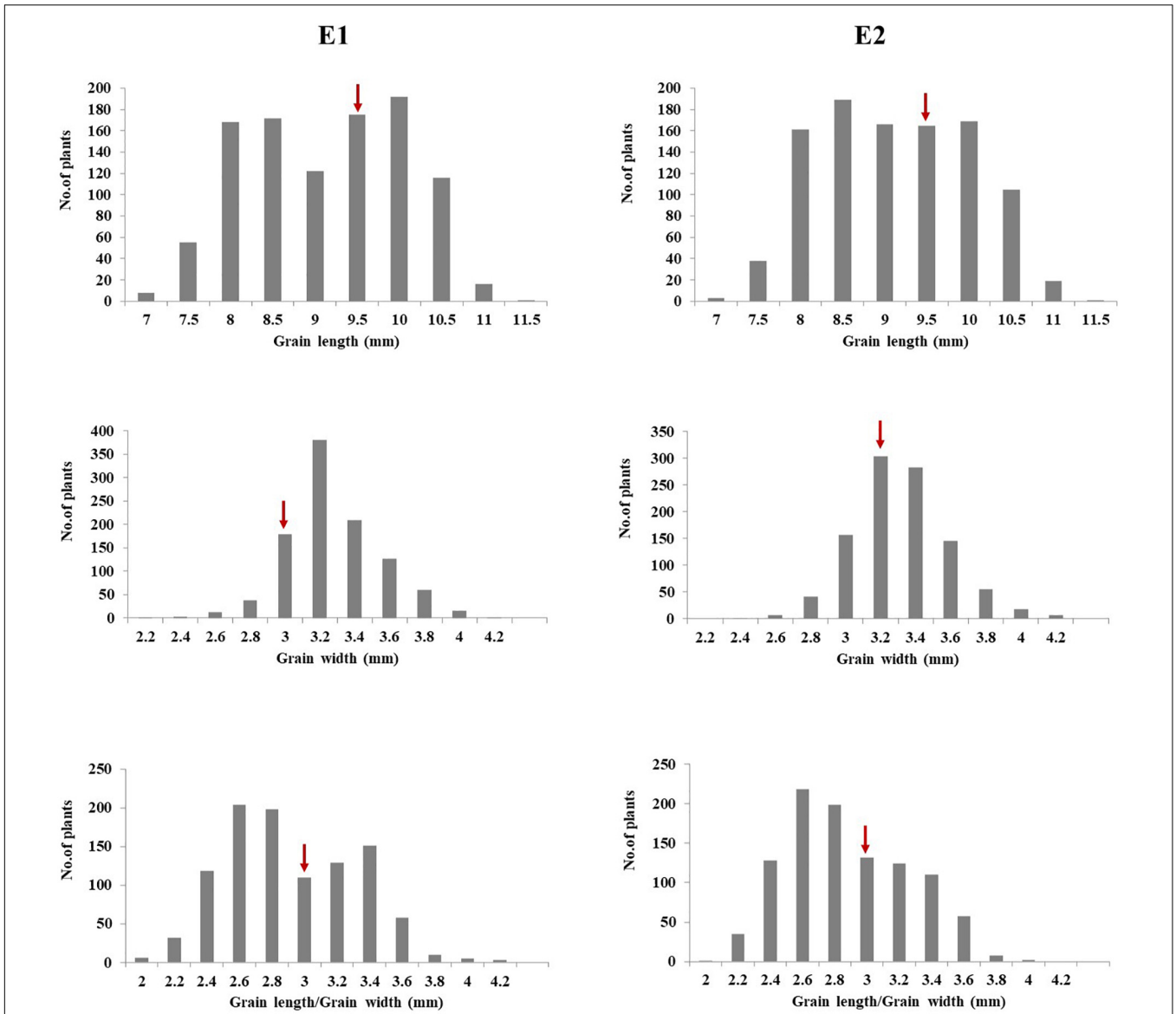


FIGURE 2 | Frequency distribution of grain size traits in introgression lines in the Dianjingyou 1 background. Arrows indicated the mean values for recurrent parent Dianjingyou 1.

haplotypes; it also indicated that agronomic IL library with the donor of 7 AA genome species was an excellent resource and tool to discover favorable allelic variations and new QTLs or genes for rice improvement.

DISCUSSION

Agronomic Introgression Line Library for All AA Genome Species Is an Important Stock for Breeding Improvement in Rice

The geographical distribution of wild relative species with the AA genome is in a wide range of environments, *O. nivara* and *O. rufipogon* mainly in Asia, *O. glaberrima*, *O. barthii*, and *O. longistaminata* in Africa, *O. meridionalis* in Australia, *O. glumaepatula* in Latin America, and wild relative species evolve a large extend on genetic differentiation and morphological variation under the different ecological environments (Vaughan et al., 2005). Constructing introgression lines are a feasible and effective approach to transferring the favorable genes from wild relative species to the cultivated varieties for improving elite cultivars. By now, more than 40 sets of ILs were developed using the AA genome wild species as the donor parents (Chen et al., 2006; Tian et al., 2006; McCouch et al., 2007; Rangel et al., 2008; Hao et al., 2009; Ali et al., 2010; Gutierrez et al., 2010; Ramos et al., 2016; Bhatia et al., 2017), but ILs were almost derived from a single accession of

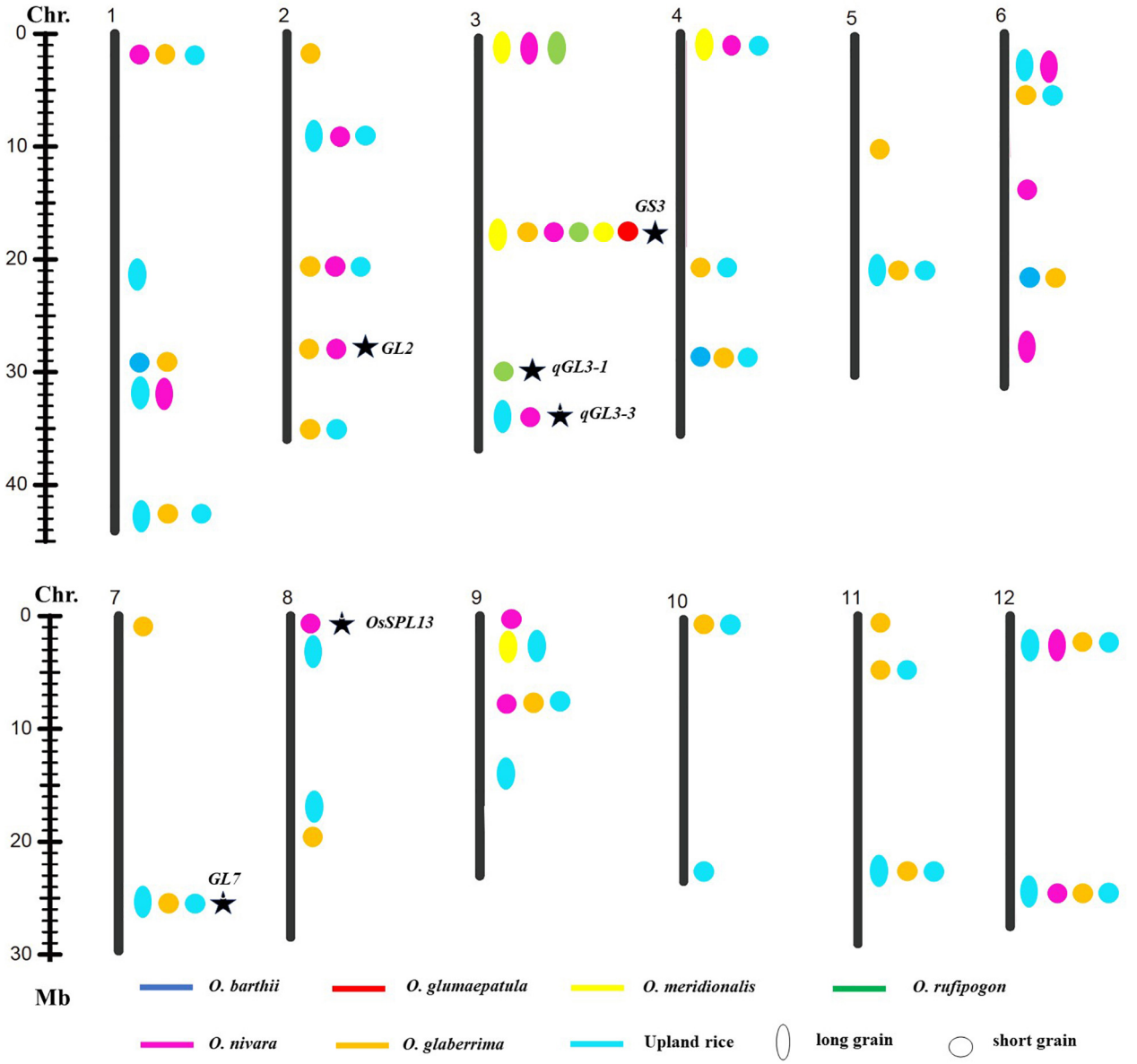


FIGURE 3 | Allelic variation linking with GL was detected from the different donors. Asterisk indicated published genes.

AA genome species in a single background, which leads to the lack of systematic utilization of favorable genes. One of the challenges of constructing interspecific introgression lines was to overcome interspecific hybrid sterility. In this study, we selected the typical 330 accessions of AA genome species distributed in the different geographical region as the donor parents to raise agronomic IL library. We observed that pollen fertility of F_1 varied from 1.92% to 93.19% dependent on the different accessions of *O. nivara* and *O. rufipogon*, whereas all the crosses with the accessions of *O. barthii*, *O. glumaepatula,* and *O. meridionalis* showed complete pollen sterility in the F_1 combinations (data not shown). When the *japonica* varieties Dianjingyou 1 and Yundao 1 used as the recurrent parents were crossed with the accession of *O. longistaminata*, the crossing was failed despite many efforts. Only the cross using an *indica* variety RD23 as the recurrent parent and *O. longistaminata* as the donor was obtained by embryo rescue. Fortunately, the female gametes from the interspecific hybrids were partially fertile, and some hybridization seeds in the different combinations could be harvested by backcrossing F_1 as the female parent with *O. sativa* as the male parent. Finally, agronomic IL library that contains 6,372 lines was developed based on the agronomic trait selection, and agronomic ILs showed a significant difference from the recurrent parents, which include spreading panicle, erect panicle, dense panicle, lax panicle, awn, prostrate growth, plant height, pericarp color, kernel color, glabrous hull, grain size, 1,000-grain

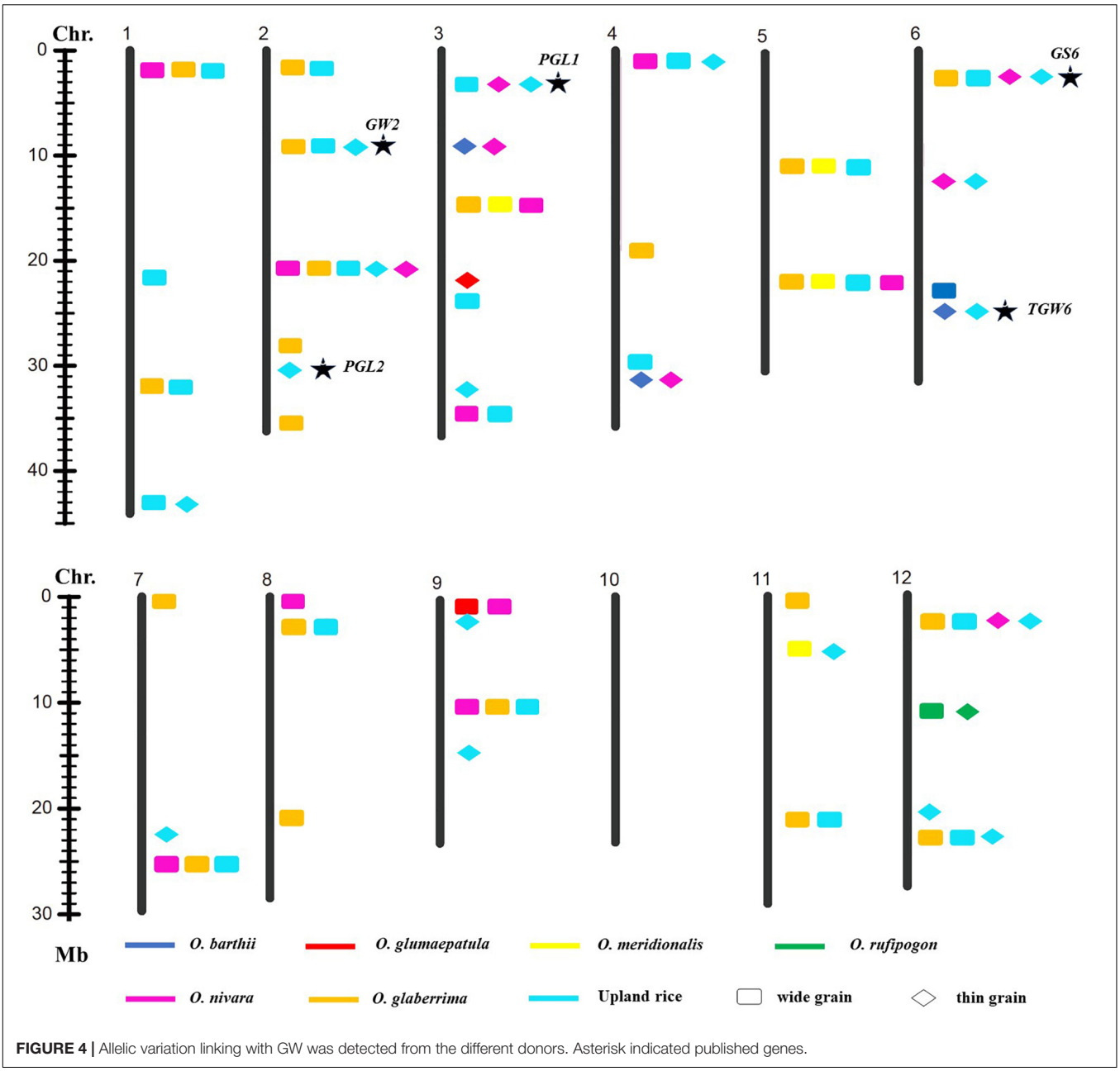


FIGURE 4 | Allelic variation linking with GW was detected from the different donors. Asterisk indicated published genes.

weight, drought resistance and aerobic adaption, and blast resistance (**Figure 1**), and this IL library help us to understand the genetic base of agronomic traits and explore the favorable genes or allelic variations, and also breeding improvement systematically and comprehensively.

Variation for agronomic traits in this IL library shows the importance of AA genome species for further breeding improvement in rice. For example, 18 ILs in Dianjingyou 1 background and 57 ILs in Yundao 1 background showed dense panicle (**Supplementary Tables 2–4**), 137 ILs registered significant improvement over the recurrent parents in grain weight (data not shown), and these ILs derived from multiple donors could contribute to the variations for yield.

The upland rice is a predominant ecotype adapted to aerobic and rain-fed conditions in the mountainous areas that have high genetic variability in the characteristics of morphology and physiology, such as glabrous hull and aerobic adaptation (Bridhikitti and Overcamp, 2012; Sandhu et al., 2013). In this study, 160 accessions of upland rice that represent an abundant genetic diversity were used to raise introgression lines. Among those agronomic ILs, 93 and 18 ILs in the Diangjingyou 1 and Yundao 1 background exhibited the glabrous hull phenotype. Those resources were not only good for breeding varieties suitable for agricultural operation, but also helpful to understand the genetic mechanism of glabrous hull development. A total of 125 ILs in aerobic adaptation were

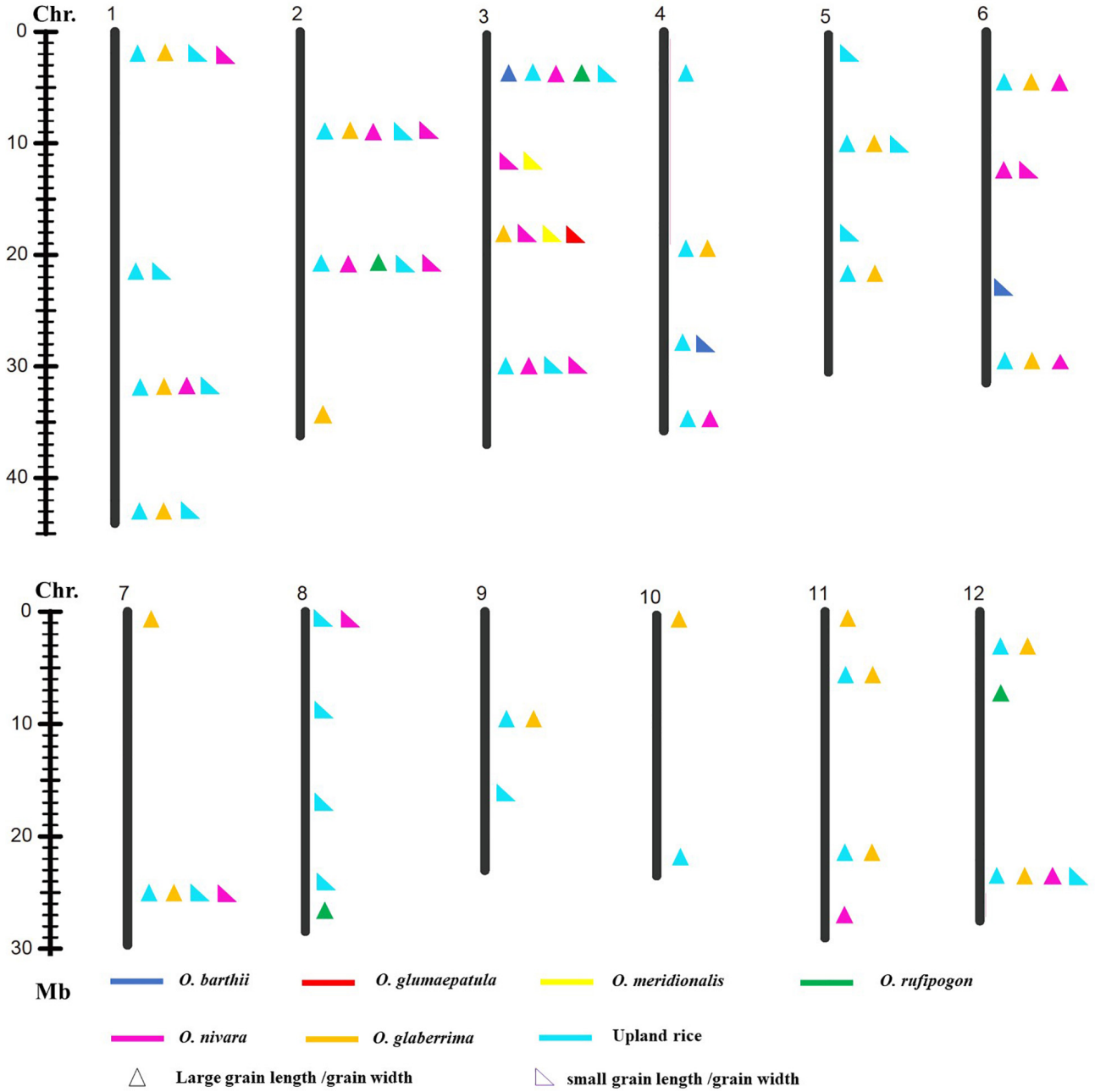


FIGURE 5 | Allelic variation linking with the RLW was detected from the different donors.

superior to the recurrent parents, which are important and environmental-friendly breeding materials to meet the need of aerobic rice development.

The effect of alleles on the agronomic traits varied with genetic background. The different recurrent parents helped us to find background-dependent useful traits or stable traits in the different backgrounds. In this study, with the donor accessions of *O. barthii*, *O. glumaepatula*, *O. meridionalis*, *O. nivara*, *O. rufipogon,* and upland rice of *O. sativa*, 62, 37, 147, 470, 122, and 710 ILs that conferred agronomic traits were found in the genetic background of the Dianjingyou 1 and Yundao 1, which suggested that the genes for agronomic traits had a stable effect on the different genetic backgrounds (**Supplementary Tables 2–4**). In addition, we found ILs that showed aerobic adaptation with the donors of *O. barthii*, *O. glumaepatula*, *O. meridionalis*, *O. nivara*, *O. rufipogon* were detected in the Yundao 1 background, rather than Dianjingyou 1 background (**Supplementary Tables 2–4**). The results would provide the

theoretical guidance of the relationship between the traits and genetic background in rice breeding.

Exploration of Natural Allelic Variations Using Agronomic Introgression Line Library

Genetic diversity and allele were lost during the domestication from the wild species of rice to the cultivated rice (Sun et al., 2002), whereas narrow genetic basis led to the yield bottleneck of Asian cultivated rice. In the recent years, mining and utilization of useful allele variation have made great progress in rice breeding. For example, the allelic variation in the *Wx* gene and *SSSI* was proved to contribute greatly to the differences in rice eating and cooking qualities (ECQs) in the two subspecies (Li et al., 2018). Allelic variation at the *E1/Ghd7* locus allowed an expansion of the rice cultivation area through adjusting heading date (Saito et al., 2019). The allele types of *BPH9* conferred varying levels of resistance to different biotypes of *BPH* and enabled rice to combat planthopper variation (Zhao et al., 2016). The allelic variation at the rice blast resistance (*R*) *Pid3* locus was analyzed based on the 3K RGP sequencing data, and different strategies were developed to apply the functional *Pid3* alleles to *indica* and *japonica* rice breeding (Lv et al., 2017). In this study, one locus for GL and one locus for GW were explored from the six and five different donor species, respectively. Two loci for GL, three loci for GW, and one locus for the RLW were detected from the donors of four species, respectively (**Figures 3–5**). Additionally, many of published genes for grain size were found based on the agronomic introgression library analysis, such as *GW2* (Song et al., 2007), *GL2* (Hu et al., 2015), *PGL2* (Heang and Sassa, 2012b), *PGL1* (Heang and Sassa, 2012a), *GL3.2/CYP78A13* (Xu F. et al., 2015), *GS3* (Fan et al., 2006; Takano-Kai et al., 2009;Mao et al., 2010), *qGL3-1* (Qi et al., 2012), *qGL3.3* (Hu et al., 2018), *GS6* (Sun et al., 2013), *TGW6* (Ishimaru et al., 2013), *GL7* (Wang et al., 2015), and *OsSPL13* (Si et al., 2016). In addition, 22 loci might be new QTLs or genes controlled for GL and GW from the different AA genome donors. Accordingly, agronomic introgression library with multiple donors from different relatives of Asian cultivated rice is a powerful resource platform to discover novel and functional allelic variations for agronomic traits. Mining the natural functional variations in the useful genes derived from the multiple donors and combing the different alleles through diversification could be useful for an accurate rice breeding program.

Therefore, agronomic IL libraries derived from the multiple donors have some advantages: (1) an abundant genetic variations were introgressed into the cultivated rice genome; (2) target genes or QTLs for the same phenotype could be validated by the different donors, and it will provide the information that these target genes or QTLs could be the same haplotype; (3) the genes or QTLs responsible for the opposite phenotypes, for example, long-grain size and short-grain size, could also be confirmed using the different populations from multiple donors, and it could be the different haplotypes. Therefore, this agronomic IL library will help us to improve rice breeding and interesting gene discovery and utilization.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YZ and DT draft the manuscript. DT designed the research. JZ, PX, WD, and XD developed the introgression lines. YZ, YYa, YYu, JL, and QP participated the genotype and phenotype evaluation. YZ performed the data analysis. All authors reviewed and approved the final manuscript.

FUNDING

This research was supported by the National Natural Science Foundation of China (Grant Nos. 31991221, 31660380, and 31201196, and 32160489), the Yunnan Provincial Science and Technology Department, China (2018FG001-086, 202101AS070036, and 202101AT070193), the Yunnan Provincial Government (YNWR-QNBJ-2018-359), and the Yunnan Seed Industrialization Laboratory Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.856514/full#supplementary-material

Supplementary Figure 1 | Physical map of the 168 SSR markers in rice used for this study.

Supplementary Figure 2 | Chromosome segments distribution and coverage degree of introgression library from the donor of *O. barthii*. Each block at the row represented introgression from a total of 29 ILs on the target chromosome, regardless of introgression segments in other chromosomes, and each column represented a molecular marker locus.

Supplementary Figure 3 | Chromosome segments distribution and coverage degree of introgression library from the donor of *O. glumaepatula*. Each block at the row represented introgression from a total of 30 ILs on the target chromosome, regardless of introgression segments in other chromosomes, and each column represented a molecular marker locus.

Supplementary Figure 4 | Chromosome segments distribution and coverage degree of introgression library from the donor of *O. meridionalis*. Each block at the row represented introgression from a total of 76 ILs on the target chromosome, regardless of introgression segments in other chromosomes, and each column represented a molecular marker locus.

Supplementary Figure 5 | Chromosome segments distribution and coverage degree of introgression library from the donor of *O. nivara*. Each block at the row represented introgression from a total of 380 ILs on the target chromosome, regardless of introgression segments in other chromosomes, and each column represented a molecular marker locus.

Supplementary Figure 6 | Chromosome segments distribution and coverage degree of introgression library from the donor of *O. rufipogon*. Each block at the row represented introgression from a total of 74 ILs on the target chromosome, regardless of introgression segments in other chromosomes, and each column represented a molecular marker locus.

Supplementary Figure 7 | Chromosome segments distribution and coverage degree of introgression library from the donor of *O. glaberrima*. Each block at the row represented introgression from a total of 81 ILs on the target chromosome, regardless of introgression segments in other chromosomes, and each column represented a molecular marker locus.

Supplementary Figure 8 | Chromosome segments distribution and coverage degree of introgression library from the donor of upland rice. Each block at the row represented introgression from a total of 731 ILs on the target chromosome, regardless of introgression segments in other chromosomes, and each column represented a molecular marker locus.

REFERENCES

Ali, M., Sanchez, P., Yu, S., Lorieux, M., and Eizenga, G. (2010). Chromosome segment substitution lines: a powerful tool for the introgression of valuable genes from Oryza wild species into cultivated rice (O. sativa). *Rice* 3, 218–234.

Ballini, E., Berruyer, R., Morel, J., Lebrun, M., Notteghem, J., and Tharreau, D. (2007). Modern elite rice varieties of the 'Green revolution' have retained a large introgression from wild rice around the Pi33 rice blast resistance locus. *New Phytol.* 175, 340–350. doi: 10.1111/j.1469-8137.2007.02105.x

Bhatia, D., Joshi, S., Das, A., Vikal, Y., Gurpreet, K., Neelam, et al. (2017). Introgression of yield component traits in rice (*Oryza sativa* ssp. indica) through interspecific hybridization. *Crop Sci.* 57, 1–17.

Bhatia, D., Wing, R. A., Yu, Y., Chougule, K., Kudrna, D., Lee, S., et al. (2018). Genotyping by sequencing of rice interspecific backcross inbred lines identifies QTLs for grain weight and grain length. *Euphytica* 214:41.

Bian, J., Jiang, L., Liu, L., Wei, X., Xiao, Y., Zhang, L., et al. (2010). Construction of a new set of rice chromosome segment substitution lines and identification of grain weight and related traits QTLs. *Breed Sci.* 60, 305–313.

Bridhikitti, A., and Overcamp, T. (2012). Estimation of Southeast Asian rice paddy areas with different ecosystems from moderate-resolution satellite imagery. *Agric. Ecosyst. Environ.* 146, 113–120.

Chen, J., Hafeez, U., Chen, D., Liu, G., Zheng, K., and Zhuang, J. (2006). Development of chromosomal segment substitution lines from a backcross recombinant inbred population of interspecific rice cross. *Rice Sci.* 13, 15–21.

Divya, B., Malathi, S., Sukumar, M., and Sarla, N. (2019). Development and use of chromosome segment substitution lines as a genetic resource for crop improvement. *Theor. Appl. Genet.* 132, 1–25. doi: 10.1007/s00122-018-3219-y

Edwards, K., Johnstone, C., and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res.* 19, 1349. doi: 10.1093/nar/19.6.1349

Eizenga, G., Agrama, H., Lee, F., and Jia, Y. (2009). Exploring genetic diversity and potential novel disease resistance genes in a collection of rice (Oryza spp.) wild relatives. *Genet. Resour. Crop Evol.* 56, 65–76.

Eshed, Y., and Zamir, D. (1995). An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* 141, 1147–1162. doi: 10.1093/genetics/141.3.1147

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., et al. (2006). GS3 a major QTL for grain length and weight and minor QTL for grain width and thickness in rice encodes a putative transmembrane protein. *Theor. Appl. Genet.* 112, 1164–1171. doi: 10.1007/s00122-006-0218-1

Gutierrez, A., Carabali, S., Giraldo, O., Martinez, C., Correa, F., Prado, G., et al. (2010). Identification of a rice stripe necrosis virus resistance locus and yield component QTLs using *Oryza sativa* x O. glaberrima introgression lines. *BMC Plant Biol.* 10:6. doi: 10.1186/1471-2229-10-6

Hao, W., Zhu, M., Gao, J., Sun, S., and Lin, H. (2009). Identification of quantitative trait loci for rice quality in a population of chromosome segment substitution lines. *J. Integr. Plant Biol.* 51, 500–512. doi: 10.1111/j.1744-7909.2009.00822.x

Heang, D., and Sassa, H. (2012a). Antagonistic actions of HLH/bHLH proteins are involved in grain length and weight in rice. *PLoS One* 7:e31325. doi: 10.1371/journal.pone.0031325

Heang, D., and Sassa, H. (2012b). An atypical bHLH protein encoded by positive regulator of grain length 2 is involved in controlling grain length and weight of rice through interaction with a typical bHLH protein APG. *Breed Sci.* 62, 133–141. doi: 10.1270/jsbbs.62.133

Hu, J., Wang, Y., Fang, Y., Zeng, L., Xu, J., and Yu, H. (2015). A Rare allele of GS2 enhances grain size and grain yield in rice. *Mol. Plant* 8, 1455–1465. doi: 10.1016/j.molp.2015.07.002

Hu, Z., Lu, S., Wang, M., He, H., Sun, L., and Wang, H. (2018). A novel QTL qTGW3 encodes the GSK3/SHAGGY-Like kinase OsGSK5/OsSK41 that interacts with OsARF4 to negatively regulate grain size and weight in rice. *Mol. Plant* 11, 736–749. doi: 10.1016/j.molp.2018.03.005

Huang, X., Kurata, N., Wei, X., Wang, Z., Wang, A., and Zhao, Q. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497–501. doi: 10.1038/nature11532

Ishimaru, K., Hirotsu, N., Madoka, Y., Murakami, N., Hara, N., and Onodera, H. (2013). Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. *Nat. Genet.* 45, 707–711.

Jin, J., Hua, L., Zhu, Z., Tan, L., Zhao, X., and Zhang, W. (2016). GAD1 encodes a secreted peptide that regulates grain number grain length and awn development in rice domestication. *Plant Cell* 28, 2453–2463. doi: 10.1105/tpc.16.00379

Khush, G. (1997). Origin dispersal cultivation and variation of rice. *Plant Mol. Biol.* 35, 25–34.

Li, Q., Liu, X., Zhang, C., Jiang, L., Jiang, M., and Zhong, M. (2018). Rice soluble starch synthase I: allelic variation, expression, function, and interaction with Waxy. *Front. Plant Sci.* 9:1591.

Lv, Q., Huang, Z., Xu, X., Tang, L., Liu, H., and Wang, C. (2017). Allelic variation of the rice blast resistance gene Pid3 in cultivated rice worldwide. *Sci. Rep.* 7:10362. doi: 10.1038/s41598-017-10617-2

Mao, H., Sun, S., Yao, J., Wang, C., Yu, S., Xu, C., et al. (2010). Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc. Natl. Acad. Sci. U S A.* 107, 19579–19584. doi: 10.1073/pnas.1014419107

McCouch, S., Sweeney, M., Li, J., Jiang, H., Thomson, M., and Septiningsih, E. (2007). Through the genetic bottleneck: O. rufipogon as a source of trait-enhancing alleles for O. sativa. *Euphytica* 154, 317–339.

McCouch, S., Teytelman, L., Xu, Y., Lobos, K., Clare, K., and Walton, M. (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.) (supplement). *DNA Res.* 9, 257–279. doi: 10.1093/dnares/9.6.257

Monna, L., Kitazawa, N., Yoshino, R., Suzuki, J., Masuda, H., and Maehara, Y. (2002). Positional cloning of rice semidwarfing gene sd-1: rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. *DNA Res.* 9, 11–17. doi: 10.1093/dnares/9.1.11

Orjuela, J., Garavito, A., Bouniol, M., Arbelaez, J., Moreno, L., and Kimball, J. (2010). A universal core genetic map for rice. *Theor. Appl. Genet.* 120, 563–572. doi: 10.1007/s00122-009-1176-1

Paterson, A., Lander, E., Hewitt, J., Peterson, S., Lincoln, S., and Tanksley, S. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335, 721–726. doi: 10.1038/335721a0

Qi, P., Lin, Y., Song, X., Shen, J., Huang, W., Shan, J., et al. (2012). The novel quantitative trait locus GL3.1 controls rice grain size and yield by regulating Cyclin-T1;3. *Cell Res.* 22, 1666–1680. doi: 10.1038/cr.2012.151

Rama, S., Singh, K., Umakanth, B., Vishalakshi, B., Renuka, P., and Vijay, S. (2015). Development and identification of novel rice blast resistant sources and their characterization using molecular markers. *Rice Sci.* 22, 300–308.

Ramos, J., Furuta, T., Kanako, U., Niwa, C., Rosalyn, B., and Angeles, S. (2016). Development of chromosome segment substitution lines (CSSLs) of *Oryza longistaminata* A., Chev., Röhr in the background of the elite japonica rice cultivar Taichung 65 and their evaluation for yield traits. *Euphytica* 210, 151–163.

Rangel, P., Brondani, R., Rangel, P., and Brondani, C. (2008). Agronomic and molecular characterization of introgression lines from the interspecific cross *Oryza sativa* (BG90-2) x Oryza glumaepatula (RS-16). *Genet. Mol. Res.* 7, 184–195. doi: 10.4238/vol7-1gmr406

Ray, D., Mueller, N., West, P., and Foley, J. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8:e66428. doi: 10.1371/journal.pone.0066428

Ren, F., Lu, B., Li, S., Huang, J., and Zhu, Y. (2003). A comparative study of genetic relationships among the AA-genome Oryza species using RAPD and SSR markers. *Theor. Appl. Genet.* 108, 113–120. doi: 10.1007/s00122-003-1414-x

Saito, H., Okumoto, Y., Tsukiyama, T., Xu, C., Teraishi, M., and Tanisaka, T. (2019). Allelic differentiation at the E1/Ghd7 Locus has allowed expansion of rice cultivation area. *Plants* 8:550. doi: 10.3390/plants8120550

Sandhu, N., Jain, S., Kumar, A., Mehla, B. S., and Jain, R. (2013). Genetic variation, linkage mapping of QTL and correlation studies for yield, root, and agronomic traits for aerobic adaptation. *BMC Genet.* 14:104. doi: 10.1186/1471-2156-14-104

Sasaki, A., Ashikari, M., Ueguchi-Tanaka, M., Itoh, H., Nishimura, A., and Swapan, D. (2002). Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature* 416, 701–702.

Si, L., Chen, J., Huang, X., Gong, H., Luo, J., and Hou, Q. (2016). OsSPL13 controls grain size in cultivated rice. *Nat. Genet.* 48, 447–456. doi: 10.1038/ng.3518

Song, X., Huang, W., Shi, M., Zhu, M., and Lin, H. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* 39, 623–630. doi: 10.1038/ng2014

Sun, L., Li, X., Fu, Y., Zhu, Z., Tan, L., and Liu, F. (2013). GS6, A member of the GRAS gene family, negatively regulates grain size in rice. *J. Integr. Plant Biol.* 55, 38–949. doi: 10.1111/jipb.12062

Sun, Q., Wang, K., Yoshimura, A., and Doi, K. (2002). Genetic differentiation for nuclear mitochondrial and chloroplast genomes in common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 104, 1335–1345. doi: 10.1007/s00122-002-0878-4

Takano-Kai, N., Jiang, H., Kubo, T., Sweeney, M., Matsumoto, T., and Kanamori, H. (2009). Evolutionary history of GS3 a gene conferring grain length in rice. *Genetics* 182, 1323–1334. doi: 10.1534/genetics.109.103002

Tanksley, S., and Mccouch, S. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277, 1063–1066. doi: 10.1126/science.277.5329.1063

Tian, F., Li, D., Fu, Q., Zhu, Z., Fu, Y., Wang, X., et al. (2006). Construction of introgression lines carrying wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield-related traits. *Theor. Appl. Genet.* 112, 570–580. doi: 10.1007/s00122-005-0165-2

Vaughan, D., Kadowaki, K., Kaga, A., and Tomooka, N. (2005). On the phlylogeny and biogeography of the genus Oryza. *Breed. Sci.* 55, 113–122. doi: 10.1016/j.simyco.2014.10.002

Virmani, S., Aquino, R., and Khush, G. (1982). Heterosis breeding in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 63, 373–380.

Wang, Y., Xiong, G., Hu, J., Jiang, L., Yu, H., and Xu, J. (2015). Copy number variation at the GL7 locus contributes to grain size diversity in rice. *Nat. Genet.* 47, 944–948. doi: 10.1038/ng.3346

Xiao, J., Li, J., Grandillo, S., Ahn, S., Yuan, L., Tanksley, S., et al. (1998). Identification of trait-improving quantitative trait loci alleles from a wild rice relative. *Oryza rufipogon*. *Genetics* 150, 899–909. doi: 10.1093/genetics/150.2.899

Xing, Y., and Zhang, Q. (2010). Genetic and molecular bases of rice yield. *Annu. Rev. Plant Biol.* 61, 421–442. doi: 10.1146/annurev-arplant-042809-112209

Xu, F., Fang, J., Ou, S., Gao, S., Zhang, F., and Du, L. (2015). Variations in CYP78A13 coding region influence grain size and yield in rice. *Plant Cell Environ.* 38, 800–811. doi: 10.1111/pce.12452

Xu, P., Zhou, J., Li, J., Hu, F., Deng, X., and Feng, S. (2014). Mapping three new interspecific hybrid sterile loci between *Oryza sativa* and O. glaberrima. *Breed. Sci.* 63, 476–482. doi: 10.1270/jsbbs.63.476

Xu, P., Zhou, J., Li, J., Zhang, Y., Hu, F., and Liu, S. (2015). Identification and mapping of a novel blast resistance gene pi57(t) in *Oryza longistaminata*. *Euphytica* 205, 95–102. doi: 10.1007/s10681-015-1402-7

Yamagata, Y., Win, K., Miyazaki, Y., Ogata, C., Yasui, H., and Yoshimura, A. (2019). Development of introgression lines of AA genome Oryza species O. glaberrima, O. rufipogon and O. nivara in the genetic background of O. sativa L. cv Taichung 65. *Breed. Sci.* 69, 359–363. doi: 10.1270/jsbbs.19002

Yang, D., Ye, X., Zheng, X., Cheng, C., Ye, N., and Huang, F. (2016). Development and evaluation of chromosome segment substitution lines carrying overlapping chromosome segments of the whole wild rice genome. *Front. Plant Sci.* 7:1737. doi: 10.3389/fpls.2016.01737

Young, N., and Tanksley, S. (1989). Restriction fragment length polymorphism maps and the concept of graphical genotypes. *Theor. Appl. Genet.* 77, 95–101.

Zhao, Y., Huang, J., Wang, Z., Jing, S., Wang, Y., and Ouyang, Y. (2016). Allelic diversity in an NLR gene BPH9 enables rice to combat planthopper variation. *Proc. Natl. Acad. Sci. U S A* 113, 12850–12855.

Zhu, Q., Zheng, X., Luo, J., Gaut, B., and Ge, S. (2007). Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Mol. Biol. Evol.* 24, 875–888. doi: 10.1093/molbev/msm005

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 04 May 2022
doi: 10.3389/fpls.2022.884632

Dissection of Structural Reorganization of Wheat 5B Chromosome Associated With Interspecies Recombination Suppression

Elena Salina[1,2], Alexander Muterko[1], Antonina Kiseleva[1,2], Zhiyong Liu[3] and Abraham Korol[4]*

[1]Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russia, [2]Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia, [3]Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China, [4]Institute of Evolution, University of Haifa, Haifa, Israel

Chromosomal rearrangements that lead to recombination suppression can have a significant impact on speciation, and they are also important for breeding. The regions of recombination suppression in wheat chromosome 5B were identified based on comparisons of the 5B map of a cross between the Chinese Spring (CS) variety of hexaploid wheat and CS-5Bdic (genotype CS with 5B substituted with its homologue from tetraploid *Triticum dicoccoides*) with several 5B maps of tetraploid and hexaploid wheat. In total, two regions were selected in which recombination suppression occurred in cross CS × CS-5Bdic when compared with other maps: one on the short arm, 5BS_RS, limited by markers BS00009810/BS00022336, and the second on the long arm, 5BL_RS, between markers Ra_c10633_2155 and BS00087043. The regions marked as 5BS_RS and 5BL_RS, with lengths of 5 Mb and 3.6 Mb, respectively, were mined from the 5B pseudomolecule of CS and compared to the homoeologous regions (7.6 and 3.8 Mb, respectively) of the 5B pseudomolecule of Zavitan (*T. dicoccoides*). It was shown that, in the case of 5BS_RS, the local heterochromatin islands determined by the satellite DNA (119.2) and transposable element arrays, as well as the dissimilarity caused by large insertions/deletions (chromosome rearrangements) between 5BSs *aestivum/dicoccoides*, are likely the key determinants of recombination suppression in the region. Two major and two minor segments with significant loss of similarity were recognized within the 5BL_RS region. It was shown that the loss of similarity, which can lead to suppression of recombination in the 5BL_RS region, is caused by chromosomal rearrangements, driven by the activity of mobile genetic elements (both DNA transposons and long terminal repeat retrotransposons) and their divergence during evolution. It was noted that the regions marked as 5BS_RS and 5BL_RS are associated with chromosomal rearrangements identified earlier by C-banding analysis of intraspecific polymorphism of tetraploid emmer wheat. The revealed divergence in 5BS_RS and 5BL_RS may be a consequence of interspecific hybridization, plant genetic adaptation, or both.

Keywords: wheat, *dicoccoides*, recombination suppression, chromosome 5B, DNA transposons, LTR retrotransposons, tandem repeats, heterochromatin

OPEN ACCESS

Edited by:
Ruslan Kalendar, University of Helsinki, Finland

Reviewed by:
Ming Hao, Sichuan Agricultural University, China
Khalil Kashkush, Ben-Gurion University of the Negev, Israel

****Correspondence:***
Elena Salina
salina@bionet.nsc.ru

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

Received: *26 February 2022*
Accepted: *08 April 2022*
Published: *04 May 2022*

Citation:
Salina E, Muterko A, Kiseleva A, Liu Z and Korol A (2022) Dissection of Structural Reorganization of Wheat 5B Chromosome Associated With Interspecies Recombination Suppression. Front. Plant Sci. 13:884632. doi: 10.3389/fpls.2022.884632

INTRODUCTION

The evolutionary history of the wheat genome has been widely discussed over the last three decades (Goncharov, 2011; Gornicki et al., 2014; El Baidouri et al., 2017; Mirzaghaderi and Mason, 2017; Pont et al., 2019; Li et al., 2022), with genomic changes occurring during both the divergence of its diploid progenitors and the formation of tetraploid and hexaploid forms of wheat and their subsequent evolution. Overall, results suggest that although structural features of the genomes of the diploid species involved in amphiploidization were retained, certain changes also occurred in each round of polyploidization (Jiang and Gill, 1994; Devos et al., 1995; Salina et al., 2006; Marcussen et al., 2014; Jorgensen et al., 2017). The main genomic changes arising during the allopolyploids' formation were connected to the need for joint "accommodation" of the different genomes in one nucleus and cytogenetic diploidization. The wheat B genome and its progenitor in the diploid species *Aegilops speltoides* are rich in tandem repeats (families pSc119.2, pAs1, GAAGAG, etc.), which are involved in the formation of heterochromatin blocks; in contrast, the A genome is quite lacking in tandem repeat families. This imbalance in the content of tandem repeats in the genomes of diploid progenitors may significantly affect meiotic synchronization in the newly developed amphiploid nucleus. Variability in telomeric heterochromatin and centromeric regions has been seen in wheat-rye hybrids and in their progeny (Fu et al., 2013; Bashir et al., 2018). In the case of the bread wheat B genome, Spelt1 repeats specific to the subtelomeric regions of the diploid progenitor *Ae. speltoides* were lost in this process, which was also experimentally confirmed using newly synthesized amphiploids (Salina et al., 2004). Thus, the imbalance between homoeologous genomes A and B in the content of tandem repeats may decrease as a result of partial deletion of tandem repeats specific to one of the parents during the formation of amphiploids.

The situation with mobile elements is somewhat different. In particular, a study of long terminal repeat (LTR) retroelements in wheats and their progenitors has shown that most transposable elements (TEs) proliferated differentially in diploid progenitors and were then inherited by allopolyploids (Bariah et al., 2020a). Presumably, amphiploidization did not affect the proliferation of retrotransposons (Charles et al., 2008; Salina et al., 2011). However, certain changes in the short interspersed nuclear element (SINE) and non-LTR-retrotransposon families in the second generation of synthetic allohexaploid wheat, and their further active methylation in the third generation, have been reported (Ben-David et al., 2013). Thus, the reorganization of mobile elements at stages following the formation of the first amphiploids warrants further analysis.

One of the most important areas in allopolyploid research is determining the activity of the genes once they undergo duplication. Various interacting mechanisms can partially switch off genes *via* methylation or mutations of duplicated genes (Feldman and Levy, 2012). The consequences of such processes are of paramount importance in allopolyploid organisms. Thus, some 45S RNA genes are inactivated by methylation in chromosomes of one of the species constituting the allopolyploid genome in order to maintain the necessary number of 45S RNA in the cells (Shcherban et al., 2008). The *Ph1* locus emerged in allopolyploid species to prohibit homoeologous pairing in allopolyploid meiosis. Candidates for the *Ph1* gene, as well as the molecular mechanisms underlying its effect, have been discussed (Griffiths et al., 2006; Bhullar et al., 2014; Fan et al., 2021). The formation of an active *Ph1* locus on chromosome 5B also reflects the consequences of interactions between homoeologous gene loci that include activation of genes on one chromosome accompanied by silencing of their orthologues on the homoeologous chromosomes. Chromosome 5B of common wheat is known for the fact that its reorganization throughout the course of evolution led to the emergence of the *Ph* locus, which then contributed to stabilization of the genome of the tetraploid wild emmer wheat *Triticum dicoccoides* and other widespread allopolyploid *Triticum* species. The emergence of *T. dicoccoides*, a wild wheat species of the emmer group, was accompanied by species-specific translocation involving the arms of chromosomes 4AL, 5AL, and 7BS, which was then transmitted to tetraploid and hexaploid species of the emmer group, including cultivated wheat (Naranjo et al., 1987; Devos et al., 1995; Nelson et al., 1995). Chromosome 5B was not involved in large translocation rearrangements during the allopolyploidization.

At the same time, the use of the LTR retrotransposon family in wheat as a genetic marker allowed to identify large-scale genomic rearrangements between the reference chromosome 5B sequences of wild emmer and bread wheat. Six cases of large-scale rearrangements were identified, including 4 cases of long deletions in bread wheat, the introduction of a new DNA fragment, and a single example of copy number variation of a long tandem repeat in chromosome 5B (Bariah et al., 2020b).

A high level of intraspecific polymorphism for chromosome 5B has been noted for the allopolyploid emmer wheat group (*T. dicoccoides, Triticum dicoccum, Triticum aestivum*). Translocations between the long arm of chromosome 7A and the short arm of chromosome 5B were detected in *T. dicoccoides* and *T. dicoccum* (Badaeva et al., 2015). There is a high level of population-specific and region-specific polymorphisms in the distribution of C-bands on the long arm of chromosome 5B in emmer wheat (Badaeva et al., 2015). A group of cultivars of the hexaploid wheat *T. aestivum* that carry 5BS:7BS and 5BL:7BL translocations resulting from intragenomic rearrangements have been bred and are still successfully grown in Western Europe (Badaeva et al., 2007). Thus, chromosome 5B has definite potential for reorganization, resulting in the emergence of intraspecific polymorphisms in common wheat that are utilized in breeding programs.

One significant consequence of intra- or interspecies chromosome rearrangements is the suppression of recombination in hybrids obtained from crossing forms that differ in such rearrangements. Chromosomal rearrangements that lead to recombination suppression can have a significant impact on genetic divergence and even speciation, but are also important for practical breeding.

In the present work, we evaluated the key points of chromosome 5B reorganization throughout evolution by identifying regions of recombination suppresssion in interspecific crosses. Sequence

analysis of 5B pseudomolecules of *T. aestivum* (2*n* = 6x) and *T. dicoccoides* (2*n* = 4x) in the studied regions on 5B allowed us to identify possible determinants of recombination suppression.

MATERIALS AND METHODS

Genetic Linkage Maps

In this work, we used the genetic linkage map of CS and a disomic substitution line in which the CS 5B chromosome is replaced with the *T. dicoccoides* 5B chromosome (CS-5Bdic; Salina et al., 2018). Tetraploid wheat maps developed with populations from a cross of *T. durum* and *T. dicoccoides* were also used. The first map, based on a Svevo × Zv recombinant inbred line (RIL) population, was constructed and published by Avni et al. (2014), and the second map was based on a Langdon × Hermon (TZ-2) RIL population, which was genotyping previously (Wang et al., 2018). Both maps were developed using MultiPoint-UltraDense software. The hexaploid wheat map was constructed using the same tools, with the genotyping data of the Chara × Glenlea doubled haploid population kindly provided by E. Akhunov (Kansas State University) and previously published (Wang et al., 2014). The information about the mapping populations used in this study is summarized in **Supplementary Table S1**.

Comparative Analysis of Genetic Maps for Chromosome 5B

To compare marker order between genetic linkage maps, The Genetic Map Comparator (Holtz et al., 2017) and BioMercator v4 (Sosnowski et al., 2012) were used. To identify chromosomal regions showing considerable recombination suppression, we compared distances between corresponding markers in CS × CS-5Bdic and the tetraploid and hexaploid wheat maps mentioned above. The regions in CS × CS-5Bdic where the distances between the markers were reduced by more than 3–3.5 times compared to the tetraploid and hexaploid wheat were selected for the following analysis. Information about markers flanking these regions of suppression is presented in **Supplementary Table S2**.

The recombination regions mapped to deletion bins were performed based on CS × CS-5Bdic genetic mapping data using SNP, SSR, ISBP (insertion site-based polymorphism) markers and localization SSR, ISBP, and SNP markers in bins (Gadaleta et al., 2014; Salina et al., 2018).

Comparative DNA Analysis for Target Regions of Chromosome 5B

Regions of chromosome 5B where recombination was suppressed according to the results of the genetic analysis were extracted from the reference pseudomolecules of *T. aestivum* cv. CS (IWGSC RefSeq v2.1)[1] and *T. dicoccoides* cv. Zv (WEWSeq v.2.0)[2] based on localization of flanking DNA markers. Genome assembly of CS line 42 (CS42, PRJNA392179) was retrieved from the NCBI server.[3] A sequence similarity search against pseudomolecules of *T. aestivum* varieties Arina (Switzerland), Jagger (United States), Julius (Germany), and Landmark (Canada) was carried out on the IPK server.[4] Multiple sequence alignment was carried out using MAFFT 7.311 software (Katoh and Standley, 2013).

Tandem repeats were identified with Tandem Repeats Finder software (Benson, 1999) using the default settings. All nested repeats were grouped. Groups separated by less than 100 bp were merged into the same cluster. The obtained clusters containing tandemly repeated DNA sequences were filtered by length ≥ 1 kb. For each cluster containing repeats of the pSc119.2 satDNA family, the sequences of the repeated motif were aligned, and a consensus sequence was obtained.

To calculate the map coverage, local DNA sequence alignments were generated using NUCmer, included in MUMmer (Kurtz et al., 2004) with the following parameters: "-mum -b 150 -c 600 -g 30 -l 24," and local alignments with ≥80% similarity to the reference sequence were studied.

DNA-sequence alignments for the collinearity analysis were obtained with BLASTn (Zhang et al., 2000; Camacho et al., 2009) and filtered by length (≥3,000 bp) and identity (≥75%). The assemblies containing non-overlapping fragments ordered on the same strand were calculated, and the most representative assembly was taken for downstream analysis.

TEs were recognized in RepeatModeler (v2.0.2, Smit et al., 2008) and RepeatMasker (v4.1.2-p1, Smit et al., 2013) using the RepBase database (RepBaseRepeatMaskerEdition- 20181026, Bao et al., 2015).

Dot plots of 5BS_RS and 5BL_RS regions were calculated in Gepard (v.1.40, Krumsiek et al., 2007) using a word length of 30 bp. Circular data visualization was performed in Circos (Krzywinski et al., 2009).

DNA Shape Analysis

Modeling of predicted DNA helical paths and calculations of DNA shape features were performed according to previously published methods (Muterko, 2017). The curvature distribution was analyzed in a sliding window of 40 bp with a 1-bp step. DFT was performed with a sampling frequency of 1,024. Three-dimensional DNA structures were visualized by the PyMOL Molecular Graphics System, version 1.7.2 Schrödinger, LLC.

Evaluation of LTR-Retrotransposon Divergence

Putative LTR retrotransposons with LTR–LTR divergence ≥60% were predicted with LTRharvest (Ellinghaus et al., 2008), and elements with a LTR length ≥ 180 bp were selected. LTR of each element was aligned, and free-end gaps of 150-bp fragments were used in computing divergence. LTR TE divergence was calculated from LTR–LTR mismatches, assuming InDels as single evolutionary events. The mean LTR–LTR

[1]https://wheat-urgi.versailles.inra.fr/Seq-Repository
[2]http://wewseq.wixsite.com/consortium
[3]ftp://ftp.ncbi.nlm.nih.gov/genomes
[4]http://webblast.ipk-gatersleben.de/wheat_ten_genomes

divergence (age) is the local maximum of the Gamma distribution, with the shape and scale parameters obtained from the negative binomial distribution fitted to the LTR–LTR mismatch distribution. Furthermore, to evaluate the fractions of LTR–LTR divergences (ages) and improve the fit to the LTR–LTR mismatch distribution, the mixture of Poisson distributions was resolved with the use of the expectation–maximization algorithm. Goodness of fit was evaluated using the coefficient of determination (r^2). LTR TE carrying a number of mismatches equal to or less than the mean LTR divergence was considered young, while all others were considered old. The over- and under-representation of young LTR elements was evaluated in a sliding window of 10 elements with a 1 element step, using hypergeometric test.

RESULTS

Comparative Analysis of Chromosome 5B Genetic Maps

To identify chromosome 5B rearrangements that have occurred during evolution, we compared chromosome 5B recombination frequency in hybridizations of tetraploid (*T. durum* × *T. dicoccoides*) and hexaploid (*T. aestivum* × *T. aestivum*) wheat and in hybrids of hexaploid wheat cv. Chinese Spring (CS) and a disomic substitution line in which CS chromosome 5B has been replaced with *T. dicoccoides* chromosome 5B (CS-5Bdic). Only populations genotyped with Illumina Infinium arrays were used in this work (**Supplementary Table S1**). All of the maps were constructed using MultiPoint-UltraDense software.

A comparison of the maps demonstrated that the used markers were arranged in the same order (**Figure 1**). Although the total number of markers in the Chara × Glenlea and Svevo × Zavitan (Zv) maps compared to that of CS × CS-5Bdic was larger, the number of skeletal markers (markers representing groups of cosegregating markers occupying separate positions on the constructed linkage map) varied to a lesser degree. A comparative analysis of the length of the 5B maps and the distances between corresponding markers in the studied tetraploid and hexaploid populations relative to CS × CS-5Bdic showed that the latter was 1.5- to 1.7-fold shorter.

The regions with suppressed recombination are evident on the 5B map of CS × CS-5Bdic (**Figure 1**). In these regions, the distances between the markers are reduced by more than 3–3.5 times, as compared to the maps constructed for tetraploid (Svevo × Zavitan and Langdon × Hermon) and hexaploid (Chara × Glenlea) wheats and exceed the whole map reduction coefficient (1.5–1.7) by more than twofold. Two regions of recombination suppression on the short arms (between markers BS00009810 and BS00022336) and long arms (between markers Ra_c10633_2155 and BS00087043) are observed in all comparisons of intra- versus interspecies chromosome maps (all markers cosegregating with these basic markers are summarized in **Supplementary Table S2**).

Recombination suppression is presumably associated with the fact that hybridization in the case of CS × CS-5Bdic involved chromosomes 5B of common wheat and the wild species *Triticum dicoccoides*, which may have discrepancies along the 5B sequence. In addition, heterochromatin and possibly associated epigenetic regulation may be involved in recombination suppression (Termolino et al., 2016) and divergence of sequences of related wheat species. For a better understanding of the potential causes of recombination suppression in specific regions of chromosome 5B, the 5B pseudomolecules of *Triticum* species need to be compared.

Comparative Analysis of *Triticum aestivum* and *Triticum dicoccoides* Chromosome 5B Regions Associated With Recombination Suppression in Interspecific Cross

To define the possible determinants of recombination suppression, we analyzed the primary structure of the two chromosome 5B regions of *T. aestivum* and *T. dicoccoides*: (a) 5BS_RS, between markers BS00009810 (same position as BS00083715: 10489937–10,490,037 bp (CS chr5B pseudomolecule), 9,941,766–9,941,866 bp (Zv chr5B pseudomolecule)) and BS00022336 (15585936–15,586,036 bp (CS), 17,514,968–17,515,068 (Zv)); and (b) 5BL_RS, between markers Ra_c10633_2155 [same position as Tdurum_contig 25513_123: 549917016–549,917,116 bp (CS), 561,877,394–561,877,494 bp (Zv)] and BS00087043 [553533144–553,533,244 bp (CS), 565,677,935–565,678,035 bp (Zv)]. The regions of recombination suppression on the short and long arms, with lengths of 5 Mb and 3.6 Mb, respectively, were mined from the 5B pseudomolecule of CS and compared to the homoeologous regions (7.6 and 3.8 Mb, respectively) of the 5B pseudomolecule of *T. dicoccoides* Zv.

5BS_RS Regions

Comparison of the 5BS_RS regions of CS and Zv revealed two areas of tandem repeat sequences and several areas that had either very low similarity or did not map to CS chromosomes at all (**Figure 2A**). The latter included different chromosomal rearrangements, with a predominance of insertions/deletions (InDels; **Figure 2B**). This also followed from the difference in length of the 5BS_RS region on pseudomolecules of CS and Zv. In particular, the 5BS_RS of Zv includes around 2 Mb of the DNA sequences, which did not map to the collinear segment of the CS chromosome, suggesting that they are the major contributors to recombination suppression in the 5BS_RS region through loss of similarity.

Extended arrays of tandemly repeated DNA sequences (satellite DNA [satDNA]) were localized to the beginning of the 5BS_RS region. A comparison of fractions containing clusters of tandem repeats with repetitive elements of different lengths, as well as an evaluation of the total length of repetitive sequences in each fraction, showed that repeated sequences within 5BS_RS are mainly represented by clusters of a 118-bp tandem unit or its dimer (235–236 bp) or trimer (353–354 bp). Primary structure analysis has shown that this highly repetitive sequence belongs to the pSc119.2 satDNA family, which is widely distributed within the *Triticeae*, as well as in some *Aveneae* species, and forms a large and evolutionarily old component of the genome (Contento et al., 2005). For a more detailed investigation of

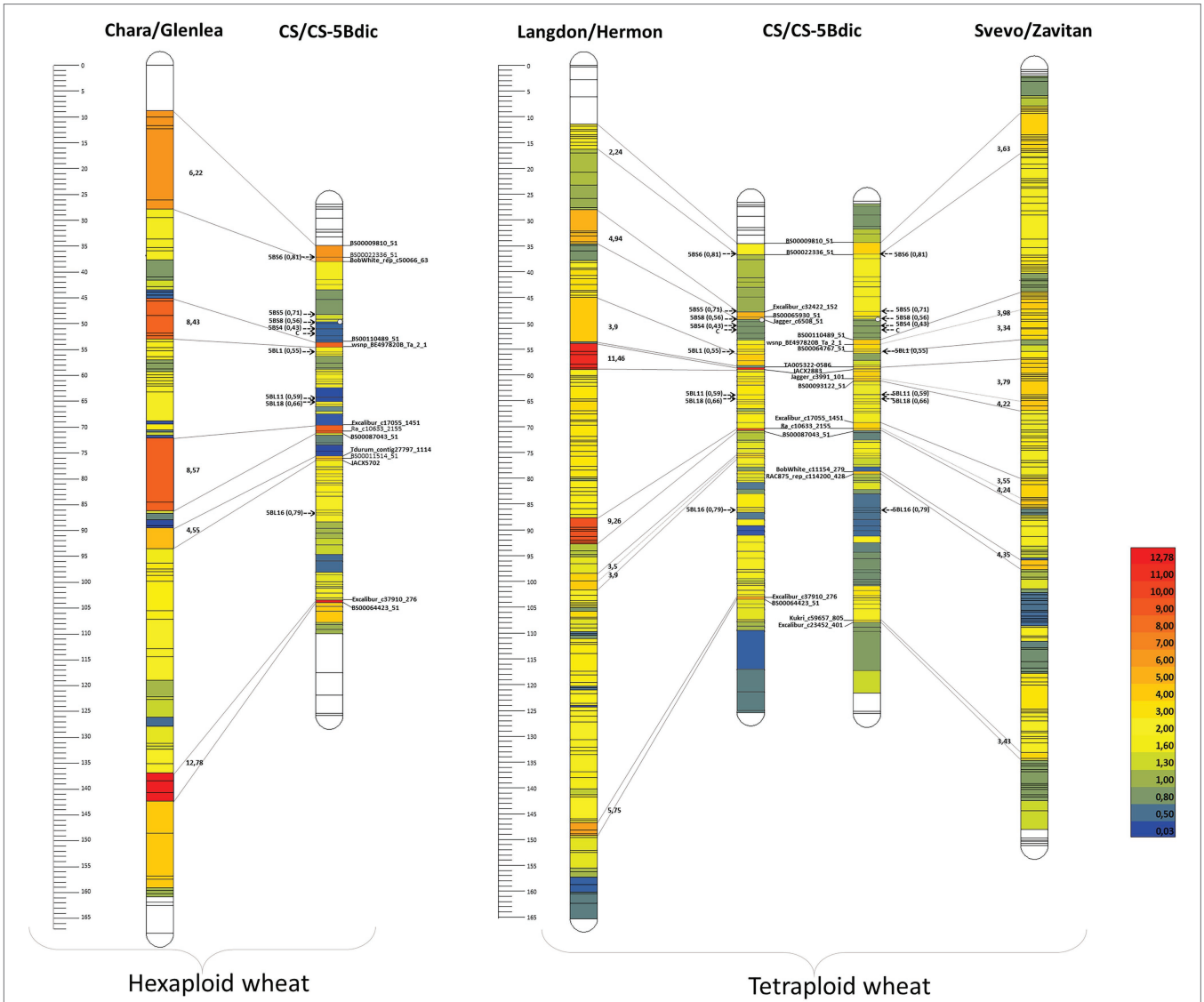


FIGURE 1 | Comparison of CS × CS-5Bdic RIL map with maps of three populations: RIL *T. durum* x *T. dicoccoides* (Svevo/Zavitan); *T. durum* x *T. dicoccoides* (Langdon/Hermon); double haploid population from crossing common wheat Chara x Glenlea. The level of suppression compared to CS × CS-5Bdic is displayed on a color scale; arrows indicate centromeres and breaking points of bin; distances on ruler are presented in cM; lines are connected to the regions with suppression of more than 3–4 times. The deletion bins are points 5BL16, 5BL18, 5BL11, 5BL1, 5BS4, 5BS8, 5BS5, and 5BS (Salina et al., 2018).

distribution and genetic diversity, the arrays of 118-bp tandem repeats of the pSc119.2 family were clustered and filtered, as described in Experimental Procedures, and a consensus sequence was deduced for each cluster. The analysis of pSc119.2 distribution identified three large pools (sized ~500, 600, and 90 kb) located on half of the 5BS_RS region (**Figures 2A,C**; **Supplementary Figure S1**). A total of 113 clusters of pSc119.2 were identified for CS and 82 for Zv (**Supplementary Figure S1**), even though both provided similar coverage (**Figure 2C**).

With the exclusion of single clusters, the 118-bp motif was highly conserved in each pool (identity of 86–93%) and between pools (94%) for both CS and Zv. Similarity between consensus DNA sequences of pSc119.2 pools of CS and Zv was 95–97%. Multiple sequence alignment found that pool 2 in CS contains an array of inverted clusters (**Supplementary Figure S1**). This is an important feature distinguishing pool 2 of the pSc119.2 clusters in CS and Zv. The total length of this inverted region was more than 100 kb. Furthermore, the chromosome region within pool 3 in both CS and Zv was inverted compared to pool 1 (as well as pool 2 of Zv and a part of pool 2 from CS). Although the presence of inverted repeats can lead to their instability in the genome, high-level identity in both the position and primary structure was revealed for the studied pSc119.2 repeats between wild tetraploid and cultivar hexaploid wheat (**Supplementary Figures S1, S2**).

Such stability of the tandem repeat areas is presumably determined by the tertiary structure of the DNA. The DNA sequence of pSc119.2 includes three A-tracts (AnTn, $n > 3$).

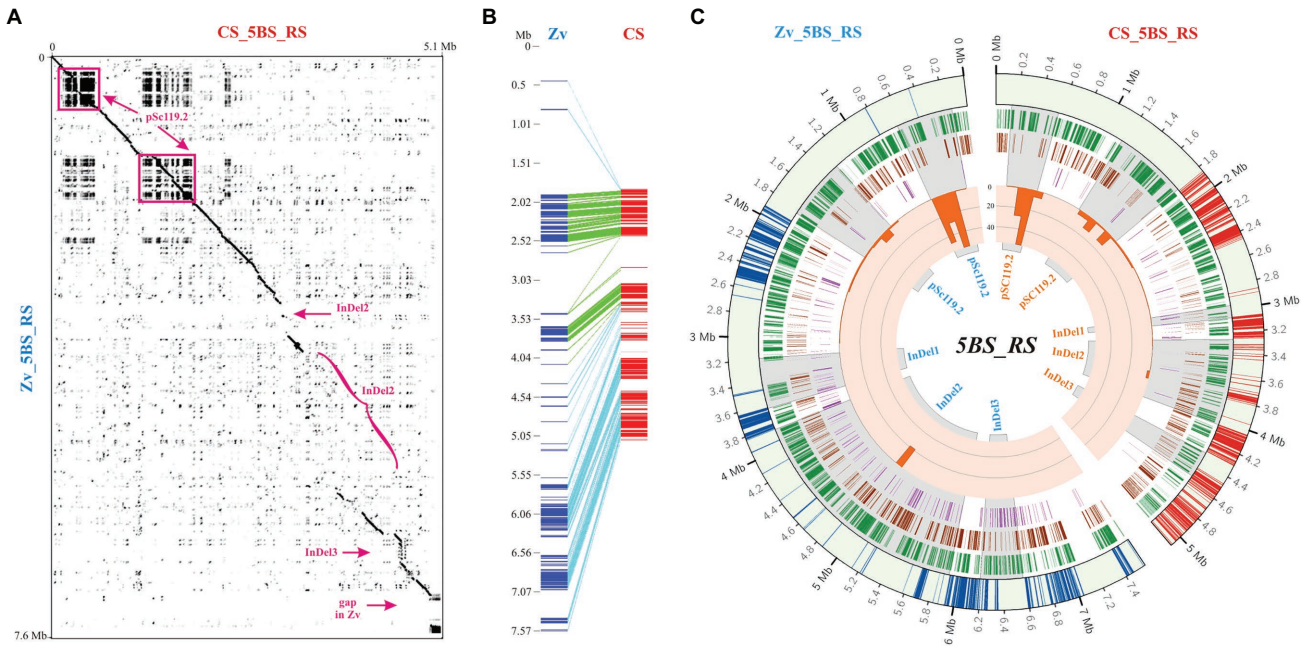


FIGURE 2 | Comparative analysis of the 5BS_RS region in Chinese Spring (CS) and Zavitan (Zv). **(A)** Dot plot of the 5BS_RS DNA sequences from the CS and Zv pseudomolecules. Localization of highly diverged regions containing insertions/deletions (InDels) and a 300-kb gap in Zv is indicated. **(B)** Collinearity between the 5BS_RS regions of CS and Zv assembled from fragments of >3 kb in length and >75% identity between species. **(C)** Distributions of TEs and satDNA for the 5BS_RS regions of CS and Zv. Four datasets are represented in the circular diagram, from outside to inside: distribution of LTR retroelements (track 1, green highlights), DNA transposons (track 2, brown highlights), unknown TEs (track 3, purple highlights); the fourth track (inside – orange) includes a histogram of the satDNA coverage distribution.

A-tracts are known as key determinants of the local bends and global curvature of the DNA molecule (Marini et al., 1982; Koo et al., 1986). This global DNA bending can increase significantly when A-tracts are repeated in phase with the helical screw (Marini et al., 1982; Koo et al., 1986; Koo and Crothers, 1988), i.e., are localized in a molecule such that the intrinsic curvature progressively increases in a certain direction. The tandemly repeated sequence containing A-tracts can easily satisfy such conditions. To evaluate the contribution of this satDNA to molecular bending, the predicted DNA helical path for the 118-bp motif of pSc119.2 was calculated and the distribution of the DNA curvature was analyzed (**Figure 3A**). We found that assuming a DNA helical period of 10 bp, the 118–120-bp periodicity of the pSc119.2 motif is in phase with the DNA helical screw. Two consecutive A-tracts within the analyzed pSc119.2 motif also alternated in phase with the DNA helix (20-bp distance), leading to an increase in the macroscopic DNA curvature. Since A-tracts and T-tracts produce bends in opposite directions, we suggest that those alternating at the same frequency and in phase with the helical screw should mutually compensate for each other. However, analysis of the discrete Fourier transform (DFT)-phase spectra of A- and T-tract alternations showed a phase shift of up to 150° (**Figure 3B**). Thus, T-tracts alternating out of phase with A-tracts, but producing a bend in the opposite direction, will increase the macroscopic curvature of the DNA molecule. This prediction also follows from the DFT power spectrum of the curvature distribution, where the two detected major frequency bins were 9 and 17, corresponding to 118- and 60-bp periodicity, respectively. One of the important features of regularity is the formation of extended homogeneous rigid structures. In the considered case, this structure is almost a straight helix with a 4.5-monomer period. To demonstrate this, the three-dimensional DNA helical path was inferred for a 90-kb fragment of pSc119.2 arrays (pool 1) flanking the common DNA sequences (**Figure 3C**). The assumption of involving of satDNA in heterochromatin formation was supported with the use of publicly available for CS the distribution of the chromatin states, determined with a multivariate hidden Markov model for different epigenetic marks (Li et al., 2019). It was shown that the pSc119.2 repeat-containing regions within 5BS_RS are associated with the E13 chromatin state, whose signature is the modified histone H3K9me2 (**Supplementary Figure S3**), found in constitutive heterochromatin of many eukaryotes. Furthermore, the DNase I profiling (Li et al., 2019) also indicates the low chromatin accessibility within these regions of 5BS_RS.

Although the presence of pSc119.2 tandem repeats providing denser DNA packaging in heterochromatin areas does not explain the observed suppression of recombination, the presence of constitutive heterochromatin can enhance the influence of adjacent regions that are associated with the recombination suppression.

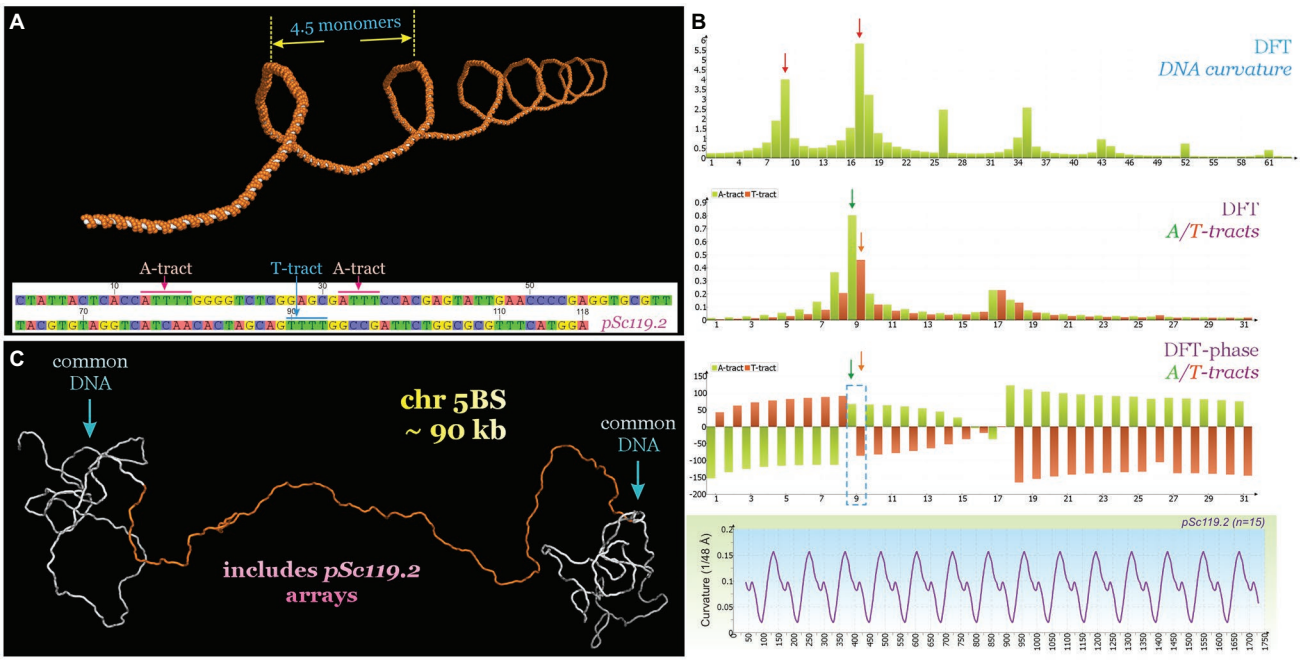


FIGURE 3 | (A) Predicted DNA helical path of the pSc119.2 monomer repeated 30 times. **(B)** Parts of DFT power spectra (sampling frequency is 1,024) of curvature and A/T-tract distributions, DFT-phase spectrum of A/T-tract distributions as well as curvature distribution of pSc119.2 array. **(C)** Predicted DNA trajectory of fragment of 5BS_RS region (704–1,012 kb) containing 90-kb pSc119.2 arrays (690–780 kb).

As noted above, a region with low similarity between the homologues can significantly contribute to recombination suppression in heterozygotes. Comparison of the 5BS_RS regions in CS and Zv revealed three large, highly diverged consecutive segments containing InDel-1: ~0.11 and 0.46 Mb, InDel-2: 0.77 and 2.1 Mb, and InDel-3: 0.25 and 0.38 Mb on CS and Zv pseudomolecules, respectively (**Figure 2A**). Detailed analysis of these regions characterized predominant insertions of TEs (**Figure 2C**). InDel-1 contained a 0.39-Mb insertion in CS and 0.28-Mb insertion in Zv. Within InDel-2, a total insertion length of 1.6 Mb was detected for Zv and 0.4 Mb for CS. InDel-3 included insertion of repeated DNA sequences in CS, most of which were nested. Zv also contained repeats in this region; however, they were not similar to those in CS and differed in lengths.

The mobile genetic elements identified within the first and second InDel regions were predominantly the TEs of the LTR order of the class retrotransposons, superfamily Gypsy, and Copia (**Supplementary Figure S2**). However, it is not clear when during the evolution of polyploid wheat these chromosomal rearrangements occurred. To shed light on this problem, the LTR-retrotransposon age distribution was analyzed along the 5BS_RS chromosome region, and the over- and under-representation of "young" LTR elements was evaluated within the segments where rearrangements in CS and Zv are localized (**Figure 4**). Analysis of LTR–LTR mismatch distribution in CS found the prevalent fraction of LTR divergence to be five mismatches (**Figure 4A**). Thus, the elements carrying five or less LTR–LTR mismatches were considered "young," whereas longer ones were regarded as "old." Assuming a substitution rate of 1.3×10^{-8} mutations per site per year for the plant LTR retrotransposons (Ma and Bennetzen, 2004), this age threshold corresponds to insertion of the LTR retrotransposons before and after 1.3 million years ago for "young" and "old" retrotransposons, respectively.

Statistically significant over-representation of "young" LTR elements ($p = 0.046$ and 0.037, respectively) was shown for two regions (300 and 500 kb), including InDel-1 in CS and the proximal part of InDel-2 in Zv (**Figure 4B**). The under-representation of young LTR elements was shown for a 1-Mb region between InDel-2 and InDel-3 in CS and Zv (**Figures 4B,C**).

Thus, given a wild emmer emergence time of about 0.5 MYA (Feldman and Levy, 2012) and according to the estimated time of movement of mobile elements, the rearrangements observed in the studied regions of chromosome 5BS took place in diploid progenitors of allopolyploid wheat and accompanied the processes of evolution and the formation of allopolyploids.

5BL_RS Regions

Despite the apparent similarity of the 5BL_RS segment between the CS and Zv genome assemblies, our analysis revealed several highly diverged regions (HDRs, **Figure 5A**). This loss of similarity was mainly caused by insertions and deletions (InDels) as it followed from the different lengths of 5BL_RS in CS and Zv pseudomolecules (3,617 and 3,802 Mb, respectively) and displacements in the distribution of collinear segments (**Figure 4B**). In particular, from 0.3 to 2 Mb (**Figure 5B**), the

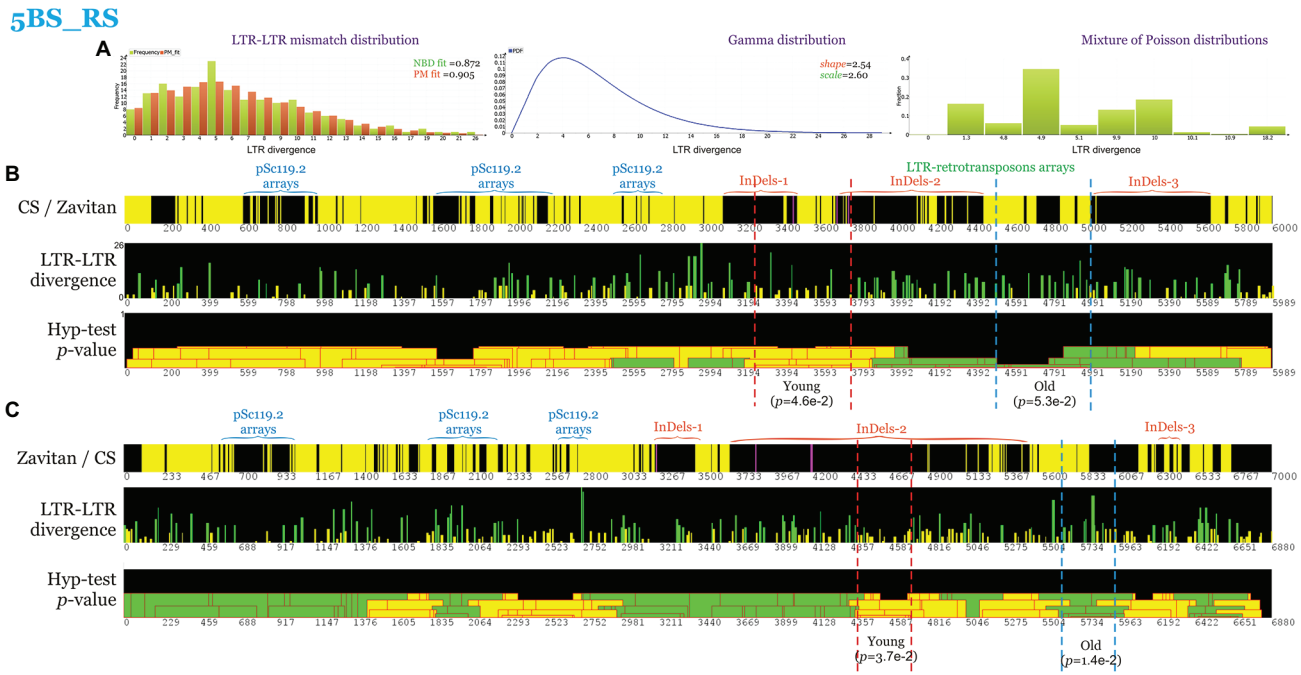


FIGURE 4 | Coverage maps and LTR-LTR divergence distribution within 5BS_RS. Positions are indicated in kb. **(A)** LTR-LTR mismatch distribution and mixture of Poisson distributions (PM) fit. Goodness of fit to negative binomial distribution and PM are indicated. Gamma distribution with the shape and scale parameters obtained from NBD. Mixture of Poisson distributions (abscissa—lambda values, ordinate—proportions). **(B)** Coverage maps and LTR–LTR divergence distribution of CS 5BS_RS region. LTR-TEs with 5 or less LTR-LTR mismatches were considered young (yellow, < 1.28 million years), while other—old (green, more than 1.28 million years). The *p*-values of the over- and under-representation of young LTR-elements are indicated. **(C)** Coverage maps and LTR-LTR divergence of Zavitan 5BS_RS region. Yellow and purple colors indicate the forward and reverse DNA strand, respectively, while black color indicates regions with similarity less than 80% or gaps.

displacement of collinear segments implied insertions in CS (or deletions in Zv), and from 2 Mb on, insertions in Zv (or deletions in CS). DNA sequence analysis of these regions revealed that in the first case, the initial offset of collinear segments is caused by the prolonged gap in CS (0.3–0.48 Mb), while all following displacements in this area, beginning from 1.16 Mb, are caused by insertions of additional sequences in the CS genome (**Figure 5B**). We conclude that in the second case, insertions of additional DNA sequences in the Zv genome beginning from a position of 2 Mb occurred at least twice, leading to the displacement of collinear segments within the alignment in the opposite direction.

Overall, two major and two minor segments with significant loss of similarity could be recognized within the 5BL_RS region (HDR1–4, **Figure 5A**). The first (HDR1) is localized proximal to the Tdurum_contig25513_123 marker (1–1.57 Mb) and includes around 550 kb on the 5B pseudomolecules of CS and Zv. The second (HDR2) begins near the center of 5BL_RS and expands to the BS00087043_51 marker (2–2.7 Mb). It encompasses almost 350 kb in CS and is twice as long in Zv (close to 700 kb), assuming insertions with a total length of 350 kb within the Zv genome. Although both of these segments are enriched with TEs, different types of TEs are not equally represented. In particular, in HDR1, the DNA transposons are prevalent over LTR retrotransposons (**Figure 5C**), and among them, elements of the DNA/CMC-EnSpm family are represented almost exclusively. Two short (around 0.1 and 0.2 Mb) distal segments (HDR3 and HDR4, respectively) with a complete absence of similarity were enriched in LTR retrotransposons (**Figure 5C**).

To investigate the possible reasons for divergence of HDRs within 5BL_RS, analysis of the LTR age distribution was carried out exactly as described above for 5BS_RS. The results obtained for Zv showed that in the HDRs of 5BL_RS, where LTR elements are few and DNA transposons are prevalent (HDR1 and HDR2), there is a statistically significant over-representation of the "old" LTR retrotransposons, whereas in the HDRs where "young" LTR elements are over-represented (HDR3 and HDR4), the DNA transposons are less frequent (**Figure 5C**, left side). Except for HDR2, opposite results were obtained for CS (**Figure 5C**, right side). In fact, this opposition is in good accordance with the occurrence of HDRs, which allows us to explain the suppression of recombination on 5BL_RS for homologous chromosomes of the considered wheat species. In particular, the over-representation of the "young" retroelements in HDR1 of CS agrees with the insertion of additional DNA sequences in this region (rather than their selective deletion in Zv) and hence contributes to the lowering of similarity with a collinear chromosomal segment in Zv for which the "old" LTR retrotransposons are over-represented on this region. Similarly, over-representation of the "young" LTR elements in HDR3 and HDR4 of Zv corresponds to over-representation

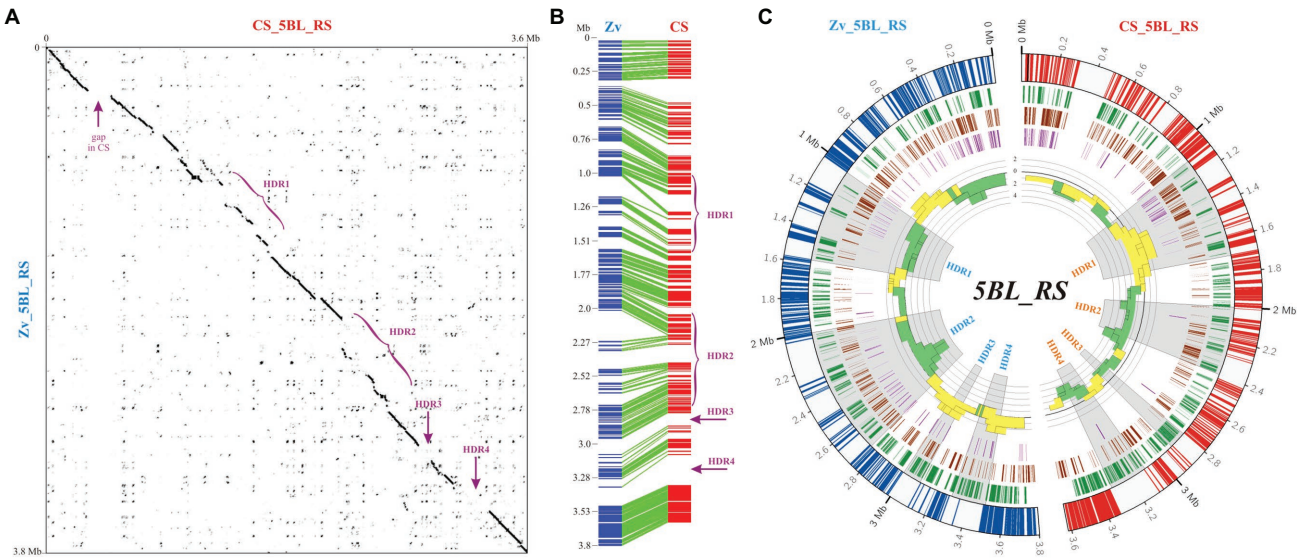


FIGURE 5 | Comparative analysis of the 5BL_RS region in Chinese Spring (CS) and Zavitan (Zv). **(A)** Dot plot of the 5BL_RS DNA sequences from CS and Zv pseudomolecules. Localization of highly diverged regions and 200-kb gap in CS is indicated. **(B)** Collinearity between the 5BL_RS regions of CS and Zv assembled from fragments of >3 kb in length and >75% identity between species. **(C)** Distributions of TEs and LTR-(retrotransposon?) age analysis for 5BL_RS regions of CS and Zv. Four datasets are represented in the circular diagram, from outside to inside: distribution of LTR retroelements (track 1, green highlights), DNA transposons (track 2, brown highlights), unknown TEs (track 3, purple highlights); the fourth track includes a histogram of the -log10 of *p*-values of the hypergeometric test for the over-representation of "young" (outside – yellow) and "old" (inside—green) LTR elements.

of the "old" LTR elements in the collinear regions of CS (**Figure 5C**).

Based on these data, we suggest that the loss of similarity, which can lead to recombination suppression in the 5BL_RS region, is caused by chromosomal rearrangements, driven by the activity of mobile genetic elements (both DNA transposons and LTR retrotransposons) and their divergence during evolution.

It should be noted that in CS, HDR2 includes a scaffold of around 400 kb (scaffold2564–2), which is flanked by a long span of simple tandem repeats (STRs) of the same type: (TCTTCT)n, $n > 500$. Since flanked STRs are known to cause problems for correct scaffold orientation during *de novo* genome assembly, the orientation of this scaffold on pseudomolecule 5B of CS (IWGSC assembly) until the appearance of version 1.2 was questionable. The direction of this genomic region in common wheat cultivars, such as *T. aestivum* varieties Arina, Jagger, Julius, and Landmark, is similar to that in Zv (Walkowiak et al., 2020). However, in all of these cases, the scaffolds break at the flanked satDNA. On the other hand, this inversion is absent in the 5B pseudomolecule of the CS42 genome assembly (Zimin et al., 2017), which, together with optical maps, was used to improve the IWGSC CS assembly during the release of RefSeq v2.1 (Zhu et al., 2021). Although these data most likely indicate the absence of an inversion within HDR2, we cannot rule out its presence on chromosomes 5B of CS and CS-5Bdic used in the cross in the present study. Inversion of a scaffold would be a good explanation for the recombination suppression on the 5BL_RS region, although we do not yet have enough evidence to confirm this assumption.

DISCUSSION

Recombination suppression is an important phenomenon involved in processes such as the formation of sex chromosomes, evolution of hybrid zones, and speciation (Mank, 2012; Charlesworth, 2016). In plant breeding, the transfer of agronomically important genes in non-recombinant linkage groups occurs when they are localized on alien translocations (Wingen et al., 2017). There are diverse causes of recombination suppression; the most commonly discussed are InDels, inversions, heterochromatin regions, and epigenetic mechanisms (Hammarlund et al., 2005; Ziolkowski et al., 2015; Guo et al., 2016).

In our case, recombination-suppression regions on wheat chromosome 5B were detected in a cross between CS and CS-5Bdic (genotype CS with 5B substitution by its homologue from tetraploid *T. dicoccoides*). It is currently not possible to use information on the 5Bdic chromosome donor to establish the possible causes of recombination suppression, although this substitution line has been actively used in various genetic studies (Faris et al., 2000; Lu et al., 2006; Salina et al., 2018). At the same time, much experimental material has been accumulated on intraspecific polymorphism of tetraploid species of the emmer group, which may be useful for identifying the causes of recombination suppression observed in CS × CS-5Bdic crosses. Thus, a comparative chromosome analysis by C-banding in 446 accessions of *T. dicoccum* and 105 accessions of *T. dicoccoides* from seven countries, covering its whole present-day natural distribution range, revealed population-specific and region-specific

polymorphisms (Badaeva et al., 2015). The following intraspecific polymorphisms were observed for chromosome 5B in *T. dicoccum* and *T. dicoccoides*: (1) a 7A:5B translocation involving the short arm of chromosome 5B (T5BL.5BS- 7AL); (2) differences in the distribution of diagnostic C-bands on the long chromosome arm. It should be noted that the 5BS_RS and 5BL_RS regions identified in our study also fall into the region of high intraspecific polymorphism. Thus, a more detailed analysis of the T5BL.5BS-7AL breakpoint on chromosome 5BS showed that it is located distal to the minor pSc119.2 site and the 5S rRNA gene locus, whereas the major pSc119.2 site was relocated to chromosome 7A (Dedkova et al., 2009; Badaeva et al., 2015). A comparison of these findings with our data on the location of the recombination-suppression region relative to the distribution of repeats on chromosome 5BS (**Supplementary Figure S2**) suggests that the region of the chromosome that we labeled 5BS_RS is involved in intraspecific rearrangement in *T. dicoccum* and *T. dicoccoides*. We cannot rule out the possibility that the breakpoint runs in close proximity to the tandem pSc119.2 repeats identified in this region (**Figure 2**).

Intraspecific polymorphism on the long arm of chromosome 5B was revealed by variation in the location of C bands, the total number of which on 5BL is eight (Badaeva et al., 2015). In *T. dicoccum* accessions, the region attributable to intraspecific polymorphism covers the interstitial and distal region of 5BL, whereas for *T. dicoccoides* accessions, this region is significantly wider and covers more than 70% of 5BL. The 5BL_RS region that we studied, located in the 5BL18 bin (**Figure 1**), also falls within the region of intraspecific polymorphism of *T. dicoccum* and *T. dicoccoides* species. In addition, as previously shown, alien translocations often occur in the 5BL18 bin of chromosome 5B during interspecific hybridization (Timonova et al., 2013). Taken together, this indicates the relative instability of the regions marked as 5BS_RS and 5BL_RS during both plant adaptation to different agroclimatic zones and interspecific hybridization.

The reasons for such instability, leading to intraspecific differences and subsequently, to the suppression of recombination during interspecific hybridization, can be clarified by a detailed analysis of the structure of the chromosome regions involved in these processes. As possible factors of suppressed recombination in 5BS_RS, 5BL_RS or both, one can consider heterochromatization associated with the presence of extended tracts of tandem repeats (119.2), large insertions, and highly divergent regions enriched with TEs.

Role of Tandem Repeats and TEs in Recombination Suppression

Tandem repeats and TEs are quite often associated with the heterochromatin regions of chromosomes. The possible participation of heterochromatin regions in recombination suppression has been discussed previously (Charlesworth et al., 1986; Kagawa et al., 2002). There is evidence that heterochromatin suppression of crossing-over is controlled in the chromatin structure (Westphal and Reuter, 2002).

The region of recombination suppression on the short arm of chromosome 5B is colocalized with a heterochromatin region at the boundary of two distal bins (5BS6–5BS5) according to physical and genetic mapping (**Figure 1**; Salina et al., 2018). Analysis of 5BS_RS revealed that nearly 2.8 Mb of this region is covered by pSc119.2 tandem repeats. The other part of the region is covered by TEs and displays very low between-species similarity due to extended InDels.

One of the most common features of satDNA, which is shared among many eukaryotic organisms, is its intrinsic curvature. In previous studies, it has been repeatedly noted that clusters of tandemly repeated nucleotide sequences probably play a key role in DNA stabilization, as well as in protein packaging and higher-order chromatin condensation (Fitzgerald et al., 1994; Tolstorukov et al., 2005; Mrazek, 2010). It is assumed that sequence-directed bends produced by the repeated A-tracts in the satDNA in phase with the helix facilitate the assembly of DNA into nucleosomes and therefore, represent an essential structural basis for heterochromatin condensation (Fitzgerald et al., 1994; Rohs et al., 2009; Segal and Widom, 2009; Struhl and Segal, 2013). The results obtained in the present study agree with the results of Vershinin and Heslop-Harrison (1998) indicating the phasing of nucleosomes along arrays of the pSc119.2 satDNA. Furthermore, the pSc119.2 repeat is known to form a major part of heterochromatin regions in the B genome of wheat (Mukai et al., 1993; Contento et al., 2005). In the present study, we demonstrated that the localization of A-tracts within the pSc119.2 motif leads to sequence-directed bending of the DNA molecule. We suggest that the observed periodicity of the DNA bend can facilitate molecular packing and may be a structural basis for local heterochromatin condensation in this region. The high degree of similarity between different clusters of pSc119.2 found herein supports the existing opinion that there is selective pressure acting on the maintenance of curved DNA (Charlesworth et al., 1994). Bearing in mind the overlapping location of the studied region with the heterochromatin stretch, we argue that large arrays of pSc119.2 satDNA within the 5BS_RS locus can lead to the formation of local islands of interstitial heterochromatin within the euchromatic chromosomal arm, contributing to recombination suppression in this and adjacent regions.

There are also other mechanisms supporting heterochromatin structure. Thus, direct or inverted repeats can lead to the creation of a heterochromatic state (Bennetzen et al., 1994; Dorer and Henikoff, 1994; Tikhonov et al., 1999). The methylation of histone H3mK9 (Jenuwein and Allis, 2001; Martienssen and Colot, 2001; Vermaak et al., 2003), DNA methylation (Martienssen and Colot, 2001; Law and Jacobsen, 2010) and histone variant H2A.W (Yelagandula et al., 2014) play an important function in regulating heterochromatin and act in synergy to maintain transposon silencing. In particular, in the present study it was shown that chromatin of the pSc119.2 repeat-containing regions within 5BS_RS is marked with histone H3K9me2. This last is the signature of repeat-rich constitutive heterochromatin for many eukaryotes (Naumann et al., 2005; Nicetto and Zaret, 2019) and is an evolutionarily conserved, specific mark of nuclear peripheral heterochromatin (Poleshko et al., 2019).

Furthermore, it has been shown that heterochromatin in *Arabidopsis* is determined by TEs and related tandem repeats under the control of the chromatin-remodeling ATPase DDM1 (Lippman et al., 2004). The enrichment of TE sequences in heterochromatin is widespread among higher eukaryotes. In plant genomes, even relatively small retrotransposon blocks are methylated (Bennetzen et al., 1994; SanMiguel et al., 1996; Tikhonov et al., 1999) and can be considered heterochromatic regions. Moreover, small tandem expansions, such as adjacent duplication of a TE from a monomer to a dimer, or an increase in the copy number within a simple sequence repeat, can lead to altered epigenetic status within and around these sequences, and these may also be considered heterochromatin (Dorer and Henikoff, 1994).

In the present study, during the analysis of the LTR–LTR divergence, we found indirect evidence for heterochromatin formation in the 5BS_RS region carrying extended TE arrays. It was shown that local heterochromatin islands, determined by the satDNA and TE arrays, as well as dissimilarity caused by the large InDels (chromosome rearrangements), are likely key determinants of recombination suppression in the 5BS_RS region.

Our results allow suggesting that mobile genetic elements (both DNA transposons and LTR retrotransposons) played a key role in the loss of similarity in the region labeled 5BL_RS. Overall, two major and two minor segments with significant loss of similarity can be recognized within the 5BL_RS region (HDR1–4, **Figure 3A**). The HDRs are enriched with TEs, but each region is characterized by a different history of integration and movement of LTR retroelements, and a different distribution ratio of DNA transposons and LTR retrotransposons. The results are consistent with recent studies indicating a negative correlation between regional genomic variation in TE density and frequency of recombination. The strength of this correlation varies depending on the TE type; TE insertion polymorphism may be an important factor determining intraspecific variations in recombination rate in the surrounding genomic regions, with all of the consequences of this event (Kent et al., 2017).

CONCLUSION

Chromosomal rearrangements that lead to recombination suppression can have a significant impact on speciation, and they are also important for breeding. In our case, recombination-suppression regions on wheat chromosome 5B (5BS_RS and 5BL_RS) were detected in a cross between CS and CS-5Bdic (genotype CS with 5B substitution by its homologue from tetraploid *T. dicoccoides)*. It was shown that the loss of similarity, which can lead to suppression of recombination in the studied regions, is caused by chromosomal rearrangements, driven by the activity of mobile genetic elements (both DNA transposons and long terminal repeat retrotransposons) and their divergence during evolution. Also, the local heterochromatin islands determined by the satellite DNA (pSc119.2), as well as the dissimilarity caused by large insertions/deletions (chromosome rearrangements) between *aestivum/dicoccoides*, are likely the key determinants of recombination suppression in the region.

It was noted that the regions marked as 5BS_RS and 5BL_RS are associated with chromosomal rearrangements identified earlier by C-banding analysis of intraspecific polymorphism of tetraploid emmer wheat. Taken together, this indicates the relative instability of the studied regions during both plant adaptation to different agroclimatic zones and interspecific hybridization.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

ES, AM, and AKo: conceptualization. AKi: formal analysis and investigation (comparative mapping data). ZL: formal analysis and investigation (Langdon × Hermon RIL mapping data). AM: formal analysis and investigation (RS regions of pseudomolecules, satDNA, TEs, large InDels, and DNA shape analyses). ES: project administration and supervision. ES, AM, and AKi: writing-original draft preparation. ES, AM, ZL, and AKo: writing, review, and editing. All authors contributed to the article and approved the submitted version.

FUNDING

Genetic linkage maps and comparative chromosome analysis were supported by RSF (Russian Science Foundation) project no. 21-76-30003. Analysis of the 5BS_RS and 5BL-RS regions was funded by the Kurchatov Genomics Center of IC&G (075-15-2019-1662). DNA shape analysis was performed within the budgetary project FWNR-2022-0017.

ACKNOWLEDGMENTS

We thank Camille Vainstein for English editing of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.884632/full#supplementary-material

Supplementary Figure S1 | (A) Scatter graph of distribution of number of pSc119.2 clusters with length not less than 1 kb, within the 5BS_RS region extracted from Cs and Zv pseudomolecules and divided on 100-kb segment. **(B)** Scatter graph of coverage by the sequences of the pSc119.2 clusters evaluated for each 100-kb segment. Arrows indicate direction (forward/inverse) of the pcS119.2 satDNA arrays.

Supplementary Figure S2 | Distribution of the pSc119.2 arrays on pseudomolecule of 5BS. Localization of telomere, centromere, and the recombination suppression region are indicated. Forward and reverse orientation of pSc119.2 is indicated by the blue and red colors, respectively. The pSc119.2 was mapped and plotted in MUMmer 3.0 (Kurtz et al., 2004).

Supplementary Figure S3 | Distribution of the E12 and E13 chromatin states within 5BS_RS and their signatures (Li et al., 2019).

Supplementary Table S2 | Markers flanking regions of suppression. The markers flanking 5BS_RS and 5BL_RS are in red.

REFERENCES

Avni, R., Nave, M., Eilam, T., Sela, H., Alekperov, C., Peleg, Z., et al. (2014). Ultra-dense genetic map of durum wheat × wild emmer wheat developed using the 90K iSelect SNP genotyping assay. *Mol. Breed.* 34, 1549–1562. doi: 10.1007/s11032-014-0176-2

Badaeva, E. D., Dedkova, O. S., Gay, G., Pukhalskyi, V. A., Zelenin, A. V., Bernard, S., et al. (2007). Chromosomal rearrangements in wheat: their types and distribution. *Genome* 50, 907–926. doi: 10.1139/g07-072

Badaeva, E. D., Keilwagen, J., Knüpffer, H., Waßermann, L., Dedkova, O. S., Mitrofanova, O. P., et al. (2015). Chromosomal passports provide new insights into diffusion of emmer wheat. *PLoS One* 10:e0128556. doi: 10.1371/journal.pone.0128556

Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6:11. doi: 10.1186/s13100-015-0041-9

Bariah, I., Keidar-Friedman, D., and Kashkush, K. (2020a). Where the wild things are: transposable elements as drivers of structural and functional variations in the wheat genome. *Front. Plant Sci.* 11:585515. doi: 10.3389/fpls.2020.585515

Bariah, I., Keidar-Friedman, D., and Kashkush, K. (2020b). Identification and characterization of largescale genomic rearrangements during wheat evolution. *PLoS One* 15:e0231323. doi: 10.1371/journal.pone.0231323

Bashir, T., Chandra Mishra, R., Hasan, M. M., Mohanta, T. K., and Bae, H. (2018). Effect of hybridization on somatic mutations and genomic rearrangements in plants. *Int. J. Mol. Sci.* 19:3758. doi: 10.3390/ijms19123758

Ben-David, S., Yaakov, B., and Kashkush, K. (2013). Genome-wide analysis of short interspersed nuclear elements SINES revealed high sequence conservation, gene association and retrotranspositional activity in wheat. *Plant J.* 76, 201–210. doi: 10.1111/tpj.12285

Bennetzen, J. L., Schrick, K., Springer, P. S., Brown, W. E., and SanMiguel, P. (1994). Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. *Genome* 37, 565–576. doi: 10.1139/g94-081

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573

Bhullar, R., Nagarajan, R., Bennypaul, H., Sidhu, G. K., Sidhu, G., Rustgi, S., et al. (2014). Silencing of a metaphase I-specific gene results in a phenotype similar to that of the pairing homoeologous 1 (*Ph1*) gene mutations. *Proc. Natl Acad. Sci.* 111, 14187–14192. doi: 10.1073/pnas.1416241111

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC* 10:421. doi: 10.1186/1471-2105-10-421

Charles, M., Belcram, H., Just, J., Huneau, C., Viollet, A., Couloux, A., et al. (2008). Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. *Genetics* 180, 1071–1086. doi: 10.1534/genetics.108.092304

Charlesworth, D. (2016). The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evol. Appl.* 9, 74–90. doi: 10.1111/eva.12291

Charlesworth, B., Langley, C. H., and Stephan, W. (1986). The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* 112, 947–962. doi: 10.1093/genetics/112.4.947

Charlesworth, B., Snegowski, P., and Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371, 215–220. doi: 10.1038/371215a0

Contento, A., Heslop-Harrison, J. S., and Schwarzacher, T. (2005). Diversity of a major repetitive DNA sequence in diploid and polyploid Triticeae. *Cytogenet. Genome Res.* 109, 34–42. doi: 10.1159/000082379

Dedkova, O. S., Badaeva, E. D., Amosova, A. V., Ruanet, V. V., Mitrofanova, O. P., and Pukhal'skiĭ, V. A. (2009). Diversity and the origin of the European population of Triticum dicoccum (Schrank) Schuebl. as revealed by chromosome analysis. *Russ. J. Genet.* 45, 1082–1091. doi: 10.1134/S1022795409090099

Devos, K. M., Dubcovsky, J., Dvořák, J., Chinoy, C. N., and Gale, M. D. (1995). Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. *Theor. Appl. Genet.* 91, 282–288. doi: 10.1007/BF00220890

Dorer, D. R., and Henikoff, S. (1994). Expansions of transgene repeats cause heterochromatin formation and gene silencing in *drosophila*. *Cell* 77, 993–1002. doi: 10.1016/0092-8674(94)90439-1

El Baidouri, M., Murat, F., Veyssiere, M., Molinier, M., Flores, R., Burlot, L., et al. (2017). Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). *New Phytol.* 213, 1477–1486. doi: 10.1111/nph.14113

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC* 9:18. doi: 10.1186/1471-2105-9-18

Fan, C., Hao, M., Jia, Z., Neri, C., Chen, X., Chen, W., et al. (2021). Some characteristics of crossing over in induced recombination between chromosomes of wheat and rye. *Plant J.* 105, 1665–1676. doi: 10.1111/tpj.15140

Faris, J. D., Haen, K. M., and Gill, B. S. (2000). Saturation mapping of a gene-rich recombination hot spot region in wheat. *Genetics* 154, 823–835. doi: 10.1093/genetics/154.2.823

Feldman, M., and Levy, A. A. (2012). Genome evolution due to allopolyploidization in wheat. *Genetics* 192, 763–774. doi: 10.1534/genetics.112.146316

Fitzgerald, D. J., Dryden, G. L., Bronson, E. C., Williams, J. S., and Anderson, J. N. (1994). Conserved patterns of bending in satellite and nucleosome positioning DNA. *J. Biol. Chem.* 269, 21303–21314. doi: 10.1016/S0021-9258(17)31963-4

Fu, S., Yang, M., Fei, Y., Tan, F., Ren, Z., Yan, B., et al. (2013). Alterations and abnormal mitosis of wheat chromosomes induced by wheat-rye monosomic addition lines. *PLoS One* 8:e70483. doi: 10.1371/journal.pone.0070483

Gadaleta, A., Giancaspro, A., Nigro, D., Giove, S. L., Incerti, O., Simeone, R., et al. (2014). A new genetic and deletion map of wheat chromosome 5A to detect candidate genes for quantitative traits. *Mol. Breed.* 34, 1599–1611. doi: 10.1007/s11032-014-0185-1

Goncharov, N. P. (2011). Genus Triticum L. taxonomy: the present and the future. *Plant Syst. Evol.* 295, 1–11. doi: 10.1007/s00606-011-0480-9

Gornicki, P., Zhu, H., Wang, J., Challa, G. S., Zhang, Z., Gill, B. S., et al. (2014). The chloroplast view of the evolution of polyploid wheat. *New Phytol.* 204, 704–714. doi: 10.1111/nph.12931

Griffiths, S., Sharp, R., Foote, T. N., Bertin, I., Wanous, M., Reader, S., et al. (2006). Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature* 439, 749–752. doi: 10.1038/nature04434

Guo, C., du, J., Wang, L., Yang, S., Mauricio, R., Tian, D., et al. (2016). Insertions/deletions-associated nucleotide polymorphism in Arabidopsis thaliana. *Front. Plant Sci.* 7:1792. doi: 10.3389/fpls.2016.01792

Hammarlund, M., Davis, M. W., Nguyen, H., Dayton, D., and Jorgensen, E. M. (2005). Heterozygous insertions alter crossover distribution but allow crossover interference in Caenorhabditis elegans. *Genetics* 171, 1047–1056. doi: 10.1534/genetics.105.044834

Holtz, Y., David, J., and Ranwez, V. (2017). The genetic map comparator: a user-friendly application to display and compare genetic maps. *Bioinformatics* 33, btw816–btw1388. doi: 10.1093/bioinformatics/btw816

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. *Science* 293, 1074–1080. doi: 10.1126/science.1063127

Jiang, J., and Gill, B. S. (1994). Different species-specific chromosome translocation in *Triticum timopheevii* and T. turgidum support diphyletic origin of polyploid wheats. *Chromosom. Res.* 2, 59–64. doi: 10.1007/BF01539455

Jorgensen, C., Luo, M. C., Ramasamy, R., Dawson, M., Gill, B. S., Korol, A. B., et al. (2017). A high-density genetic map of wild emmer wheat from the Karaca dağ region provides new evidence on the structure and evolution of wheat chromosomes. *Front. Plant Sci.* 8:1798. doi: 10.3389/fpls.2017.01798

Kagawa, N., Nagaki, K., and Tsujimoto, H. (2002). Tetrad-fish analysis reveals recombination suppression by interstitial heterochromatin sequences in rye (secale cereale). *Mol. Gen.* 267, 10–15. doi: 10.1007/s00438-001-0634-5

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kent, T. V., Uzunović, J., and Wright, S. I. (2017). Coevolution between transposable elements and recombination. *Phil. Trans. Roy. Soc. Lond B Biol. Sci.* 372:20160458. doi: 10.1098/rstb.2016.0458

Koo, H. S., and Crothers, D. M. (1988). Calibration of DNA curvature and a unified description of sequence-directed bending. *Proc. Natl. Acad. Sci. U. S. A.* 85, 1763–1767. doi: 10.1073/pnas.85.6.1763

Koo, H. S., Wu, H. M., and Crothers, D. M. (1986). DNA bending at adenine-thymine tracts. *Nature* 320, 501–506. doi: 10.1038/320501a0

Krumsiek, J., Arnold, R., and Rattei, T. (2007). Gepard: A rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23, 1026–1028. doi: 10.1093/bioinformatics/btm039

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12

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

Li, Z., Wang, M., Lin, K., Xie, Y., Guo, J., Ye, L., et al. (2019). The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. *Genome Biol.* 20:139. doi: 10.1186/s13059-019-1746-8

Li, L. F., Zhang, Z. B., Wang, Z. H., Li, N., Sha, Y., Wang, X. F., et al. (2022). Genome sequences of five Sitopsis species of Aegilops and the origin of polyploid wheat B subgenome. *Mol. Plant* 15, 488–503. doi: 10.1016/j.molp.2021.12.019

Lippman, Z., Gendrel, A. V., Black, M., Vaughn, M. W., Dedhia, N., Richard McCombie, W., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430, 471–476. doi: 10.1038/nature02651

Lu, H.-J., Fellers, J. P., Friesen, T. L., Meinhardt, S. W., and Faris, J. D. (2006). Genomic analysis and marker development for the Tsn1 locus in wheat using bin-mapped ESTs and flanking BAC contigs. *Theor. Appl. Genet.* 112, 1132–1142. doi: 10.1007/s00122-006-0215-4

Ma, J., and Bennetzen, J. L. (2004). Rapid recent growth and divergence of rice nuclear genomes. *Proc. Natl. Acad. Sci.* 101, 12404–12410. doi: 10.1073/pnas.0403715101

Mank, J. E. (2012). Small but mighty: the evolutionary dynamics of W and Y sex chromosomes. *Chromosome Res.* 20, 21–33. doi: 10.1007/s10577-011-9251-2

Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., The International Wheat Genome Sequencing Consortiumet al. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345:1250092. doi: 10.1126/science.1250092

Marini, J. C., Levene, S. D., Crothers, D. M., and Englund, P. T. (1982). Bent helical structure in kinetoplast DNA. *Proc. Natl. Acad. Sci. U. S. A.* 79, 7664–7668. doi: 10.1073/pnas.79.24.7664

Martienssen, R. A., and Colot, V. (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* 293, 1070–1074. doi: 10.1126/science.293.5532.1070

Mirzaghaderi, G., and Mason, A. S. (2017). Revisiting pivotal-differential genome evolution in wheat. *Trends Plant Sci.* 22, 674–684. doi: 10.1016/J.TPLANTS.2017.06.003

Mrazek, J. (2010). Comparative analysis of sequence periodicity among prokaryotic genomes points to differences in nucleoid structure and a relationship to gene expression. *J. Bacteriol.* 192, 3763–3772. doi: 10.1128/JB.00149-10

Mukai, Y., Nakahara, Y., and Yamamoto, M. (1993). Simultaneous discrimination of the three genomes in hexaploid wheat by multicolor fluorescence in situ hybridization using total genomic and highly repeated DNA probes. *Genome* 36, 489–494. doi: 10.1139/g93-067

Muterko, A. F. (2017). Quaternion modeling of the helical path for analysis of the shape of the DNA molecule. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov. *J. Genet. Breed.* 21, 878–886. doi: 10.18699/VJ17.308

Naranjo, T., Roca, A., Goicoecha, P. G., and Giraldez, R. (1987). Arm homoeology of wheat and rye chromosomes. *Genome* 29, 873–882. doi: 10.1139/g87-149

Naumann, K., Fischer, A., Hofmann, I., Krauss, V., Phalke, S., Irmler, K., et al. (2005). Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in Arabidopsis. *EMBO J.* 24, 1418–1429. doi: 10.1038/sj.emboj.7600604

Nelson, J. C., Sorrells, M. E., Van Deynze, A. E., Lu, Y. H., Atkinson, M., Bernard, M., et al. (1995). Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5, and 7. *Genetics* 141, 721–731. doi: 10.1093/genetics/141.2.721

Nicetto, D., and Zaret, K. S. (2019). Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. *Curr. Opin. Genet. Dev.* 55, 1–10. doi: 10.1016/j.gde.2019.04.013

Poleshko, A., Smith, C. L., Nguyen, S. C., Sivaramakrishnan, P., Wong, K. G., Murray, J. I., et al. (2019). H3K9me2 orchestrates inheritance of spatial positioning of peripheral heterochromatin through mitosis. *elife* 8:e49278. doi: 10.7554/eLife.49278

Pont, C., Leroy, T., Seidel, M., Tondelli, A., Duchemin, W., Armisen, D., et al. (2019). Tracing the ancestry of modern bread wheats. *Nat. Genet.* 51, 905–911. doi: 10.1038/s41588-019-0393-z

Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., and Honig, B. (2009). The role of DNA shape in protein-DNA recognition. *Nature* 461, 1248–1253. doi: 10.1038/nature08473

Salina, E. A., Lim, K. Y., Badaeva, E. D., Shcherban, A. B., Adonina, I. G., Amosova, A. V., et al. (2006). Phylogenetic reconstruction of Aegilops section Sitopsis and the evolution of tandem repeats in the diploids and derived wheat polyploids. *Genome* 49, 1023–1035. doi: 10.1139/g06-050

Salina, E. A., Nesterov, M. A., Frenkel, Z., Kiseleva, A. A., Timonova, E. M., Magni, F., et al. (2018). Features of the organization of bread wheat chromosome 5BS based on physical mapping. *BMC Genomics* 19:80. doi: 10.1186/s12864-018-4470-y

Salina, E. A., Numerova, O. M., Ozkan, H., and Feldman, M. (2004). Alterations in subtelomeric tandem repeats during early stages of allopolyploidy in wheat. *Genome* 47, 860–867. doi: 10.1139/g04-044

Salina, E. A., Sergeeva, E. M., Adonina, I. G., Shcherban, A. B., Belcram, H., Huneau, C., et al. (2011). The impact of Ty3-gypsy group LTR retrotransposons Fatima on B-genome specificity of polyploid wheats. *BMC Plant Biol.* 11:99. doi: 10.1186/1471-2229-11-99

SanMiguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., et al. (1996). Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274, 765–768. doi: 10.1126/science.274.5288.765

Segal, E., and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* 19, 65–71. doi: 10.1016/j.sbi.2009.01.00

Shcherban, A. B., Sergeeva, E. M., Badaeva, E. D., and Salina, E. A. (2008). Analysis of 5S rDNA changes in synthetic allopolyploids Triticum×Aegilops. *Mol. Biol.* 42, 536–542. doi: 10.1134/S0026893308040080

Smit, A. F. A., Hubley, R., and Green, P. (2008). RepeatModeler Open-1.0. Seattle, USA: Institute for Systems Biology. Available at: http://www.repeatmasker.org

Smit, A. F. A., Hubley, R., and Green, P. (2013). RepeatMasker Open-4.0. Available at: http://www.repeatmasker.org

Sosnowski, O., Charcosset, A., and Joets, J. (2012). Biomercator V3: An upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms. *Bioinformatics* 28, 2082–2083. doi: 10.1093/bioinformatics/bts313

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* 20, 267–273. doi: 10.1038/nsmb.2506

Termolino, P., Cremona, G., Consiglio, M. F., and Conicella, C. (2016). Insights into epigenetic landscape of recombination-free regions. *Chromosoma* 125, 301–308. doi: 10.1007/s00412-016-0574-9

Tikhonov, A. P., SanMiguel, P. J., Nakajima, Y., Gorenstein, N. M., Bennetzen, J. L., and Avramova, Z. (1999). Colinearity and its exceptions in orthologous adh regions of maize and sorghum. *Proc. Natl. Acad. Sci. U. S. A.* 96, 7409–7414. doi: 10.1073/pnas.96.13.7409

Timonova, E. M., Dobrovol'skaya, O. B., Sergeeva, E. M., Bildanova, L. L., Sourdille, P., Feuillet, C., et al. (2013). A comparative genetic and cytogenetic mapping of wheat chromosome 5B using introgression lines. *Russ. J. Genet.* 49, 1200–1206. doi: 10.1134/S1022795413120132

Tolstorukov, M. Y., Virnik, K. M., Adhya, S., and Zhurkin, V. B. (2005). A-tract clusters may facilitate DNA packaging in bacterial nucleoid. *Nucleic Acids Res.* 33, 3907–3918. doi: 10.1093/nar/gki699

Vermaak, D., Ahmad, K., and Henikoff, S. (2003). Maintenance of chromatin states: an open-and- shut case. *Curr. Opin. Cell Biol.* 15, 266–274. doi: 10.1016/s0955-0674(03)00043-7

Vershinin, A. V., and Heslop-Harrison, J. S. (1998). Comparative analysis of the nucleosomal structure of rye, wheat and their relatives. *Plant Mol. Biol.* 36, 149–161. doi: 10.1023/a:1005912822671

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton, J., et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature* 588, 277–283. doi: 10.1038/s41586-020-2961-x

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796. doi: 10.1111/pbi.12183

Wang, Z. Z., Xie, J. Z., Guo, L., Zhang, D. Y., Li, G. Q., Fang, T. L., et al. (2018). Molecular mapping of YrTZ2, a stripe rust resistance gene in wild emmer accession TZ-2 and its comparative analyses with Aegilops tauschii. *J. Integr. Agric.* 17, 1267–1275. doi: 10.1016/S2095-3119(17)61846-X

Westphal, T., and Reuter, G. (2002). Recombinogenic effects of suppressors of position-effect variegation in drosophila. *Genetics* 160, 609–621. doi: 10.1093/genetics/160.2.609

Wingen, L. U., West, C., Leverington-Waite, M., Collier, S., Orford, S., Goram, R., et al. (2017). Wheat landrace genome diversity. *Genetics* 205, 1657–1676. doi: 10.1534/genetics

Yelagandula, R., Stroud, H., Holec, S., Zhou, K., Feng, S., Zhong, X., et al. (2014). The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. *Cell* 158, 98–109. doi: 10.1016/j.cell.2014.06.006

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* 7, 203–214. doi: 10.1089/10665270050081478

Zhu, T., Wang, L., Rimbert, H., Rodriguez, J. C., Deal, K. R., de Oliveira, R., et al. (2021). Optical maps refine the bread wheat Triticum aestivum cv. *Chinese Spr. Gen. Assembly. Plant J* 107, 303–314. doi: 10.1111/tpj.15289

Zimin, A. V., Puiu, D., Hall, R., Kingan, S., Clavijo, B. J., and Salzberg, S. L. (2017). The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *Gigascience* 6, 1–7. doi: 10.1093/gigascience/gix097

Ziolkowski, P. A., Berchowitz, L. E., Lambing, C., Yelina, N. E., Zhao, X., Kelly, K. A., et al. (2015). Juxtaposition of heterozygous and homozygous regions causes reciprocal crossover remodelling via interference during Arabidopsis meiosis. *Elife* 4:e03708. doi: 10.7554/eLife.03708

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 19 May 2022
doi: 10.3389/fpls.2022.899206

Parental Genome Imbalance Causes Hybrid Seed Lethality as Well as Ovary Abscission in Interspecific and Interploidy Crosses in *Nicotiana*

*Hai He[1†], Kumi Sadahisa[1], Shuji Yokoi[1,2,3,4] and Takahiro Tezuka[1,2,3]**

[1]*Laboratory of Plant Breeding and Propagation, Graduate School of Life and Environmental Sciences, Osaka Prefecture University, Sakai, Japan,* [2]*Laboratory of Breeding and Genetics, Graduate School of Agriculture, Osaka Metropolitan University, Sakai, Japan,* [3]*Education and Research Field, School of Agriculture, Osaka Metropolitan University, Sakai, Japan,* [4]*Bioeconomy Research Institute, Research Center for the 21st Century, Osaka Metropolitan University, Sakai, Japan*

OPEN ACCESS

Edited by:
Dayun Tao,
Yunnan Academy of Agricultural Sciences, China

Reviewed by:
Thomas Städler,
ETH Zürich, Switzerland
Jugou Liao,
Yunnan University, China

***Correspondence:**
Takahiro Tezuka
tezuka@omu.ac.jp

†Present address:
Hai He,
FAFU-UCR Joint Center for Horticultural Biology and Metabolomics, Haixia Institute of Science and Technology, Fujian Agriculture and Forestry University, Fuzhou, China

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

Received: *18 March 2022*
Accepted: *29 April 2022*
Published: *19 May 2022*

Citation:
He H, Sadahisa K, Yokoi S and Tezuka T (2022) Parental Genome Imbalance Causes Hybrid Seed Lethality as Well as Ovary Abscission in Interspecific and Interploidy Crosses in Nicotiana. Front. Plant Sci. 13:899206. doi: 10.3389/fpls.2022.899206

Enhanced ovary abscission after pollination and hybrid seed lethality result in post-zygotic reproductive isolation in plant interspecific crosses. However, the connection between these barriers remains unclear. Here, we report that an imbalance in parental genomes or endosperm balance number (EBN) causes hybrid seed lethality and ovary abscission in both interspecific and intraspecific-interploidy crosses in the genus *Nicotiana*. Auxin treatment suppressed ovary abscission, but not hybrid seed lethality, in an interspecific cross between *Nicotiana suaveolens* and *N. tabacum*, suggesting that ovary abscission-related genes are located downstream of those involved in hybrid seed lethality. We performed interploidy crosses among *N. suaveolens* tetraploids, octoploids, and neopolyploids and revealed hybrid seed lethality and ovary abscission in interploid crosses. Furthermore, a higher maternal EBN than paternal EBN caused these barriers, as previously observed in *N. suaveolens* × *N. tabacum* crosses. Altogether, these results suggest that maternal excess of EBN causes hybrid seed lethality, which in turn leads to ovary abscission through the same mechanism in both interspecific and interploidy crosses.

Keywords: auxin treatment, endosperm balance number, hybrid seed lethality, interploidy cross, interspecific cross, ovary abscission, tobacco

INTRODUCTION

Speciation is highly dependent on the evolution of reproductive isolation by the accumulation of barriers to gene exchange (Kulmuni et al., 2020). Reproductive isolation involves various pre-mating and post-mating prezygotic and postzygotic isolating barriers in animals and plants (Coyne and Orr, 2004; Rieseberg and Willis, 2007; Rieseberg and Blackman, 2010). However, post-mating isolation barriers in plants can be obstacles to breeding by interspecific crossings. Conspecific pollen precedence and gametic incompatibility are examples of post-mating prezygotic isolating barriers (Rieseberg and Willis, 2007). In turn, postzygotic isolation barriers include enlarged ovary (immature fruit) abscission (Gupta et al., 1996; He et al., 2019), seed abortion or hybrid seed lethality (Coughlan et al., 2020; Dziasek et al., 2021), hybrid weakness (Shiragaki et al., 2019, 2020), hybrid lethality (Kawaguchi et al., 2021; Si et al., 2021; Tezuka et al., 2021;

Mino et al., 2022), and hybrid sterility (Li et al., 2020) in seedlings of the F_1 generation, as well as hybrid breakdown recognized in generations after the F_1 (Matsubara et al., 2007; Zhang et al., 2021). These isolating barriers may be observed independently or combined, even in a single cross-combination.

Hybrid seed lethality has been observed in both intraspecific interploidy and interspecific crosses. While the endosperm is generally a triploid tissue and an important component of seeds that supports embryo development and germination, abnormal endosperm development is considered to cause hybrid seed lethality (Oneal et al., 2016; Dziasek et al., 2021; Köhler et al., 2021; Städler et al., 2021), which is often explained by effective ploidy or endosperm balance number (EBN). Further, normal development of the endosperm requires a 2:1 maternal:paternal EBN ratio, a deviation from which results in endosperm developmental failure (Johnston et al., 1980; Coughlan et al., 2020; Städler et al., 2021). In plants with nuclear-type endosperm, inviable hybrid seeds show a disturbance in the timing of endosperm cellularization, which is an important developmental transition for embryo development in this type of endosperm (Ishikawa et al., 2011; Sekine et al., 2013; Lafon-Placette et al., 2017; İltaş et al., 2021). Although few examples of hybrid seed lethality have been reported in plant species with *ab initio* cellular-type endosperm, where karyokinesis and cytokinesis occur simultaneously without syncytial phase (Vijayaraghavan and Prabhakar, 1984; Floyd and Friedman, 2000), inviable hybrid seeds show impaired endosperm proliferation (Oneal et al., 2016; Roth et al., 2018).

Post-pollination ovary or pod abscission has been reported in several interspecific crosses of the genera *Cicer* (Mallikarjuna, 1999), *Lupinus* (Gupta et al., 1996), *Phaseolus* (Mok et al., 1978), and *Vigna* (Barone et al., 1992), in Fabaceae. In these cases, ovarian abscission was accompanied by hybrid seed lethality. Rabakoarihanta et al. (1979) presumed that a severe delay in embryo and endosperm divisions might be the major cause of ovary abscission in *Phaseolus* interspecific crosses, although no direct evidence has been provided.

Recently, we reported ovary abscission occurring in interspecific crosses between the octoploid *Nicotiana suaveolens* accession PI 555565 (♀) and allotetraploid *N. tabacum* (♂), in Solanaceae (He et al., 2019). In this cross, type II seed lethality was also observed with characteristics of precocious developmental transition of the endosperm, subsequent narrowing of the endosperm region as if pressed by surrounding cells, and developmental arrest of embryos in the early globular stage (He et al., 2020). In contrast, the other two accessions of *N. suaveolens* yielded different results after crossing with *N. tabacum*: thus, tetraploid *N. suaveolens* PI 555568 produced normal seeds, whereas octoploid *N. suaveolens* PI 555561 produced abnormal seeds, showing type I seed lethality characterized by precocious developmental transition and subsequent developmental arrest of the endosperm, and abnormal hypertrophy of the embryo during the globular state (the main differences between type I and II seed lethality were the size of the endosperm and the embryo at the globular state). Further, although ovary abscission was not observed in either cross, successive increases in maternal ploidy using ploidy-manipulated lines of PI 555568 and PI 555561 resulted in successive type I and type II seed lethality, and the latter was accompanied by ovary abscission. Therefore, it was suggested that a high maternal-genome excess cross might cause severe seed developmental defects and ovarian abscission (He et al., 2020). However, because this study was based on interspecific crosses, it is unclear whether the cause of abnormal seed development and ovary abscission was only the difference in parental ploidy levels. Furthermore, the cause and result relationship between hybrid seed lethality and ovary abscission also remains unclear.

In this study, we treated peduncles with auxin after pollination to suppress ovary abscission and observed seed development in the interspecific cross between *N. suaveolens* PI 555565 and *N. tabacum* to determine whether ovary abscission is caused by hybrid seed lethality or vice versa, or completely independent isolating barriers occur, because auxin is known to usually inhibit organ abscission (Nakano and Ito, 2013). Additionally, intraspecific-interploidy crosses were conducted using *N. suaveolens* accessions with or without ploidy manipulation to obtain further insight into hybrid seed lethality and ovary abscission in *Nicotiana* interspecific crosses. Our findings demonstrated that maternal genome excess over the paternal genome causes hybrid seed lethality, thereby leading to ovary abscission in both interspecific and interploid crosses.

MATERIALS AND METHODS

Plant Materials

Three *N. suaveolens* accessions, PI 555561 (2n = 8x = 64), PI 555565 (2n = 8x = 64), and PI 555568 (2n = 4x = 32; He et al., 2019), provided by the United States *Nicotiana* Germplasm Collection (Lewis and Nicholson, 2007) and *N. tabacum* 'Red Russian' (2n = 4x = 48) provided by the Leaf Tobacco Research Center, Japan Tobacco Inc., Oyama, Japan, were used. We also used chromosome-doubled plants (neopolyploids) of PI 555561 (2n = 16x = 128) and PI 555568 (2n = 8x = 64) developed in a previous study (He et al., 2020), and the PI 555565 (2n = 16x) plants from this study. As previously described (He et al., 2020), chromosome doubling of PI 555565 (8x) was induced using colchicine and was confirmed by flow cytometry using external standardization (Hendrix and Stewart, 2005; Dolezel et al., 2007), where PI 555565 (8x) was used as the standard, against which each sample was analyzed. Next, nuclei were isolated from leaves using Otto buffer (Otto, 1990), stained with 4′,6-diamidino-2-phenylindole (DAPI), and analyzed using a flow cytometer CyFlow Space (Sysmex Partec, Görlitz, Germany; purchased from CytoTechs, Tsuchiura, Japan) and WinMDI 2.9 software.[1] All plants used for crossing experiments were grown under fluorescent lamps (FL40S·BRN; Toshiba Lighting and Technology Corp., Yokosuka, Japan; approximately 70 μmol m^{-2} s^{-1}) in a cultivation room under a 16:8 h light/dark photoperiod regime, at 25°C.

[1]http://www.cyto.purdue.edu/flowcyt/software/Winmdi.htm

Crossing Experiments

Conventional crossing was performed as follows: flowers of plants used as maternal parents were emasculated 1 day before anthesis and pollinated with the pollen of the paternal plants. For auxin treatments in crosses between PI 555685 (8x) and *N. tabacum*, lanolin paste containing 0, 10, 100, or 1,000 μM of indole-3-acetic acid (IAA) or 1-naphthaleneacetic acid (NAA) was applied to the peduncles of PI 555565 (8x) at 7 days after pollination (DAP), because the precocious developmental transition of the endosperm was observed at least before 6 DAP and enlarged ovaries of PI 555685 (8x) dropped at 12–17 DAP (He et al., 2019; He et al., 2020); auxin treatments were not conducted before 7 DAP, to avoid possible phytotoxicity expressed as browning of peduncles and ovaries.

One hundred thirty seed weights were determined for interspecific crosses with auxin treatments and intraspecific-interploidy crosses, respectively. An analytical balance (AB54; Mettler Toledo, Greifensee, Switzerland) was used, with three capsules (three replicates) for each cross and seed weights were expressed as single seed weights. For seeds obtained after performing the crosses with auxin treatment, the surface area of seeds was calculated based on photographs of seeds using ImageJ software (Schneider et al., 2012). Seed germinability was evaluated using *in vitro* sowing. Briefly, seeds were soaked in 0.5% gibberellic acid (GA_3) solution for 30 min and sterilized with 5% sodium hypochlorite for 15 min. Sterilized seeds were sown in Petri dishes (90 mm diameter, 17 mm deep) containing 25 ml of half-strength MS medium (Murashige and Skoog, 1962) supplemented with 1% sucrose, solidified with 0.8% agar (pH 5.8), and then cultured at 28°C for 30 days under continuous illumination (approximately 150 μmol m^{-2} s^{-1}).

Histological Observation

Histological analyses of seeds after pollination were conducted as follows: collected samples were fixed in formalin–acetic acid–alcohol (FAA), after which air in the tissue was extracted using a vacuum pump prior to storing the samples at room temperature until further use. After fixing, samples were dehydrated in an ethanol and t-butyl alcohol series (ethanol:t-butyl alcohol:water = 4:1:5, 5:2:3, 10:7:3, 9:11:0, 1:3:0, and 0:10:0). The t-butyl alcohol was gradually replaced with paraffin at 63°C, over 1 week, in an open bottle to evaporate traces of t-butyl alcohol and then they were embedded in paraffin. Embedded samples were then cut into 10–12-μm-thick sections using a microtome (PR-50; Yamato Kohki Industrial, Asaka, Japan). The tissue slices were placed on glass slides with distilled water (DW) and dried overnight at 50°C on a warming plate. The slides were deparaffinized twice in xylene for 30 min (twice) and hydrated in a graded ethanol series (100%, 95%, 85%, 70%, and 50% in DW). All sections were first treated with 3% iron alum and then stained with 1% fast green (90% ethanol) for 1 min at room temperature. The sections were observed under a microscope (BX50; Olympus) under conventional bright-field illumination. The area of the endosperm was calculated based on photographs of sections using ImageJ software.

Statistical Analysis

Data were analyzed using the SPSS Statistics software (version 22; IBM, Armonk, NY, United States). Seed weight and surface area were compared using Tukey's multiple comparison test.

RESULTS

Suppression of Ovary Abscission by Auxin Treatment

To investigate whether ovary abscission was caused by hybrid seed lethality or vice versa in the cross PI 555565 (8x) × *N. tabacum*, peduncles were treated with lanolin paste containing IAA or NAA after pollination. No difference was observed between self-crosses of PI 555565 (8x) treated with or without lanolin paste alone, indicating that lanolin itself did not affect the crossing results (**Figure 1**; **Table 1**). In auxin untreated peduncles of the cross PI 555565 (8x) × *N. tabacum*, ovary abscission occurred as expected (**Table 1**). All IAA and NAA concentrations suppressed ovary abscission and seeds were obtained (**Figure 2**; **Table 1**), although they appeared to be empty (**Figure 1A**); the seeds from the cross PI 555565 (8x) × *N. tabacum* were significantly lighter (33–39 μg) and generally showed a smaller surface area (0.33–0.42 mm^2) than the seeds of self-pollinated plants of PI 555565 (8x; 107–110 μg and 0.54–0.57 mm^2 respectively), and never germinated (**Figure 1B**; **Table 1**). Furthermore, characteristics of type II seed lethality were observed in the seeds from the cross PI 555565 (8x) × *N. tabacum*; the endosperm region narrowed as if pressed by surrounding cells, and embryos did not develop (**Figure 3**).

Seed Development Observed in Crosses Between 8x and 4x of *Nicotiana suaveolens*

Previously, we reported that self-crosses of PI 555568 (4x), PI 555561 (8x), and PI 555565 (8x) yielded capsules at high rates (94.1–95.2%), while PI 555568 (8x) yielded capsules at only 4.2% (He et al., 2020; **Table 2**). When the three octoploid lines, PI 555561 (8x), PI 555565 (8x), and PI 555568 (8x) were crossed with PI 555568 (4x), abscission of the enlarged ovary was not observed, and 100%, 91.7%, and 59.3% of the flowers produced capsules with seeds in the crosses PI 555561 (8x) × PI 555568 (4x), PI 555565 (8x) × PI 555568 (4x), and PI 555568 (8x) × PI 555568 (4x), respectively (**Figure 4A**; **Table 2**).

Seeds obtained from the three 8x × 4x crosses were significantly lighter than those of the respective maternal parents (138.9 μg in the interploid cross using PI 555561 vs. 217.8 μg in PI 555561, 93.3 μg in the interploid cross using PI 555565 vs. 183.3 μg in PI 555565, and 92.2 μg in the interploid cross using PI 555568 vs. 171 μg in PI 555568), although seed weights of the 8x × 4x crosses were the same as or higher than those of the paternal parent PI 555568 (4x; 94 μg; **Figure 4B**). Additionally, the germination rates of the seeds from the three 8x × 4x crosses

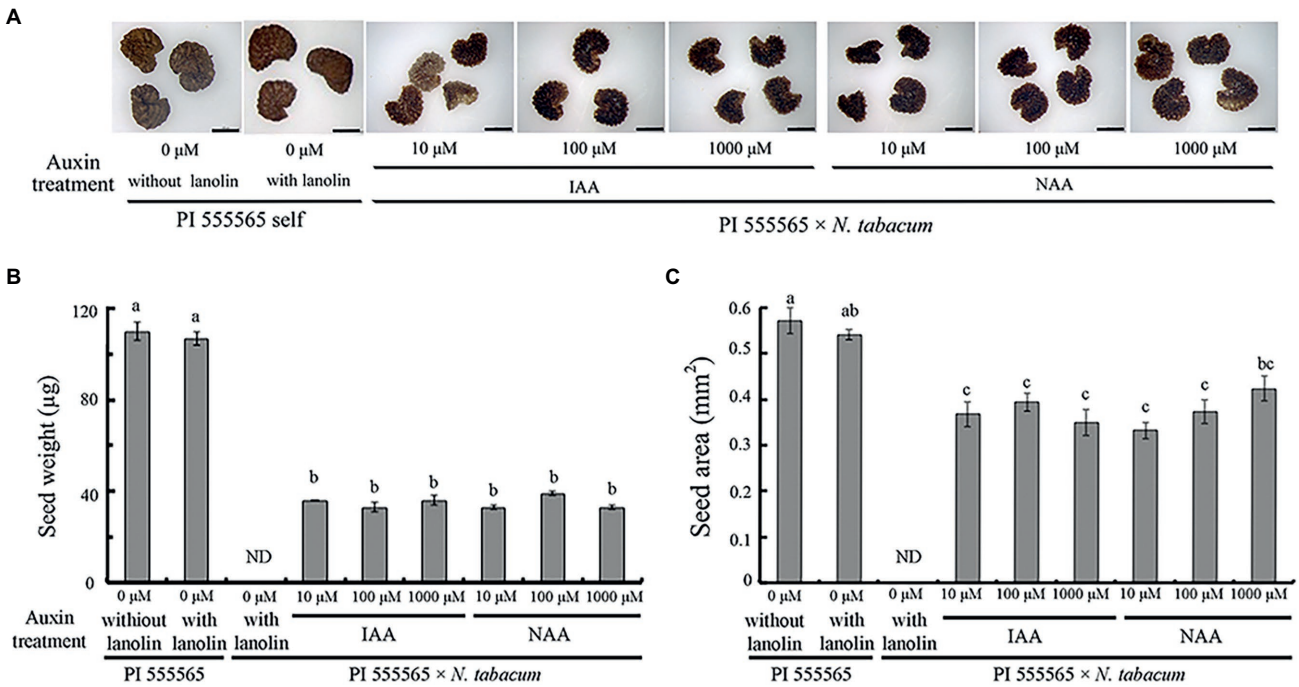


FIGURE 1 | Aborted seeds after auxin treatment in the cross *Nicotiana suaveolens* PI 555565 (8x) × *Nicotiana tabacum*. **(A)** Appearance of normal seeds produced by the self-pollination of PI 555565 and aborted seeds from the interspecific cross. Scale bars = 500 μm. **(B)** Single seed weight. **(C)** Surface area of seeds. Bars represent standard error. Different lowercase letters indicate significant differences at the 5% level as per Tukey's test. ND, no data.

TABLE 1 | Effect of auxin treatment on ovary abscission and seed viability in the cross between *Nicotiana suaveolens* PI 555565 (8x) and *Nicotiana tabacum*.

Cross combination	Auxin treatment	No. of flowers pollinated	No. of ovaries		No. of capsules		No. of seeds sown	No. of plants obtained
			Enlarged	Dropped after enlarging	Produced	Showing phytotoxicity		
PI 555565 (8x)	0 μM (without lanolin paste)[a]	21	20	0	20	0	59	48 (81.4%[b])
	0 μM (with lanolin paste)	20	19	0	19	0	104	85 (81.7%)
PI 555565 (8x) × *Nicotiana tabacum*	0 μM (with lanolin paste)	11	11	11	0	0	–	–
	10 μM IAA	10	10	0	10	0	101	0 (0%)
	100 μM IAA	20	20	0	20	0	82	0 (0%)
	1,000 μM IAA	15	15	0	15	0	120	0 (0%)
	10 μM NAA	13	13	0	13	0	235	0 (0%)
	100 μM NAA	10	10	0	10	0	131	0 (0%)
	1,000 μM NAA	22	22	0	19	3	130	0 (0%)

[a]*The data were previously reported by* He et al. (2020).
[b]*Percentage of plants obtained (seed germination percentage).*

were lower than those of the respective parental lines; 67.2% in the interploid cross using PI 555561 vs. 93.8% in PI 555561, 17.6% in the interploid cross using PI 555565 vs. 81.4% in PI 555565, 36.0% in the interploid cross using PI 555568 vs. 100% in PI 555568 (8x), and the percentages of all interploid crosses vs. 87.5% in PI 555568 (4x; **Figure 4C**; **Table 2**). Further, histological observations showed that an earlier developmental transition of the endosperm occurred at least before 6 DAP in the cross PI 555561 (8x) × PI 555568 (4x), compared with both parents, as well as in the cross PI 555568 (8x) × PI 555568 (4x), compared with the self-cross of PI 555568 (4x) reported by He et al. (2020) (**Figure 5A**). Moreover, the endosperm region at 6 DAP in the PI 555565 (8x) × PI 555568 (4x) was narrower than that of both parents; 0.23 mm^2 in the cross PI 555565 (8x) × PI 555568 (4x), whereas 0.45 mm^2 in PI 555565 (8x) and 0.56 mm^2 in PI 555568 (4x; **Figure 5B**). Nevertheless, endosperm and embryo development appeared normal in the three 8x × 4x crosses, i.e., successive stages of embryogenesis, globular, heart-shaped, and torpedo-shaped embryos were observed, as in the cases of self-crosses of the parental lines PI 555568 (4x), PI 555561 (8x), and PI 555565 (8x; **Figure 5**; **Table 2**).

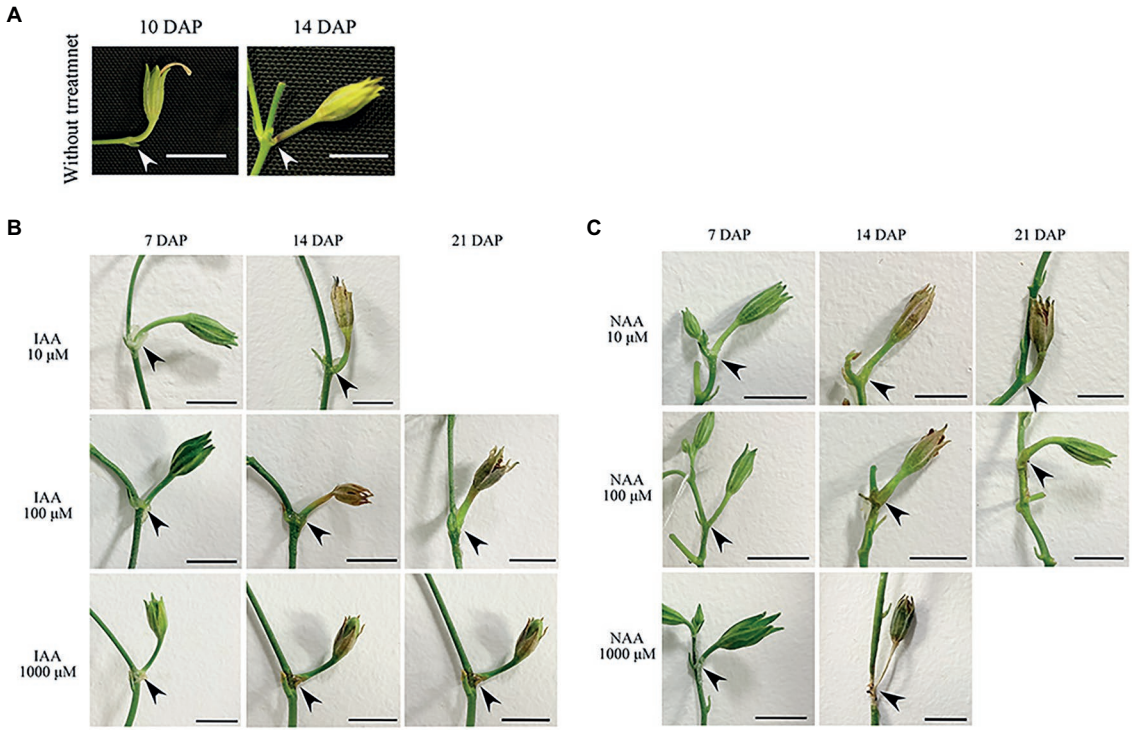


FIGURE 2 | Suppression of ovary abscission by auxin treatment in the cross *Nicotiana suaveolens* PI 555565 (8x) × *Nicotiana tabacum*. **(A)** Ovaries treated without lanolin paste. Enlarged ovaries dropped at 12–17 DAP. **(B)** Ovaries treated with IAA. **(C)** Ovaries treated with NAA. Arrowheads indicate the abscission zone at the intersection between the peduncle and branch. Scale bars = 1 cm. DAP, days after pollination.

Hybrid Seed Lethality and Ovary Abscission in Crosses Between 16x and 4x *Nicotiana suaveolens* Parents

Neopolyploids (16x) were developed from PI 555565 (8x; **Supplementary Figure 1**). Using the PI 555565 (16x) and PI 555561 (16x) developed in a previous study (He et al., 2020), crosses between 16x and 4x parents were conducted to investigate the effect of increasing maternal ploidy level on seed development and ovary abscission. In the case of self-pollination, 35.3% of the flowers produced capsules with seeds in PI 555561 (16x; He et al., 2020), whereas none of the flowers produced capsules in PI 555565 (16x; **Figure 4A**; **Table 2**). In turn, in the cross PI 555561 (16x) × PI 555568 (4x), 34.1% of flowers produced capsules with seeds. Seed weight was 72.2 μg and significantly lower than that in the self-cross of PI 555561 (16x; 212.2 μg) and crosses PI 555561 (8x) × PI 555568 (4x; **Figure 4B**). The germination rate of the seeds was 2.7%, which was also lower than that in self-crosses of both parents and the cross PI 555561 (8x) × PI 555568 (4x; **Figure 4C**; **Table 2**). As for the cross PI 555565 (16x) × PI 555568 (4x), ovaries were enlarged at a high rate (90.9%) after pollination; however, all enlarged ovaries dropped at 8–14 DAP (**Figure 4A**; **Table 2**).

Apparent abnormal seed development was observed histologically in two 16x × 4x crosses (**Figure 5**). In the cross PI 555561 (16x) × PI 555568 (4x), endosperm development was arrested at 8 DAP, and abnormal hypertrophy of embryos in the globular state as well as void spaces between the seed coat and endosperm were observed from 8 DAP (**Figure 5**), all of which are characteristics of type I seed lethality. Furthermore, in the cross PI 555565 (16x) × PI 555568 (4x), the endosperm region narrowed as if pressed by surrounding cells, and embryo development was not observed (**Figure 5**), which are characteristics of type II seed lethality.

DISCUSSION

Auxin Suppresses Ovary Abscission but Not Hybrid Seed Lethality

Auxins inhibit the abscission of several types of organs by rendering abscission zone cells insensitive to ethylene, which promotes abscission (van Overbeek, 1952; Nakano and Ito, 2013). Plant growth regulators have been used in Fabaceae to prevent pod abscission in interspecific crosses. Thus, for example, in interspecific crossings of *Phaseolus*, application of NAA alone or in combination with gibberellic acid to the pedicels of pollinated flowers stimulated pod growth and delayed embryo abortion (Al-Yasiri and Coyne, 1964). Similarly, the application of a mixture of gibberellic acid, NAA, and kinetin to the pedicels of developing buds after pollination delayed the abscission of pods and sometimes

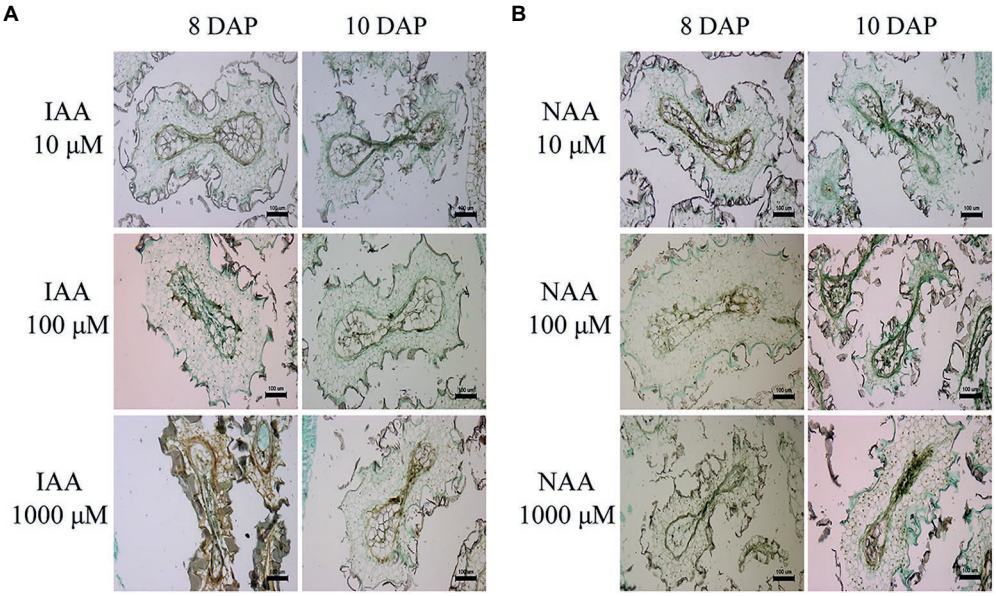


FIGURE 3 | Abnormal endosperm development and characteristics of type II hybrid seed lethality in the cross *Nicotiana suaveolens* PI 555565 (8x) × *Nicotiana tabacum* upon auxin treatment. **(A)** Seeds produced after IAA treatment. **(B)** Seeds produced after NAA treatment. Scale bars = 100 μm. DAP, days after pollination.

TABLE 2 | Interploidy crosses of *Nicotiana suaveolens*.

Cross combination	No. of flowers pollinated	No. of ovaries		No. of capsules produced	No. of seeds sown	No. of plants obtained	Seed development[c]
		Enlarged	Dropped after enlarging				
PI 555561 (8x) × PI 555568 (4x)	17	17	0	17	131	88 (67.2%[b])	Normal
PI 555565 (8x) × PI 555568 (4x)	36	33	0	33	142	25 (17.6%)	Normal
PI 555568 (8x) × PI 555568 (4x)	27	16	0	16	100	36 (36.0%)	Normal
PI 555561 (16x) × PI 555568 (4x)	44	15	0	15	112	3 (2.7%)	Type I seed abortion
PI 555565 (16x) × PI 555568 (4x)	11	10	10	0	–	–	Type II seed abortion
PI 555568 (4x)[a]	17	16	0	16	48	42 (87.5%)	Normal
PI 555561 (8x)[a]	19	18	0	18	48	45 (93.8%)	Normal
PI 555565 (8x)[a]	21	20	0	20	59	48 (81.4%)	Normal
PI 555568 (8x)[a]	48	2	0	2	45	45 (100%)	ND
PI 555561 (16x)[a]	34	12	0	12	50	48 (96.0%)	ND
PI 555565 (16x)	17	0	–	–	–	–	–

[a]*The data were previously reported by He et al. (2020).*
[b]*Percentage of plants obtained (seed germination percentage).*
[c]*Seed development was judged by histological observation.*

prevented the abortion of hybrid embryos in interspecific crosses of *Cicer* and *Vigna* (Gosal and Bajaj, 1983; Chen et al., 1990; Mallikarjuna, 1999). In this study, the application of IAA or NAA suppressed ovary abscission in interspecific crosses in *Nicotiana*, confirming the general function of auxin in preventing organ abscission, whether it occurs physiologically or due to an interspecific cross. Previously, we reported that ovarian abscission in interspecific crosses involves a mechanism similar to that of abscission in other organs (He et al., 2020).

Because both ovary abscission and hybrid seed lethality were observed in the cross PI 555565 (8x) × *N. tabacum*, the question arouse as to the cause and the result, or whether completely independent isolating barriers occurred. In a previous study, successive increases in maternal ploidy using ploidy-manipulated lines resulted in successive type I and type II hybrid seed lethality, and the latter was accompanied by ovary abscission. Therefore, we hypothesized that type II seed lethality might lead to ovary abscission (He et al., 2020). Now, this study provided more direct evidence for this hypothesis. Ovary abscission, but not hybrid seed lethality, was effectively suppressed by auxin treatment, strongly indicating that the latter caused ovary abscission in interspecific crosses.

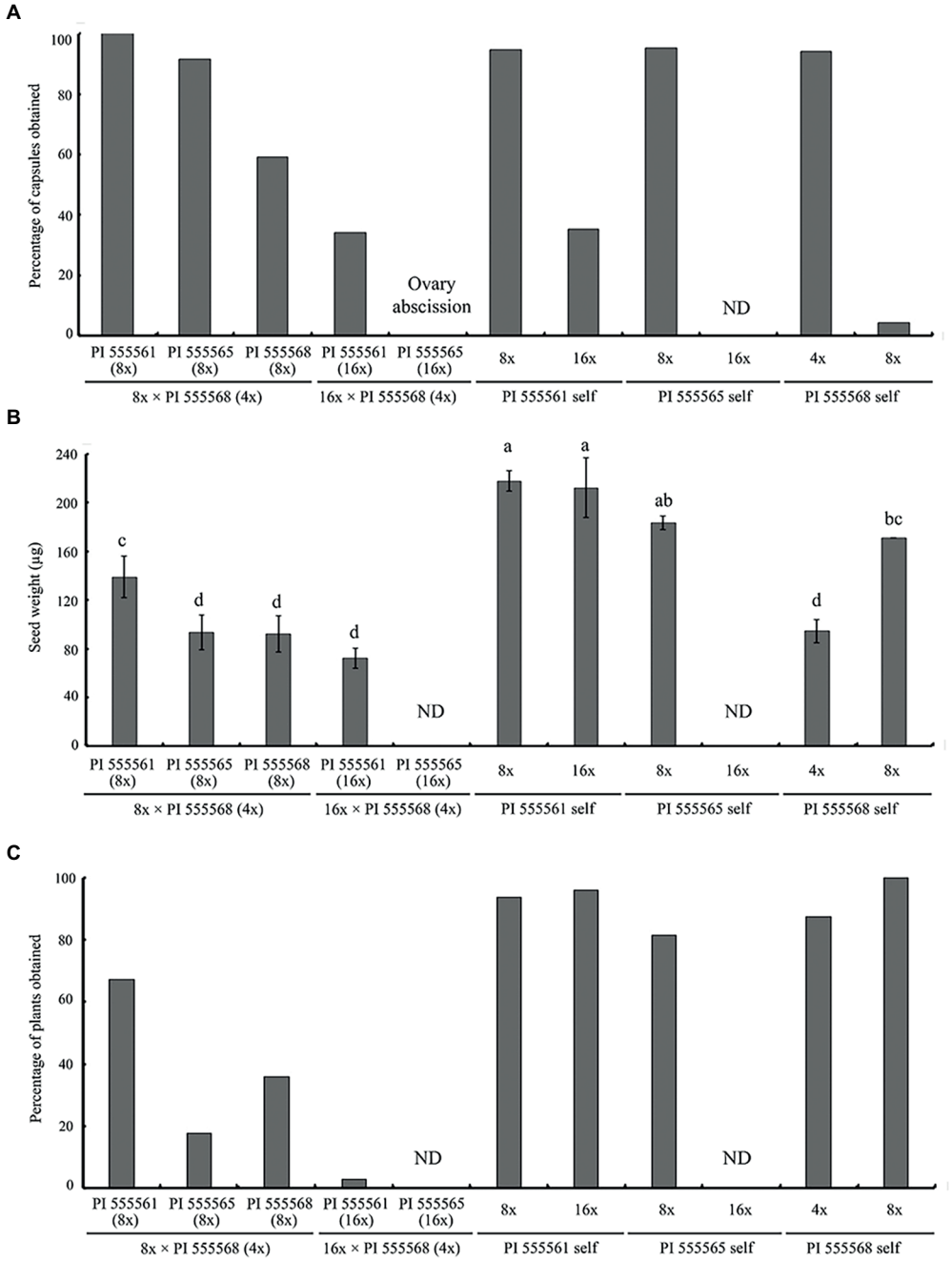


FIGURE 4 | Ovary abscission and seed abortion were observed in intraspecific-interploidy crosses of *Nicotiana suaveolens*. **(A)** Percentages of capsules produced after pollination. **(B)** Single seed weight. Bars represent standard error. Different lowercase letters indicate significant differences at the 5% level as per Tukey's test. **(C)** Percentages of plants obtained after seed sowing. ND, no data. The data for self-pollinations other than PI 555565 (16x) have been reported previous (He et al., 2020) and are included here for comparison.

Hybrid Seed Lethality Leading to Ovary Abscission Is Observed in Interploidy as Well as Interspecific Crosses

Ovary abscission has been reported in interspecific crosses of several genera in Fabaceae (Mok et al., 1978; Rabakoarihanta et al., 1979; Barone et al., 1992; Gupta et al., 1996; Mallikarjuna, 1999) and *Nicotiana* in Solanaceae (He et al., 2020). However, no studies had been conducted on ovary abscission by using intraspecific-interploidy crosses. Additionally, our previous study using *Nicotiana* suggested that a high maternal genome excess might cause type II hybrid seed lethality and ovary abscission (He et al., 2020). Nonetheless, whether these were caused solely

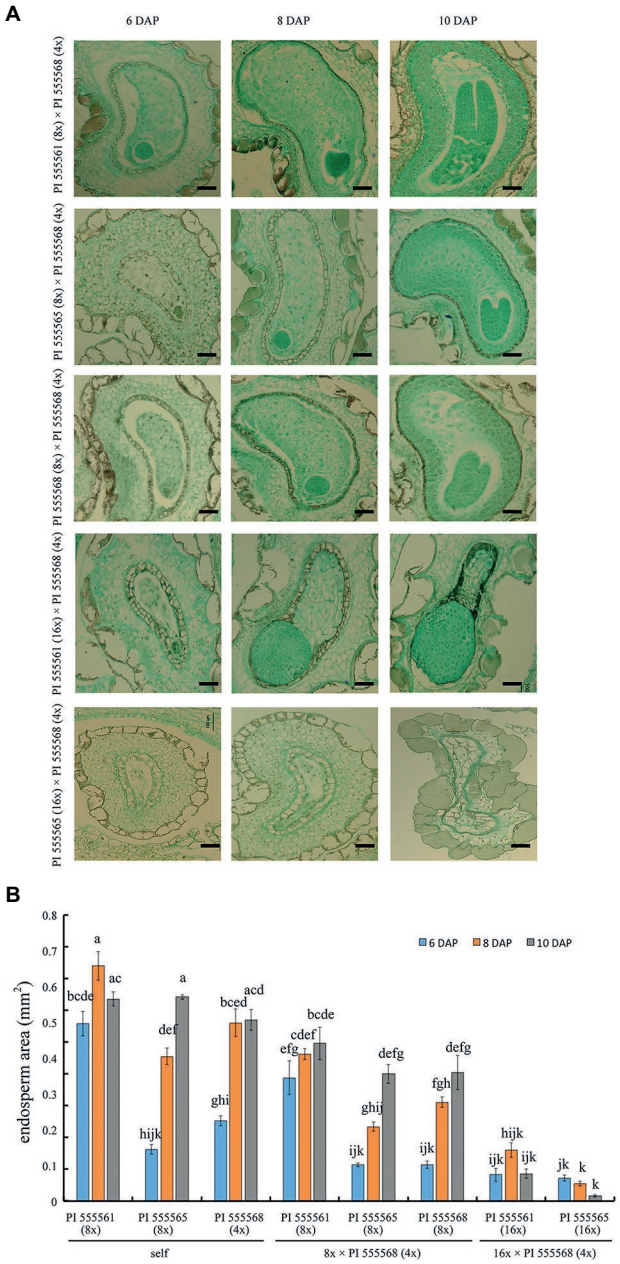


FIGURE 5 | Seed development in intraspecific-interploidy crosses of *Nicotiana suaveolens*. **(A)** Abnormal endosperm and embryo development. Normal endosperm and embryo development were observed in 8x × 4x crosses, while type I and type II hybrid seed lethality were observed in crosses between PI 555561 (16x) or PI 555565 (16x) and PI 555568 (4x), respectively. Scale bars = 200 μm. **(B)** The surface area of the endosperm. Bars represent standard error ($n = 5$ at 6 and 8 DAP, and $n = 3$ at 10 DAP). Different lowercase letters indicate significant differences at the 5% level as per Tukey's test.

due to differences in parental ploidy levels remained unclear. The results of this study on interploid crosses using neopolyploids clearly demonstrated that increases in the ploidy levels of PI 555561 and PI 555565 from 8x to 16x caused type I seed lethality and type II seed lethality, leading to ovary abscission upon crossing with PI 555568 (4x). Thus, the magnitude of the difference in ploidy level between the two parents determines the fate of seeds, namely, whether they will undergo normal development, type I seed lethality, or type II seed lethality, leading to ovary abscission, in interploidy and interspecific crosses of *Nicotiana*.

The histological differences between type I and II seed lethality were in the degree of endosperm and embryo development. For embryos, abnormal hypertrophy and no development at the globular state were observed in type I and II seed lethality, respectively; furthermore, no transition to the heart stage was observed in this case. As embryo development depends on endosperm development (Hehenberger et al., 2012; Lafon-Placette

and Köhler, 2014), differences may be caused by deviations from normal progression. Endosperm developmental failure in type I seed lethality occurs before the critical stage for the transition to a heart-shaped embryo. Moreover, whether ovary abscission occurs may also depend on the degree of seed development, as ovary abscission does not occur when the endosperm and/or embryo develops to the level of type I seed lethality, whereas it does occur at the level of type II seed lethality.

The Severity of Hybrid Seed Lethality in *Nicotiana* Is Explained by the EBN Hypothesis

Endosperm developmental failure responsible for hybrid seed lethality has been well studied in plant genera undergoing nuclear-type endosperm development, such as *Arabidopsis* and *Oryza*. In this developmental mode, a disturbance in the timing of endosperm cellularization is the primary cause observed histologically (Ishikawa et al., 2011; Sekine et al., 2013; Lafon-Placette et al., 2017; İltaş et al., 2021; Köhler et al., 2021). In contrast, endosperm development in *Nicotiana* is an *ab initio* cellular type (Sehgal and Gifford, 1979). The cellular-type endosperm is considered the ancestral type of endosperm development (Floyd and Friedman, 2000). The results of this, and previous studies (He et al., 2020), are consistent with other studies reporting that, although nuclear- and cellular-type endosperms exhibit different developmental abnormalities, endosperm disruption is likely to affect seed development at similar early stages, when embryo development depends on nutrient supply from the endosperm (Roth et al., 2018; Städler et al., 2021).

The EBN hypothesis has been proposed to conceptualize the function of the endosperm in interploidy and interspecific crosses (Johnston et al., 1980). EBN is not directly related to ploidy, and the genome of each plant shows a specific EBN, where a 2:1 maternal: paternal EBN ratio is required for normal endosperm development. According to the EBN hypothesis, the EBN of *N. suaveolens* accessions ranked PI 555565 (8x) > PI 555561 (8x) > PI 555568 (4x). Additionally, the EBN of *N. suaveolens* neopolyploids was as follows: PI 555565 (16x) was the largest, PI 555561 (16x) was the second-largest, and PI 555568 (8x) was similar to or somewhat larger than PI 555561 (8x). Meanwhile, *N. tabacum* would have the same EBN as PI 555568 (4x).

Interestingly, among *N. suaveolens* lineages (this species is endemic to Australia), PI 555565 exhibits a higher EBN than PI 555561, although both accessions are octoploids. Furthermore, although a few anomalies were observed, seed development in 8x × 4x crosses of *N. suaveolens* appeared histologically normal, suggesting flexibility in *Nicotiana* seed development to overcome dosage imbalances. Possibly, evolutionary changes such as sequence changes or duplication of genes might be related to these cases.

Johnston et al. (1980) reported that the EBN of a species may be determined by a few genes rather than the whole genome. To date, many imprinted genes that are expressed in a parent-of-origin-specific manner have been identified as related to hybrid seed lethality (Josefsson et al., 2006; Erilova et al., 2009; Wolff et al., 2015; Florez-Rueda et al., 2016; Garner et al., 2016; Lafon-Placette et al., 2018; Wang et al., 2018; Kinser et al., 2021). Furthermore, transposable elements (Josefsson et al., 2006; Borges et al., 2018) and small RNAs (Erdmann et al., 2017; Martinez et al., 2018; Wang et al., 2018; Satyaki and Gehring, 2019) are reportedly involved in hybrid seed lethality. As for plants undergoing cellular-type endosperm development, a candidate group of genes that may underlie EBN differences was reported in *Solanum* (Roth et al., 2019). However, despite these extensive studies, underlying molecular mechanisms of EBN differences are largely unknown. Therefore, it is an interesting challenge to identify the factors underlying EBN in *Nicotiana* species, which may cause hybrid seed lethality in a dose-dependent manner in interspecific and interploid crosses.

Parental conflict can arise in non-monogamous systems because the interests of maternal and paternal parents can be expected to differ with respect to the amount of maternal resource allocation to the offspring. According to the parental conflict hypothesis, the maternal parent is equally related to all of their progeny and thus should allocate equally, while the paternal parents are only related to their own progeny but not to the competing half-siblings, and thus should somehow direct the maternal parent to allocate differentially (Haig and Westoby, 1989, 1991; Coughlan et al., 2020; Köhler et al., 2021). Thus, co-evolutionary arms race can occur for resource-acquiring paternal alleles and resource-repressive maternal alleles (Coughlan et al., 2020), and reproductive isolation can be established when differences in endosperm or seed development, which are possibly fueled by differences in levels of parental conflict between diverging lineages, reach a critical level. Recent studies have suggested that endosperm-based hybridization barriers are rapidly evolving system of reproductive isolation (Lafon-Placette and Köhler, 2016; Städler et al., 2021). The weak inbreeder/strong outbreeder hypothesis posits that parental conflict is less intense in self-pollinating plants than in outcrossing plants, and thus changes in mating system can change levels of parental conflict (Brandvain and Haig, 2005). Extending this hypothesis, recent studies imply that several factors, such as demographic history, population subdivision, and persistent soil seed banks, as well as mating system changes, can modify effective population size, leading to divergence in EBN (Roth et al., 2019; Coughlan et al., 2020; Städler et al., 2021). Further studies will be needed to verify whether these factors are involved in hybrid seed lethality in *Nicotiana*.

EBN might be related to parental conflict, because opposite phenotypes were observed between seeds obtained from reciprocal crosses of plants with different EBN in *Capsella*, *Mimulus*, and *Solanum*; high EBN × low EBN crosses produce smaller seeds than those in the reciprocal crosses (Lafon-Placette et al., 2018; Roth et al., 2018, 2019; Coughlan et al., 2020). In this and previous studies (He et al., 2020), we conducted maternal excess crosses, but not the reciprocal crosses. For interspecific crosses, this was because fertilization did not occur by conventional cross-pollination, and test-tube pollination in combination with ovule culture was necessary to obtain seeds from crosses between *N. tabacum* and *N. suaveolens* (Tezuka and Marubashi, 2004). However, seeds might be produced by conventional cross-pollination in interploidy paternal excess crosses of *N. suaveolens*, and this should be analyzed in future studies.

CONCLUSION

Maternal excess causes hybrid seed lethality based on EBN in *Nicotiana* interspecific and interploid crosses. Ovary abscission occurs depending on the severity of hybrid seed lethality. The endosperm plays an important role in establishing reproductive isolation in angiosperms. *Nicotiana suaveolens* PI 555565 evolved a higher EBN than *N. suaveolens* PI 555561, although both accessions are octoploids. Determining the factors that cause the difference between the two accessions will help to elucidate endosperm-based postzygotic hybridization barriers.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

HH and TT conceived and designed the research and wrote the manuscript. HH and KS conducted the experiments. HH, KS, and TT analyzed the data. SY and TT supervised the study. HH prepared the figures. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partly supported by JSPS KAKENHI grant numbers JP17K15224 and JP20K05988 from the Japan Society for the Promotion of Science (to TT), and Sasakawa Scientific Research grant numbers 2018-5034 and 2019-5018 from the Japan Science Society (to HH).

ACKNOWLEDGMENTS

The authors thank the Leaf Tobacco Research Center, Japan Tobacco Inc., Oyama, Japan for providing seeds of *Nicotiana tabacum*, and the United States *Nicotiana* Germplasm Collection for providing seeds of *Nicotiana suaveolens*.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.899206/full#supplementary-material

REFERENCES

Al-Yasiri, S., and Coyne, D. (1964). Effect of growth regulators in delaying pod abscission and embryo abortion in the interspecific cross *Phaseolus vulgaris* × *P. acutifolius*. *Crop Sci.* 4, 433–435. doi: 10.2135/cropsci1964.0011183X000400040032x

Barone, A., Del Giudice, A., and Ng, N. Q. (1992). Barriers to interspecific hybridization between *Vigna unguiculata* and *Vigna vexillata*. *Sex. Plant Reprod.* 5, 195–200. doi: 10.1007/BF00189811

Borges, F., Parent, J. S., Van Ex, F., Wolff, P., Martínez, G., Köhler, C., et al. (2018). Transposon-derived small RNAs triggered by mi R845 mediate genome dosage response in *Arabidopsis*. *Nat. Genet.* 50, 186–192. doi: 10.1038/s41588-017-0032-5

Brandvain, Y., and Haig, D. (2005). Divergent mating systems and parental conflict as a barrier to hybridization in flowering plants. *Am. Nat.* 166, 330–338. doi: 10.1086/432036

Chen, H. K., Mok, M. C., and Mok, D. W. (1990). Somatic embryogenesis and shoot organogenesis from interspecific hybrid embryos of *Vigna glabrescens* and *V. radiata*. *Plant Cell Rep.* 9, 77–79. doi: 10.1007/bf00231553

Coughlan, J. M., Wilson Brown, M., and Willis, J. H. (2020). Patterns of hybrid seed inviability in the *Mimulus guttatus* sp. complex reveal a potential role of parental conflict in reproductive isolation. *Curr. Biol.* 30, 83–93.e5. doi: 10.1016/j.cub.2019.11.023

Coyne, J. A., and Orr, H. A. (2004). *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Dolezel, J., Greilhuber, J., and Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* 2, 2233–2244. doi: 10.1038/nprot.2007.310

Dziasek, K., Simon, L., Lafon-Placette, C., Laenen, B., Wärdig, C., Santos-González, J., et al. (2021). Hybrid seed incompatibility in *Capsella* is connected to chromatin condensation defects in the endosperm. *PLoS Genet.* 17:e1009370. doi: 10.1371/journal.pgen.1009370

Erdmann, R. M., Satyaki, P. R. V., Klosinska, M., and Gehring, M. (2017). A small RNA pathway mediates allelic dosage in endosperm. *Cell Rep.* 21, 3364–3372. doi: 10.1016/j.celrep.2017.11.078

Erilova, A., Brownfield, L., Exner, V., Rosa, M., Twell, D., Mittelsten Scheid, O., et al. (2009). Imprinting of the polycomb group gene *MEDEA* serves as a ploidy sensor in Arabidopsis. *PLoS Genet.* 5:e1000663. doi: 10.1371/journal.pgen.1000663

Florez-Rueda, A. M., Paris, M., Schmidt, A., Widmer, A., Grossniklaus, U., and Städler, T. (2016). Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. *Mol. Biol. Evol.* 33, 2935–2946. doi: 10.1093/molbev/msw175

Floyd, S. K., and Friedman, W. E. (2000). Evolution of endosperm developmental patterns among basal flowering plants. *Int. J. Plant Sci.* 161, S57–S81. doi: 10.1086/317579

Garner, A. G., Kenney, A. M., Fishman, L., and Sweigart, A. L. (2016). Genetic loci with parent-of-origin effects cause hybrid seed lethality in crosses between *Mimulus* species. *New Phytol.* 211, 319–331. doi: 10.1111/nph.13897

Gosal, S., and Bajaj, Y. (1983). Interspecific hybridization between *Vigna mungo* and *Vigna radiata* through embryo culture. *Euphytica* 32, 129–137. doi: 10.1007/BF00036873

Gupta, S., Buirchell, B., and Cowling, W. (1996). Interspecific reproductive barriers and genomic similarity among the rough-seeded *Lupinus* species. *Plant Breed.* 115, 123–127. doi: 10.1111/j.1439-0523.1996.tb00886.x

Haig, D., and Westoby, M. (1989). Parent-specific gene expression and the triploid endosperm. *Am. Nat.* 134, 147–155. doi: 10.1086/284971

Haig, D., and Westoby, M. (1991). Genomic imprinting in endosperm: its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. *Philos. Trans. R. Soc. Lond. B* 333, 1–13. doi: 10.1098/rstb.1991.0057

He, H., Iizuka, T., Maekawa, M., Sadahisa, K., Morikawa, T., Yanase, M., et al. (2019). *Nicotiana suaveolens* accessions with different ploidy levels exhibit different reproductive isolation mechanisms in interspecific crosses with *Nicotiana tabacum*. *J. Plant Res.* 132, 461–471. doi: 10.1007/s10265-019-01114-w

He, H., Yokoi, S., and Tezuka, T. (2020). A high maternal genome excess causes severe seed abortion leading to ovary abscission in *Nicotiana* interploidy-interspecific crosses. *Plant Direct* 4:e00257. doi: 10.1002/pld3.257

Hehenberger, E., Kradolfer, D., and Köhler, C. (2012). Endosperm cellularization defines an important developmental transition for embryo development. *Development* 139, 2031–2039. doi: 10.1242/dev.077057

Hendrix, B., and Stewart, J. M. (2005). Estimation of the nuclear DNA content of *Gossypium* species. *Ann. Bot.* 95, 789–797. doi: 10.1093/aob/mci078

İltaş, Ö., Svitok, M., Cornille, A., Schmickl, R., and Lafon Placette, C. (2021). Early evolution of reproductive isolation: a case of weak inbreeder/strong outbreeder leads to an intraspecific hybridization barrier in *Arabidopsis lyrata*. *Evolution* 75, 1466–1476. doi: 10.1111/evo.14240

Ishikawa, R., Ohnishi, T., Kinoshita, Y., Eiguchi, M., Kurata, N., and Kinoshita, T. (2011). Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. *Plant J.* 65, 798–806. doi: 10.1111/j.1365-313X.2010.04466.x

Johnston, S. A., Den Nijs, T. P., Peloquin, S. J., and Hanneman, R. E. Jr. (1980). The significance of genic balance to endosperm development in interspecific crosses. *Theor. Appl. Genet.* 57, 5–9. doi: 10.1007/bf00276002

Josefsson, C., Dilkes, B., and Comai, L. (2006). Parent-dependent loss of gene silencing during interspecies hybridization. *Curr. Biol.* 16, 1322–1328. doi: 10.1016/j.cub.2006.05.045

Kawaguchi, K., Ohya, Y., Maekawa, M., Iizuka, T., Hasegawa, A., Shiragaki, K., et al. (2021). Two *Nicotiana occidentalis* accessions enable gene identification for type II hybrid lethality by the cross to *N. sylvestris*. *Sci. Rep.* 11:17093. doi: 10.1038/s41598-021-96482-6

Kinser, T. J., Smith, R. D., Lawrence, A. H., Cooley, A. M., Vallejo-Marín, M., Conradi Smith, G. D., et al. (2021). Endosperm-based incompatibilities in hybrid monkeyflowers. *Plant Cell* 33, 2235–2257. doi: 10.1093/plcell/koab117

Köhler, C., Dziasek, K., and Del Toro-De León, G. (2021). Postzygotic reproductive isolation established in the endosperm: mechanisms, drivers and relevance. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 376:20200118. doi: 10.1098/rstb.2020.0118

Kulmuni, J., Butlin, R. K., Lucek, K., Savolainen, V., and Westram, A. M. (2020). Towards the completion of speciation: the evolution of reproductive isolation beyond the first barriers. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 375:20190528. doi: 10.1098/rstb.2019.0528

Lafon-Placette, C., Hatorangan, M. R., Steige, K. A., Cornille, A., Lascoux, M., Slotte, T., et al. (2018). Paternally expressed imprinted genes associate with hybridization barriers in *Capsella*. *Nat. Plants* 4, 352–357. doi: 10.1038/s41477-018-0161-6

Lafon-Placette, C., Johannessen, I. M., Hornslien, K. S., Ali, M. F., Bjerkan, K. N., Bramsiepe, J., et al. (2017). Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe. *Proc. Natl. Acad. Sci. U. S. A.* 114, E1027–E1035. doi: 10.1073/pnas.1615123114

Lafon-Placette, C., and Köhler, C. (2014). Embryo and endosperm, partners in seed development. *Curr. Opin. Plant Biol.* 17, 64–69. doi: 10.1016/j.pbi.2013.11.008

Lafon-Placette, C., and Köhler, C. (2016). Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. *Mol. Ecol.* 25, 2620–2629. doi: 10.1111/mec.13552

Lewis, R. S., and Nicholson, J. S. (2007). Aspects of the evolution of *Nicotiana tabacum* L. and the status of the United States *Nicotiana* Germplasm collection. *Genet. Resour. Crop. Evol.* 54, 727–740. doi: 10.1007/s10722-006-0024-2

Li, J., Zhou, J., Zhang, Y., Yang, Y., Pu, Q., and Tao, D. (2020). New insights into the nature of interspecific hybrid sterility in rice. *Front. Plant Sci.* 11:555572. doi: 10.3389/fpls.2020.555572

Mallikarjuna, N. (1999). Ovule and embryo culture to obtain hybrids from interspecific incompatible pollinations in chickpea. *Euphytica* 110, 1–6. doi: 10.1023/A:1003621908663

Martinez, G., Wolff, P., Wang, Z., Moreno-Romero, J., Santos-González, J., Conze, L. L., et al. (2018). Paternal easi RNAs regulate parental genome dosage in *Arabidopsis*. *Nat. Genet.* 50, 193–198. doi: 10.1038/s41588-017-0033-4

Matsubara, K., Ando, T., Mizubayashi, T., Ito, S., and Yano, M. (2007). Identification and linkage mapping of complementary recessive genes causing hybrid breakdown in an intraspecific rice cross. *Theor. Appl. Genet.* 115, 179–186. doi: 10.1007/s00122-007-0553-x

Mino, M., Tezuka, T., and Shomura, S. (2022). The hybrid lethality of interspecific F_1 hybrids of *Nicotiana*: a clue to understanding hybrid inviability—a major obstacle to wide hybridization and introgression breeding of plants. *Mol. Breed.* 42:10. doi: 10.1007/s11032-022-01279-8

Mok, D. W., Mok, M. C., and Rabakoarihanta, A. (1978). Interspecific hybridization of *Phaseolus vulgaris* with *P. lunatus* and *P. acutifolius*. *Theor. Appl. Genet.* 52, 209–215. doi: 10.1007/bf00273891

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

Nakano, T., and Ito, Y. (2013). Molecular mechanisms controlling plant organ abscission. *Plant Biotech.* 30, 209–216. doi: 10.5511/plantbiotechnology.13.0318a

Oneal, E., Willis, J. H., and Franks, R. G. (2016). Disruption of endosperm development is a major cause of hybrid seed inviability between *Mimulus guttatus* and *Mimulus nudatus*. *New Phytol.* 210, 1107–1120. doi: 10.1111/nph.13842

Otto, F. (1990). DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. *Methods Cell Biol.* 33, 105–110. doi: 10.1016/S0091-679X(08)60516-6

Rabakoarihanta, A., Mok, D. W., and Mok, M. C. (1979). Fertilization and early embryo development in reciprocal interspecific crosses of *Phaseolus*. *Theor. Appl. Genet.* 54, 55–59. doi: 10.1007/bf00265469

Rieseberg, L. H., and Blackman, B. K. (2010). Speciation genes in plants. *Ann. Bot.* 106, 439–455. doi: 10.1093/aob/mcq126

Rieseberg, L. H., and Willis, J. H. (2007). Plant speciation. *Science* 317, 910–914. doi: 10.1126/science.1137729

Roth, M., Florez-Rueda, A. M., Griesser, S., Paris, M., and Städler, T. (2018). Incidence and developmental timing of endosperm failure in post-zygotic isolation between wild tomato lineages. *Ann. Bot.* 121, 107–118. doi: 10.1093/aob/mcx133

Roth, M., Florez-Rueda, A. M., and Städler, T. (2019). Differences in effective ploidy drive genome-wide endosperm expression polarization and seed failure in wild tomato hybrids. *Genetics* 212, 141–152. doi: 10.1534/genetics.119.302056

Satyaki, P. R. V., and Gehring, M. (2019). Paternally acting canonical RNA-directed DNA methylation pathway genes sensitize Arabidopsis endosperm to paternal genome dosage. *Plant Cell* 31, 1563–1578. doi: 10.1105/tpc.19.00047

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089

Sehgal, C., and Gifford, E. Jr. (1979). Developmental and histochemical studies of the ovules of *Nicotiana rustica* L. *Bot. Gaz.* 140, 180–188. doi: 10.1086/337074

Sekine, D., Ohnishi, T., Furuumi, H., Ono, A., Yamada, T., Kurata, N., et al. (2013). Dissection of two major components of the post-zygotic hybridization barrier in rice endosperm. *Plant J.* 76, 792–799. doi: 10.1111/tpj.12333

Shiragaki, K., Iizuka, T., Ichitani, K., Kuboyama, T., Morikawa, T., Oda, M., et al. (2019). *HWA1*-and *HWA2*-mediated hybrid weakness in rice involves cell death, reactive oxygen species accumulation, and disease resistance-related gene upregulation. *Plants* 8:450. doi: 10.3390/plants8110450

Shiragaki, K., Yokoi, S., and Tezuka, T. (2020). A hypersensitive response-like reaction is involved in hybrid weakness in F_1 plants of the cross *Capsicum annuum* × *Capsicum chinense*. *Breed. Sci.* 70, 430–437. doi: 10.1270/jsbbs.19137

Si, Y., Zheng, S., Niu, J., Tian, S., Gu, M., Lu, Q., et al. (2021). *Ne2*, a typical CC-NBS-LRR-type gene, is responsible for hybrid necrosis in wheat. *New Phytol.* 232, 279–289. doi: 10.1111/nph.17575

Städler, T., Florez-Rueda, A. M., and Roth, M. (2021). A revival of effective ploidy: the asymmetry of parental roles in endosperm-based hybridization barriers. *Curr. Opin. Plant Biol.* 61:102015. doi: 10.1016/j.pbi.2021.102015

Tezuka, T., Kitamura, N., Imagawa, S., Hasegawa, A., Shiragaki, K., He, H., et al. (2021). Genetic mapping of the *HLA1* locus causing hybrid lethality in *Nicotiana* interspecific hybrids. *Plants* 10:2062. doi: 10.3390/plants10102062

Tezuka, T., and Marubashi, W. (2004). Apoptotic cell death observed during the expression of hybrid lethality in interspecific hybrids between *Nicotiana tabacum* and *N. suaveolens*. *Breed. Sci.* 54, 59–66. doi: 10.1270/jsbbs.54.59

Van Overbeek, J. (1952). Agricultural application of growth regulators and their physiological basis. *Annu. Rev. Plant Physiol.* 3, 87–108. doi: 10.1146/annurev.pp.03.060152.000511

Vijayaraghavan, M. R., and Prabhakar, K. (1984). "The endosperm" in *Embryology of Angiosperms*. ed. B. M. Johri (Berlin, Heidelberg: Springer), 319–376.

Wang, L., Yuan, J., Ma, Y., Jiao, W., Ye, W., Yang, D. L., et al. (2018). Rice interploidy crosses disrupt epigenetic regulation, gene expression, and seed development. *Mol. Plant* 11, 300–314. doi: 10.1016/j.molp.2017.12.006

Wolff, P., Jiang, H., Wang, G., Santos-González, J., and Köhler, C. (2015). Paternally expressed imprinted genes establish postzygotic hybridization barriers in *Arabidopsis thaliana*. *eLife* 4:e10074. doi: 10.7554/eLife.10074

Zhang, M., Wei, H., Liu, J., Bian, Y., Ma, Q., Mao, G., et al. (2021). Non-functional *GoFLA19s* are responsible for the male sterility caused by hybrid breakdown in cotton (*Gossypium* spp.). *Plant J.* 107, 1198–1212. doi: 10.1111/tpj.15378

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 07 June 2022
doi: 10.3389/fpls.2022.921700

Reconstruction of the High Stigma Exsertion Rate Trait in Rice by Pyramiding Multiple QTLs

Quanya Tan[1,2†], Suhong Bu[1,2†], Guodong Chen[1,2], Zhenguang Yan[1,2], Zengyuan Chang[1,2], Haitao Zhu[1,2], Weifeng Yang[1,2], Penglin Zhan[1,2], Shaojun Lin[1,2], Liang Xiong[1,2], Songliang Chen[1,2], Guifu Liu[1,2], Zupei Liu[1,2], Shaokui Wang[1,2] and Guiquan Zhang[1,2*]*

[1] Guangdong Provincial Key Laboratory of Plant Molecular Breeding, State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University, Guangzhou, China, [2] Guangdong Laboratory for Lingnan Modern Agriculture, South China Agricultural University, Guangzhou, China

OPEN ACCESS

Edited by:
Dayun Tao,
Yunnan Academy of Agricultural Sciences, China

Reviewed by:
Ryo Ishikawa,
Kobe University, Japan
Jiawu Zhou,
Yunnan Academy of Agricultural Sciences, China

***Correspondence:**
Shaokui Wang
shaokuiwang@scau.edu.cn
Guiquan Zhang
gqzhang@scau.edu.cn

[†] These authors have contributed equally to this work

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

Received: 16 April 2022
Accepted: 05 May 2022
Published: 07 June 2022

Citation:
Tan Q, Bu S, Chen G, Yan Z, Chang Z, Zhu H, Yang W, Zhan P, Lin S, Xiong L, Chen S, Liu G, Liu Z, Wang S and Zhang G (2022) Reconstruction of the High Stigma Exsertion Rate Trait in Rice by Pyramiding Multiple QTLs. Front. Plant Sci. 13:921700. doi: 10.3389/fpls.2022.921700

Asian cultivated rice is a self-pollinating crop, which has already lost some traits of natural outcrossing in the process of domestication. However, male sterility lines (MSLs) need to have a strong outcrossing ability to produce hybrid seeds by outcrossing with restorer lines of male parents in hybrid rice seed production. Stigma exsertion rate (SER) is a trait related to outcrossing ability. Reconstruction of the high-SER trait is essential in the MSL breeding of rice. In previous studies, we detected eighteen quantitative trait loci (QTLs) for SER from *Oryza sativa*, *Oryza glaberrima*, and *Oryza glumaepatula* using single-segment substitution lines (SSSLs) in the genetic background of Huajingxian 74 (HJX74). In this study, eleven of the QTLs were used to develop pyramiding lines. A total of 29 pyramiding lines with 2–6 QTLs were developed from 10 SSSLs carrying QTLs for SER in the HJX74 genetic background. The results showed that the SER increased with increasing QTLs in the pyramiding lines. The SER in the lines with 5–6 QTLs was as high as wild rice with strong outcrossing ability. The epistasis of additive by additive interaction between QTLs in the pyramiding lines was less-than-additive or negative effect. One QTL, *qSER3a-sat*, showed minor-effect epistasis and increased higher SER than other QTLs in pyramiding lines. The detection of epistasis of QTLs on SER uncovered the genetic architecture of SER, which provides a basis for using these QTLs to improve SER levels in MSL breeding. The reconstruction of the high-SER trait will help to develop the MSLs with strong outcrossing ability in rice.

Keywords: outcrossing, stigma exsertion, QTL-pyramiding, epistasis, trait reconstruction, rice

INTRODUCTION

Every cultivated crop was once wild. In the process of domestication, crop species increased their productivity and narrowed their genetic base. The result is that many crops contain only a small fraction of the genetic variation of their wild relatives (Zamir, 2001). Asian cultivated rice (*Oryza sativa*) is a self-pollinating crop with <1% natural cross-pollination (Virmani and Athwal, 1973). During domestication, the homozygosity of varieties led to a steady decline in the outcrossing ability of cultivated rice. In contrast, African cultivated rice (*Oryza glaberrima*), which

has been domesticated for a shorter period of time, has more wild characteristics in floral traits and outcrossing ability (Marathi and Jena, 2015; Marathi et al., 2015). Wild *Oryza* species were more outcrossed due to their larger stigmas, longer styles, higher stigma exsertion, and longer spikelet opening period, with perennial species being more outcrossed than annual species (Oka and Morishima, 1967; Parmar et al., 1979; Marathi and Jena, 2015; Marathi et al., 2015). Therefore, the cultivated rice has lost some traits of natural outcrossing during domestication (Parmar et al., 1979).

Hybrid rice was successfully applied in China in the 1970s. Since then, hybrid rice has been widely planted in China and around the world (Yuan and Virmani, 1988). The hybrid rice exhibits strong heterosis, which plays an important role in improving rice yield (Cheng et al., 2007; Zhang et al., 2021; Zhang, 2022). Since male sterility lines (MSLs) are used to produce hybrid seeds by outcrossing with restorer lines of male parents, the outcrossing ability of MSLs is a key factor in improving hybrid seed yield. Therefore, the improvement of outcrossing ability is an important goal of MSL breeding. Stigma exsertion rate (SER) is an evaluating indicator of outcrossing ability and is controlled by quantitative trait loci (QTLs). Over the past decades, many QTLs controlling SER and related traits from different genetic resources have been located in the rice genome (Marathi and Jena, 2015; Zhou et al., 2017a; Liu et al., 2019; Tan et al., 2022). Some of the QTLs were detected in relative species of the cultivated rice, such as *O. glaberrima* (Tan et al., 2022), *Oryza rufipogon* (Xiong et al., 1999; Li et al., 2001; Uga et al., 2003a; Huang et al., 2012; Bakti and Tanaka, 2019; Zou et al., 2020), *Oryza barthii* and *Oryza meridionalis* (Zou et al., 2020), *Oryza longistaminata* (Li et al., 2010), and *Oryza glumaepatula* (Tan et al., 2020). The detection of QTLs for SER laid a foundation for the improvement of outcrossing ability of MSLs.

Understanding the genetic architecture of traits is a prerequisite for trait reconstruction. Genetic architecture of quantitative traits includes the numbers and genome locations of genes affecting a trait, the magnitude of their effects, and the relative contributions of additive, dominant, and epistatic gene effects (Kroymann and Mitchell-Olds, 2005; Holland, 2007). Epistasis is defined as interactions between genome-wide loci, impacting target traits either additively increasing each other's phenotypic effects or non-additively counteracting the effects of the major loci, *via* other loci that interact epistatically (Mackay, 2014; Misra et al., 2021). Epistasis complicates the genotype–phenotype relationship because it causes hidden quantitative genetic variation in natural populations and may lead to the small additive effects (Holland, 2007; Mackay, 2014). The role of epistasis in the genetic architecture of quantitative traits is controversial due to the fact that most genetic variation for quantitative traits is additive (Mackay, 2014). In QTL mapping approaches, first-order effects are commonly fitted before second- and higher-order (epistatic) effects, which makes it statistically difficult to detect epistatic effects (Gaertner et al., 2012; Doust et al., 2014). The studies in *Arabidopsis* and rice suggested that epistatic QTL effects were more important than additive QTL for fitness traits (Malmberg et al., 2005; Mei et al., 2005). By contrast, the studies designed to explicitly model epistatic interactions in maize revealed that epistasis was of less or only moderate importance for quantitative traits (Schon et al., 2004; Mihaljevic et al., 2005; Blanc et al., 2006; Holland, 2007).

Like near isogenic lines (NILs), single-segment substitution lines (SSSLs) carry only one chromosome substitution segment from donors in a recipient genetic background (Zhang et al., 2004; Keurentjes et al., 2007; Zhang, 2021). We have developed an SSSL library, which includes 2,360 SSSLs derived from 43 donors of 7 species of rice AA genome in the Huajingxian 74 (HJX74) genetic background (Zhang et al., 2004; Xi et al., 2006; Zhang, 2021). The HJX74-SSSLs were widely used to detect QTLs for traits of agronomic importance (Zhang et al., 2012; Zhu et al., 2014, 2018; Yang et al., 2016, 2021a,b; Zhou et al., 2017b; Pan et al., 2021; Zhan et al., 2021; Fu et al., 2022), to clone genes of functional importance, and to mine alleles of functional variants (Zeng et al., 2006; Teng et al., 2012; Wang et al., 2012; Fang et al., 2019; Sui et al., 2019; Gao et al., 2021; Zhan et al., 2022). Using the HJX74-SSSL library as a platform for rice design, a series of cytoplasmic male sterility (CMS), maintainer and restorer lines, and wide-compatible *indica* lines (WCILs) were developed (Dai et al., 2015, 2016; Luan et al., 2019; Guo et al., 2022). Recently, the SSSLs were used to detect QTLs controlling SER in rice. Eighteen QTLs for SER from *O. sativa*, *O. glaberrima*, and *O. glumaepatula* were detected in the HJX74-SSSLs (Tan et al., 2020, 2021, 2022). In this study, eleven of the QTLs were used to develop 2- to 6-QTL pyramiding lines in the HJX74 genetic background. The results showed that the SER increased with increasing QTLs in the pyramiding lines. The epistasis of additive by additive interaction between QTLs in 2- to 6-QTL lines was less-than-additive or negative effect. The pyramiding lines carrying 5–6 QTLs for SER showed as high SER as wild rice. The reconstruction of the high SER trait will help to develop MSLs with strong outcrossing ability in rice.

MATERIALS AND METHODS

Plant Materials

In total, eleven of the QTLs for SER were used to develop QTL-pyramiding lines. Among the 11 QTLs, four were from *O. sativa* (Tan et al., 2021), four were from *O. glaberrima* (Tan et al., 2022), and three were from *O. glumaepatula* (Tan et al., 2020). The QTLs were located in the substitution segments of SSSLs in the HJX74 genetic background. In total, twenty-one SSSLs carrying the 11 QTLs were used in this study, 10 of which were used as parents to develop pyramiding lines (**Supplementary Table 1**, **Supplementary Figure 1**).

Field Experiments

All plant materials were planted in the farm of South China Agricultural University, Guangzhou (23°07′N, 113°15′E). The materials were planted in 2015–2020, two cropping seasons per year. The first cropping season (FCS) was from late February to mid-July, and the second cropping season (SCS) was from late July to mid-November. Germinated seeds were sown in a seedling bed, and seedlings were transplanted to the paddy field in a single seedling. Field cultivation and controlling of diseases and insect pests were followed by conventional methods in the local area.

TABLE 1 | SER-QTL combinations in pyramiding lines.

		QTL										
Group	**Pyramiding line**	***qSER1a-gla***	***qSER1b-gla***	***qSER1b-glu***	***qSER2a-sat***	***qSER2b-sat***	***qSER3a-sat***	***qSER3b-sat***	***qSER3b-glu***	***qSER5-glu***	***qSER8b-gla***	***qSER12-gla***
2-QTL line	A35	−	−	−	+	+	−	−	−	−	−	−
	A88	−	−	−	−	−	+	+	−	−	−	−
	2QL-1	−	+	−	−	−	−	−	−	−	−	+
	2QL-2	−	+	−	−	−	−	−	−	−	+	−
	2QL-3	−	−	−	−	−	−	−	−	−	+	+
	2QL-4	−	−	−	−	−	+	−	−	−	+	−
	2QL-5	−	−	−	−	−	+	−	−	−	−	+
3-QTL line	3QL-1	+	+	−	−	−	−	−	−	−	−	+
	3QL-2	+	+	−	−	−	−	−	−	−	+	−
	3QL-3	−	+	−	−	−	−	−	−	−	+	+
	3QL-4	+	−	−	−	−	−	−	−	−	+	+
	3QL-5	−	+	−	−	−	+	+	−	−	−	−
	3QL-6	−	−	−	−	−	+	+	−	−	+	−
	3QL-7	−	−	−	−	−	+	−	+	−	+	−
	3QL-8	−	−	+	−	−	+	−	−	−	+	−
	3QL-9	−	−	−	−	−	+	−	−	+	+	−
	3QL-10	−	−	−	+	+	+	−	−	−	−	−
4-QTL line	4QL-1	−	+	−	+	+	−	−	−	−	−	+
	4QL-2	−	+	−	+	+	−	−	−	−	+	−
	4QL-3	−	−	−	+	+	−	−	−	−	+	+
	4QL-4	−	+	−	−	−	+	+	−	−	−	+
	4QL-5	−	−	−	−	−	+	+	−	−	+	+
	4QL-6	−	+	−	−	−	+	+	−	−	+	−
	4QL-7	+	−	−	−	−	+	+	−	−	−	+
5-QTL line	5QL-1	−	+	−	−	−	+	+	−	−	+	+
	5QL-2	−	−	−	+	+	+	+	−	−	−	+
	5QL-3	−	−	−	−	−	+	+	+	+	+	−
	5QL-4	−	+	−	−	−	+	+	−	+	+	−
	5QL-5	−	−	−	+	+	+	+	−	−	+	−
6-QTL line	6QL-1	−	−	−	+	+	+	+	−	+	+	−
	6QL-2	−	+	−	+	+	+	+	−	−	+	−

"+," Presence; "–," Absence; SER, stigma exsertion rate. A35 and A88 are the single-segment substitution lines (SSSLs) carrying two QTLs for SER in the substitution segments.

Genotyping of the QTLs for SER

All markers in the substitution segments carrying QTLs for SER were selected from the previous QTL mapping studies. The substitution segments and the QTLs for SER in SSSLs and pyramiding lines were detected following the previous studies (Tan et al., 2020, 2021, 2022). To distinguish different loci with the same name and to show the allelic origin of the QTLs, the first three letters of the donor species name were added to the suffix of the QTL names (**Supplementary Table 1**, **Supplementary Figure 1**).

Phenotyping and Statistical Analysis

The SER and agronomic traits were investigated following the previous studies (Tan et al., 2020, 2021, 2022). For the statistical analysis, data of percentages were converted to the arcsine square root. The student's *t*-test was used to compare the two sets of data. Dunnett's *t*-test was used for multiple group comparison with the control group. The least significance range (LSR) was used for the multiple range test among multiple groups (Duncan, 1955). SPSS statistics 23.0 and OriginPro 9.0 were used for the data analysis and figure making (https://www.originlab.com).

Estimation of Additive Effects and Epistatic Effects of QTLs

The additive effect of a QTL for SER was defined as the genotypic value of the homozygous genotype of a QTL in the HJX74 genetic background. Therefore, the additive effect of each locus was the difference in mean values of SER between HJX74 and

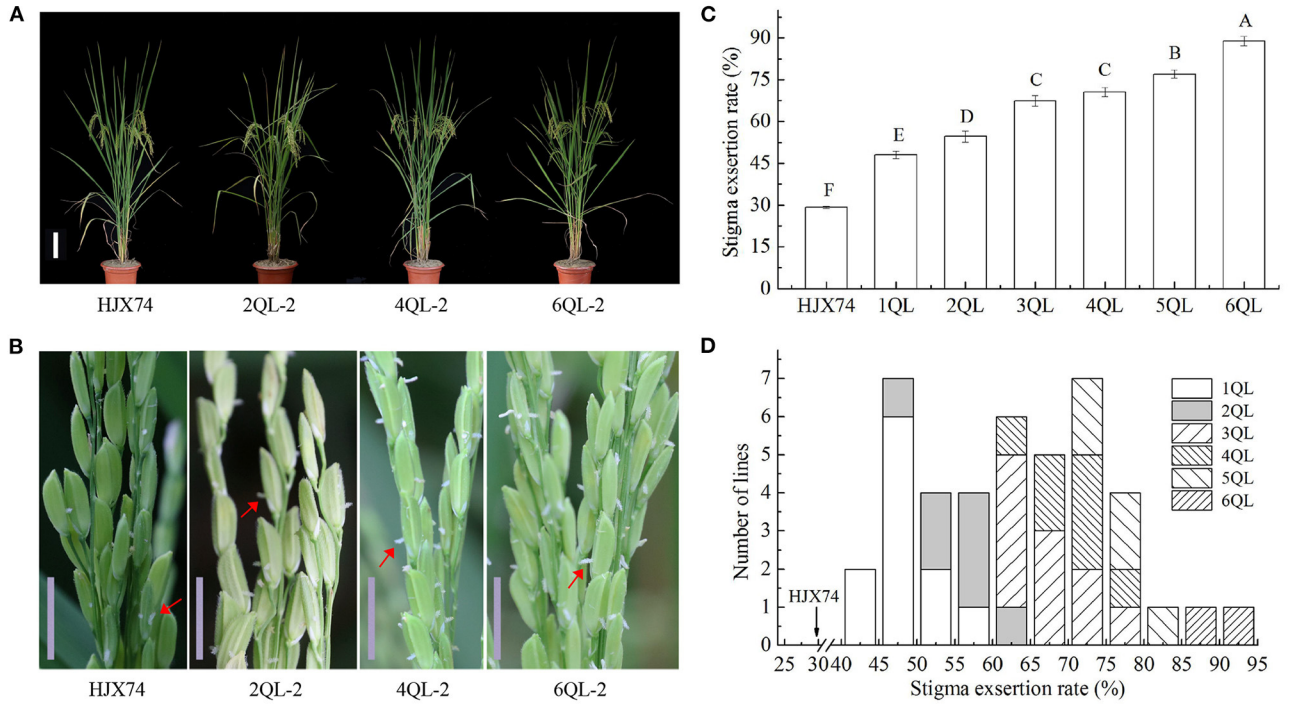


FIGURE 1 | Stigma exsertion rate (SER) of 1- to 6-QTL lines. **(A)** Plant types in the 2QL-2, 4QL-2, and 6QL-2 lines and HJX74. Scale bar, 15 cm. **(B)** The phenotypes of stigma exsertion in the 2QL-2, 4QL-2, and 6QL-2 lines and HJX74. Scale bar, 1 cm. The red arrows point to the exposed stigma. **(C)** SER of 1- to 6-QTL lines. HJX74 as a control. Data are shown as the mean ± S.E. of three cropping seasons. Capital letters indicate the significance of differences at the 1% levels. **(D)** Frequency distributions of SER in 1- to 6-QTL lines.

an SSSL carrying a target QTL, and the additive effects of two- or multiple-locus genotypes were the difference in mean values of SER between HJX74 and pyramiding lines carrying two or multiple QTLs in the HJX74 genetic background. Epistasis of additive by additive interactions among QTLs for SER in 2- to 6-QTL lines was detected using homozygous genotypes. Epistatic effects among QTLs were estimated by the formula, $i = (P_n - P_0) - \sum_{i=1}^{n}(a_i)$, where i is an epistasis among the pyramided QTLs, P_n is a phenotype of a pyramiding line harboring n of QTLs, P_0 is a phenotype of HJX74, a_i ($1 \leq i \leq n$) is an additive effect of a single QTL at the ith QTL. Epistatic effects among QTLs in pyramiding lines were tested in Student's t-test under null hypothesis (H_0) $i = 0$.

RESULTS

Genotypes in the Lines With 1- to 6-QTLs for SER

In total, eleven QTLs for SER were used to develop QTL-pyramiding lines. The QTLs were located in the substitution segments of SSSLs with the HJX74 genetic background. Three SSSLs, namely, A35, A42, and A88, each carried two SER-QTLs in a substitution segment, while the other 18 SSSLs all carried one SER-QTL in their substitution segments (**Supplementary Table 1**, **Supplementary Figure 1**). The 1-QTL lines (1QLs), which carried 1 QTL for SER in a substitution segment, were mutually crossed to develop 2-QTL lines (2QLs). The 2QLs were crossed with 1QLs to develop 3-QTL lines (3QLs) or crossed with other 2QLs to develop 4-QTL lines (4QLs). In this way, 5-QTL lines (5QLs) and 6-QTL lines (6QLs) were developed (**Supplementary Figure 2**).

A total of 29 QTL-pyramiding lines were developed from the SSSLs, which carried different combinations of QTLs for SER in the HJX74 genetic background (**Table 1**, **Supplementary Table 2**), including five 2QLs (**Supplementary Figure 3**), ten 3QLs (**Supplementary Figure 4**), seven 4QLs (**Supplementary Figure 5**), five 5QLs (**Supplementary Figure 6**), and two 6QLs (**Supplementary Figure 7**).

Phenotypes in the Lines With 1- to 6-QTLs for SER

The SER in the lines with 1- to 6-QTLs were investigated in three cropping seasons (**Figures 1A–D**). The average SER for 1-QTL lines (1QLs) was 48.1%, ranging from 42.5 to 56.5%, which was significantly higher than the 29.2% of HJX74 (**Figures 1C,D**, **Supplementary Table 3**). In seven 2QLs including two SSSLs each carrying 2 QTLs in a single substitution segment and five QTL-pyramiding lines, the average SER was 54.6%, ranging from 46.9 to 63.2%, which was 6.5% higher than the SER in 1QLs. In ten 3QLs, the average SER was 67.4%, ranging from 60.0 to 78.1%, which was 12.8% higher than the SER in 2QLs. In seven 4QLs,

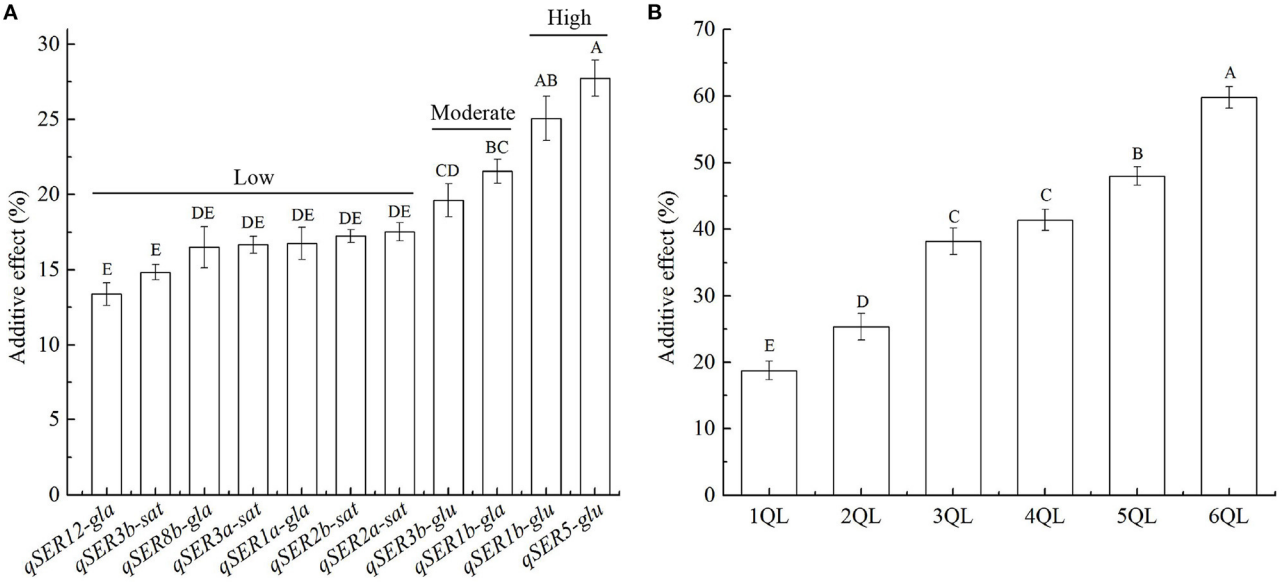


FIGURE 2 | Additive effects of QTLs on stigma exsertion rate (SER). **(A)** Additive effects of 11 QTLs on SER. **(B)** Additive effects of QTL combinations on SER in 1- to 6-QTL lines. Data are shown as the mean ± S.E. of three cropping seasons. Capital letters indicate the significance of differences at the 1% levels.

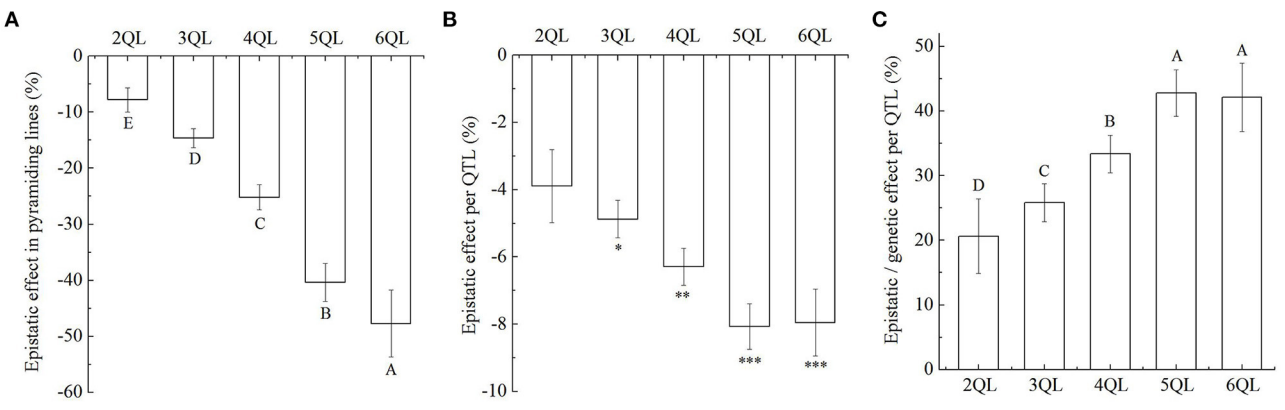


FIGURE 3 | Epistatic effects of QTLs on stigma exsertion rate (SER). **(A)** The total epistatic effects of QTL combinations on SER in 2- to 6-QTL lines. **(B)** The average epistatic effects per QTL on SER in 2- to 6-QTL lines. $^{*}P \leq 0.05$; $^{**}P \leq 0.01$; $^{***}P \leq 0.001$. **(C)** Percentage of epistatic effects in genetic effects for each QTL in 2- to 6-QTL lines. Data are shown as the mean ± S.E. of three cropping seasons. Capital letters indicate the significance of differences at the 1% levels.

the average SER was 70.6%, ranging from 63.4 to 76.5%, which was 3.2% higher than the SER in 3QLs. In five 5QLs, the average SER was 77.1%, ranging from 73.3 to 80.9%, which was 6.5% higher than the SER in 4QLs. In two 6QLs, the average SER was 88.8%, ranging from 86.0 to 91.7%, which was 11.7% higher than the SER in 5QLs (**Figures 1C,D**, **Supplementary Table 4**). These results showed that SERs increased with the increase of QTLs in pyramiding lines.

Phenotypes of eight agronomic traits in 2- to 6-QTL lines were investigated. Most of the traits in these lines, such as heading date, plant height, grain number per panicle, panicle length, and seed setting rate, were not significantly different from those of HJX74. In grain length, 13 lines were significantly different from HJX74, while the other 18 lines were not significantly different. In grain width, 14 lines were significantly different from HJX74, while the other 17 lines were not significantly different (**Supplementary Table 5**). These results showed that the pyramiding lines had a similar genetic background to HJX74, except for the genotype of SER. The grain size, including grain length and grain width, had no significant effect on SER.

Additive Effects of QTLs for SER in 1- to 6-QTL Lines

The additive effects of 11 QTLs on SER varied greatly. The average additive effect of 11 QTLs on SER was 18.9%, ranging from 13.3 to 27.3%. Based on the additive effects, 11 QTLs could

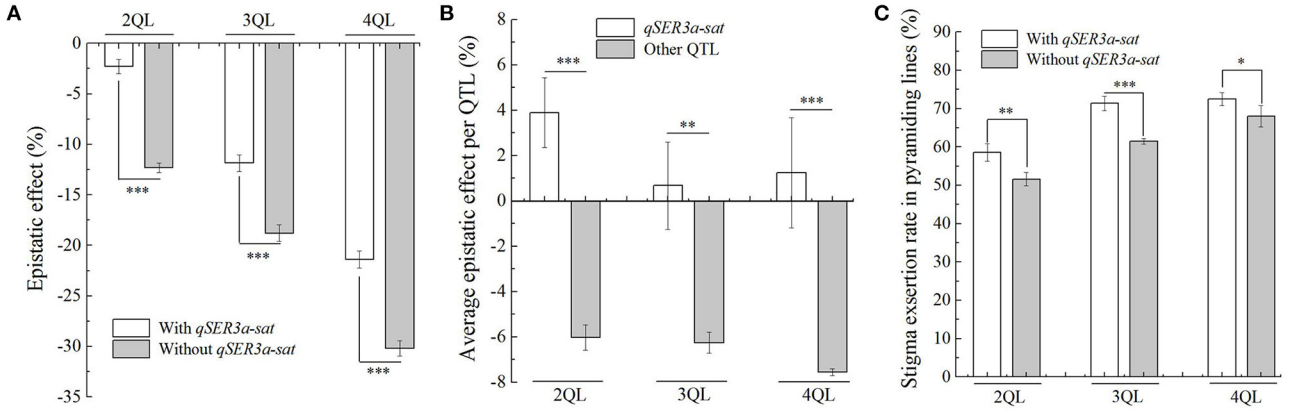


FIGURE 4 | Epistatic effect of *qSER3a-sat* on stigma exsertion rate (SER). **(A)** Epistatic effects of the QTLs in the 2- to 4-QTL lines with and without *qSER3a-sat*. **(B)** Epistatic effects of a single QTL of *qSER3a-sat* and other QTLs in 2- to 4-QTL lines. **(C)** SER in the 2- to 4-QTL lines with and without *qSER3a-sat*. Data are shown as the mean ± S.E. of three cropping seasons. $^{*}P \leq 0.05$; $^{**}P \leq 0.01$; $^{***}P \leq 0.001$.

be divided into three levels. There were 7 QTLs at the low-effect level with the average additive effect of 16.2%, ranging from 13.3 to 17.5%. There were two QTLs at the moderate-effect level with the average additive effect of 20.8%, ranging from 20.2 to 21.4%. At the high-effect level, the additive effects of two QTLs were 25.5 and 27.3%, respectively, and 26.4% on average (**Figure 2A**, **Supplementary Table 6**). The 11 QTLs were from three donors. Four QTLs from *O. sativa* showed the low additive effects with the average additive effect of 16.6%, ranging from 14.8 to 17.5%. The average additive effect of four QTLs from *O. glaberrima* was 17.1%, ranging from 13.3 to 21.4% with three low-effect QTLs and one moderate-effect QTL. The average additive effect of three QTLs from *O. glumaepatula* was 24.3%, ranging from 20.2 to 27.3% with two high-effect QTLs and one moderate-effect QTL (**Figure 2A**). The magnitude of additive effects of QTLs for SER was consistent with the evolutionary levels of species.

Additive effects of 2- to 6-QTLs were calculated from the SER in 2- to 6-QTL lines. The average additive effect of 2-QTL combinations on SER was 25.4% in seven 2QLs, ranging from 17.7 to 34.1%. The average additive effect of 3-QTL combinations on SER was 38.3% in ten 3QLs, ranging from 30.9 to 48.9%. The average additive effect of 4-QTL combinations on SER was 41.4% in seven 4QLs, ranging from 34.3 to 47.3%. The average additive effect of 5-QTL combinations on SER was 47.9% in five 5QLs, ranging from 44.2 to 51.8%. The average additive effect of 6-QTL combinations on SER was 59.7% in two 6QLs, ranging from 56.8 to 62.6% (**Figure 2B**, **Supplementary Table 7**). These results showed that the total additive effects of QTLs increased with increasing QTLs in 1- to 6-QTL lines.

Epistasis of Additive by Additive Interaction Among QTLs on SER in 2- to 6-QTL Lines

Although the additive effects of QTLs on SER were increasing in 2- to 6-QTL lines, they were generally below the expected values (**Figure 2B**, **Supplementary Table 4**). It suggested that there were epistatic effects of QTLs in the 2- to 6-QTL lines. Among the seven 2QLs, the average epistatic effect per QTL combination was −7.8%, ranging from 0.9 to −14.1%, where the average epistatic effect per QTL was −3.9%, but the epistatic effect was not significant. Among the ten 3QLs, the average epistatic effect per QTL combination was −14.6%, ranging from −3.9 to −21.4%, where the average epistatic effect per QTL was −4.9%, and the epistatic effect was significant at the 0.05 level. Among the seven 4QLs, the average epistatic effect per QTL combination was −25.2%, ranging from −17.0 to −31.2%, where the average epistatic effect per QTL was −6.3%, and the epistatic effect was significant at the 0.01 level. Among the five 5QLs, the average epistatic effect per QTL combination was −40.4%, ranging from −32.4 to −49.0%, where the average epistatic effect per QTL was −8.1%, and the epistatic effect was significant at the 0.001 level. Among the two 6QLs, the average epistatic effect per QTL combination was −47.7%, ranging from −41.7 to −53.7%, where the average epistatic effect per QTL was −8.0%, and the epistatic effect was significant at the 0.001 level (**Figures 3A,B**, **Supplementary Table 8**). These results showed that the additive effect of multiple QTLs was smaller than the sum of the additive effects of the corresponding single QTL. The epistatic interactions between QTLs in pyramiding lines reduced the total additive effect of QTL combinations. The epistatic effects of QTLs increased with increasing QTLs in 2- to 6-QTL lines. Therefore, the epistasis was less-than-additive or negative effect.

In 2- to 6-QTL lines, the percentage of epistatic effects in genetic effects for each QTL were 20.7, 26.0, 33.4, 42.9, and 42.4%, respectively (**Figure 3C**). The results showed that although there was epistasis, additivity was the major component of QTL interaction.

QTL *qSER3a-sat* Showed a Minor Epistasis Effect

In the 2- to 4-QTL lines, there were two genotypes with and without *qSER3a-sat*. In 2QLs, the average epistatic effect of three

lines with *qSER3a-sat* was −2.1%, while that of four lines without *qSER3a-sat* was −12.0%. In 3QLs, the average epistatic effect of six lines with *qSER3a-sat* was −11.9%, while that of four lines without *qSER3a-sat* was −18.8%. In 4QLs, the average epistatic effect of four lines with *qSER3a-sat* was −21.4%, while that of three lines without *qSER3a-sat* was −30.4%. In the three sets of QTL lines, the average epistatic effects in the lines with *qSER3a-sat* were significantly less than those without *qSER3a-sat* (**Figure 4A**, **Supplementary Tables 8**, **9**). In 2- to 4-QTL lines, the epistatic effects of *qSER3a-sat* in the *qSER3a-sat* groups were 3.9, 0.7, and 1.4%, respectively, while the average epistatic effects per QTL in the *qSER3a-sat*-free groups were −6.0, −6.3, and −7.6%, respectively (**Figure 4B**, **Supplementary Table 9**). Differing from other QTLs, the epistatic effects of *qSER3a-sat* on SER were minor and positive. As a result, the lines with *qSER3a-sat* had greater SER than those without *qSER3a-sat*. In 2QLs, the average SER of the lines with *qSER3a-sat* was 58.6%, while that of the lines without *qSER3a-sat* was 51.6%. In 3QLs, the average SER of the lines with *qSER3a-sat* was 71.4%, while that of the lines without *qSER3a-sat* was 61.5%. In 4QLs, the average SER of the lines with *qSER3a-sat* was 72.5%, while that of the lines without *qSER3a-sat* was 68.0% (**Figure 4C**, **Supplementary Table 4**). These results showed that *qSER3a-sat* significantly increased SER than other QTLs in the pyramiding lines due to its minor epistasis effect.

DISCUSSION

The Reduced Additive Effect of QTLs on SER May Be Responsible for the Degradation of Outcrossing Ability in Cultivated Rice

The emergence of cultivated crops is the result of natural and artificial selection. During the domestication of cultivated rice, the selection of varieties led to homozygosity of genotypes. At the same time, genotypic homozygosity reinforced the inbreeding of varieties, leading to a gradual decline in their outcrossing ability. Therefore, the degradation of outcrossing ability is an inevitable outcome of the domestication of cultivated rice. Carter et al. (2005) studied the effects of different forms of epistasis on the response to directional selection using a multilinear epistatic model and showed that the main factor determining the evolution of the additive variance, and thus the evolvability, is directional epistasis. Doust et al. (2014) argued that domestication is a multifaceted evolutionary process, involving changes in individual genes, gene interactions, and emergent phenotypes, and found that alleles favored during domestication tend to have larger phenotypic effects and are relatively insensitive to genetic background and environmental influences compared with wild progenitor alleles. In this study, the additive effects of 11 QTLs could be divided into three levels. The four QTLs from *O. sativa*, an Asian cultivated rice, were the low additive effects. *O. glaberrima*, the recently domesticated African cultivated rice, had three low-effect QTLs and one moderate-effect QTL. Among the three QTLs from *O. glumaepatula*, a wild species, two were high-effect QTLs and one was moderate-effect QTL (**Figure 2A**). The magnitude of additive effects of QTLs for SER was consistent with the evolutionary levels of species. In the process of domestication, the additive effects of QTLs on SER were reduced in cultivated rice. Therefore, the reduction in the additive effect of QTLs on SER during domestication may be an important factor in the degradation of the outcrossing ability of cultivated rice.

The Epistasis of Additive by Additive Interaction Between QTLs on SER Was Less-Than-Additive

Mapping epistatic interactions is experimentally, statistically, and computationally challenging. The classical populations used to map QTLs are poorly efficient to detect epistasis. The use of secondary mapping populations, such as chromosome segment substitution lines (CSSLs), introgression lines (ILs), and NILs, in which a region containing the QTL is introgressed into the isogenic background of one of the parental lines, and the QTLs are narrowed down to a small genomic interval by recombination in successive generations, facilitates the analysis of epistasis between naturally occurring variants (Causse et al., 2007; Mackay, 2014). Eshed and Zamir (1996) analyzed interactions between individual *Lycopersicon pennellii* chromosome segments introgressed into an otherwise homogeneous genetic background of *L. esculentum* in a half diallele scheme. Of the 180 tested interactions, 28% were epistatic on both linear and geometric scales, and the detected epistasis was predominately less-than-additive. Although the frequency of epistasis was high, additivity was the major component in the interaction of pairs of QTLs. Kroymann and Mitchell-Olds (2005) reported that a small chromosome interval with no effect on the growth rate of *A. thaliana* NILs contained two epistatically interacting QTLs affecting growth, for one of which the effect on growth was in opposite directions in the different genetic backgrounds. Secondary mapping populations were also used to detect epistasis of QTLs in tomato (Causse et al., 2007) and rice (Uwatoko et al., 2008; Chen et al., 2014; Qin et al., 2015; Yang et al., 2018; Misra et al., 2021). In this study, we used pyramiding lines with 2- to 6-QTLs for SER developed by SSSLs to detect epistasis at two or more loci. Due to the SSSLs and pyramiding lines being homozygous genotypes in the HJX74 genetic background, only the epistasis of additive by additive interaction was detected. The results showed that while the epistasis of QTLs increased with increasing QTLs in 2- to 6-QTL lines, the total additive effects of QTLs increased (**Figures 2B**, **3**). These results indicated that the epistatic interactions of the QTLs only changed the magnitude of effects, in which the phenotype of one locus is suppressed by genotypes at the other locus. Therefore, the epistasis of additive by additive interaction among QTLs in 2- to 6-QTL lines was less-than-additive or negative effect. The results also suggest that detecting epistasis of additive-additive interactions in pyramiding lines developed from SSSLs of the same genetic background is a simpler and more effective scheme.

The High-SER Trait Lost in the Domestication of Cultivated Rice Can Be Reconstructed

In the past decades, great progress has been made in the heterosis utilization of hybrid rice (Yuan, 2017; Zhang, 2022). Improving the outcrossing ability of MSLs is very important for hybrid rice seed production. In the past two decades, a larger number of QTLs controlling SER and related traits have been identified on 12 chromosomes of the rice genome from various genetic resources (Marathi and Jena, 2015; Zhou et al., 2017a; Liu et al., 2019). Recently, eighteen QTLs for SER from *O. sativa*, *O. glaberrima*, and *O. glumaepatula* were detected in the SSSLs in the HJX74 genetic background (Tan et al., 2020, 2021, 2022). It laid a foundation for improving the outcrossing ability of MSLs of cultivated rice using the QTLs for SER. Most MSLs and wild rice were found to have the SER of 70% or more (Virmani and Athwal, 1973; Ying and Zhang, 1989; Uga et al., 2003b; Li et al., 2014; Lou et al., 2014; Guo et al., 2017; Zou et al., 2020). In this study, we found that despite the presence of widespread epistasis, additivity remained a major component of QTL interactions, which resulted in an increase in the SER with increasing QTLs in the pyramiding lines. When carrying 5–6 QTLs, the pyramiding lines had the SER level of most MSLs and wild rice (**Figure 1**). In addition, it was found that *qSER3a-sat* had a minor epistasis, which was more effective for improving the SER (**Figure 4**). Our study suggests that it is possible to reconstruct the high-SER trait, which was lost during the domestication of cultivated rice. The understanding of the genetic architecture of SER in rice lays the foundation for reconstructing the high-SER trait in MSL breeding.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

GZ and SW designed and supervised the work. QT and SB performed most of the experiments and prepared the experimental data. GC, ZY, ZC, WY, PZ, SL, LX, and SC conducted a part of the experiments. HZ, GL, and ZL prepared the experimental materials and assisted in the experiments. GZ analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by grants from the Major Program of Transgenic New Variety Breeding of China (2014ZX08009-037B) and from the National Natural Science Foundation of China (91435207 and 91735304).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.921700/full#supplementary-material

REFERENCES

Bakti, C., and Tanaka, J. (2019). Detection of dominant QTLs for stigma exsertion ratio in rice derived from *Oryza rufipogon* accession 'W0120'. *Breed. Sci.* 69, 143–150. doi: 10.1270/jsbbs.18139

Blanc, G., Charcosset, A., Mangin, B., Gallais, A., and Moreau, L. (2006). Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. *Theor. Appl. Genet.* 113, 206–224. doi: 10.1007/s00122-006-0287-1

Carter, A. J., Hermisson, J., and Hansen, T. F. (2005). The role of epistatic gene interactions in the response to selection and the evolution of evolvability. *Theor. Popul. Biol.* 68, 179–196. doi: 10.1016/j.tpb.2005.05.002

Causse, M., Chaib, J., Lecomte, L., Buret, M., and Hospital, F. (2007). Both additivity and epistasis control the genetic variation for fruit quality traits in tomato. *Theor. Appl. Genet.* 115, 429–442. doi: 10.1007/s00122-007-0578-1

Chen, J., Li, X., Cheng, C., Wang, Y., Qin, M., Zhu, H., et al. (2014). Characterization of epistatic interaction of QTLs *LH8* and *EH3* controlling heading date in rice. *Sci. Rep.* 4, 4263. doi: 10.1038/srep04263

Cheng, S., Cao, L., Zhuang, J., Chen, S., Zhan, X., Fan, Y., et al. (2007). Super hybrid rice breeding in China: achievements and prospects. *J. Integr. Plant Biol.* 49, 805–810. doi: 10.1111/j.1744-7909.2007.00514.x

Dai, Z., Lu, Q., Luan, X., Cai, J., Zhu, H., Liu, Z., et al. (2015). Development of a platform for breeding by design of CMS lines based on an SSSL library in rice (*Oryza sativa* L.). *Euphytica* 205, 63–72. doi: 10.1007/s10681-015-1384-5

Dai, Z., Lu, Q., Luan, X., Ouyang, L., Guo, J., Liang, J., et al. (2016). Development of a platform for breeding by design of CMS restorer lines based on an SSSL library in rice (*Oryza sativa* L.). *Breed. Sci.* 66, 768–775. doi: 10.1270/jsbbs.16044

Doust, A. N., Lukens, L., Olsen, K. M., Mauro-Herrera, M., Meyer, A., and Rogers, K. (2014). Beyond the single gene: how epistasis and gene-by-environment effects influence crop domestication. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6178–6183. doi: 10.1073/pnas.1308940110

Duncan, D. B. (1955). Multiple range and multiple *F* tests. *Biometrics* 11, 1–42. doi: 10.2307/3001478

Eshed, Y., and Zamir, D. (1996). Less-than-additive epistatic interactions of quantitative trait loci in tomato. *Genetics* 143, 1807–1817. doi: 10.1093/genetics/143.4.1807

Fang, C., Li, L., He, R., Wang, D., Wang, M., Hu, Q., et al. (2019). Identification of *S23* causing both interspecific hybrid male sterility and environment-conditioned male sterility in rice. *Rice* 12, 10. doi: 10.1186/s12284-019-0271-4

Fu, Y., Zhao, H., Huang, J., Zhu, H., Luan, X., Bu, S., et al. (2022). Dynamic analysis of QTLs on plant height with single segment substitution lines in rice. *Sci. Rep.* 12, 5465. doi: 10.1038/s41598-022-09536-8

Gaertner, B. E., Parmenter, M. D., Rockman, M.V., Kruglyak, L., and Phillips, P. C. (2012). More than the sum of its parts: a complex epistatic network underlies natural variation in thermal preference behavior in *Caenorhabditis elegans*. *Genetics* 192, 1533. doi: 10.1534/genetics.112.142877

Gao, M., He, Y., Yin, X., Zhong, X., Yan, B., Wu, Y., et al. (2021). Ca^{2+} sensor-mediated ROS scavenging suppresses rice immunity and is exploited by a fungal effector. *Cell* 184, 5391. doi: 10.1016/j.cell.2021.09.009

Guo, J., Li, Y., Xiong, L., Yan, T., Zou, J., Dai, Z., et al. (2022). Development of wide-compatible *indica* lines (WCILs) by pyramiding multiple neutral alleles of *indica-japonica* hybrid sterility loci. *Front. Plant Sci.* 13, 890568. doi: 10.3389/fpls.2022.890568

Guo, L., Qiu, F., Gandhi, H., Kadaru, S., De Asis, E. J., Zhuang, J., et al. (2017). Genome-wide association study of outcrossing in cytoplasmic male sterile lines of rice. *Sci. Rep.* 7, 3223. doi: 10.1038/s41598-017-03358-9

Holland, J. B. (2007). Genetic architecture of complex traits in plants. *Curr. Opin. Plant Biol.* 10, 156–161. doi: 10.1016/j.pbi.2007.01.003

Huang, X., Kurata, N., Wei, X., Wang, Z., Wang, A., Zhao, Q., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497–501. doi: 10.1038/nature11532

Keurentjes, J. J. B., Bentsink, L., Alonso-Blanco, C., Hanhart, C. J., Vries, H. B., Effgen, S., et al. (2007). Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics* 175, 891–905. doi: 10.1534/genetics.106.066423

Kroymann, J., and Mitchell-Olds, T. (2005). Epistasis and balanced polymorphism influencing complex trait variation. *Nature* 435, 95–98. doi: 10.1038/nature03480

Li, C., Sun, C., Mu, P., Chen, L., and Wang, X. (2001). QTL analysis of anther length and ratio of stigma exsertion, two key traits of classification for cultivated rice (*Oryza sativa* L.) and common wild rice (*O. rufipogon* Griff.). *Acta Genet. Sin.* 28, 746–751.

Li, H., Gao, F., Zeng, L., Li, Q., Lu, X., Li, Z., et al. (2010). QTL analysis of rice stigma morphology using an introgression line from *Oryza longistaminata* L. *Mol. Plant Breed.* 8, 1082–1089. doi: 10.3969/mpb.008.001082

Li, P., Feng, F., Zhang, Q., Chao, Y., Gao, G., and He, Y. (2014). Genetic mapping and validation of quantitative trait loci for stigma exsertion rate in rice. *Mol. Breed.* 34, 2131–2138. doi: 10.1007/s11032-014-0168-2

Liu, Y., Zhang, A., Wang, F., Kong, D., Li, M., Bi, J., et al. (2019). Fine mapping a quantitative trait locus, *qSER-7*, that controls stigma exsertion rate in rice (*Oryza sativa* L.). *Rice* 12, 46. doi: 10.1186/s12284-019-0304-z

Lou, J., Yue, G. H., Yang, W. Q., Mei, H. W., Luo, L. J., and Lu, H. J. (2014). Mapping QTLs influencing stigma exertion in rice. *Bulg. J. Agric. Sci.* 20, 1450–1456.

Luan, X., Dai, Z., Yang, W., Tan, Q., Lu, Q., Guo, J., et al. (2019). Breeding by design of CMS lines on the platform of SSSL library in rice. *Mol. Breed.* 39, 126. doi: 10.1007/s11032-019-1028-x

Mackay, T. F. C. (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat. Rev. Genet.* 15, 22–33. doi: 10.1038/nrg3627

Malmberg, R. L., Held, S., Waits, A., and Mauricio, R. (2005). Epistasis for fitness-related quantitative traits in *Arabidopsis thaliana* grown in the field and in the greenhouse. *Genetics* 171, 2013–2027. doi: 10.1534/genetics.105.046078

Marathi, B., and Jena, K. K. (2015). Floral traits to enhance outcrossing for higher hybrid seed production in rice: present status and future prospects. *Euphytica* 201, 1–14. doi: 10.1007/s10681-014-1251-9

Marathi, B., Ramos, J., Hechanova, S. L., Oane, R. H., and Jena, K. K. (2015). SNP genotyping and characterization of pistil traits revealing a distinct phylogenetic relationship among the species of *Oryza*. *Euphytica* 201, 131–148. doi: 10.1007/s10681-014-1213-2

Mei, H., Li, Z., Shu, Q., Guo, L., Wang, Y., Yu, X., et al. (2005). Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two backcross populations. *Theor. Appl. Genet.* 110, 649–659. doi: 10.1007/s00122-004-1890-7

Mihaljevic, R., Utz, H. F., and Melchinger, A. E. (2005). No evidence for epistasis in hybrid and *per se* performance of elite European flint maize inbreds from generation means and QTL analyses. *Crop Sci.* 45, 2605–2613. doi: 10.2135/cropsci2004.0760

Misra, G., Badoni, S., Parween, S., Singh, R. K., Leung, H., Ladejobi, O., et al. (2021). Genome-wide association coupled gene to gene interaction studies unveil novel epistatic targets among major effect loci impacting rice grain chalkiness. *Plant Biotechnol. J.* 19, 910–925. doi: 10.1111/pbi.13516

Oka, H., and Morishima, H. (1967). Variations in the breeding systems of a wild rice, *Oryza perennis*. *Evolution* 21, 249–258. doi: 10.1111/j.1558-5646.1967.tb00153.x

Pan, Z., Tan, B., Cao, G., Zheng, R., Liu, M., Zeng, R., et al. (2021). Integrative QTL identification, fine mapping and candidate gene analysis of a major locus *qLTG3a* for seed low-temperature germinability in rice. *Rice* 14, 103. doi: 10.1186/s12284-021-00544-2

Parmar, K. S., Siddiq, E. A., and Swaminathan, M. S. (1979). Variation in components of flowering behaviour of rice. *Indian J. Genet. Plant Breed.* 39, 542–550.

Qin, M., Zhao, X., Ru, J., Zhang, G., and Ye, G. (2015). Bigenic epistasis between QTLs for heading date in rice analyzed using single segment substitution lines. *Field Crop. Res.* 178, 16–25. doi: 10.1016/j.fcr.2015.03.020

Schon, C. C., Utz, H. F., Groh, S., Truberg, B., Openshaw, S., and Melchinger, A. E. (2004). Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* 167, 485–498. doi: 10.1534/genetics.167.1.485

Sui, F., Zhao, D., Zhu, H., Gong, Y., Tang, Z., Huang, X., et al. (2019). Map-based cloning of a new total loss-of-function allele of *OsHMA3* causes high cadmium accumulation in rice grain. *J. Exp. Bot.* 70, 2857–2871. doi: 10.1093/jxb/erz093

Tan, Q., Wang, C., Luan, X., Zheng, L., Ni, Y., Yang, W., et al. (2021). Dissection of closely linked QTLs controlling stigma exsertion rate in rice by substitution mapping. *Theor. Appl. Genet.* 134, 1253–1262. doi: 10.1007/s00122-021-03771-9

Tan, Q., Zhu, H., Liu, H., Ni, Y., Wu, S., Luan, X., et al. (2022). Fine mapping of QTLs for stigma exsertion rate from *Oryza glaberrima* by chromosome segment substitution. *Rice Sci.* 29, 55–66. doi: 10.1016/j.rsci.2021.12.005

Tan, Q., Zou, T., Zheng, M., Ni, Y., Luan, X., Li, X., et al. (2020). Substitution mapping of the major quantitative trait loci controlling stigma exsertion rate from *Oryza glumaepatula*. *Rice* 13, 37. doi: 10.1186/s12284-020-00397-1

Teng, B., Zeng, R., Wang, Y., Liu, Z., Zhang, Z., Zhu, H., et al. (2012). Detection of allelic variation at the *Wx* locus with single-segment substitution lines in rice (*Oryza sativa* L.). *Mol. Breed.* 30, 583–595. doi: 10.1007/s11032-011-9647-x

Uga, Y., Fukuta, Y., Cai, H.W., Iwata, H., Ohsawa, R., Morishima, H., et al. (2003a). Mapping QTLs influencing rice floral morphology using recombinant inbred lines derived from a cross between *Oryza sativa* L. and *Oryza rufipogon* Griff. *Theor. Appl. Genet.* 107, 218–226. doi: 10.1007/s00122-003-1227-y

Uga, Y., Fukuta, Y., Ohsawa, R., and Fujimura, T. (2003b). Variations of floral traits in Asian cultivated rice (*Oryza sativa* L.) and its wild relatives (*O. rufipogon* Griff.). *Breed. Sci.* 53, 345–352. doi: 10.1270/jsbbs.53.345

Uwatoko, N., Onishi, A., Ikeda, Y., Kontani, M., Sasaki, A., Matsubara, K., et al. (2008). Epistasis among the three major flowering time genes in rice: coordinate changes of photoperiod sensitivity, basic vegetative growth and optimum photoperiod. *Euphytica* 163, 167–175. doi: 10.1007/s10681-007-9584-2

Virmani, S. S., and Athwal, D. S. (1973). Genetic variability in floral characteristics influencing outcrossing in *Oryza sativa* L. *Crop Sci.* 13, 66–67. doi: 10.2135/cropsci1973.0011183X001300010019x

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

Xi, Z., He, F., Zeng, R., Zhang, Z., Ding, X., Li, W., et al. (2006). Development of a wide population of chromosome single-segment substitution lines in the genetic background of an elite cultivar of rice (*Oryza sativa* L.). *Genome* 49, 476–484. doi: 10.1139/g06-005

Xiong, L., Liu, K., Dai, X., Xu, C., and Zhang, Q. (1999). Identification of genetic factors controlling domestication-related traits of rice using an F_2 population of a cross between *Oryza sativa* and *O. rufipogon*. *Theor. Appl. Genet.* 98, 243–251. doi: 10.1007/s001220051064

Yang, T., Zhang, S., Zhao, J., Liu, Q., Huang, Z., Mao, X., et al. (2016). Identification and pyramiding of QTLs for cold tolerance at the bud bursting and the seedling stages by use of single segment substitution lines in rice (*Oryza sativa* L.). *Mol. Breed.* 36, 96. doi: 10.1007/s11032-016-0520-9

Yang, W., Liang, J., Hao, Q., Luan, X., Tan, Q., Lin, S., et al. (2021a). Fine mapping of two grain chalkiness QTLs sensitive to high temperature in rice. *Rice* 14, 33. doi: 10.1186/s12284-021-00476-x

Yang, W., Xiong, L., Liang, J., Hao, Q., Luan, X., Tan, Q., et al. (2021b). Substitution mapping of two closely linked QTLs on chromosome 8 controlling grain chalkiness in rice. *Rice* 14, 85. doi: 10.1186/s12284-021-00526-4

Yang, Z., Jin, L., Zhu, H., Wang, S., Zhang, G., and Liu, G. (2018). Analysis of epistasis among QTLs on heading date based on single segment substitution lines in rice. *Sci. Rep.* 8, 3059. doi: 10.1038/s41598-018-20690-w

Ying, C., and Zhang, Q. (1989) Studies on the character of stigma exsertion among some of *Oryza* species. *Chin. J. Rice Sci.* 3, 62–66.

Yuan, L. (2017). Progress in super-hybrid rice breeding. *Crop J.* 5, 100–102. doi: 10.1016/j.cj.2017.02.001

Yuan, L., and Virmani, S. (1988). "Status of hybrid rice research and development," in *Hybrid Rice*. eds Smith, W. H., Bostian, L. R., and Cervantes, E. (Manila: Internationa Rice Research Institute), 7–24.

Zamir, D. (2001). Improving plant breeding with exotic genetic libraries. *Nat. Rev. Genet.* 2, 983–989. doi: 10.1038/35103590

Zeng, R., Zhang, Z., He, F., Xi, Z., Akshay, T., Shi, J., et al. (2006). Identification of multiple alleles at the *Wx* locus and development of single segment substitution lines for the alleles in rice. *Rice Sci.* 13, 9–14. doi: 10.1360/aps040074

Zhan, P., Ma, S., Xiao, Z., Li, F., Wei, X., Lin, S., et al. (2022). Natural variations in *grain length 10* (*GL10*) regulate rice grain size. *J. Genet. Genomics*. 49, 405–413. doi: 10.1016/j.jgg.2022.01.008

Zhan, P., Wei, X., Xiao, Z., Wang, X., Ma, S., Lin, S., et al. (2021). *GW10*, a member of P450 subfamily regulates grain size and grain number in rice. *Theor. Appl. Genet.* 134, 3941–3950. doi: 10.1007/s00122-021-03939-3

Zhang, G. (2021). Target chromosome-segment substitution: a way to breeding by design in rice. *Crop J.* 9, 658–668. doi: 10.1016/j.cj.2021.03.001

Zhang, G. (2022). The next generation of rice: inter-subspecific *indica-japonica* hybrid rice. *Front. Plant Sci.* 13, 857896. doi: 10.3389/fpls.2022.857896

Zhang, G., Zeng, R., Zhang, Z., Ding, X., Li, W., Liu, G., et al. (2004). The construction of a library of single segment substitution lines in rice (*Oryza sativa* L.). *Rice Genet. Newsl.* 21, 85–87.

Zhang, S., Huang, X., and Han, B. (2021). Understanding the genetic basis of rice heterosis: advances and prospects. *Crop J.* 9, 688–692. doi: 10.1016/j.cj.2021.03.01

Zhang, Y., Yang, J., Shan, Z., Chen, S., Qiao, W., Zhu, X., et al. (2012). Substitution mapping of QTLs for blast resistance with SSSLs in rice (*Oryza sativa* L.). *Euphytica* 184, 141–150. doi: 10.1007/s10681-011-0601-0

Zhou, H., Li, P., Xie, W., Hussain, S., Li, Y., Xia, D., et al. (2017a). Genome-wide association analyses reveal the genetic basis of stigma exsertion in rice. *Mol. Plant* 10, 634–644. doi: 10.1016/j.molp.2017.01.001

Zhou, Y., Xie, Y., Cai, J., Liu, C., Zhu, H., Jiang, R., et al. (2017b). Substitution mapping of QTLs controlling seed dormancy using single segment substitution lines derived from multiple cultivated rice donors in seven cropping seasons. *Theor. Appl. Genet.* 130, 1191–1205. doi: 10.1007/s00122-017-2881-9

Zhu, H., Li, Y., Liang, J., Luan, X., Xu, P., Wang, S., et al. (2018). Analysis of QTLs on heading date based on single segment substitution lines in rice (*Oryza sativa* L.). *Sci. Rep.* 8, 13232. doi: 10.1038/s41598-018-31377-7

Zhu, Y., Zuo, S., Chen, Z., Chen, X., Li, G., Zhang, Y., et al. (2014). Identification of two major rice sheath blight resistance QTLs, $qSB1\text{-}1^{HJX74}$ and $qSB11^{HJX74}$, in field trials using chromosome segment substitution lines. *Plant Dis.* 98, 1112–1121. doi: 10.1094/PDIS-10-13-1095-RE

Zou, T., Zhao, H., Li, X., Zheng, M., Zhang, S., Sun, L., et al. (2020). QTLs detection and pyramiding for stigma exsertion rate in wild rice species by using the single-segment substitution lines. *Mol. Breed.* 40, 74. doi: 10.1007/s11032-020-01157-1

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

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

frontiers | Frontiers in Plant Science

MINI REVIEW
published: 23 June 2022
doi: 10.3389/fpls.2022.907363

The Role of Interspecific Hybridisation in Adaptation and Speciation: Insights From Studies in *Senecio*

Edgar L. Y. Wong[1*†], *Simon J. Hiscock*[1,2] *and Dmitry A. Filatov*[1]

[1]*Department of Plant Sciences, University of Oxford, Oxford, United Kingdom,* [2]*Oxford Botanic Garden and Arboretum, Oxford, United Kingdom*

Hybridisation is well documented in many species, especially plants. Although hybrid populations might be short-lived and do not evolve into new lineages, hybridisaiton could lead to evolutionary novelty, promoting adaptation and speciation. The genus *Senecio* (Asteraceae) has been actively used to unravel the role of hybridisation in adaptation and speciation. In this article, we first briefly describe the process of hybridisation and the state of hybridisation research over the years. We then discuss various roles of hybridisation in plant adaptation and speciation illustrated with examples from different *Senecio* species, but also mention other groups of organisms whenever necessary. In particular, we focus on the genomic and transcriptomic consequences of hybridisation, as well as the ecological and physiological aspects from the hybrids' point of view. Overall, this article aims to showcase the roles of hybridisation in speciation and adaptation, and the research potential of *Senecio*, which is part of the ecologically and economically important family, Asteraceae.

Keywords: hybridisation, speciation, adaptation, *Senecio aethnensis*, *Senecio chrysanthemifolius*, Mount Etna

OPEN ACCESS

Edited by:
Andrew H. Paterson, University of Georgia, United States

Reviewed by:
Petr Smýkal, Palacký University, Olomouc, Czechia

****Correspondence:***
Edgar L. Y. Wong
edgar.wong@zoo.ox.ac.uk

†Present address:
Edgar L. Y. Wong, Department of Zoology, University of Oxford, Oxford, United Kingdom

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

Received: *29 March 2022*
Accepted: *03 June 2022*
Published: *23 June 2022*

Citation:
Wong ELY, Hiscock SJ and Filatov DA (2022) The Role of Interspecific Hybridisation in Adaptation and Speciation: Insights From Studies in Senecio. Front. Plant Sci. 13:907363. doi: 10.3389/fpls.2022.907363

INTRODUCTION

Understanding the evolutionary genetic processes that underpin phenotypic adaptation and speciation is fundamental for understanding the process of Darwinian evolution. It has been more than 160 years since Darwin described how species adapt and evolve through the force of natural selection, but despite the subsequent advances in population genetics and evolutionary theory, our understanding of adaptation and speciation is still far from complete (Coyne and Orr, 2004; Rieseberg and Willis, 2007; Abbott et al., 2009). Speciation is one of the oldest problems in evolutionary biology, which has successfully resisted the efforts of generations of evolutionary biologists (e.g., Coyne and Orr, 1989). The advance in molecular genetics techniques in the last 15 years or so resulted in the reincarnation of the field which became one of the hottest topics of evolutionary biology (e.g., Ravinet et al., 2017; Campbell et al., 2018; Becraft and Moya, 2021). The role of interspecific hybridisation in adaptation and speciation is actively debated in the literature and its importance becomes more apparent (e.g., Ebersbach et al., 2020; Nevado et al., 2020; Wong et al., 2020; Hobbs et al., 2021; Bush, 2022).

Plant speciation (or at least the literature on plant speciation) differs substantially from that in animals. Plant literature often focuses on species hybridisation and introgression during

speciation, rather than on reproductive isolation (reviewed in Abbott, 1992). Historically, the animal-focused researchers considered hybridisation an evolutionary dead-end (Mayr, 1963) since it homogenises the diverging genomes and prevents speciation. However, plant biologists (Anderson, 1948; Anderson and Stebbins, 1954; Grant, 1972) have long considered hybridisation as an important force in adaptation and speciation. Indeed, hybridisation is widespread in plants (e.g., Grant, 1972; Mallet, 2001; Rieseberg et al., 2004), and it may play a substantial role in the adaptation and speciation of plant populations (Barton, 2001; Rieseberg et al., 2003). Recent studies have shown that hybridisation can have more complex outcomes than just homogenisation of diverging genomes. For example, it could lead to extinction of hybrid lineages, evolution of new species (hybrid speciation), and introgression of adaptive alleles, leading to faster adaptation. While the importance of interspecific hybridisation in evolution is becoming more apparent, the extent (and the role) of gene exchange during hybridisation of plant and animal species is not entirely clear.

Many partially isolated species are known to form hybrid zones. A *Helianthus* hybrid zone was demonstrated to be 'semi-permeable', meaning that while there was a barrier to gene flow of some genomic regions, the majority of the genome can introgress freely (Rieseberg et al., 1999). On the other hand, the two *Senecio* species forming an elevational hybrid zone on Mount Etna, Sicily, are fully compatible, though the evidence for numerous interspecific incompatibilities between these species is starting to emerge (Brennan et al., 2014, 2019; Chapman et al., 2016). Hybrid zones can be considered as 'windows on the evolutionary process' (Harrison, 1993) and they represent 'evolutionary laboratories' providing the researchers an opportunity to analyse and dissect the role of hybridisation in speciation and adaptation. In particular, the analyses of hybrid zones inform the debate whether hybrids are an evolutionary dead-end (Mayr, 1963) or play a more creative role in adaptation and speciation (Anderson, 1948; Arnold, 1997; Rieseberg, 1997).

There are several ways in which hybridisation could promote speciation and adaptation (Seehausen, 2013; Vallejo-Marín and Hiscock, 2016). Hybridisation can either act to transfer adaptive alleles between lineages to aid adaptation, or result in hybrid speciation with or without polyploidisation. Allopolyploid hybrid speciation occurs when two parental lineages with different chromosome ploidies hybridise. This may result in hybrids with an odd (often sterile) or even number of chromosome sets. Some of the sterile hybrid populations may persist without reproduction due to constant hybridisation events, while some can reproduce asexually. Others may undergo further genome duplications to overcome genomic conflicts such as chromosomal pairing during meiosis (e.g., *Mimulus peregrinus*; Vallejo-Marín, 2012; Vallejo-Marín et al., 2015). Homoploid hybrid speciation occurs when both parental lineages have the same chromosome number (e.g., Italian Sparrow, *Passer iltaliae*; Hermansen et al., 2011; and Oxford ragwort, *Senecio squalidus*; James and Abbott, 2005). Depending on their origin and genomic structure, hybrids have different obstacles to overcome (such as problems in meiosis and gene regulation) and different evolutionary pathways to eventually become a reproductively isolated taxon. Even if hybridisation does not lead to speciation, it can provide opportunities for adaptation. Following hybridisation, hybrid lineages often experience tremendous changes compared to their parents. Instead of detailing all the consequences of hybridisation on the phenotypic and genomic level, this short review focuses on the ones that are potentially beneficial for adaptation and speciation. Examples from different *Senecio* species will be used to illustrate the role of hybridisation in adaptation and speciation.

Senecio L. is a genus of herbaceous plants, shrubs, small trees and climbers in the Asteraceae family. The genus has a worldwide distribution, containing at least 1,400 described species (Royal Botanic Gardens Kew, 2022), many of which are cultivated extensively. Alongside other genera such as *Artemisia*, *Cynara*, *Echinacea*, *Helianthus*, *Lactuca*, *Tragopogon*, the Asteraceae family presents huge economic values, with numerous species being used in food, medicine, and horticulture. The '*Senecio* system' is also rapidly becoming recognised as one of the most tractable plant models in which to study the process of speciation at a genetic, genomic, and ecological level (Abbott and Rieseberg, 2012; Gross, 2012; Walter et al., 2020). The fact that speciation events in the genus have occurred relatively recently, and involve examples of both ecological speciation and hybrid speciation (homoploid and allopolyploid; Abbott and Rieseberg, 2012; Hegarty et al., 2012), make *Senecio* a unique alternative to more conventional plant models, such as *Arabidopsis*, for studies of plant evolution in action.

Natural hybridisation and stable hybrid zones present natural experiments that can be dissected at the molecular level to identify genomic factors associated with local adaptation and the maintenance of species differences and boundaries (Grant, 1981; Rieseberg, 1997; Arnold, 2006; Lexer and Widmer, 2008). A classic example of natural hybridisation is found on Mount Etna, Sicily. Here, two species of *Senecio*, *S. aethnensis* and *S. chrysanthemifolius*, which are locally adapted to high- and low-elevation conditions respectively, form a stable hybrid zone at the boundaries of their respective ecological ranges mid-way up the volcano. *S. aethnensis* populations are found at high elevations [>2,000 meters above sea level (masl)] and *S. chrysanthemifolius* are found at low elevations (<1,000 masl; Brennan et al., 2009; Muir et al., 2013). The two species are distinguishable through an array of phenotypic (such as leaf dissection: **Figure 1**; James and Abbott, 2005; Brennan et al., 2009; Wong et al., 2020), physiological (such as seed germination temperature: Ross, 2010), and ecological differences (such as flowering time). Significant differences between these species have also been observed at the level of gene expression (Hegarty et al., 2009; Muir et al., 2013; Chapman et al., 2016), and candidate genes identified in these studies are predicted to be adaptive (Wong et al., 2020). For instance, genes predicted to be involved in adaptation to high light intensity, UV stress, sulphur metabolism, dehydration and cold stress are highly expressed in *S. aethnensis* compared to *S. chrysanthemifolius* (Hegarty et al., 2009). The two species maintain a hybrid zone at intermediate elevations, where hybrids display intermediate

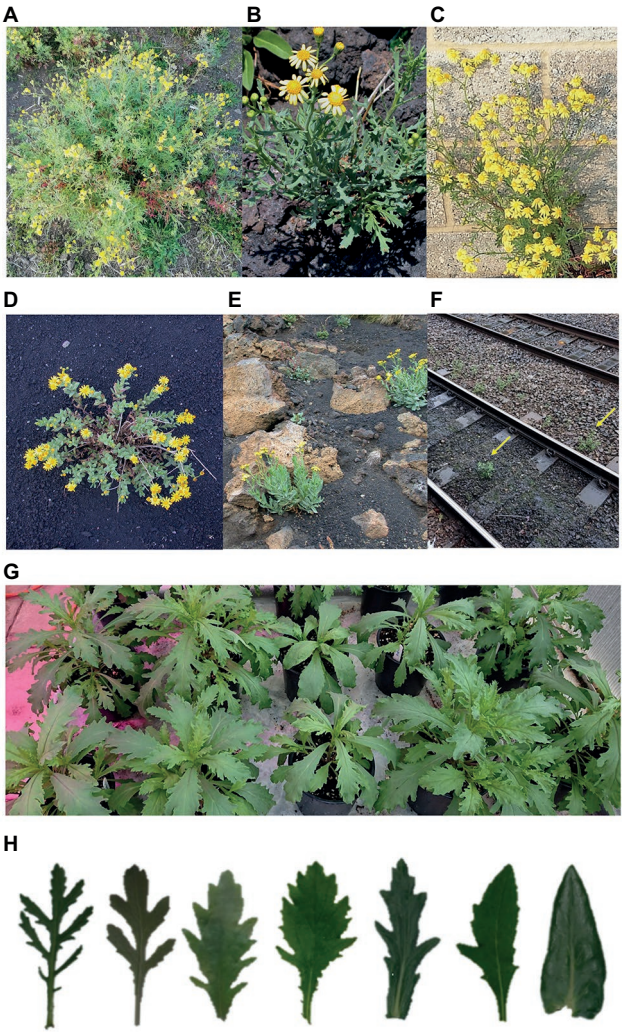


FIGURE 1 | Examples of phenotypes and habitats of *Senecio aethnensis*, *Senecio chrysanthemifolius*, their natural hybrids and *Senecio squalidus*. **(A)** *S. chrysanthemifolius* from Mount Etna. **(B)** Natural hybrid on Mount Etna. **(C)** Homoploid hybrid species *S. squalidus* from Oxford, United Kingdom. **(D)** *S. aethnensis* from Mount Etna. **(E)** Natural habitat of *Senecio* on Mount Etna. **(F)** *S. squalidus* found on railways in the United Kingdom (identified by yellow arrows). **(G)** Variation in leaf phenotypes in one F_2 synthetic hybrid (between *S. aethnensis* and *S. chrysanthemifolius*) family. **(H)** Leaf shape variation in greenhouse-grown *Senecio*. Left most: *S. chrysanthemifolius*; Right most: *S. aethnensis*; All in-between: synthetic F_2 hybrids.

phenotypes (James and Abbott, 2005; Brennan et al., 2009). Despite recent divergence (<200 KYA; Chapman et al., 2013; Muir et al., 2013; Osborne et al., 2013) and regular gene flow, *S. aethnensis* and *S. chrysanthemifolius* evolved as distinct species and maintain very different phenotypes, with leaf shape showing the most extreme differences at the phenotypic level (**Figure 1**). How the species identity is maintained despite the on-going gene flow remains unclear, but it was suggested that multiple factors act together to keep the species identity, including genetic incompatibilities (Brennan et al., 2014, 2016, 2019; Chapman et al., 2016), strong divergent selection and selection against hybrids (Brennan et al., 2009; Wong et al., 2020). For instance, transmission ratio distortion (TRD) was also identified in this system (34 out of 127 marker loci in Brennan et al., 2014; three regions in Chapman et al., 2016; 2.9%–26.8% of loci in Brennan et al., 2019), with pre-zygotic events (such as gametophytic selection), cytonuclear incompatibility, Bateson–Dobzhansky–Muller incompatibility and potentially underdominance contributing to these TRFs (Brennan et al., 2014). Hybrid breakdown as a consequence of genetic incompatibilities was also observed in synthetic hybrids. Some of the breakdown traits include low germination and albinism (Hegarty et al., 2009), necrotic growth (Chapman et al., 2016) and early mortality (Brennan et al., 2014). Thus, current data strongly suggest that *S. aethnensis* and *S. chrysanthemifolius* represent a clear-cut case of ecological speciation driven by adaptation to contrasting conditions of high- and low-elevation. There are relatively few studied cases of ecological speciation (Counterman et al., 2010; Martínez-Fernández et al., 2010; Papadopulos et al., 2011), making *Senecio* a particularly valuable model system for research in adaptation and speciation.

Another attractive feature of this study system is a case of rapid recent (<300 years) homoploid speciation of *S. squalidus* in the United Kingdom. This speciation occurred following the introduction of *Senecio* plants from Mount Etna to England some 300 years ago (Nevado et al., 2020). This case of speciation is relatively well documented because it occurred in Oxford Botanic Garden, hence the common name of *Oxford Ragwort*. Originating from hybridisation an English garden and a period of sustained cultivation in Oxford, *S. squalidus* has now spread to the majority of the United Kingdom as far north as Scotland, and was found to hybridise with native species, such as Groundsel (*Senecio vulgaris*) leading to the origin of two new allopolyploid species, the allohexaploid *S. cambrensis* and the allotetraploid *S. eboracensis* (Lowe and Abbott, 2004). This system presents an exciting model for studying speciation, adaptation, invasion and hybridisation (e.g., Nevado et al., 2020; Walter et al., 2020). It also presents an excellent example of how hybridisation can lead to speciation.

ROLES OF HYBRIDISATION IN ADAPTATION AND SPECIATION

Transcriptome Shock

Studies have shown altered gene expressions in hybrids compared to parental lineages, a process known as 'transcriptome shock' (Lee and Chen, 2001; Shaked et al., 2001; Adams et al., 2003; Adams and Wendel, 2005; Comai, 2005; Madlung et al., 2005; term first used in Hegarty et al., 2006). It is worth stressing that although transcriptome shock is often observed in polyploid hybrids, it is an outcome of hybridisation, rather than genome duplication (Wang et al., 2006). The alterations to gene expressions are found to be nonadditive (Wang et al., 2006; Hegarty et al., 2011), immediate in F_1 hybrids but stable in subsequent hybrid generations (Comai et al., 2000; Adams et al., 2003; Hegarty et al., 2006, 2009; Wang et al., 2006). Studying the triploid hybrids (*S. x baxteri*) between the tetraploid *S. vulgaris* and

diploid *S. squalidus*, and hexaploid allopolyploid (*S. cambrensis*) arisen from the triploid hybrid, Hegarty et al. (2006) showed that transcriptome shock was evident in *S. x baxteri* and that the shock was 'ameliorated' after genome duplication in the *S. cambrensis*. It could manifest in mechanisms involving gene silencing, regulatory networks, chromatin remodelling and DNA methylation (Shaked et al., 2001; Madlung et al., 2005). Although this epigenetic instability could be disadvantageous, it could serve as a target for selection to act on to subsequently facilitate adaptation and speciation in the hybrid lineage (Hegarty et al., 2011).

In the homoploid hybrid species *S. squalidus*, two genes, ATP-sulfurylase precursor and glutathione-S-transferase, were found to have transgressive up-regulation compared to the midpoint of the parental species (Hegarty et al., 2009). These two genes are likely up-regulated in response to deficiency in sulphur (Xiang and Oliver, 1998; Harada et al., 2002), as most of United Kingdom soils contain much less sulphur (<20 kg/ha/year; Brown et al., 2000) than soil on Mount Etna where the parental species live (>40 kg/ha/year in quiescent period between 1997 and 2001, and much more following volcanic eruption; Aiuppa et al., 2006). Research has shown that the hybridisation event leading to speciation of *S. squalidus* most likely happened after parental plants were brought to the United Kingdom, instead of hybrid material arriving in the United Kingdom from Mount Etna, as thought previously (Nevado et al., 2020). Hence the altered gene expression observed in *S. squalidus* likely evolved due to hybridisation but not pre-adaptation. This is a good example of how transcriptome shock can facilitate adaptation in hybrid lineages in a novel environment which is drastically different than the parental ones (in this case includes sulphur level).

Genome Reorganisation

It is not uncommon for hybrid lineages to experience genome reorganisation ('genome shock': e.g., Rieseberg, 2001; Chen and Ni, 2006), such as chromosomal rearrangements, translocations, and movement of transposable elements. These rearrangements may not be involved in adaptation to new environments, but they often serve as a form of reproductive isolation from parental lineages through restricting backcrossing (Rieseberg et al., 2003; Coghlan et al., 2005; Hegarty and Hiscock, 2005; Paun et al., 2007), an important step in speciation. Genomic restructuring is also commonly observed in new, successful hybrid lineages alongside other ecological and spatial divergence from progenitors (Buerkle et al., 2000; Baack and Rieseberg, 2007; Karrenberg et al., 2007; Brennan et al., 2019), and might be crucial to restore nucleocytoplasmic compatibility (Soltis and Soltis, 1999).

Hybridisation-induced chromosomal rearrangements have been documented in a few allopolyploid species such as *Triticum* (Levy and Feldman, 2004), *Nicotiana* (Lim et al., 2004) and *Arabidopsis* (Pontes et al., 2004); as well as homoploid species such as *Helianthus* (e.g., Burke et al., 2004; Lai et al., 2005), *Iris* (Tang et al., 2010; Taylor et al., 2013), *Agrodiaetus* (Lukhtanov et al., 2015), and our focal group *Senecio* (Brennan et al., 2019).

Comparing the genome structure of *S. squalidus* and its progenitors *S. aethnensis* and *S. chrysanthemifolius* using genetic mapping, it was found that there are indeed differences in genomic architecture between the latter two and this led to the inheritance of some of this genetic incompatibility in *S. squalidus* (Brennan et al., 2019). Comparison between genetic maps of F_2 mapping families with either parent also revealed genomic reorganisation between maps in half of the linkage groups (Brennan et al., 2019). They also showed evidence for colocation between transmission ratio distortion loci and genomic rearrangements. These rearranged regions were hypothesised to contribute to incompatibilities and reproductive isolation, and where divergent selection acts to promote adaptation and speciation. This hypothesis can be tested with the *S. squalidus* genome that will soon be available.

Increased Heterozygosity, Heterosis and Transgressive Segregation

Another opportunity for adaptation and speciation in both homoploid and polyploid hybrid lineages is heterosis, in which the hybrid lineages express more vigorous phenotypes compared to parental lineages due to increased heterozygosity; and transgressive segregation, in which extreme phenotypes (positive or negative) are formed. Because of recombination and transgressive effects, hybrids usually possess higher level of variation compared to parental lineages, which creates vast potential for novel evolutionary trajectories (Arnold, 2006; Abbott and Brennan, 2014).

In *Senecio*, heterosis was observed in the F_1 hybrids between *Senecio jacobaea* and *S. aquaticus*. Hybrids had superior fitness, and they were found to possess adaptations such as drought and flooding resistance, inherited from either parent, respectively, (Kirk et al., 2005). These features would allow the hybrid lineage to expand and occupy niches outside of their parental ones. Unlike homoploid hybrids whose heterozygosity is expected to decline over generations due to recombination, the enforced pairing of homologous chromosomes in polyploid hybrids inhibits intergenomic recombination, thus conserving the high level of heterozygosity through generations (Comai, 2005). An exceptional example of the role of polyploid hybridisation in adaption and speciation is the arctic flora. Research has suggested that polyploid lineages are better at colonising after deglaciation compared to diploid lineages, and that the polyploid lineages' fixed-heterozygosity prevented the disadvantageous effects of inbreeding and loss of heterozygosity caused by genetic drift (Brochmann et al., 2004).

Transgressive segregation is very commonly applied in crop breeding, but it can also be found in the wild species. For example, the homoploid species, *Helianthus anomalus*, *H. paradoxus*, and *H. deserticola* (all originated from the same pair of parental species), occupy different habitats and also exhibit adaptive traits not seen in the parental species (Schwarzbach and Rieseberg, 2002; Welch and Rieseberg, 2002; Gross et al., 2003; Gross and Rieseberg, 2005; Lai et al., 2005). This demonstrates that hybridisation is able to generate novelty in terms of morphology, anatomy, life history and

physiology which in turn allows for adaptation and speciation (Abbott et al., 2010). Transgressive up-regulation of genes were also observed in *Senecio* (discussed above: Hegarty et al., 2009).

Change in Mating System and Reproductive Traits

Hybridisation and polyploidisation can sometimes lead to a different mating system in the hybrids. For instance, it is well-known that allopolyploidy is frequently associated with a shift from self-incompatibility (in the parental species) to self-compatibility in the hybrid polyploid (Entani et al., 1999; Miller and Venable, 2000; Nasrallah et al., 2000; Brennan and Hiscock, 2010); and be associated with asexual reproduction, both vegetative and agamospermy (Otto and Whitton, 2000; Janko et al., 2012). In the arctic flora, numerous diploid taxa of hybrid origin are self-compatible or clonal, making them successful in the arctic environment where pollinators are scarce (Brochmann et al., 2004). Having a different mating system would allow these hybrid taxa to occupy new niches and/or perpetuate in smaller populations since there is reduced reliance on pollinators and mating partners.

Hybridisation between the tetraploid *S. vulgaris* and diploid *S. squalidus* in the United Kingdom has also resulted in two hybrid species with varying reproductive traits and mating system. *S. vulgaris* is self-compatible with capitula that are rayless; whereas *S. squalidus* is self-incompatible with capitula showing a mix of ray and disc florets. Their hybridisation led to the evolution of an allohexaploid species, *S. cambrensis*, and tetraploid species, *S. eboracensis* (Lowe and Abbott, 2004; Brennan and Hiscock, 2010; Hegarty et al., 2012). Both hybrid species possess self-compatibility of *S. vulgaris* and ray florets from *S. squalidus*. Although some *S. cambrensis* were found to be self-sterile (Brennan and Hiscock, 2010). Compared to the tetraploid parent *S. vulgaris*, *S. eboracensis* was also found to have more stigmatic papillae that facilitates pollen capture (Richards, 1986), and higher production of pollen grains which are the main food source of its pollinators (Gilbert, 1986). These changes in reproductive traits (especially self-compatibility) are crucial, especially to new hybrid lineages, to sustain their initial small populations.

Adaptive Introgression

As reproductive isolation of closely related species is often incomplete, mutations may traverse species boundaries. Low levels of gene flow due to rare interspecific hybridisation may have little effect on neutral diversity within species, but it may be extremely important for genes under positive selection, which can spread across a subdivided 'population' (i.e., several hybridising species) with very little gene flow (Slatkin, 1976; Slatkin and Wiehe, 1998). Natural selection may substantially accelerate the transfer of genes between the species (reviewed in Barton, 2001), and horizontal gene transfer (HGT) in bacteria is often detected for genes conferring advantage to their hosts (Ochman et al., 2000; Bapteste et al., 2004), such as antibiotic resistance, or a 'widespread colonization island' locus that is involved in adherence and colonisation of surfaces (Planet et al., 2003). The extent and the role of HGT in non-microbial organisms is less clear.

Sharing of adaptive mutations may significantly accelerate adaptation process, as species do not have to 'wait' for an adaptive mutation to arise *de novo* (Seehausen, 2004). Sharing of adaptive mutations between species is likely to be particularly important for species with small population sizes, such as endemic adaptive radiations on islands, while species with large population sizes may have sufficient standing variation to make sharing of adaptive mutations unimportant. However, this conjecture remains to be tested. While the number of examples of adaptive gene sharing is growing (e.g., Kapralov and Filatov, 2006; Meier et al., 2017; Richards and Martin, 2017) the role of adaptive allele sharing in adaptation and speciation is still far from clear.

Previous studies have identified multiple cases of cytonuclear phylogenetic discordance (e.g., Shaw, 2002), suggesting introgression of chloroplast or mitochondrial DNA, but cytoplasmic DNA may be particularly prone to interspecific introgression (Tsitrone et al., 2003) and may not reflect the situation with nuclear genes. The literature survey of F_{st} values and selection gradients and differentials in phenotypic traits suggested that 'collective evolution' of species exchanging adaptive alleles may be fairly widespread (Morjan and Rieseberg, 2004), but more work is needed to clarify the importance of this factor in evolution.

An excellent example of adaptive introgression is the one responsible for adaptation to serpentine soils in *Arabidopsis* (Arnold et al., 2016) and wing colours in mimic *Heliconius* (Pardo-Diaz et al., 2012). Another example of apparently adaptive introgression was also reported for two *Senecio* species in the United Kingdom (Kim et al., 2008). *Senecio vulgaris* that normally does not have ray florets on the capitula, evolved a variety, *S. vulgaris* var. *hibernicus*, which possess rayed capitula like *S. squalidus* following introgression from the latter species (Abbott et al., 1992; Kim et al., 2008). The production of ray florets in this variety of *S. vulgaris* involves the expression of various cycloidea (CYC)-like genes (Kim et al., 2008; Garcês et al., 2016), and was proven to enhance pollination attraction (Abbott and Irwin, 1988) and maternal outcrossing (Marshall and Abbott, 1982, 1984) compared to the non-introgressed *S. vulgaris*. In another pair of *Senecio* species in the Bavarian Forest National Park, Germany, low-elevation *S. ovatus* has benefitted from adaptive introgression from the high-elevation *S. hercynicus*, with introgressed traits related to climatic conditions at high elevations and also shorter vegetative phases as *S. ovatus* spreads towards higher elevations (Bog et al., 2017).

The spread of globally adaptive mutations across several species should result in the loss of species divergence, the loss of intraspecific polymorphism and a characteristic bias in the frequency spectrum of mutations towards rare alleles for the region adjacent to the advantageous gene (Braverman et al., 1995). On the other hand, diversifying selection is expected to reduce gene flow and inflate species differentiation for genes involved in traits that have differing adaptive significance in

the two species. Coupled with the effects of adaptive gene sharing, diversifying and adaptive selection are expected to create a mosaic genome, with some parts of the genome having very little divergence between species, while other parts may show strong interspecific differentiation, so called genomic 'speciation islands'. Such islands were reported in several animal (Duranton et al., 2018; Irwin et al., 2018; Hejase et al., 2020; Zhang et al., 2021) and plant (Renaut et al., 2013; Tavares et al., 2018; Papadopulos et al., 2019) species, including the high- and low-elevation *Senecio* species on Mount Etna, where the genes with high interspecific differentiation clustered around the regions with quantitative trait loci responsible for phenotypic differences between the species (Chapman et al., 2016). However, how much adaptive gene sharing occurs in this *Senecio* hybrid zone remains to be tested.

Evolution of Novel Compounds

Besides gaining adaptive advantages from mixing parental genomes, hybridisation can also drive the evolution of novel compounds that neither parent can produce, such as secondary metabolites in plants (Rieseberg et al., 1993; Orians, 2000). This is likely due to new selective pressures experienced by the putative hybrids, especially when they occupy novel habitats. Novel compounds can be synthesised by a number of mechanisms, including inhibition or re-direction of biochemical pathways, change in regulatory genes hence gene expression, and segregation of alleles (Orians, 2000). One example is the evolution of a novel methylated luteolin derivatives (flavonoids) in hybrids between *Salix viminalis* and *S. dasyclados*, which are involved in resistance against the lead beetle *Phratora vulgatissima* (Torp et al., 2013).

A novel pyrrolizidine alkaloid, florosenine, that is potentially involved in resistance against thrip species was also discovered in synthetic and natural hybrids between *S. jacobaea* and *S. aquaticus* (Kirk et al., 2010). Although florosenine has been found in other *Senecio* species in other areas (Mendez et al., 1990; Reina et al., 1993; Pelser et al., 2005), it has never been reported for the two species in the studied population and other European popoulations except for one *S. jacobaea* individual with trace amount, likely due to introgression (Kirk et al., 2004, 2010). This suggests the novelty of florosenine in *S. jacobaea* and *S. aquaticus*, although further confirmation is required (Kirk et al., 2010).

Gene Redundancy

Another potential for evolution lies within redundant genes in the duplicated genomes in auto- and allopolyploids. There are many outstanding questions regarding gene redundancy, such as how it varies among species and its relationship with genome architecture (Barghi et al., 2019). Many duplicated genes are inactivated due to accumulation of mutations (Parisod et al., 2009). They could also be eliminated in the hybrid genomes (e.g., in wheat: Shaked et al., 2001; in maize: Lai et al., 2004; in *Tragopogon miscellus*: Tate et al., 2006). A consequence of sequence elimination is divergence of homoeologous chromosomes preventing their meiotic pairing. Polyploid hybrid lineages have also been shown to purge redundant genomic regions that are far from adaptive optimum as they progress to behave in a more diploidised way (Wu et al., 2006), potentially allowing for better adaptation in novel habitats (Paun et al., 2007).

Nonetheless, there is some empirical evidence hinting on the role of redundant genes in adaptation and speciation. There are many possible reasons why these genes are not purged, for example due to gene balance (Freeling and Thomas, 2006; Birchler and Veitia, 2007) or dosage balance (Conant and Wolfe, 2008). In the early stages of possessing gene redundancy (such as soon after polyploidisation), the hybrid lineage also has a lower chance of creating homozygous recessive genotypes (Comai, 2005). Selection can act on the redundant genes that are not inactivated or purged to diversity gene function (Comai, 2005). They could either evolve new (neofunctionalisation) or complementary functions (subfunctionalisation; Lynch et al., 2001; Parisod et al., 2010). For example, gene redundancy has been suggested to be the basis of polygenic adaptation to new temperature regimes in *Drosophila simulans* (Barghi et al., 2019).

Similarly, in an experiment using *Senecio lautus* it was found that replicate populations of the same ecotype showed parallel evolution of similar phenotypes through different mixtures of adaptive alleles or different mutations in different genes that underlie the same biological functions (James et al., 2021). Most SNPs and genes studied in the divergence between the dune and headland ecotypes were not shared (non-parallel evolution); among all the candidate outlier SNPs, only five were shared across the whole system (James et al., 2021). These indicate that there is plenty of genetic redundancy underlying each biological function in the species (James et al., 2021).

CONCLUDING REMARKS

Hybridisation may not always allow for adaptation and speciation. There is a trade-off between the advantages and disadvantaging of combining divergent genomes. For example, hybrids could obtain the advantageous, higher environmental tolerance, while possessing intermediate traits between the parents that are disadvantageous for surviving in parental habitats (Shimizu-Inatsugi et al., 2017). The successful establishment of hybrids depends on a complex interplay of many evolutionary mechanisms, some of which were discussed in this article. The research in the genus *Senecio*, especially the work focused on the *S. aethnensis*—*S. chrysanthemifolius*—*S. squalidus* system, has significantly advanced our understanding of adaptation and speciation. In particular, the work in this system revealed some of the roles hybridisation could play in evolution, including transcriptome shock (e.g., up-regulation of genes linked to sulphur deficiency), genome reorganisation (e.g., between *S. aethnensis* and *S. chrysanthemifolius*, and inherited in *S. squalidus*), change in mating system and reproductive traits (e.g., self-compatibility and gain of ray florets in *S. cambrenisis* and *S. eboracensis*, hybrid species involving self-incompatible *S. squalidus* and self-compatible *S. vulgaris*), and adaptive introgression (e.g., gain in ray florets in *S. vulgaris* through hybridisation with *S. squalidus*). Other aspects, such as evolution of novel compounds, gene redundancy, and the extent of

adaptive allele sharing, have been explored in other *Senecio* species (e.g., novel florosenine in *S. jacobaea*, heterosis in *S. jacobaea x S. aquaticus* hybrids), but remain to be explored in the *S. aethnensis—S. chrysanthemifolius—S. squalidus* system. This showcases the research potential of *Senecio* as a whole to not only study the role of hybridisation in speciation and adaptation, but also other questions in evolutionary biology and ecology (reviewed in Walter et al., 2020). With the worldwide distribution of a vast number of species and ease of cultivation, *Senecio* offers great potential for evolutionary biologists to address outstanding questions regarding the role of hybridisation in adaptation and speciation. Specifically, how do hybridising species maintain their identity despite their gene pools being homogenised by hybridisation and interspecific gene flow? How do hybridising (sub)species diverge and evolve reproductive isolation? How strong and widespread in the genome diversifying selection should be to drive speciation of actively hybridising (sub)species? Under what conditions (e.g., large versus small populations) interspecific hybridisation plays more important in adaptation and speciation processes? The upcoming *S. squalidus* genome will help to address these questions using the *S. aethnensis—S. chrysanthemifolius—S. squalidus* system.

AUTHOR CONTRIBUTIONS

ELYW and DAF came up with the concept of this mini-review. ELYW wrote the initial draft and all authors contributed to editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by NERC (NE/P002145/1) projects to SJH and DAF as well as by a BBSRC grant (BB/P009808/1) to DAF.

REFERENCES

Abbott, R. J. (1992). Plant invasions, interspecific hybridization and the evolution of new plant taxa. *Trends Ecol. Evol.* 7, 401–405. doi: 10.1016/0169-5347(92)90020-C

Abbott, R. J., Ashton, P. A., and Forbes, D. G. (1992). Introgressive origin of the radiate groundsel, *Senecio vulgaris L. var. hibernicus* Syme: Aat-3 evidence. *Heredity* 68, 425–435. doi: 10.1038/hdy.1992.62

Abbott, R. J., and Brennan, A. C. (2014). Altitudinal gradients, plant hybrid zones and evolutionary novelty. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 369:20130346. doi: 10.1098/rstb.2013.0346

Abbott, R. J., Brennan, A. C., James, J. K., Forbes, D. G., Hegarty, M. J., and Hiscock, S. J. (2009). Recent hybrid origin and invasion of the British Isles by a self-incompatible species, Oxford ragwort (*Senecio squalidus* L., Asteraceae). *Biol. Invasions* 11, 1145–1158. doi: 10.1007/s10530-008-9382-3

Abbott, R. J., Hegarty, M. J., Hiscock, S. J., and Brennan, A. C. (2010). Homoploid hybrid speciation in action. *Taxon* 59, 1375–1386. doi: 10.1002/tax.595005

Abbott, R. J., and Irwin, J. A. (1988). Pollinator movements and the polymorphism for outcrossing rate at the ray floret locus in groundsel, *Senecio vulgaris* L. *Heredity* 60, 295–298. doi: 10.1038/hdy.1988.45

Abbott, R. J., and Rieseberg, L. H. (2012). "Hybrid speciation," in *Encyclopedia of Life Sciences (eLS)*. John Wiley & Sons, Ltd.

Adams, K. L., Cronn, R., Percifield, R., and Wendel, J. F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci.* 100, 4649–4654. doi: 10.1073/pnas.0630618100

Adams, K. L., and Wendel, J. F. (2005). Novel patterns of gene expression in polyploid plants. *Trends Genet.* 21, 539–543. doi: 10.1016/j.tig.2005.07.009

Aiuppa, A., Bellomo, S., Brusca, L., d'Alessandro, W., Di Paola, R., and Longo, M. (2006). Major-ion bulk deposition around an active volcano (Mt. Etna, Italy). *Bull. Volcanol.* 68, 255–265. doi: 10.1007/s00445-005-0005-x

Anderson, E. (1948). Hybridization of the habitat. *Evolution* 2, 1–9. doi: 10.1111/j.1558-5646.1948.tb02726.x

Anderson, E., and Stebbins, G. L. (1954). Hybridization as an evolutionary stimulus. *Evolution* 8, 378–388. doi: 10.1111/j.1558-5646.1954.tb01504.x

Arnold, M. L. (1997). *Natural Hybridization and Evolution*. Oxford, UK: Oxford University Press.

Arnold, M. L. (2006). *Evolution Through Genetic Exchange*. Oxford, UK: Oxford University Press.

Arnold, B. J., Lahner, B., DaCosta, J. M., Weisman, C. M., Hollister, J. D., Salt, D. E., et al. (2016). Borrowed alleles and convergence in serpentine adaptation. *Proc. Natl. Acad. Sci.* 113, 8320–8325. doi: 10.1073/pnas.1600405113

Baack, E. J., and Rieseberg, L. H. (2007). A genomic view of introgression and hybrid speciation. *Curr. Opin. Genet. Dev.* 17, 513–518. doi: 10.1016/j.gde.2007.09.001

Bapteste, E., Boucher, Y., Leigh, J., and Doolittle, W. F. (2004). Phylogenetic reconstruction and lateral gene transfer. *Trends Microbiol.* 12, 406–411. doi: 10.1016/j.tim.2004.07.002

Barghi, N., Tobler, R., Nolte, V., Jakšić, A. M., Mallard, F., Otte, K. A., et al. (2019). Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLoS Biol.* 17:e3000128. doi: 10.1371/journal.pbio.3000128

Barton, N. H. (2001). The role of hybridization in evolution. *Mol. Ecol.* 10, 551–568. doi: 10.1046/j.1365-294x.2001.01216.x

Becraft, E. D., and Moya, A. (2021). Searching for the boundaries of microbial speciation in a rapidly evolving world. *Front. Microbiol.* 12:808595. doi: 10.3389/fmicb.2021.808595

Birchler, J. A., and Veitia, R. A. (2007). The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* 19, 395–402. doi: 10.1105/tpc.106.049338

Bog, M., Bässler, C., and Oberprieler, C. (2017). Lost in the hybridisation vortex: high-elevation *Senecio hercynicus* (Compositae, Senecioneae) is genetically swamped by its congener *S. ovatus* in the Bavarian Forest National Park (SE Germany). *Evol. Ecol.* 31, 401–420. doi: 10.1007/s10682-017-9890-7

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H., and Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* 140, 783–796. doi: 10.1093/genetics/140.2.783

Brennan, A. C., Bridle, J. R., Wang, A. L., Hiscock, S. J., and Abbott, R. J. (2009). Adaptation and selection in the *Senecio* (Asteraceae) hybrid zone on Mount Etna, Sicily. *New Phytol.* 183, 702–717. doi: 10.1111/j.1469-8137.2009.02944.x

Brennan, A. C., and Hiscock, S. J. (2010). Expression and inheritance of sporophytic self-incompatibility in synthetic allohexaploid *Senecio cambrensis* (Asteraceae). *New Phytol.* 186, 251–261. doi: 10.1111/j.1469-8137.2009.03082.x

Brennan, A. C., Hiscock, S. J., and Abbott, R. J. (2014). Interspecific crossing and genetic mapping reveal intrinsic genomic incompatibility between two *Senecio* species that form a hybrid zone on Mount Etna, Sicily. *Heredity* 113, 195–204. doi: 10.1038/hdy.2014.14

Brennan, A. C., Hiscock, S. J., and Abbott, R. J. (2016). Genomic architecture of phenotypic divergence between two hybridizing plant species along an elevational gradient. *AoB Plants* 8:plw022. doi: 10.1093/aobpla/plw022

Brennan, A. C., Hiscock, S. J., and Abbott, R. J. (2019). Completing the hybridization triangle: the inheritance of genetic incompatibilities during homoploid hybrid speciation in ragworts (*Senecio*). *AoB Plants* 11:ply078. doi: 10.1093/aobpla/ply078

Brochmann, C., Brysting, A. K., Alsos, I. G., Borgen, L., Grundt, H. H., Scheen, A. C., et al. (2004). Polyploidy in arctic plants. *Biol. J. Linn. Soc.* 82, 521–536. doi: 10.1111/j.1095-8312.2004.00337.x

Brown, L., Scholefield, D., Jewkes, E. C., Preedy, N., Wadge, K., and Butler, M. (2000). The effect of Sulphur application on the efficiency of nitrogen use in two contrasting grassland soils. *J. Agric. Sci.* 135, 131–138. doi: 10.1017/S0021859699008072

Buerkle, C. A., Morris, R. J., Asmussen, M. A., and Rieseberg, L. H. (2000). The likelihood of homoploid hybrid speciation. *Heredity* 84, 441–451. doi: 10.1046/j.1365-2540.2000.00680.x

Burke, J. M., Lai, Z., Salmaso, M., Nakazato, T., Tang, S., Heesacker, A., et al. (2004). Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. *Genetics* 167, 449–457. doi: 10.1534/genetics.167.1.449

Bush, D. (2022). Long-term research reveals potential role of hybrids in climate-change adaptation. A commentary on 'expansion of the rare *Eucalyptus risdonii* under climate change through hybridisation with a closely related species despite hybrid inferiority'. *Ann. Bot.* 129, i–iii. doi: 10.1093/aob/mcab085

Campbell, C. R., Poelstra, J. W., and Yoder, A. D. (2018). What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biol. J. Linn. Soc.* 124, 561–583. doi: 10.1093/biolinnean/bly063

Chapman, M. A., Hiscock, S. J., and Filatov, D. A. (2013). Genomic divergence during speciation driven by adaptation to altitude. *Mol. Biol. Evol.* 30, 2553–2567. doi: 10.1093/molbev/mst168

Chapman, M. A., Hiscock, S. J., and Filatov, D. A. (2016). The genomic bases of morphological divergence and reproductive isolation driven by ecological speciation in *Senecio* (Asteraceae). *J. Evol. Biol.* 29, 98–113. doi: 10.1111/jeb.12765

Chen, Z. J., and Ni, Z. (2006). Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays* 28, 240–252. doi: 10.1002/bies.20374

Coghlan, A., Eichler, E. E., Oliver, S. G., Paterson, A. H., and Stein, L. (2005). Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends Genet.* 21, 673–682. doi: 10.1016/j.tig.2005.09.009

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711

Comai, L., Tyagi, A. P., Winter, K., Holmes-Davis, R., Reynolds, S. H., Stevens, Y., et al. (2000). Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* 12, 1551–1567. doi: 10.1105/tpc.12.9.1551

Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950. doi: 10.1038/nrg2482

Counterman, B. A., Araujo-Perez, F., Hines, H. M., Baxter, S. W., Morrison, C. M., Lindstrom, D. P., et al. (2010). Genomic hotspots for adaptation: the population genetics of Müllerian mimicry in *Heliconius erato*. *PLoS Genet.* 6:e1000796. doi: 10.1371/journal.pgen.1000796

Coyne, J. A., and Orr, H. A. (1989). Patterns of speciation in *Drosophila*. *Evolution* 43, 362–381. doi: 10.1111/j.1558-5646.1989.tb04233.x

Coyne, J. A., and Orr, H. A. (2004). *Speciation* (Vol. *37*). Sunderland, MA: Sinauer Associates.

Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., and Gagnaire, P. A. (2018). The origin and remolding of genomic islands of differentiation in the European sea bass. *Nat. Commun.* 9:2518. doi: 10.1038/s41467-018-04963-6

Ebersbach, J., Tkach, N., Röser, M., and Favre, A. (2020). The role of hybridisation in the making of the species-rich arctic-alpine genus *Saxifraga* (Saxifragaceae). *Diversity* 12:440. doi: 10.3390/d12110440

Entani, T., Takayama, S., Iwano, M., Shiba, H., Che, F. S., and Isogai, A. (1999). Relationship between polyploidy and pollen self-incompatibility phenotype in *Petunia hybrida* Vilm. *Biosci. Biotechnol. Biochem.* 63, 1882–1888.

Freeling, M., and Thomas, B. C. (2006). Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res.* 16, 805–814. doi: 10.1101/gr.3681406

Garcês, H. M. P., Spencer, V. M., and Kim, M. (2016). Control of floret symmetry by RAY3, SvDIV1B, and SvRAD in the capitulum of *Senecio vulgaris*. *Plant Physiol.* 171, 2055–2068. doi: 10.1104/pp.16.00395

Gilbert, F.S. (1986). *Hoverflies. Naturalist's Handbooks 5*. Cambridge Press: Cambridge.

Grant, P. R. (1972). Convergent and divergent character displacement. *Biol. J. Linn. Soc.* 4, 39–68. doi: 10.1111/j.1095-8312.1972.tb00690.x

Grant, V. (1981). *Plant Speciation*. New York, US: Columbia University Press.

Gross, B. L. (2012). Genetic and phenotypic divergence of homoploid hybrid species from parental species. *Heredity* 108:157. doi: 10.1038/hdy.2011.80

Gross, B. L., and Rieseberg, L. (2005). The ecological genetics of homoploid hybrid speciation. *J. Hered.* 96, 241–252. doi: 10.1093/jhered/esi026

Gross, B. L., Schwarzbach, A. E., and Rieseberg, L. H. (2003). Origin(s) of the diploid hybrid species *Helianthus deserticola* (Asteraceae). *Am. J. Bot.* 90, 1708–1719. doi: 10.3732/ajb.90.12.1708

Harada, E., Yamaguchi, Y., Koizumi, N., and Hiroshi, S. (2002). Cadmium stress induces production of thiol compounds and transcripts for enzymes involved in sulfur assimilation pathways in *Arabidopsis*. *J. Plant Physiol.* 159, 445–448. doi: 10.1078/0176-1617-00733

Harrison, R. G. (1993). *Hybrids and Hybrid Zones: Historical Perspective. Hybrid Zones and the Evolutionary Process*. New York, NY: Oxford University Press. 3–12.

Hegarty, M. J., Abbott, R. J., and Hiscock, S. J. (2012). "Allopolyploid speciation in action: the origins and evolution of *Senecio cambrensis*," in *Polyploidy and Genome Evolution*. eds. P. S. Soltis and D. E. Soltis (Berlin, Heidelberg: Springer), 245–270.

Hegarty, M. J., Barker, G. L., Brennan, A. C., Edwards, K. J., Abbott, R. J., and Hiscock, S. J. (2009). Extreme changes to gene expression associated with homoploid hybrid speciation. *Mol. Ecol.* 18, 877–889. doi: 10.1111/j.1365-294X.2008.04054.x

Hegarty, M. J., Barker, G. L., Wilson, I. D., Abbott, R. J., Edwards, K. J., and Hiscock, S. J. (2006). Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Curr. Biol.* 16, 1652–1659. doi: 10.1016/j.cub.2006.06.071

Hegarty, M. J., Batstone, T. O. M., Barker, G. L., Edwards, K. J., Abbott, R. J., and Hiscock, S. J. (2011). Nonadditive changes to cytosine methylation as a consequence of hybridization and genome duplication in *Senecio* (Asteraceae). *Mol. Ecol.* 20, 105–113. doi: 10.1111/j.1365-294X.2010.04926.x

Hegarty, M. J., and Hiscock, S. J. (2005). Hybrid speciation in plants: new insights from molecular studies. *New Phytol.* 165, 411–423. doi: 10.1111/j.1469-8137.2004.01253.x

Hejase, H. A., Salman-Minkov, A., Campagna, L., Hubisz, M. J., Lovette, I. J., Gronau, I., et al. (2020). Genomic islands of differentiation in a rapid avian radiation have been driven by recent selective sweeps. *Proc. Natl. Acad. Sci.* 117, 30554–30565. doi: 10.1073/pnas.2015987117

Hermansen, J. S., Sæther, S. A., Elgvin, T. O., Borge, T., Hjelle, E., and Sætre, G. P. (2011). Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. *Mol. Ecol.* 20, 3812–3822. doi: 10.1111/j.1365-294X.2011.05183.x

Hobbs, J. P. A., Richards, Z. T., Popovic, I., Lei, C., Staeudle, T. M., Montanari, S. R., et al. (2021). Hybridisation and the evolution of coral reef biodiversity. *Coral Reefs* 41, 535–549. doi: 10.1007/s00338-021-02193-9

Irwin, D. E., Milá, B., Toews, D. P., Brelsford, A., Kenyon, H. L., Porter, A. N., et al. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. *Mol. Ecol.* 27, 4839–4855. doi: 10.1111/mec.14858

James, J. K., and Abbott, R. J. (2005). Recent, allopatric, homoploid hybrid speciation: the origin of *Senecio squalidus* (Asteraceae) in the British Isles from a hybrid zone on Mount Etna, Sicily. *Evolution* 59, 2533–2547. doi: 10.1111/j.0014-3820.2005.tb00967.x

James, M. E., Wilkinson, M. J., Bernal, D. M., Liu, H., North, H. L., Engelstädter, J., et al. (2021). Phenotypic and genotypic parallel evolution in parapatric ecotypes of *Senecio*. *Evolution* 75, 3115–3131. doi: 10.1111/evo.14387

Janko, K., Kotusz, J., De Gelas, K., Šlechtová, V., Opoldusova, Z., Drozd, P., et al. (2012). Dynamic formation of asexual diploid and polyploid lineages: multilocus analysis of Cobitis reveals the mechanisms maintaining the diversity of clones. *PLoS One* 7:e45384. doi: 10.1371/journal.pone.0045384

Kapralov, M. V., and Filatov, D. A. (2006). Molecular adaptation during adaptive radiation in the Hawaiian endemic genus Schiedea. *PLoS One* 1:e8. doi: 10.1371/journal.pone.0000008

Karrenberg, S., Lexer, C., and Rieseberg, L. H. (2007). Reconstructing the history of selection during homoploid hybrid speciation. *Am. Nat.* 169, 725–737. doi: 10.1086/516758

Kim, M., Cui, M. L., Cubas, P., Gillies, A., Lee, K., Chapman, M. A., et al. (2008). Regulatory genes control a key morphological and ecological trait

transferred between species. *Science* 322, 1116–1119. doi: 10.1126/science.1164371

Kirk, H., Máčel, M., Klinkhamer, P. G., and Vrieling, K. (2004). Natural hybridization between *Senecio jacobaea* and *Senecio aquaticus*: molecular and chemical evidence. *Mol. Ecol.* 13, 2267–2274. doi: 10.1111/j.1365-294X.2004.02235.x

Kirk, H., Vrieling, K., and Klinkhamer, P. G. (2005). Maternal effects and heterosis influence the fitness of plant hybrids. *New Phytol.* 166, 685–694. doi: 10.1111/j.1469-8137.2005.01370.x

Kirk, H., Vrieling, K., Van Der Meijden, E., and Klinkhamer, P. G. (2010). Species by environment interactions affect pyrrolizidine alkaloid expression in *Senecio jacobaea*, *Senecio aquaticus*, and their hybrids. *J. Chem. Ecol.* 36, 378–387. doi: 10.1007/s10886-010-9772-8

Lai, J., Ma, J., Swigoňová, Z., Ramakrishna, W., Linton, E., Llaca, V., et al. (2004). Gene loss and movement in the maize genome. *Genome Res.* 14, 1924–1931. doi: 10.1101/gr.2701104

Lai, Z., Nakazato, T., Salmaso, M., Burke, J. M., Tang, S., Knapp, S. J., et al. (2005). Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics* 171, 291–303. doi: 10.1534/genetics.105.042242

Lee, H. S., and Chen, Z. J. (2001). Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. *Proc. Natl. Acad. Sci.* 98, 6753–6758. doi: 10.1073/pnas.121064698

Levy, A. A., and Feldman, M. (2004). Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. *Biol. J. Linn. Soc.* 82, 607–613. doi: 10.1111/j.1095-8312.2004.00346.x

Lexer, C., and Widmer, A. (2008). The genic view of plant speciation: recent progress and emerging questions. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 363, 3023–3036. doi: 10.1098/rstb.2008.0078

Lim, K. Y., Matyasek, R., Kovarik, A., and Leitch, A. R. (2004). Genome evolution in allotetraploid *Nicotiana*. *Biol. J. Linn. Soc.* 82, 599–606. doi: 10.1111/j.1095-8312.2004.00344.x

Lowe, A. J., and Abbott, R. J. (2004). Reproductive isolation of a new hybrid species, *Senecio eboracensis* Abbott & Lowe (Asteraceae). *Heredity* 92, 386–395. doi: 10.1038/sj.hdy.6800432

Lukhtanov, V. A., Shapoval, N. A., Anokhin, B. A., Saifitdinova, A. F., and Kuznetsova, V. G. (2015). Homoploid hybrid speciation and genome evolution via chromosome sorting. *Proc. R. Soc. B Biol. Sci.* 282:20150157. doi: 10.1098/rspb.2015.0157

Lynch, M., O'Hely, M., Walsh, B., and Force, A. (2001). The probability of preservation of a newly arisen gene duplicate. *Genetics* 159, 1789–1804. doi: 10.1093/genetics/159.4.1789

Madlung, A., Tyagi, A. P., Watson, B., Jiang, H., Kagochi, T., Doerge, R. W., et al. (2005). Genomic changes in synthetic *Arabidopsis* polyploids. *Plant J.* 41, 221–230. doi: 10.1111/j.1365-313X.2004.02297.x

Mallet, J. (2001). "Concepts of species," in *Encyclopedia of Biodiversity*. *Vol*. 5. ed. S. A. Levin (Massachusetts, US: Academic Press), 427–440.

Marshall, D. F., and Abbott, R. J. (1982). Polymorphism for outcrossing frequency at the ray floret locus in *Senecio vulgaris* LI evidence. *Heredity* 48, 227–235. doi: 10.1038/hdy.1982.28

Marshall, D. F., and Abbott, R. J. (1984). Polymorphism for outcrossing frequency at the ray floret locus in *Senecio vulgaris* L. III. Causes. *Heredity* 53, 145–149. doi: 10.1038/hdy.1984.70

Martínez-Fernández, M., Bernatchez, L., Rolán-Alvarez, E., and Quesada, H. (2010). Insights into the role of differential gene expression on the ecological adaptation of the snail *Littorina saxatilis*. *BMC Evol. Biol.* 10:356. doi: 10.1186/1471-2148-10-356

Mayr, E. (1963). *Animal Species and Evolution (No. 591.38)*. Massachusetts, US: Harvard University Press.

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., and Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 8:14363. doi: 10.1038/ncomms14363

Mendez, M. D. C., Riet-Correa, F., Schild, A. L., and Martz, W. (1990). Experimental poisoning of cattle and chicks by five *Senecio* species. *Pesqui. Vet. Bras.* 10, 63–69.

Miller, J. S., and Venable, D. L. (2000). Polyploidy and the evolution of gender dimorphism in plants. *Science* 289, 2335–2338. doi: 10.1126/science.289.5488.2335

Morjan, C. L., and Rieseberg, L. H. (2004). How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Mol. Ecol.* 13, 1341–1356. doi: 10.1111/j.1365-294X.2004.02164.x

Muir, G., Osborne, O. G., Sarasa, J., Hiscock, S. J., and Filatov, D. A. (2013). Recent ecological selection on regulatory divergence is shaping clinal variation in *Senecio* on Mount Etna. *Evolution* 67, 3032–3042. doi: 10.1111/evo.12157

Nasrallah, M. E., Yogeeswaran, K., Snyder, S., and Nasrallah, J. B. (2000). *Arabidopsis* species hybrids in the study of species differences and evolution of amphiploidy in plants. *Plant Physiol.* 124, 1605–1614. doi: 10.1104/pp.124.4.1605

Nevado, B., Harris, S. A., Beaumont, M. A., and Hiscock, S. J. (2020). Rapid homoploid hybrid speciation in British gardens: the origin of Oxford ragwort (*Senecio squalidus*). *Mol. Ecol.* 29, 4221–4233. doi: 10.1111/mec.15630

Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304. doi: 10.1038/35012500

Orians, C. M. (2000). The effects of hybridization in plants on secondary chemistry: implications for the ecology and evolution of plant–herbivore interactions. *Am. J. Bot.* 87, 1749–1756. doi: 10.2307/2656824

Osborne, O. G., Batstone, T. E., Hiscock, S. J., and Filatov, D. A. (2013). Rapid speciation with gene flow following the formation of Mt. Etna. *Genome Biol. Evol.* 5, 1704–1715. doi: 10.1093/gbe/evt127

Otto, S. P., and Whitton, J. (2000). Polyploid incidence and evolution. *Annu. Rev. Genet.* 34, 401–437. doi: 10.1146/annurev.genet.34.1.401

Papadopulos, A. S., Baker, W. J., Crayn, D., Butlin, R. K., Kynast, R. G., Hutton, I., et al. (2011). Speciation with gene flow on Lord Howe Island. *Proc. Natl. Acad. Sci.* 108, 13188–13193. doi: 10.1073/pnas.1106085108

Papadopulos, A. S., Igea, J., Dunning, L. T., Osborne, O. G., Quan, X., Pellicer, J., et al. (2019). Ecological speciation in sympatric palms: 3. Genetic map reveals genomic islands underlying species divergence in Howea. *Evolution* 73, 1986–1995. doi: 10.1111/evo.13796

Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., et al. (2012). Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genet.* 8:e1002752. doi: 10.1371/journal.pgen.1002752

Parisod, C., Salmon, A., Zerjal, T., Tenaillon, M., Grandbastien, M. A., Ainouche, M. (2009). Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytolog.* 184, 1003–1015. doi: 10.1111/j.1469-8137.2009.03029.x

Parisod, C., Holderegger, R., and Brochmann, C. (2010). Evolutionary consequences of autopolyploidy. *New Phytol.* 186, 5–17. doi: 10.1111/j.1469-8137.2009.03142.x

Paun, O., Fay, M. F., Soltis, D. E., and Chase, M. W. (2007). Genetic and epigenetic alterations after hybridization and genome doubling. *Taxon* 56, 649–656. doi: 10.2307/25065849

Pelser, P. B., de Vos, H., Theuring, C., Beuerle, T., Vrieling, K., and Hartmann, T. (2005). Frequent gain and loss of pyrrolizidine alkaloids in the evolution of *Senecio* section Jacobaea (Asteraceae). *Phytochemistry* 66, 1285–1295. doi: 10.1016/j.phytochem.2005.04.015

Planet, P. J., Kachlany, S. C., Fine, D. H., DeSalle, R., and Figurski, D. H. (2003). The widespread colonization island of *Actinobacillus actinomycetemcomitans*. *Nat. Genet.* 34, 193–198. doi: 10.1038/ng1154

Pontes, O., Neves, N., Silva, M., Lewis, M. S., Madlung, A., Comai, L., et al. (2004). Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid *Arabidopsis suecica* genome. *Proc. Natl. Acad. Sci.* 101, 18240–18245. doi: 10.1073/pnas.0407258102

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., et al. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J. Evol. Biol.* 30, 1450–1477. doi: 10.1111/jeb.13047

Reina, M., Delafuente, G., Villarroel, L., and Torres, R. (1993). Pyrrolizidine alkaloids from *Senecio erraticus*, *S. glaber*, and *S. microphyllus*. *An. Quím.* 89, 387–390.

Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Kane, N. C., et al. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nat. Commun.* 4:1827. doi: 10.1038/ncomms2833

Richards, A. J. (1986). *Plant Breeding Systems*. London, UK: George Allen & Unwin.

Richards, E. J., and Martin, C. H. (2017). Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PLoS Genet.* 13:e1006919. doi: 10.1371/journal.pgen.1006919

Rieseberg, L. H. (1997). Hybrid origins of plant species. *Annu. Rev. Ecol. Syst.* 28, 359–389. doi: 10.1146/annurev.ecolsys.28.1.359

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* 16, 351–358. doi: 10.1016/S0169-5347(01)02187-5

Rieseberg, L. H., Church, S. A., and Morjan, C. L. (2004). Integration of populations and differentiation of species. *New Phytol.* 161, 59–69. doi: 10.1046/j.1469-8137.2003.00933.x

Rieseberg, L. H., Ellstrand, N. C., and Arnold, M. (1993). What can molecular and morphological markers tell us about plant hybridization? *Crit. Rev. Plant Sci.* 12, 213–241.

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., et al. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301, 1211–1216. doi: 10.1126/science.1086949

Rieseberg, L. H., Whitton, J., and Gardner, K. (1999). Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152, 713–727. doi: 10.1093/genetics/152.2.713

Rieseberg, L. H., and Willis, J. H. (2007). Plant speciation. *Science* 317, 910–914. doi: 10.1126/science.1137729

Ross, R. I. C. (2010). Local adaptation and adaptive divergence in a hybrid species complex in *Senecio*. (Doctoral Dissertation, University of Oxford).

Royal Botanic Gardens Kew (2022). *Senecio* L. Available at: https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:325904-2#sources (Accessed February, 2022).

Schwarzbach, A. E., and Rieseberg, L. H. (2002). Likely multiple origins of a diploid hybrid sunflower species. *Mol. Ecol.* 11, 1703–1715. doi: 10.1046/j.1365-294X.2002.01557.x

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends Ecol. Evol.* 19, 198–207. doi: 10.1016/j.tree.2004.01.003

Seehausen, O. (2013). Conditions when hybridization might predispose populations for adaptive radiation. *J. Evol. Biol.* 26, 279–281. doi: 10.1111/jeb.12026

Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., and Levy, A. A. (2001). Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* 13, 1749–1759. doi: 10.1105/TPC.010083

Shaw, K. L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proc. Natl. Acad. Sci.* 99, 16122–16127. doi: 10.1073/pnas.242585899

Shimizu-Inatsugi, R., Terada, A., Hirose, K., Kudoh, H., Sese, J., and Shimizu, K. K. (2017). Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. *Mol. Ecol.* 26, 193–207. doi: 10.1111/mec.13738

Slatkin, M. (1976). "The rate of spread of an advantageous allele in a subdivided population," in *Population Genetics and Ecology*. eds. S. Karlin and E. Nevo (New York, US: Academic Press), 767–780.

Slatkin, M., and Wiehe, T. (1998). Genetic hitch-hiking in a subdivided population. *Genet. Res.* 71, 155–160. doi: 10.1017/S001667239800319X

Soltis, D. E., and Soltis, P. S. (1999). Polyploidy: recurrent formation and genome evolution. *Trends Ecol. Evol.* 14, 348–352. doi: 10.1016/S0169-5347(99)01638-9

Tang, S., Okashah, R. A., Knapp, S. J., Arnold, M. L., and Martin, N. H. (2010). Transmission ratio distortion results in asymmetric introgression in Louisiana *Iris*. *BMC Plant Biol.* 10:48. doi: 10.1186/1471-2229-10-48

Tate, J. A., Ni, Z., Scheen, A. C., Koh, J., Gilbert, C. A., Lefkowitz, D., et al. (2006). Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* 173, 1599–1611. doi: 10.1534/genetics.106.057646

Tavares, H., Whibley, A., Field, D. L., Bradley, D., Couchman, M., Copsey, L., et al. (2018). Selection and gene flow shape genomic islands that control floral guides. *Proc. Natl. Acad. Sci.* 115, 11006–11011. doi: 10.1073/pnas.1801832115

Taylor, S. J., Rojas, L. D., Ho, S. W., and Martin, N. H. (2013). Genomic collinearity and the genetic architecture of floral differences between the homoploid hybrid species *Iris nelsonii* and one of its progenitors, *Iris hexagona*. *Heredity* 110, 63–70. doi: 10.1038/hdy.2012.62

Torp, M., Lehrman, A., Stenberg, J. A., Julkunen-Tiitto, R., and Björkman, C. (2013). Performance of an herbivorous leaf beetle (*Phratora vulgatissima*) on Salix F2 hybrids: the importance of phenolics. *J. Chem. Ecol.* 39, 516–524. doi: 10.1007/s10886-013-0266-3

Tsitrone, A., Kirkpatrick, M., and Levin, D. A. (2003). A model for chloroplast capture. *Evolution* 57, 1776–1782. doi: 10.1111/j.0014-3820.2003.tb00585.x

Vallejo-Marín, M. (2012). *Mimulus peregrinus* (Phrymaceae): a new British allopolyploid species. *PhytoKeys* 14:1. doi: 10.3897/phytokeys.14.3305

Vallejo-Marín, M., Buggs, R. J., Cooley, A. M., and Puzey, J. R. (2015). Speciation by genome duplication: repeated origins and genomic composition of the recently formed allopolyploid species *Mimulus peregrinus*. *Evolution* 69, 1487–1500. doi: 10.1111/evo.12678

Vallejo-Marín, M., and Hiscock, S. J. (2016). Hybridization and hybrid speciation under global change. *New Phytol.* 211, 1170–1187. doi: 10.1111/nph.14004

Walter, G. M., Abbott, R. J., Brennan, A. C., Bridle, J. R., Chapman, M., Clark, J., et al. (2020). *Senecio* as a model system for integrating studies of genotype, phenotype and fitness. *New Phytol.* 226, 326–344. doi: 10.1111/nph.16434

Wang, J., Tian, L., Lee, H. S., Wei, N. E., Jiang, H., Watson, B., et al. (2006). Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172, 507–517. doi: 10.1534/genetics.105.047894

Welch, M. E., and Rieseberg, L. H. (2002). Patterns of genetic variation suggest a single, ancient origin for the diploid hybrid species *Helianthus paradoxus*. *Evolution* 56, 2126–2137. doi: 10.1111/j.0014-3820.2002.tb00138.x

Wong, E. L., Nevado, B., Osborne, O. G., Papadopulos, A. S., Bridle, J. R., Hiscock, S. J., et al. (2020). Strong divergent selection at multiple loci in two closely related species of ragworts adapted to high and low elevations on Mount Etna. *Mol. Ecol.* 29, 394–412. doi: 10.1111/mec.15319

Wu, J., Hettenhausen, C., and Baldwin, I. T. (2006). Evolution of proteinase inhibitor defenses in North American allopolyploid species of *Nicotiana*. *Planta* 224, 750–760. doi: 10.1007/s00425-006-0256-6

Xiang, C., and Oliver, D. J. (1998). Glutathione metabolic genes coordinately respond to heavy metals and jasmonic acid in *Arabidopsis*. *Plant Cell* 10, 1539–1550. doi: 10.1105/tpc.10.9.1539

Zhang, Y., Teng, D., Lu, W., Liu, M., Zeng, H., Cao, L., et al. (2021). A widely diverged locus involved in locomotor adaptation in *Heliconius* butterflies. *Sci. Adv.* 7:eabh2340. doi: 10.1126/sciadv.abj9066

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 29 June 2022
doi: 10.3389/fpls.2022.914671

Identification of a Male Sterile Candidate Gene in *Lilium* x *formolongi* and Transfer of the Gene to Easter Lily (*L. longiflorum*) via Hybridization

Takahiro Moriyama[1†], *Daniel John Shea*[1†], *Naoto Yokoi*[2], *Seiro Imakiire*[3], *Takaaki Saito*[2], *Hikaru Ohshima*[1], *Hina Saito*[1], *Satoru Okamoto*[1], *Eigo Fukai*[1] *and Keiichi Okazaki*[1*]

[1] *Laboratory Plant Breeding, Graduate School of Science and Technology, Niigata University, Niigata, Japan,* [2] *Akita Prefectural Agriculture, Forestry and Fisheries Research Center, Agriculture Experimental Station, Akita, Japan,* [3] *Fruit Tree and Flower Division, Kagoshima Prefectural Institute for Agricultural Development, Kagoshima, Japan*

Pollen-free varieties are advantageous in promoting cut-flower production. In this study, we identified a candidate mutation which is responsible for pollen sterility in a strain of *Lilium* × *formolongi,* which was originally identified as a naturally occurred male-sterile plant in a seedling population. The pollen sterility occurred due to the degradation of pollen mother cells (PMCs) before meiotic cell division. Genetic analysis suggested that the male-sterile phenotype is attributed to one recessive locus. Transcriptome comparison between anthers of sterile and fertile plants in a segregated population identified a transcript that was expressed only in pollen-fertile plants, which is homologous to *TDF1* (*DEFECTIVE in TAPETAL DEVELOPMENT and FUNCTION1*) in Arabidopsis, a gene encoding a transcription factor AtMYB35 that is known as a key regulator of pollen development. Since *tdf1* mutant shows male sterility, we assumed that the absence transcript of the *TDF1*-like gene, named as *LflTDF1,* is the reason for pollen sterility observed in the mutant. A 30 kbp-long nanopore sequence read containing *LflTDF1* was obtained from a pollen-fertile accession. PCR analyses using primers designed from the sequence suggested that at least a 30kbp-long region containing *LflTDF1* was deleted or replaced by unknown sequence in the pollen-sterile mutant. Since the cross between *L.* × *formolongi* and Easter lily (*L. longiflorum*) is compatible, we successfully introgressed the male-sterile allele, designated as *lfltdf1*, to Easter lily. To our knowledge, this is the first report of molecular identification of a pollen-sterile candidate gene in lily. The identification and marker development of *LflTDF1* gene will assist pollen-free lily breeding of Easter lilies and other lilies.

Keywords: Easter lily, hybridization, male sterility, marker assisted selection, RNA-seq

OPEN ACCESS

Edited by:
Andrew H. Paterson,
University of Georgia, United States

Reviewed by:
Tomohiko Kazama,
Kyushu University, Japan
Kinya Toriyama,
Tohoku University, Japan

***Correspondence:**
Keiichi Okazaki
okazaki@agr.niigata-u.ac.jp

[†] These authors have contributed equally to this work

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

Received: *07 April 2022*
Accepted: *18 May 2022*
Published: *29 June 2022*

Citation:
Moriyama T, Shea DJ, Yokoi N, Imakiire S, Saito T, Ohshima H, Saito H, Okamoto S, Fukai E and Okazaki K (2022) Identification of a Male Sterile Candidate Gene in Lilium x formolongi and Transfer of the Gene to Easter Lily (L. longiflorum) via Hybridization. Front. Plant Sci. 13:914671. doi: 10.3389/fpls.2022.914671

INTRODUCTION

The genus *Lilium* of the Liliaceae family includes more than 100 species among which several species are important ornamental plants (Nishikawa et al., 1999, 2001; Marasek-Ciolakowska et al., 2018). *L. longiflorum* is distributed in the Amami Islands, the Ryukyu Archipelago in Japan, and in the main island in Taiwan (Hiramatsu et al., 2001; Sakazono et al., 2009). The intraspecies

crossbreeding program of this species has developed many cultivars with trumpet-shaped pure white flowers, and consequently, those cultivars became an indispensable flower for ceremonial occasions, so called Trumpet Lily or Easter Lily. *L.* × *formolongi* has been developed by the interspecific hybridization between the *L. formosanum* and *L. longiflorum* followed by continuous backcrossing with *L. longiflorum*. As a result, the progeny successfully combined broad leaves like *L. longiflorum* and the characteristics of *L. formosanum* blooming within a year after sowing the seeds (Shimizu, 1987). Commercial cultivars of *L.* × *formolongi* and *L. longiflorum* are propagated by seeds or bulbs, respectively. For the year-round cut-flower production, *L.* × *formolongi* is employed in July to October and *L. longiflorum* in November to June in Japan.

Male sterility (pollen-free) benefits the ornamental crops (Smith et al., 2004; García-Sogo et al., 2010; Roque et al., 2019). Pollen is problematic in the case of flower arrangement since it stains perianths, clothes, etc. Therefore, florists have to manually remove the anthers immediately after flowering. Male-sterile flowers, on the contrary, do not cause seed formation by self-pollination, saving energy and resources for fruit production. This will extend the life of the flower. In addition, pollen sterility helps eliminate pollen allergens and avoid gene transfer from genetically modified crops into the ecosystem. Thus, pollen-free varieties are advantageous in promoting ornamental flower production. Recently, several pollen-free lilies are distributed into the commercial market, but as far as we know, pollen-free Easter lily cultivars are not available.

In plants, the pollen formation process is controlled by many genes, suggesting that the process can be easily impaired by a small number of mutations. Indeed, many mutants which have defects in microsporogenesis have been identified. Previous studies using those mutants in Arabidopsis revealed a gene network involving five transcription factors (TFs), DYSFUNCTIONAL TAPETUM1 (DYT1) (Zhang et al., 2006) and ABORTED MICROSPORES (AMS) (Sorensen et al., 2003), DEFECTIVE in TAPETAL DEVELOPMENT and FUNCTION1 (TDF1) (Zhu et al., 2008), MS188 (Zhang et al., 2007), and MALE STERILITY1 (MS1) (Ito et al., 2007). The interactive order in the regulatory network, DYT1–TDF1–AMS–MS188–MS1, is likely to be conserved in different species (Zhu et al., 2008; Cai et al., 2015; Gómez et al., 2015). Thus far, however, functional analysis of TDF1 has been limited to the three plant species including Arabidopsis, as follows. Cai et al. (2015) demonstrated that *OsTDF1* in rice compensates the pollen sterility when transformed into Arabidopsis *tdf1* mutant, suggesting that they are functionally orthologous. Recent studies reported that a deletion of the *TDF1* homolog (*AoMYB35*) in asparagus causes female flower due to defects in the anther development (Tsugama et al., 2017; Harkess et al., 2020).

The mutations of the TFs related to pollen development are classified into genic male sterility (GMS) and recessively inherited in most cases. Gametophytic GMS provides 50% male sterility by crossing a male-sterile line (rr, recessive homozygous male-sterile genotype) with a heterozygous individual (Rr). In many ornamental plants, since they can be clonally propagated, GMS is a useful source of male sterility. However, since the inheritance of male-sterile lilies has not been analyzed, it is difficult to effectively breed pollen-free varieties based on the inheritance manner. In this study, we report the identification of a male-sterile mutant and the responsible gene candidate in a pollen-free *L.* × *formolongi* plant AR01, later registered as a cultivar "Akita Kiyohime," that naturally occurred in a seedling population.

MATERIALS AND METHODS

Plant Materials

Several F_1 plants were made from the cross between *L.* × *formolongi* cv. Hatsuki and cv. Raizan 2go (**Supplementary Figure 1**). *L.* × *formolongi* is self-compatible, but after the sib cross of F_1 plants (22-1 and 23-1), a pollen-sterile plant was selected in the next generation and named PL01. PL01 was propagated by bulbs and repeatedly used as the founder of the subsequent pollen-sterile lines. PL01 was crossed with the selected pollen-fertile sib lines (18-3, 42-3, and 97-1) to obtain the segregating populations of pollen-fertile and sterile segregants. Each individual in the segregating population was cloned by bulblet propagation and thereafter used for RNA seq, genetic analysis, and anatomical observation. In the subsequent progeny, a pollen-sterile plant, AR01, was selected (**Figure 1**) and later registered as a cultivar "Akita Kiyohime." F_1 plants were made from the cross of AR01 and *L. longiflorum* cultivars (Cristal Horn, White Fox, and Pure Horn) and sibling cross-population made between F_1 plants were used for DNA marker validation.

Pollen-sterile lines #72 and #318 developed by gamma-irradiated breeding were also used for paraffin section observation to compare pollen degradation processes between different pollen-sterile genotypes. In addition, sequence analysis of *Lilium TDF1* orthologs was conducted using a single plant of the following accessions, a pollen-fertile segregant of *L.* × *formolongi*, *L. longiflorum* cv. Hinomoto, Oriental hybrid lily cv. Siberia (*Lilium* spp.), and *L. formosanum* collected in a habitat of Niigata prefecture.

Genetic Analysis

PL1607 population was made by crossing PL01 as female with pollen-fertile line 97-1 (inferred to be heterozygous at the locus responsible for the pollen phenotype) (**Supplementary Figure 1**). PL1620 population was made by the cross between the pollen-fertile lines, 83-2 × 220-1, where both lines were inferred to be heterozygous at the locus responsible for the pollen phenotype. The obtained seeds were planted in a cell tray containing regular soil, subjected to low-temperature treatment at 4°C for 2 weeks, and then raised in a greenhouse. After the seedlings grew at the 3–4 leaf stages, they were transplanted to a field at the end of April and the phenotypic separation of the presence or absence of pollen at the time of flowering was examined.

Anatomical Observation of Anthers

Flower buds (13 to 26 mm in length) were collected from pollen-sterile and pollen-fertile segregants in the segregating population, 26–47. Microsporogenesis of sterile lines, #72 and #318, developed by the gamma irradiation breeding, was also

FIGURE 1 | Appearance of pollen-fertile and sterile *L.* × *formolongi* plants. **(A)** Flower buds of a pollen-fertile plant (left) and the pollen-sterile AR01 plant (right). **(B)** An opened flower of AR01. **(C)** Close up of sterile anthers of AR01. Bar = 1 cm.

collected. The collected buds containing all six anthers were immersed in Carnoy's solution (ethanol: acetic acid = 3: 1) at room temperature for 12 h. Then, the samples were dehydrated using alcohol series. They were stored in chloroform and embedded in paraffin. Transverse sections were made at a thickness of ∼12 μm and deparaffinized with Histo-clear and ethanol of gradient concentration. The tissues were stained with toluidine blue (0.05 %) and dehydrated in ethanol series (50–100 %) and Histo-clear. Finally, the sections were mounted in Canada balsam and observed by light microscopy (BX-60; SZX7, Olympus, Tokyo, Japan) and photographed using a CCD camera (DP60, Olympus, Tokyo, Japan).

Transcriptome and GO-Enrichment Analysis

For RNA sequencing, anthers were collected from pollen-sterile and pollen-fertile segregants of the segregating population, 26–47, where the flower buds of lengths 11, 12, and 14 mm were collected from three different plants of pollen-sterile and pollen-fertile segregants, respectively. The anthers were then excised from the three different-sized buds and mixed separately with pollen-sterile and pollen-fertile bulk sampling. The anthers were instantly frozen in liquid nitrogen and stored in the deep freezer till use. The samples were crushed with a pestle, and RNA was extracted using RNeasy Plant Mini Kit (QIAGEN) according to the protocol. RNA samples were sequenced by GeneBay, Inc., Japan, using Illumina HiSeq 2000 platform and dUTP-based directional sequencing method. The obtained 150 bp-long paired-end reads from the two cDNA libraries were mixed and *de novo* assembled with TRINITY, to obtain reference transcript sequences (TCs). A *de novo* assembly was performed using the Trinity pipeline and transcript quantification was conducted as described by Haas et al. (2013). Then, the reads were separately mapped in the two cDNA libraries of the pollen-sterile and pollen-fertile plants. Fragments Per Kilobase of transcript per Million (FPKM) mapped reads were used to indicate the expression abundance of respective unigene, and those with a transcription amount of $|\log FC| > 10$ at the level of FDR < 0.05 were defined as differentially expressed transcripts (DETs). The most homologous protein in Arabidopsis to each DET was determined by BLASTX using BLAST+ (ncbi-blast-2.9.0+) and protein sequences in Arabidopsis as database (Araport11_genes.201606.pep.fasta, https://www.arabidopsis.org/download/index-auto.jsp?dir=

%2Fdownload_files%2FSequences%2FAraport11_blastsets%2FArchived%2FJune_2021). GO-enrichment analysis was performed using the list of homologous Arabidopsis gene names as a query on https://www.arabidopsis.org/tools/go_term_enrichment.jsp.

RT-PCR and qRT-PCR

Flower buds of lengths 10–30 mm of pollen-fertile and sterile segregants were collected from the segregating population, 26–47. Then, anthers, stigmas, and perianths were separated from each bud, instantly frozen in liquid nitrogen, and stored at −80°C until use. Leaves were collected from adult plants. The total RNA was extracted using RNeasy Plant Mini Kit and RNA-Free DNase Kit (QIAGEN) according to the manufacturer's protocol. Then, cDNA was synthesized using PrimeScript 1st strand cDNA Synthesis Kit (Takara Bio, Shiga, Japan). The RT-PCR products were electrophoresed on a 1% agarose gel, stained with ethidium bromide.

qRT-PCR mixture was prepared according to the protocol of iTaq ™ Universal SYBR Green One-Step Kit (Bio-Rad). For the relative quantification of the *LflTDF1* transcripts, qRT-PCR reaction was performed with LightCycler Nano (Roche Diagnostics, Basel, Switzerland). According to the published protocol, the first step was performed at 50°C for 10 min, the second step was performed at 90°C for 1 min, and the three-step amplification reaction was performed at 95°C for 15 s and 60°C for 60 s for 45 cycles. In the melting reaction, the temperature was raised from 60°C to 97°C at a pace of 0.1°C per second, and the PCR results were analyzed using LightCycler Nano Software.

Primers used to detect *LflTDF1* transcripts were LflTDF1-f2 and LflTDF1- r2. *Lilium* actin gene was used as the internal control. Primer pair used to detect *Lilium* actin transcripts was LhACTIN (Sakai et al., 2019). Sequences of the primers are listed in **Supplementary Table 1**.

Genomic DNA Extraction and Marker Assay

Leaf tissue of about 1 cm in diameter was collected, placed in a 1.5-ml plastic tube, and frozen with liquid nitrogen for DNA extraction. DNA extraction for genotyping was followed by the CTAB method in accordance with Sato et al. (2019). For DNA sequencing, plant DNA was isolated using NucleoSpin Plant II Kit (Takara Bio, Shiga, Japan).

DNA fragments of *TDF1* allele and the flanking region were amplified with 100 ng of template DNA in a 10 μl reaction volume containing 2 pmol of each primer and 1× EmeraldAmp PCR Master Mix (Takara Bio, Shiga, Japan). Thermal cycling conditions included denaturation at 95°C for 2 min, followed by 35 cycles of denaturation at 95°C for 30 s, annealing at 53–60°C (depending on the melting temperature of primers) for 30 s, and extension at 72°C for 1 min, and a final extension at 72°C for 5 min. The allele-specific primers and the sequence primers are listed in **Supplementary Table 1**. Unigene (15801) primer pair was used for a positive control (Lang et al., 2015). PCR products were separated by electrophoresis on 1.0% agarose gel and visualized under a UV imaging system. The F_1 plants were made by crossing AR01 with three *L. longiflorum* cultivars, Cristal Horn, White Fox, and Pure Horn, and then, sibling crosses among F_1 plants were conducted to produce the subsequent progenies. A total of 31 pollen-sterile plants and seven pollen-fertile plants in those progenies were used for genotyping. For the genotyping of *L.* × *formolongi*, the segregating population of PL1607 and PL1620 was used. In addition, 24 pollen-sterile plants and seven pollen-fertile plants were selected from the segregating population, 26–47.

DNA Sequencing

For sequence analysis of the *TDF1* homolog gene of *Lilium* species, after fractionating the PCR products by 1% agarose gel electrophoresis, the portion of the target band was cut out and purified using FastGene Gel/PCR Extraction Kit (NIPPON Genetics, Tokyo, Japan). The purified PCR products were cloned into T-Vector pMD20 (Takara Bio Inc., Shiga, Japan). Three clones per each sample were sequenced by Fasmac Co. Ltd. (Kanagawa, Japan). The obtained sequences were analyzed using sequence analysis software, GENETYX v.12 (Genetyx Corporation, Tokyo, Japan) and Sequencher v 2.0 (Hitachi Software, Tokyo, Japan). The genomic sequences of the alleles collected in this study were aligned using MEGA 7 (Kumar et al., 2016). The phylogenetic relationship was inferred by using the maximum likelihood method (Hasegawa et al., 1985) added to MEGA 7.

Nanopore Sequencing and Genomic PCR in the *LflTDF1* Region

Nanopore sequencing was conducted in a pollen-fertile plant, N-1, selected from a seedling population of *L.* × *formolongi*. DNA was isolated by NucleoBond HMW DNA isolation kit (TaKaRa Bio Inc., Shiga, Japan). The DNA sequencing was conducted by GeneBay Co., Ltd., (Kanagawa, Japan) using PromethION. The nanopore read sequence harboring *LflTDF1* was corrected by the short read of *L.* × *formolongi* cv. Augusta obtained by Hiseq 2000 device. The PCR experiment using six primer pairs designed based on the nanopore sequence (contig #1) encompassing the *LflTDF1* region was conducted in five pollen-fertile genotypes including N-1, the pollen-fertile segregant (63-1 and 97-2), the originator's line of AR01 (Raizan 2go) and cv. Augusta, as well as in the pollen-sterile AR01.

RESULTS

Genetic Analysis

A sibling cross of F_1 plants made from the cross between *L.* × *formolongi* cv. Hatsuki and cv. Raizan 2go produced the pollen-sterile line, PL01. Then, we made the segregating populations, PL1620 and PL1607 using PL01 and its progeny, and investigated the inheritance manner of the pollen-sterile gene (**Supplementary Figure 1**). The PL1620 population was made by the cross between the pollen-fertile lines, 83-2 × 220-1, where both lines were inferred to be heterozygous at the locus responsible for the pollen phenotype. The phenotype assay of PL1620 revealed that the phenotyping in the progeny fit to 3

TABLE 1 | Chi-square test for segregation of pollen-fertile and sterile plants in the segregating populations.

Population	Crossing (genotype tested)	No. of plants	No. of pollen fertile plants	No. of pollen sterile plants	Chi-square vale (probability)
PL1620	Fertile × Fertile (*LflTDF1*/*lfltdf1* × *LflTDF1*/*lfltdf1*)	37	30	7	3.2 (0.3<p<0.5)
PL1607	Sterile × Fertile (*lfltdf1*/*lfltdf1* × *LflTDF1*/*lfltdf1*)	46	28	18	2.1 (0.1<p<0.2)

(fertile):1 (sterile) ratio by chi-square test (**Table 1**), indicating that both parents are heterozygous (*LflTDF1*/*lfltdf1*).

The PL1607 population was made from crossing the pollen-sterile PL01 line with the pollen fertile 97-1 inferred to be heterozygous at the locus responsible for the pollen phenotype. The seed set rate of PL01 plant was high (305 seeds/pod), indicating that the female fertility of the pollen-sterile mutant is normal. In this population, the phenotyping data matched the 1: 1 ratio by chi-square test, which fit to a single recessive gene model (**Table 1**). Thus, since the genetic analysis in the two segregating populations suggested that the genetic control of the male-sterile phenotype is under a single recessive gene, we applied RNA-sequencing technology to identify a presence/absence gene expression between pollen-fertile and sterile plants.

Anatomy of Microsporogenesis

Paraffin section observation revealed that in pollen-fertile segregants of the segregating population (26–47), pollen developmental stages were associated with the size of flower buds: 10–18 mm in the PMC proliferation stage, 20–22 mm in the premeiotic stage, 23–28 mm in the meiotic cell division stage, and >29 mm in the uninuclear and binuclear microspore stage (**Figures 2A–E**). In the pollen-sterile segregants, PMCs were normally proliferated but could not start meiotic division in PMCs in the flower bud length 20 mm, and then those PMCs were degraded in the premeiotic stage (**Figures 2F–J**). At this stage, the middle layers of the anther of the pollen-sterile segregants were densely stained and became swollen (**Figures 2H,I**). Paraffin section observation of the *lfltdf1* mutant confirmed no production of pollen debris in the anther. This can be explainable for the appearance of perfect pollen-free anthers in the blooming flowers of *lfltdf1* mutant (**Figure 1**).

The pollen-sterile lines, #72 and #318, developed by the gamma-irradiated breeding normally underwent the microsporogenesis up to the premeiotic stage where the PMCs became free due to callose degeneration in the locule (**Supplementary Figures 2A,C**) like the pollen-fertile plant (**Figure 2D**), but failed meiotic cell division (**Supplementary Figures 2B,D**). The aberration in the middle stage of meiosis produced many debris in the anther locule. This pollen debris remained in some anthers attached to the blooming flowers (**Supplementary Figures 2E,F**).

RNA-Seq and GO Analysis

Illumina sequencing of RNA extracted from anthers of the pollen-sterile and pollen-fertile segregants in the segregating population, 26–47, and the following Trinity assembly obtained 56,576 transcripts (isoforms). Of them, 253 were identified as differentially expressed transcripts (DETs) between pollen fertile (p) and sterile (np) by statistical analysis using edgeR of which FDR was <0.05, with 97 DETs in p>np (<-10 logFC) and 156 DETs in p<np (>10 logFC), respectively (**Table 2** and **Supplementary Table 2**). A blastx search using the 253 TCs as query and Arabidopsis protein sequences as a database revealed that 11 DETs (five in p>np and six in p<np) did not match any protein sequences in the Arabidopsis. On the contrary, there were 242 DETs with hits (92 in p>np and 150 in p<np). The list of identified Arabidopsis homologs was subjected to GO-enrichment analysis using the Arabidopsis platform available at TAIR website (https://www.arabidopsis.org/tools/go_term_enrichment.jsp). The "gene expression" was detected as an over-represented GO term in the list of Arabidopsis homologs of p<np DETs. On the contrary, six over-represented GO terms, which were all related to sexual reproduction in the biological process, were found in the list of Arabidopsis homologs of p>np DETs (**Table 3** and **Supplementary Table 3**).

One of the DETs included in the over-represented GO terms, which was expressed only in the pollen-fertile plants but not in the pollen-sterile plants, was identified as a homolog of *Arabidopsis thaliana TDF1* (AtMYB35) which belongs to the member of the R2R3 factor gene family. Since Arabidopsis *TDF1* mutant shows male sterility, we further analyzed the identified DET as a candidate in *L.* × *formolongi*, designated as *LflTDF1*.

RT-PCR Analysis

RT-PCR was performed using anthers, stigmas, perianths, and leaves of two pollen-fertile segregants in the segregating population (26–47) (**Figure 3A**). As a result, RT-PCR revealed that *LflTDF1* mRNA was expressed only in anthers but not in the pistils, perianths, and leaves. The anther-specific *LflTDF1* mRNA was expressed only in pollen-fertile segregants but not in pollen-sterile segregants (**Figure 3B**), which is consistent with the RNA-seq data. The RT-PCR of *LflTDF1* mRNA using the developing anthers with different pollen developmental stages revealed that the transcription levels were too low to detect at 30 mm bud length, while the expression started from a flower bud length of 11 mm and tended to reach a peak at the meiotic division stage of flower bud length of 19.0–22.9 mm (**Figure 3C**). Therefore, we

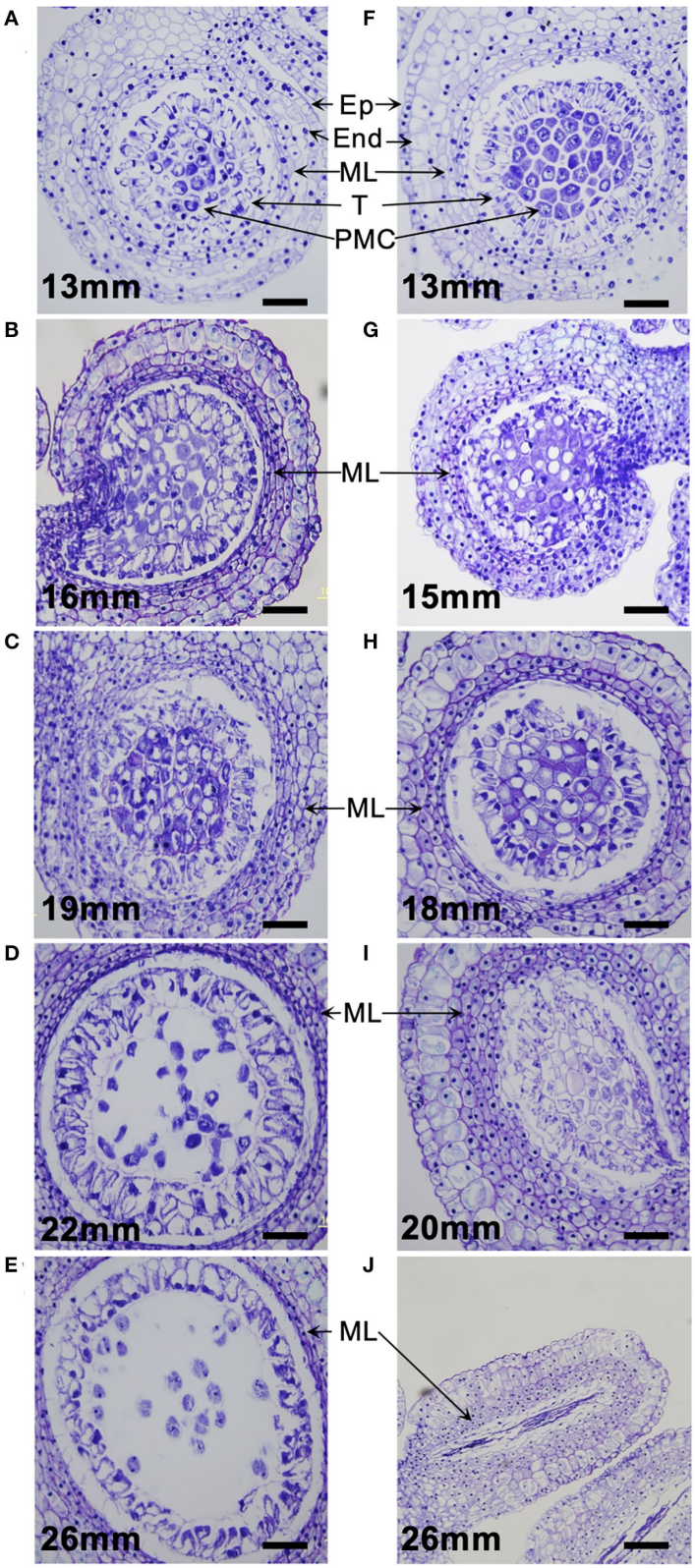


FIGURE 2 | Microsporogenesis of *L.* × *formolongi*. **(A–E)** Pollen-fertile segregants. **(F–J)** Pollen-sterile segregants. Ep, epidermis; End, endothecium; ML, middle layers; T, tapetum; PMC, pollen mother cells. Bar = 100 μm.

TABLE 2 | Differential expression analysis in the anthers (collected from the bud size, 11–14 mm) of pollen-sterile (np) and fertile (p) plants with the criteria of FDR < 0.05 and | logFC|>10, best-hit blastx homology search, and GO analysis.

DETs category	No. of DETs	Blastx		Significantly enriched GO term
		Homolog	No homolog	
p>np	97	92	5	7
p<np	156	150	6	1
total	253	242	11	8

TABLE 3 | Significantly enriched GO terms in differentially expressed transcripts (DETs) identified in the anthers (collected from the bud size, 11–14 mm) of pollen-sterile (np) and fertile (p) plants.

DETs category	GO Type	GO Term	Gene number	Relative expression	P-value
p>np	BP	Floral organ development	8	10.17	4.46E-03
		Reproductive structure development	15	3.87	2.06E-02
		Reproductive system development	15	3.86	2.10E-02
		Developmental process involved in reproduction	16	3.44	4.12E-02
		Reproductive process	18	3.22	2.70E-02
		Reproduction	18	3.2	2.91E-02
	MF	Protein binding	36	2.06	1.01E-02
p < np	BP	Gene expression	15	4.17	5.62E-03

concluded that the *LflTDF1* allele of the pollen-sterile line has lost its transcription activity in the anthers. Thus, we designated this pollen sterility allele as *lfltdf1*.

DNA Sequence Analysis of *LflTDF1*

LflTDF1 transcript and the corresponding genomic locus were sequenced in a pollen-fertile plant in the segregating population (26–47). Comparison of the transcript sequences and genomic sequences confirmed the location of four exons and three introns in the *LflTDF1* gene where an exon locates in the 5'-UTR and three exons in the CDS (**Figure 4A**). *TDF1* orthologs were amplified from a single plant of *L.* × *formolongi*, *L. longiflorum* cv. Hinomoto, *L. formosanum*, and Oriental hybrid lily cv. Siberia to determine the nucleotide sequences. As a result, the alignment of 2287bp including 5'- and 3'-UTR identified a single allele (homozygous) in each *L.* × *formolongi* and *L. formosanum,* and two alleles (heterozygous) in each cv. Hinomoto and cv. Siberia (**Supplementary Table 4**). Therefore, the two alleles of *TDF1* of Easter lily and Oriental hybrid lily were designated as *TDF1a* and *TDF1b* (**Figure 4D**). *LflTDF1* is composed of 311 amino acids (aa), and the N-terminal region containing the myb-type helix-turn-helix (HTH) domain was conserved among the distantly related species such as *Oryza rufipogon* and *A. thaliana* (**Figure 4B**), suggesting that the *LflTDF1* acts as a MYB transcription factor. The aa sequences after the conserved region showed low similarity among the different species (**Supplementary Figure 3**). Since the pollen fertility of the wild and mutant type is determined by the presence/absence of *LflTDF1* gene, we designed some intragenic markers to detect *LflTDF1* gene in the segregating population (**Figure 4A**). *LflTDF1* intragenic dominant marker, 1f4r, clearly detected the presence/absence of *LflTDF1* gene in the segregating population (26–47) (**Figure 4C**). Phylogenic analysis revealed that *L. longiflorum* and *L. formosanum* are monophyletic and Oriental hybrid lily cv. Siberia was included in the other clade (**Figure 4D**). The *LflTDF1* showed higher identity with the orthologous *LfsTDF1* gene of *L. formosanum*, a parental species of *L.* × *formolongi*, than that of *L. longiflorum* (Easter lily), the other parental species. The result suggests that *LflTDF1* originated from *L. formosanum* in the interspecific breeding process.

Correlation of Phenotypes and Genotypes

As described above, the genetic analysis in the segregating populations of PL1620 and PL1607 suggested that the genetic control of the male-sterile phenotype is under a single recessive gene. The genotype of cross combination of PL1620 and PL1607 suggested to be *LflTDF1*/*lfltdf1* × *LflTDF1*/*lfltdf1* and *lfltdf1*/*lfltdf1* × *LflTDF1*/*lfltdf1*, respectively (**Table 4**). An example of genotyping of LflTDF1 intragenic dominant marker, f4r4, using the segregating population is shown in **Supplementary Figure 4**. In the PL1620 population, the genotyping using the dominant marker LflTDF1-f4r4 was perfectly matched with the phenotyping data of the progeny. Genotyping in the PL1607 population was almost perfectly matched with the phenotyping data of the progeny, but there was one exceptional plant, which was pollen-sterile but produced the *LflTDF1* specific marker band in the PCR test. We could not re-examine this exceptional plant in the next growing season due to the rotting of the bulb.

Since the cross between *L. longiflorum* and *L.* × *formolongi* is compatible, the F_1 plants were made by crossing AR01 with three *L. longiflorum* cultivars, Cristal Horn, White Fox, and Pure Horn. Then, sibling crosses among F_1 plants produced the subsequent progenies, of which 31 pollen-sterile plants and seven pollen-fertile plants were used for genotyping. In addition, 24 pollen-sterile plants and seven pollen-fertile plants were selected from the segregating population (26–47) of *L.* × *formolongi*. In the selected plants, the dominant marker LflTDF1-f4r4 was perfectly matched with the phenotyping data of the progeny (**Table 4**).

Nanopore Sequencing of *LflTDF1* Locus

The nanopore sequencing technology was applied to collect the *LflTDF1* harboring genomic region sequences in a pollen-fertile plant of a seedling population of *L.* × *formolongi*. The read yield was 68.74GB (>1 kbp). The number of reads and the average length were 4,520,860 (> 1 kbp) and 15.21 kbp, respectively. The *LflTDF1* region was only detected in the

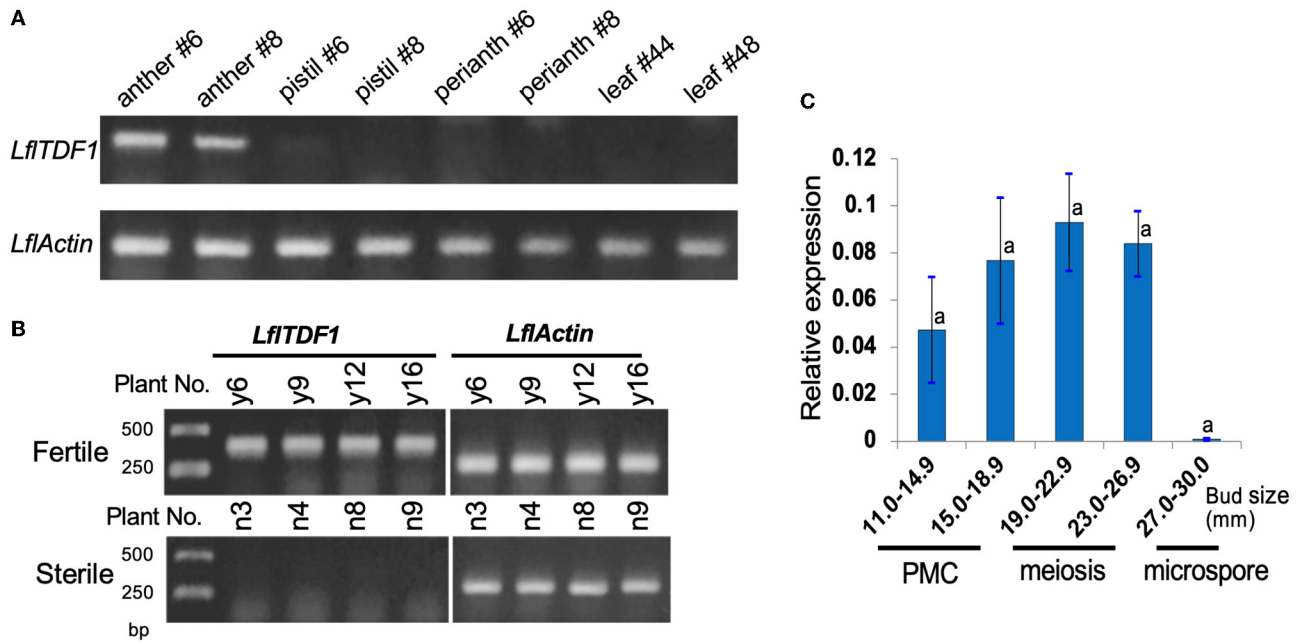


FIGURE 3 | Expression analysis of *LflTDF1* gene by RT-PCR and qRT-PCR. The samples were collected from the population derived from the cross of pollen-sterile and pollen-fertile segregants. Actin gene of *L.* × *formolongi* was used as the internal control. **(A)** Tissue-specific expression of *LflTDF1* in the pollen-fertile segregants. The samples were collected from the two different plants indicated by #. **(B)** The expression of *LflTDF1* in the anthers (collected from the bud size 14–26 mm) of the pollen-fertile and sterile segregants. Four different samples of the pollen-sterile and pollen-fertile segregants were shown. **(C)** qRT-PCR detected the expression of *LflTDF1* in the anthers collected from the buds of pollen-fertile segregants having different sizes. The pollen developmental stages are referred in **Figure 2**. Error bar represents standard deviation, and means were tested by Tukey's test at P = 0.05.

sequence (designated as contig #1), and no other sequences containing *LflTDF1* were found. Contig #1 overlapped contig #2 at the left end and repetitive sequences at right end, respectively (**Figure 5** and **Supplementary Figure 5**). The homology of the overlapping region of contig #1 and contig #2 was not high (**Supplementary Figure 5**), indicating that the location of contig #2 is tentative.

The PCR experiment using six primer pairs in the genomic region encompassing the contig #1 was conducted in five pollen-fertile genotypes including the pollen-fertile segregants and the pollen-sterile AR01 (**Figure 5**). As a result, the primer pairs, c to f, located from the left end to the right end of contig #1, amplified the same sized fragment in all pollen-fertile genotypes but not in the pollen-sterile AR01. On the contrary, the primer pairs g and h were able to amplify the fragments in both pollen-fertile and sterile genotypes. Since those primer pairs were designed in the repetitive sequence located in the right end of contig #1, we confirmed whether the primer pair g can specifically amplify the linked region of *LflTDF1* gene using the segregation population, PL3105. The primer pair g PCR amplicons of the pollen-sterile AR01 and pollen-fertile lines were sequenced, so that there was an SNP at the *Eco*RV recognition site between the two lines and the segregating population (PL3105), that is, AR01 (*lfltdf1*/*lfltdf1*) × 26–47 (*LflTDF1*/*lfltdf1*) was evaluated for genotyping experiment. Segregation profile of *TDF1* and *tdf1* genotypes in this population using the intragenic dominant markers (f4r4) and the codominant CAPS linkage marker g was perfectly matched (**Supplementary Figure 6**). This result indicates that the linkage marker g could specifically amplify the flanking region of *LfTDF1*gene. Therefore, this data suggested that the chromosomal breaking point inferred from the PCR amplification of each primer set was between the two regions amplified by the primer pairs, f and g, respectively. Overall, the corresponding region in the pollen-sterile plants is likely deleted at least 30 kbp corresponding to almost whole length of contig #1 or replaced by an unknown sequence.

DISCUSSION

The transcriptome comparison using the developing anthers (10–14 mm) identified one of the DETs expressed only in pollen-fertile plants. This DET was identified as a homolog of *Arabidopsis thaliana TDF1* (*AtMYB35*). Genetic analysis in the segregating population showed that the pollen sterility of PL01 was inherited in a single recessive manner, and the dominant marker LflTDF1-f4r4 was almost perfectly matched with the phenotyping data of the progeny. The one exceptional plant was positive for the dominant marker LflTDF1-f4r4 test but was pollen sterile. This inconsistency may be due to the pollen sterility caused by physiological disorder that occasionally occurs under unsuitable circumstances. Alternatively, it is due to the sampling miss. The nanopore sequence indicates that AR01 is

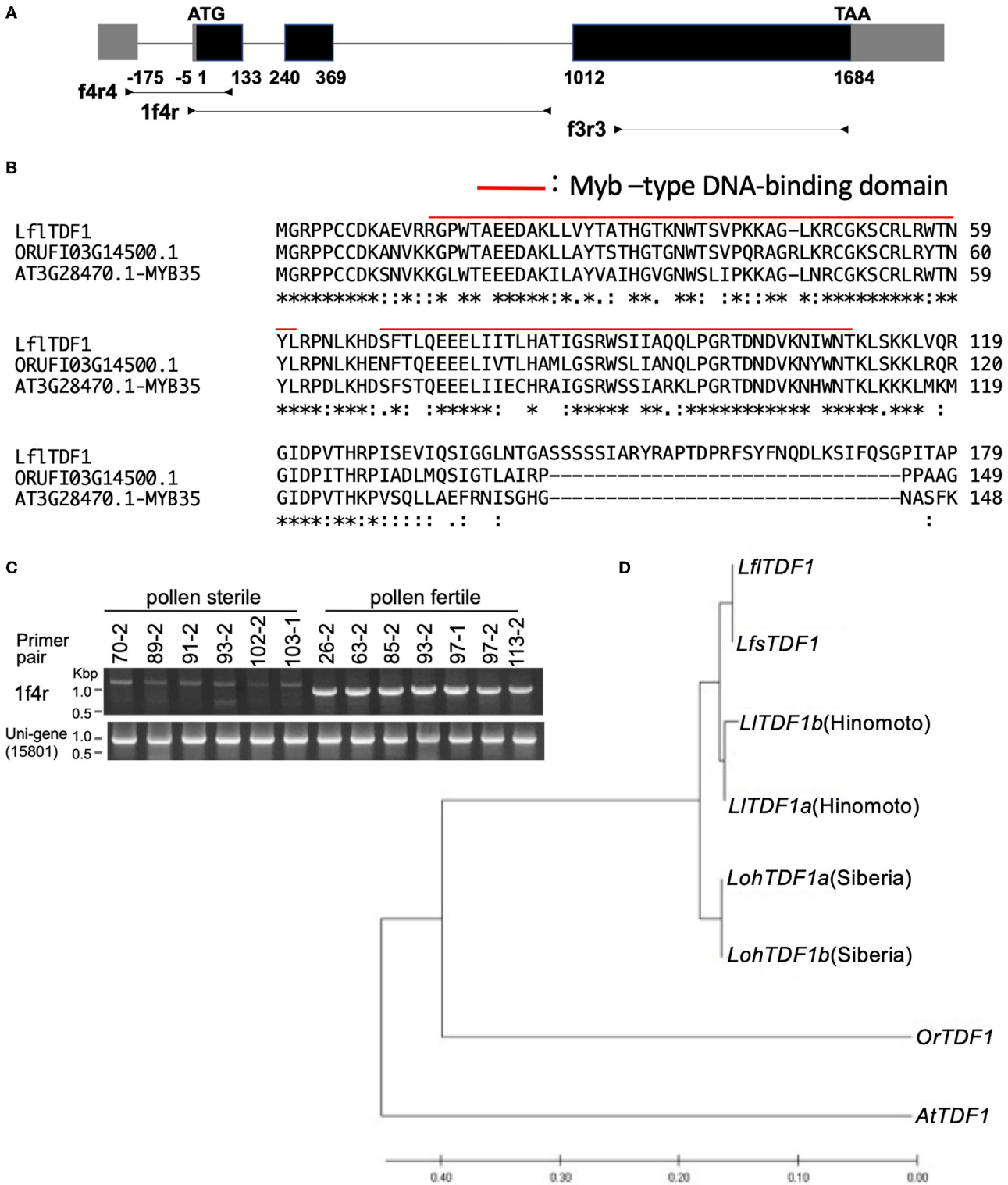


FIGURE 4 | (A) Schematic diagram of the *LflTDF1* gene structure including exons (boxes) and introns (horizontal line). Gray and black boxes indicate untranslated and CDS region, respectively. Horizontal lines with attached arrowheads indicate PCR amplicon regions for genotyping with the intragenic markers. **(B)** Comparison of partial amino acid sequences of *TDF1* orthologs in *L.* × *formolongi*, *Oryza rufipogon* (ORUFI03G14500.1), and *Arabidopsis thaliana* (AT3G28470.1), respectively. Red line indicates Myb–type DNA-binding domain predicted by Pfam. **(C)** An example of genotyping of *LflTDF1* intragenic dominant marker, 1f4r, using the segregating

(Continued)

FIGURE 4 | population (26–47). Unigene (15801) primer pair was used for a positive control. **(D)** Phylogenic relationships of *TDF1* orthologs in *L.* × *formolongi* (*LflTDF1*), *L. formosanum* (*LfsTDF1*), *L. longiflorum* cv. Hinomoto (*LlTDF1*), *L.* spp. Oriental hybrid cv. Siberia (*LohTDF1*), *Oryza rufipogon* (*OrTDF1,* ORUFI03G1 4500.1), and *A. thaliana* (*AtTDF1*, AT3G28470.1). The phylogenetic tree was inferred by using the maximum likelihood method added to MEGA 7.

TABLE 4 | Phenotyping and genotyping in the segregating populations and the selected breeding lines.

	Population (genotype tested)	No. of plants	Phenotyping	Genotyping with the dominant marker LflTDF1-f4r4		
			No. of fertile plants	No. of sterile plants	Presence of the TDF1 marker band	Absence of the TDF1 marker band
Segregating population	PL1620 (*LflTDF1*/*lfltdf1* × *LfTDF1*/*lfltdf1*)	37	30	7	30	7
Segregating population	PL1607 (*lfltdf1*/*lfltdf1* × *LflTDF1*/*lfltdf1*)	46	28	18	29	17
Selected breeding lines	Progeny of PL01 × *L. longiflorum*	38	7	31	7	31
Selected breeding lines	Progeny of PL01 × *L.* × *formolongi*	31	7	24	7	24

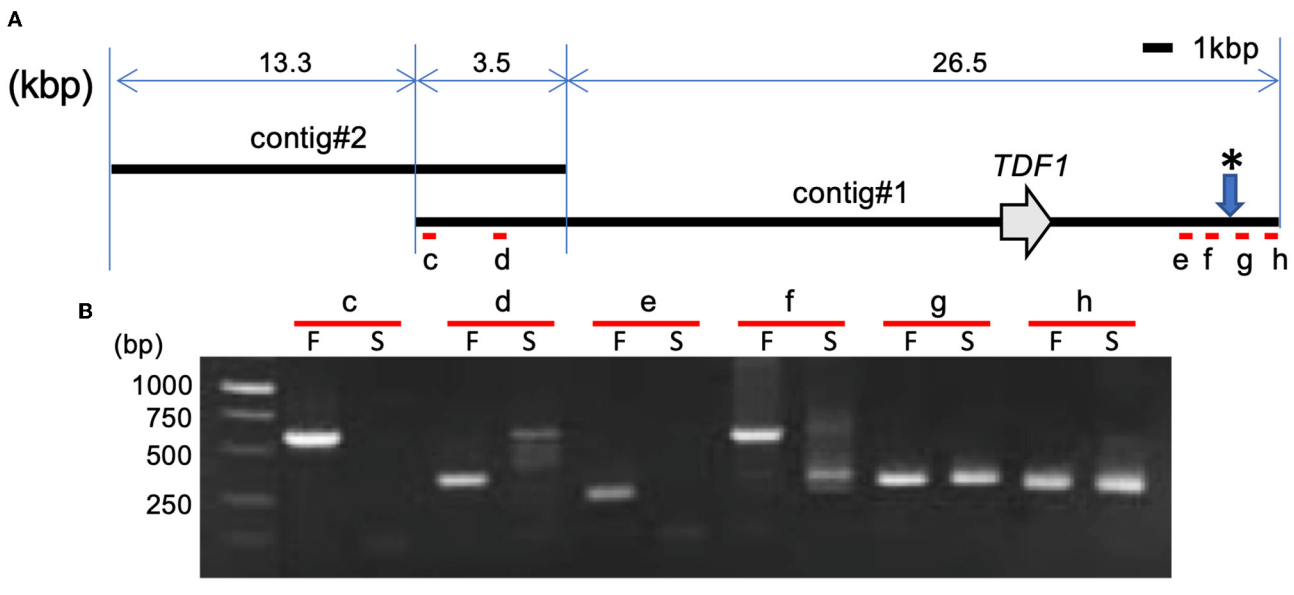


FIGURE 5 | Structural analysis of *LflTDF1* flanking region. **(A)** Location of *LflTDF1* and nanopore sequencing contigs obtained from a pollen-fertile line. **(B)** PCR amplification in the pollen-fertile genotype (F) and sterile AR01 (S) using the primer pairs (c to h) that were designed within the contig #1. Since five pollen-fertile genotypes amplified the same sized amplicon in the respective primer set, the amplicon of one genotype is represented as an example. * indicates the chromosome breaking points inferred from the PCR amplification of each primer set.

lacking *LflTDF1* region spanning at least 30 kbp or is replaced by unknown sequence. In addition to this evidence, since Arabidopsis *TDF1* mutant shows male sterility and *TDF1* is a key regulator of tapetal development and microsporogenesis, we identified *LflTDF1* as a candidate gene for the pollen sterility of AR01. One of the parental materials, cv. Hatsuki and cv. Raizan 2go (both male fertile cultivars), is probably heterozygous in the *LflTDF1* allele. However, it has not yet been determined which parent provided the mutant allele to the progeny.

The pollen sterility of Arabidopsis *tdf1* mutant is caused by dysfunction of anther wall layers including endothecium, middle

layer, and tapetum (Zhu et al., 2008). Zhu et al. (2008) reported that in the *attdf1* mutant, cell vacuolation in both the epidermis and the endothecium started in the early anther development stage appearing pollen mother cells, followed by vacuolation of the middle layer at the premeiotic stage, and thereafter, the tapetum was vacuolated, hypertrophic, and multilayered. Finally, the swollen tapetum cells crush microspore. The similar anatomical feature of anther development of *ostdf1* mutant was reported by Cai et al. (2015). In contrast to *attdf1* and *ostdf1* mutant, the *lftdf1* genotype did not show vacuolation of the epidermis, endothecium, and middle layer in the early stage of PMC proliferation and showed no enlargement of the tapetum cells. On the contrary, the middle layer of the *lftdf1* genotype became dense and swollen, compared to the wild type, and then, the thick middle layer remained visible till the degradation of PMCs. It is known that *Lilium* anther wall layers preserve many starch grains consumed in the microspore formation (Clément et al., 1994; Clément and Audran, 1995). Therefore, the degradation of PMCs and tapetum in the meiotic stage prevents the transfer of starch degradation products to the tapetum, causing the deposition of the excessive glucose or its derivatives in the anther wall layer. This may be one of the reasons why the dense thick middle layers appeared in AR01. In both *attdf1* and rice *ostdf1* mutants, the tetrads were formed by meiosis, but the resulting tetrads were abnormally unreleased and not free in the locules. In contrast, the *lfltdf1* mutant never enters meiosis and the PMCs are degraded prior to meiosis, indicating that the time of appearance of anther development defects in the *tdf1* mutants may vary from species to species.

In pollen-sterile mutants #72 and #318 produced by gamma irradiation, meiosis began but was incomplete, so some anthers attached some pollen debris during flowering. These male meiotic mutants released PMCs in the anther locule like wild type (**Supplementary Figures 2A,C**) but the meiotic cell division failed, and the resulting aberrant PMC products might be harder, probably due to the deposition of sporopollenin like substance. In contrast, the pollen sterility caused by *lfltdf1* mutation never produces pollen debris during flowering. This is owing to the earlier degradation of PMCs prior to meiotic division in *lfltdf1* plant, where the aberrant premeiotic product completely disappeared in the anther locules. In this regard, the *lfltdf1* mutant is a superior breeding material for lily pollen-free breeding.

GO analysis performed with the p>np set of DET on the Arabidopsis basis identified 18 Arabidopsis genes such as U-box domain-containing protein 4 (PUB4) ortholog in biological process ontology (**Supplementary Table 3**). PUB4 plays an important role for controlling tapetum abortion (Wang et al., 2013), suggesting that many genes including microsporogenesis-related genes were directly/indirectly affected by lack of *LflTDF1* gene. In addition, this aberrant regulatory process in the mutant began early in the development of PMCs in which sample RNA was collected. This is quite before the pollen abortion stage (meiotic stage) of the mutant. Further studies using appropriate logFC criteria instead of ± 10 log FC used in this study are needed to learn more about the gene expression profile in the mutant, which helps to better understand the pollen development process of lilies.

We found two alleles of *TDF1* homologs in a single plant of each species, *L. longiflorum* cv. Hinomoto and an Oriental hybrid cv. Siberia. *L. longiflorum* and an Oriental hybrid lily cv. Siberia are heterozygous because of outcrossing due to self-incompatible nature, and therefore, the two polymorphic *TDF1* homologs come from outcrossing. On the contrary, a single *TDF1* sequence was identified in a single plant of the respective *L. formosanum* and *L.* × *formolongi*. This may be due to the self-compatible nature of the two species, which produce more homozygosity of *LflTDF1* locus. *LflTDF1* of *L.* × *formolongi* showed higher identity with the orthologous gene of *TDF1* of *L. formosanum*, a parental species of *L.* × *formolongi*, than that of *L. longiflorum*, the other parental species, suggesting *LflTDF1* originated from *L. formosanum*. The Oriental hybrid lilies were derived from complex interspecific crosses using several species such as *L. auratum*, *L. nobilisimum*, *L. japonicum*, and *L. speciosum* (Marasek-Ciolakowska et al., 2018). Therefore, it has not yet been determined which parental species gave Siberia two *TDF1* homologs.

In this study, we successfully identified the candidate gene of the pollen sterility of *L.* × *formolongi* using RNA-seq and genetic analysis in the segregating population. Since the cross between *L.* × *formolongi* and Easter lily (*L. longiflorum*) is compatible, we successfully developed several pollen-free promising clones of Easter lily in the progenies derived from the sib cross of F_1 plants made by crossing AR01 with Easter lily. The marker development of *TDF1* gene will assist pollen-free lily breeding of Easter lilies. In addition, we showed that the *lfltdf1* mutant never produces the pollen debris, indicating that *TDF1* orthologs are excellent target genes for producing pollen-free ornamental plants through genome editing.

DATA AVAILABILITY STATEMENT

Partial mRNA sequence of LflTDF1 was deposited to DDBJ and the accession numbers, LC704438 was assigned. Genomic sequences of Lilium TDF1 orthologs were assigned as follows: LC704885 (LflTDF1), LC704888 (LfsTDF1), LC704887 (LlTDF1a), LC704886 (LlTDF1b), LC704883 (LohTDF1a), and LC704884 (LohTDF1b).

AUTHOR CONTRIBUTIONS

NY, SI, and KO conceptualized the study and were involved in funding acquisition. TM, DS, NY, SI, TS, HO, HS, and SO designed methodology. TM, DS, NY, SI, TS, HO, HS, SO, and EF investigated the study. NY, SI, TS, and KO took a leading role in project administration. DS, EF, and KO reviewed and edited the

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

FUNDING

This research was supported by grants from the Project of the NARO Bio-oriented Technology Research Advancement Institution (Research program on development of innovative technology) (No. 28036C).

ACKNOWLEDGMENTS

The authors would like to thank H. Onda for providing technical assistants for gene sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.914671/full#supplementary-material

REFERENCES

Cai, C. F., Zhu, J., Lou, Y., Guo, Z. L., Xiong, S. X., Wang, K., et al. (2015). The functional analysis of *OsTDF1* reveals a conserved genetic pathway for tapetal development between rice and Arabidopsis. *Sci. Bull.* 60, 1073–1082. doi: 10.1007/s11434-015-0810-3

Clément, C., and Audran, J. C. (1995). Anther wall layers control pollen sugar nutrition in *Lilium*. *Protoplasma.* 187, 172–181. doi: 10.1007/BF01280246

Clément, C., Chavant, L., Burrus, M., and Audran, J. C. (1994). Anther starch variations in Lilium during pollen development. *Sexual Plant Reprod.* 7, 347–356. doi: 10.1007/BF00230513

García-Sogo, B., Pineda, B., Castelblanque, L., Antón, T., Medina, M., Roque, E., et al. (2010). Efficient transformation of *Kalanchoe blossfeldiana* and production of male-sterile plants by engineered anther ablation. *Plant Cell Rep.* 29, 61–77. doi: 10.1007/s00299-009-0798-8

Gómez, J. F., Talle, B., and Wilson, Z. A. (2015). Anther and pollen development: a conserved developmental pathway. *J. Integr. Plant Biol.* 57, 876–891. doi: 10.1111/jipb.12425

Haas, B., Papanicolaou, A., Yassour, M., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Harkess, A., Huang, K., van der Hulst, R., Tissen, B., Caplan, J. L., Koppula, A., et al. (2020). Sex determination by two Y-linked genes in garden asparagus. *Plant Cell.* 32, 1790–1796. doi: 10.1105/tpc.19.00859

Hasegawa, M., Kishino, H., and Yano, T. A. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22, 160–174. doi: 10.1007/BF02101694

Hiramatsu, M., Ii, K., Okubo, H., Huang, K. L., and Huang, C. W. (2001). Biogeography and origin of *Lilium longiflorum* and *L. formosanum* (Liliaceae) endemic to the Ryukyu Archipelago and Taiwan as determined by allozyme diversity. *Amer. J. Bot.* 88, 1230–1239. doi: 10.2307/3558334

Ito, T., Nagata, N., Yoshiba, Y., Ohme-Takagi, M., Ma, H., and Shinozaki, K. (2007). Arabidopsis MALE STERILITY1 encodes a PHD-type transcription factor and regulates pollen and tapetum development. *Plant Cell.* 19, 3549–3562. doi: 10.1105/tpc.107.054536

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

Lang, V., Usadel, B., and Obermeyer, G. (2015). De novo sequencing and analysis of the lily pollen transcriptome: an open access data source for an orphan plant species. *Plant Mol. Biol.* 87, 69–80. doi: 10.1007/s11103-014-0261-2

Marasek-Ciolakowska, A., Nishikawa, T., Shea, D. J., and Okazaki, K. (2018). Breeding of lilies and tulips—Interspecific hybridization and genetic background. *Breed. Sci.* 68, 35–52. doi: 10.1270/jsbbs.17097

Nishikawa, T., Okazaki, K., Arakawa, K., and Nagamine, T. (2001). Phylogenetic analysis of section Sinomartagon in genus *Lilium* using sequences of the internal transcribed spacer region in nuclear ribosomal DNA. *Breed. Sci.* 51, 39–46. doi: 10.1270/jsbbs.51.39

Nishikawa, T., Okazaki, K., T., Uchino, Arakawa, K., and Nagamine, T. (1999). A molecular phylogeny of *Lilium* in the internal transcribed spacer region of nuclear ribosomal DNA. *J. Mol. Evol.* 49, 238–249. doi: 10.1007/PL00006546

Roque, E., Gómez-Mena, C., Hamza, R., Beltrán, J. P., and Cañas, L. A. (2019). Engineered male sterility by early anther ablation using the pea anther-specific promoter PsEND1. *Front. Plant Sci.* 10, 819. doi: 10.3389/fpls.2019.00819

Sakai, M., Yamagishi, M., and Matsuyama, K. (2019). Repression of anthocyanin biosynthesis by R3-MYB transcription factors in lily (*Lilium* spp.). *Plant Cell Rep.* 38, 609–622. doi: 10.1007/s00299-019-02391-4

Sakazono, S., Hiramatsu, M., Okubo, H., Huang, K. L., and Huang, C. L. (2009). Origin of *Lilium longiflorum* thunb. 'Hinomoto'. *J. Japan. Soc. Hort. Sci.* 78, 231–235. doi: 10.2503/jjshs1.78.231

Sato, M., Shimizu, M., Shea, D. J., Hoque, M., Kawanabe, T., Miyaji, N., et al. (2019). Allele specific DNA marker for fusarium resistance gene *FocBo1* in *Brassica oleracea*. *Breed. Sci.* 69, 308–315. doi: 10.1270/jsbbs.18156

Shimizu, M. (1987). *Lilies in Japan*. (In Japanese). Japan: Seibundo Shinkosha Publishing Co., Ltd.

Smith, A. G., Gardner, N., and Zimmermann, E. (2004). Increased flower longevity in petunia with male sterility. *Hortscience.* 39, 822. doi: 10.21273/HORTSCI.39.4.822B

Sorensen, A. M., Kröber, S., Unte, U. S., Huijser, P., Dekker, K., and Saedler, H. (2003). The *Arabidopsis ABORTED MICROSPORES* (*AMS*) gene encodes a MYC class transcription factor. *Plant J.* 33, 413–423. doi: 10.1046/j.1365-313X.2003.01644.x

Tsugama, D., Matsuyama, K., Ide, M., Hayashi, M., Fujino, K., and Masuda, K. (2017). A putative MYB35 ortholog is a candidate for the sex-determining genes in *Asparagus officinalis*. *Sci. Rep.* 7, 1–9. doi: 10.1038/srep41497

Wang, H., Lu, Y., Jiang, T., Berg, H., Li, C., and Xia, Y. (2013). The Arabidopsis U–box/ARM repeat E3 ligase AtPUB4 influences growth and degeneration of tapetal cells, and its mutation leads to conditional male sterility. *Plant J.* 74, 511–523. doi: 10.1111/tpj.12146

Zhang, W., Sun, Y., Timofejeva, L., Chen, C., Grossniklaus, U., and Ma, H. (2006). Regulation of Arabidopsis tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. *Development.* 133, 3085–3095. doi: 10.1242/dev.02463

Zhang, Z. B., Zhu, J., Gao, J. F., Wang, C., Li, H., Li, H., et al. (2007). Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in *Arabidopsis*. *Plant J.* 52, 528–538. doi: 10.1111/j.1365-313X.2007.03254.x

Zhu, J., Chen, H., Li, H., Gao, J. F., Jiang, H., Wang, C., et al. (2008). *Defective in Tapetal Development and Function 1* is essential for anther development and tapetal function for microspore maturation

in Arabidopsis. *Plant J.* 55, 266–277. doi: 10.1111/j.1365-313X.2008.03500.x

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 30 June 2022
doi: 10.3389/fpls.2022.923069

Taxonomically Restricted Genes Are Associated With Responses to Biotic and Abiotic Stresses in Sugarcane (*Saccharum* spp.)

Cláudio Benício Cardoso-Silva[1,2†], Alexandre Hild Aono[1], Melina Cristina Mancini[1], Danilo Augusto Sforça[1], Carla Cristina da Silva[1,3], Luciana Rossini Pinto[4], Keith L. Adams[2] and Anete Pereira de Souza[1,5]*

[1]Center of Molecular Biology and Genetic Engineering (CBMEG), University of Campinas (UNICAMP), Campinas, Brazil, [2]Department of Botany, University of British Columbia, Vancouver, BC, Canada, [3]Agronomy Department, Federal University of Viçosa (UFV), Viçosa, Brazil, [4]Sugarcane Research Advanced Centre, Agronomic Institute of Campinas (IAC/APTA), Ribeirão Preto, Brazil, [5]Institute of Biology, University of Campinas (UNICAMP), Campinas, Brazil

Orphan genes (OGs) are protein-coding genes that are restricted to particular clades or species and lack homology with genes from other organisms, making their biological functions difficult to predict. OGs can rapidly originate and become functional; consequently, they may support rapid adaptation to environmental changes. Extensive spread of mobile elements and whole-genome duplication occurred in the *Saccharum* group, which may have contributed to the origin and diversification of OGs in the sugarcane genome. Here, we identified and characterized OGs in sugarcane, examined their expression profiles across tissues and genotypes, and investigated their regulation under varying conditions. We identified 319 OGs in the *Saccharum spontaneum* genome without detected homology to protein-coding genes in green plants, except those belonging to Saccharinae. Transcriptomic analysis revealed 288 sugarcane OGs with detectable expression levels in at least one tissue or genotype. We observed similar expression patterns of OGs in sugarcane genotypes originating from the closest geographical locations. We also observed tissue-specific expression of some OGs, possibly indicating a complex regulatory process for maintaining diverse functional activity of these genes across sugarcane tissues and genotypes. Sixty-six OGs were differentially expressed under stress conditions, especially cold and osmotic stresses. Gene co-expression network and functional enrichment analyses suggested that sugarcane OGs are involved in several biological mechanisms, including stimulus response and defence mechanisms. These findings provide a valuable genomic resource for sugarcane researchers, especially those interested in selecting stress-responsive genes.

Keywords: orphan genes, sugarcane hybrid, stress condition, RNA-Seq, gene expression

OPEN ACCESS

Edited by:
Andrew H. Paterson, University of Georgia, United States

Reviewed by:
Igor Fesenko, Institute of Bioorganic Chemistry (RAS), Russia
Youxiong Que, Fujian Agriculture and Forestry University, China
Rasappa Viswanathan, Indian Council of Agricultural Research (ICAR), India

***Correspondence:**
Anete Pereira de Souza
anete@unicamp.br

†Present address:
Cláudio Benício Cardoso-Silva, Laboratório de Química e Função de Proteínas e Peptídeos, Universidade Estadual do Norte Fluminense, Rio de Janeiro, Brazil

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

Received: *18 April 2022*
Accepted: *13 June 2022*
Published: *30 June 2022*

Citation:
Cardoso-Silva CB, Aono AH, Mancini MC, Sforça DA, da Silva CC, Pinto LR, Adams KL and de Souza AP (2022) Taxonomically Restricted Genes Are Associated With Responses to Biotic and Abiotic Stresses in Sugarcane (Saccharum spp.). Front. Plant Sci. 13:923069. doi: 10.3389/fpls.2022.923069

INTRODUCTION

Recent advances in sugarcane genomics have created opportunities to systematically reveal the evolutionary history and diversification of the *Saccharum* group. However, the complexity of the sugarcane genome, mainly due to its size, ploidy level, and large number of mobile elements (Thirugnanasambandam et al., 2018), has hindered advances in the genomics of this important crop species. Despite the economic importance of sugarcane due to its use as a source of sugar, biofuel, and fibre, its reference genomes, including a chromosome-level *Saccharum spontaneum* genome (Zhang et al., 2018), a monoploid genome from the R570 variety (Garsmeur et al., 2018), and an SP80-3280 hybrid genome (Souza et al., 2019), have been only recently reported.

Two events of whole-genome duplication (WGD) are thought to have occurred during the evolution of the *Saccharum* group (Ming et al., 1998; Paterson et al., 2012). WGD is a major mechanism responsible for species diversification and adaptation (Soltis et al., 2009; Renny-Byfield and Wendel, 2014). These recent events of polyploidization occurring within the Saccharinae group provide an opportunity to investigate the fate of duplicated genes. Genome duplication initially results in gene duplication and gene redundancy. After duplication, some gene copies preserve their original function, while most of them are eliminated through negative selection (Tautz and Domazet-Lošo, 2011). However, some copies under positive selection, after sequence diversification, may acquire a new biological function (Van de Peer et al., 2009). Divergence of pre-existing genes is one of the mechanisms underlying the emergence of new genes (Tautz and Domazet-Lošo, 2011). An alternative origin has been proposed: new genes originate from a non-coding sequence (Singh and Syrkin Wurtele, 2020; Vakirlis et al., 2020).

Taxonomically restricted, lineage-specific or orphan genes (OGs), which have no homology to genes in other taxa, may contribute to evolutionary novelties and might be responsible for some lineage-specific trait origins (Wilson et al., 2005; Khalturin et al., 2009; Tautz and Domazet-Lošo, 2011). Even though we have not given sufficient attention to these genes, comparative genomic studies have estimated that OGs constitute at least 1% of the total genes in a genome, depending on the alignment rate and taxonomic level considered (Khalturin et al., 2009; Arendsee et al., 2014; Prabh and Rödelsperger, 2016). Several studies have been carried out to characterize OGs in plants at the species level: Arabidopsis (Li and Wurtele, 2014), sweet orange (Xu et al., 2015), rice (Guo et al., 2007), moso bamboo (Zhang et al., 2022), and at the family level: Brassicaceae (Donoghue et al., 2011) and Poaceae (Campbell et al., 2007). However, there is limited information about the function of most of these OGs, as they lack recognizable domains and functional motifs.

OGs are known to play a role in primary metabolism and response to environmental changes in plants. By establishing a gene-editing system, Jiang et al. (2020) revealed that an orphan gene (*BrOGs*) in *Brassica napus* plays a vital role in soluble sugar metabolism. In *Arabidopsis*, a functional analysis of the well-studied orphan gene *QQS* (*qua quine starch*) indicated that it could act in regulation of nitrogen allocation, affecting the protein content (O'Conner et al., 2018). Additionally, there are several works showing that OGs are regulated in response to biotic and abiotic stresses (Beike et al., 2014; Giarola et al., 2014; Khraiwesh et al., 2015; Schlötterer, 2015; Kaur et al., 2017). For example, an OG named *TaFROG* enhanced wheat resistance to Fusarium head blight (Perochon et al., 2015) and an OG in *Vigna unguiculata* (*UP12_8740*) increased plant tolerance to osmotic stresses and soil drought (Li et al., 2019). A *Physcomitrium patens* OG (*PpARDT*) was functionally characterized, and knockout mutant displayed reduced drought tolerance (Dong et al., 2022).

Despite the biological relevance of these taxonomically restricted genes, no previous reports described their occurrence and expression profile in the *Saccharum* complex. To advance our knowledge about OGs in sugarcane, a comparative genomic approach is needed for their identification, followed by regulatory inference based on gene expression analysis. In this study, we identified and characterized sugarcane OGs and their expression patterns across tissues and genotypes. Additionally, we analysed expression data from different conditions to identify those under which these genes are positively or negatively regulated in sugarcane.

MATERIALS AND METHODS

Orphan Gene Identification

A phylostratigraphic approach based on a sequence homology search was used to identify OGs in the sugarcane genome (Domazet-Lošo et al., 2007; McLysaght and Hurst, 2016). These analyses rely on sequence alignments to detect genes that lack homology in a focal species in comparison with a target clade. For this analysis, we used the gene model from *S. spontaneum* (Zhang et al., 2018) as a reference. First, the protein and coding DNA sequence (CDS) files containing the set of sugarcane genes were filtered using the CD-HIT package v.4.8.1 (Fu et al., 2012); a similarity threshold of 90% was applied for both the CD-HIT and CD-HIT-EST algorithms, which were employed for the protein and CDS files, respectively. This step was performed to remove redundancies in the dataset once all the homologous genes and duplications were included in the annotated sugarcane genome. Subsequently, a series of local alignments using both the sugarcane proteome and CDSs were performed to remove genes with homology in other species. First, to reduce the subset of candidate genes, we filtered out all sugarcane genes with detected homology to genes annotated in 13 angiosperm species including seven Poaceae species (*Arabidopsis thaliana* TAIR10, *Brachypodium distachyon* v3.1, *Citrus sinensis* v3.1, *Eucalyptus grandis* v2.0, *Miscanthus sinensis* 7.1, *Oryza sativa* 7.0, *Phaseolus vulgaris* v2.1, *Panicum virgatum* v4.1, *Setaria italica* v2.2, *Solanum lycopersicum* ITAG3.2, *Sorghum bicolor* v3.1.1, *Oropetium thomaeum* v1.0, and *Zea mays* 284 v6). The annotated sequences were downloaded from Phytozome v.13 (Goodstein et al., 2011) and converted into a database. The remaining subset of sugarcane genes was aligned to

non-redundant proteins (NR) and nucleotides (NT) from the National Center for Biotechnology Information (NCBI) database, and genes without homology in previous filtering steps were discarded. In all filtering steps, we used BLASTp and BLASTn to detect homology at the protein and nucleotide levels, respectively. All results were generated using a permissive E-value cut-off ≤10–6, allowing more relationships to be detected and increasing the chance of selecting a real OG. To discard the hypothesis that predicted OGs were missing from the genome annotation, we mapped each OG back on the chromosomes of seven species downloaded from Phytozome v.13 (*A. thaliana*, *Panicum hallii*, *S. italica*, *Z. mays*, *S. bicolor*, *S. spontaneum*, and *M. sinensis*) using sim4 software, which employs a splice-aware alignment method (Florea et al., 1998).

Orphan Gene Characteristic Features and Sequence Homology

The FASTA files containing sugarcane chromosome information and gff3 files were used for manual curation of the OGs. The position of each OG exon was used as a starting point to check intron/exon boundaries as well as the presence of start and stop codons using Artemis software v.18.0 (Carver et al., 2011). To characterize the physical and chemical properties of sugarcane genes (OGs and non-OGs), we calculated protein parameters [protein length, molecular weight, the instability index, hydrophobicity, the isoelectric point, and the grand average of hydropathicity (GRAVY)] using ProtParam tools implemented in the Bio.SeqUtils package, a Biopython module (Cock et al., 2009). To assess the protein-coding potential of these genes, we estimated the probability of each OG being a coding RNA using Coding Potential Calculator (CPC2; Kang et al., 2017). Additionally, all predicted OGs were aligned to the non-coding RNA databases derived from Rfam, Tair, Ensemble long non-coding RNAs (lncRNAs), CANTATA (Szcześniak et al., 2015), and GreeNC (Paytuví Gallart et al., 2015) using nhmmer with an E-value parameter ≤0.001 (Wheeler and Eddy, 2013). We checked for transposable element (TE) insertion in the OG sequences by using a reference collection of transposons from the Repbase database (Bao et al., 2015) using CENSOR (Jurka et al., 1996). To search for homologues within the Saccharinae subtribe, we aligned predicted proteins of OGs to annotated proteins from the draft genomes of *Saccharum* hybrids SP803280 (Souza et al., 2019) and R570 (Garsmeur et al., 2018) and complete genomes from *M. sinensis* (Mitros et al., 2020) and *S. spontaneum* (Zhang et al., 2018) using BLASTp with an E-value $\leq 10^{-6}$.

RNA-Seq Experimental Data: Retrieval and Pre-processing

An extensive search for papers reporting RNA-Seq data in sugarcane was performed, followed by a search of the NCBI Sequence Read Archive (SRA) repository (**Supplementary Table 1**). The selected RNA-Seq samples were retrieved using the 'fastq-dump' program from the SRA toolkit (version 8.22), and SRA files were converted to fastq-format files. The raw reads were subjected to quality control using Trimmomatic v0.36 (Bolger et al., 2014) to remove adapter and low-quality sequences. Three reference transcriptomes, including two full-length transcriptomes, were also selected to confirm the selected OGs being transcribed. In the first set of IsoSeq data, RNA samples were extracted from the top and bottom internodes of 22 genotypes (Hoang et al., 2017), and in the second set, RNA samples were obtained from leaves of a commercial sugarcane variety from Thailand (Piriyapongsa et al., 2018). The third transcriptome, which was *de novo* assembled from short reads, was extracted from the leaves of six sugarcane hybrids (Cardoso-Silva et al., 2014). A local alignment using the BLASTn program was performed with an e-value cut-off $\leq 1 \times 10^{-6}$ to infer the homology of putative OGs to sugarcane transcripts.

Orphan Gene Expression Profile and Differential Expression

RNA-Seq libraries were constructed for two purposes: (i) to unveil the expression patterns of OGs across sugarcane tissues and genotypes and (ii) to identify DE OGs, especially under stress conditions. The expression level of each gene was estimated by mapping the transcriptomes against the whole gene set of sugarcane using Salmon (Patro et al., 2017). The expression level of a given gene was calculated by the log transformation method implemented in Salmon [transcripts per million (TPM)], which represents the relative abundance of a transcript among a population of transcripts. Heatmaps representing the expression levels of OGs in experiments testing for differential gene expression and evaluating expression across tissues/genotypes were created using the 'superheat' R package (Barter and Yu, 2018).

To investigate whether OGs were DE, we designed RNA-Seq experiments including biotic and abiotic stresses, developmental stages, and sucrose accumulation. Cleaned reads from each library originating from experiments with biological replicates were mapped to the complete set of sugarcane genes (CDS FASTA format) using Salmon v.0.12.0 to quantify transcript abundance (Patro et al., 2017). The DESeq2 package v.3.9 (Love et al., 2014) was used to predict DEGs in each experiment using the raw read counts as input data. In cases where the same sample was sequenced in multiple runs, the technical replicates were collapsed before starting the DEG analysis. To minimize quantification biases, genes with fewer than 10 reads mapped per sample were filtered out before the gene expression analysis. The DEGs were estimated assuming a negative binomial distribution for each gene, applying a function that estimates the size factor and reducing bias caused by library size (normalization by the median ratio; gene count divided by the sample size). A value of $p < 0.05$ and an absolute log2 fold change ≥2 were used as thresholds for determining whether genes were apparently DE. For each predicted OG, hypothesis testing was performed, in which the null hypothesis was no differences between the control and treatment groups, thus supporting the assumption that any difference in gene expression occurred merely by chance.

Orphan Gene Co-expression Network and Functional Annotation

The expression matrix of the 218 samples across sugarcane tissues and genotypes was used to build a co-expression network with the 'WGCNA' R package (Langfelder and Horvath, 2008). A weighted adjacency matrix was constructed using pairwise Pearson's correlation coefficient measures and an estimated power threshold for scale-free independence ($R^2 > 0.8$ and largest mean connectivity). Subsequently, the calculated matrix was converted to a topological overlap matrix (TOM), which evaluates gene pair correlations and the degree of agreement with other genes in the matrix (Yip and Horvath, 2007). After that, we inferred network functional modules by employing average linkage hierarchical clustering in accordance with the TOM-based dissimilarity measure. We used a soft threshold power of 7 (R^2 of 0.81 and mean connectivity of 164) to calculate the TOM. Clusters were defined according to a hierarchical dendrogram using adaptive branch pruning as implemented in the 'dynamicTreeCut' R package (Langfelder et al., 2007). We conducted functional enrichment analysis of the modules containing OGs based on Gene Ontology terms. Then, we assumed a guilty-by-association approach to obtain some insight into OG functionality. Additionally, we checked for OG similarity with known protein domains using hmmscan from the HMMER3 suite (Finn et al., 2011), aligning the domains to the Pfam v35 database (Mistry et al., 2020).

RESULTS

Identification and Characterization of Sugarcane Orphan Genes

After removing redundancies in the sugarcane gene model, which contained 83,826 genes from *S. spontaneum* (Zhang et al., 2018), we obtained a total of 51,675 NR protein-coding genes. These sugarcane genes were aligned to the proteomes of 13 angiosperm species, including seven Poaceae species, represented by 435,957 proteins. This alignment returned 1,536 sugarcane genes with no homology with any protein represented. Next, these subsets of 'no-hit' genes were aligned to the NR protein database, and 442 genes with no homology were found. Finally, these remaining sets were aligned to the NT database. As a result, a total of 335 genes were identified as sugarcane OGs due to their lack of homology to other genes.

Homology searches may fail to detect homologues in other species, resulting in spurious OG prediction. To minimize this effect, we did not rely only on homology searches but also mapped OG CDSs onto each chromosome of seven grass genomes (*S. spontaneum*, *M. sinensis*, *S. bicolor*, *P. hallii*, *Z. mays*, *S. italica*, and *O. sativa*) and the genome of *A. thaliana*. Although OGs were not annotated as complete genes, except in the *Saccharum* group, we found vestiges of exons of these genes in all grass chromosomes. However, we did not detect OG vestiges in *A. thaliana* chromosomes (**Supplementary Table 2**).

To better understand the distribution of these putative OGs across grass genomes, we assessed chromosome regions in which dispersed fragments of the OGs aligned with at least 10% of their length. Intriguingly, the closer the phylogenetic relationship with *Saccharum* was, the higher the number of OG fragments observed in grass genomes (**Figure 1**; **Supplementary Table 2**). The number of OG fragments ranged from 114 in the *O. sativa* genome to 91,379 in the *M. sinensis* genome. If we consider the *S. spontaneum* genome, the number of OG fragments is even larger. Curiously, there are 25 times more OG fragments in *S. spontaneum* and *M. sinensis* chromosomes than in the sorghum genome, which is the closest relative of these Saccharinae species.

To verify whether OGs are evolutionarily conserved within the Saccharinae subtribe, we searched for OGs homologs in the *Saccharum* spp., including the annotated genome of *M. sinensis*. We found 201 OGs with similarity to other sequences predicted to be protein-coding genes in Saccharinae genomes. Most homologs were present in the *Saccharum* hybrid genome; however, we also found 39 homologs genes in the *M. sinensis* genome. Furthermore, 127 OGs have duplicate copies in *S. spontaneum*, annotated as alleles or paralogues (**Supplementary Figure 1**; **Supplementary Table 3**). Additionally, because many OGs were not predicted as coding genes in other *Saccharum* genomes, we aligned them to non-coding RNA sequences. However, we did not detect similarity of any OGs with ncRNAs deposited in public databases, including lncRNAs.

Using a vector machine-based classifier named CPC2, 152 OGs were classified as ncRNAs, and 167 OGs were classified as coding RNAs (**Supplementary Figure 2**; **Supplementary Table 4**). A total of 89 OGs were predicted as protein-coding genes with a high probability ($>=0.9$), while 47 others were classified as lncRNAs. However, short OGs were more likely to be classified as ncRNAs ($R^2 = 0.73$, value of $p < 2.2e^{-16}$; **Supplementary Figure 3**). A search for protein domains within OG sequences revealed that most OGs do not show similarity with any functional domains deposited in the Pfam database. Based on these searches, we detected only partial local alignment for 30 OGs (BLAST searches produced no significant alignments to domains; E-value $\leq$1E-3), of which nine were predicted to be domains of unknown function (**Supplementary Table 4**).

Physical and chemical analyses revealed that OGs and non-OGs were significantly different in all parameters except GC content (**Supplementary Figure 4**; **Supplementary Table 4**). OGs are shorter than non-OGs, with average protein lengths of 136 and 450, respectively. The number of exons in OGs varied from 1 to 21 ($\bar{x} = 3.54$; $\sigma = 2.37$), and 57.9% had three exons or fewer (**Supplementary Table 4**). The average GC contents of orphan and non-OGs were almost identical, at 56.7 and 56.5%, respectively. Most sugarcane genes had a negative GRAVY (grand average of hydropathy) value, i.e., ~81% of the OGs and ~80% of the non-OGs, which supports the protein being hydrophilic.

TE Fragments in the OGs

The large number of gene fragments found in the Saccharinae genomes may indicate that some of these genes originated from TE duplication. We aligned all the predicted OGs to TE sequences to test the hypothesis that some of the OGs are TEs and to verify whether some of the OGs were derived

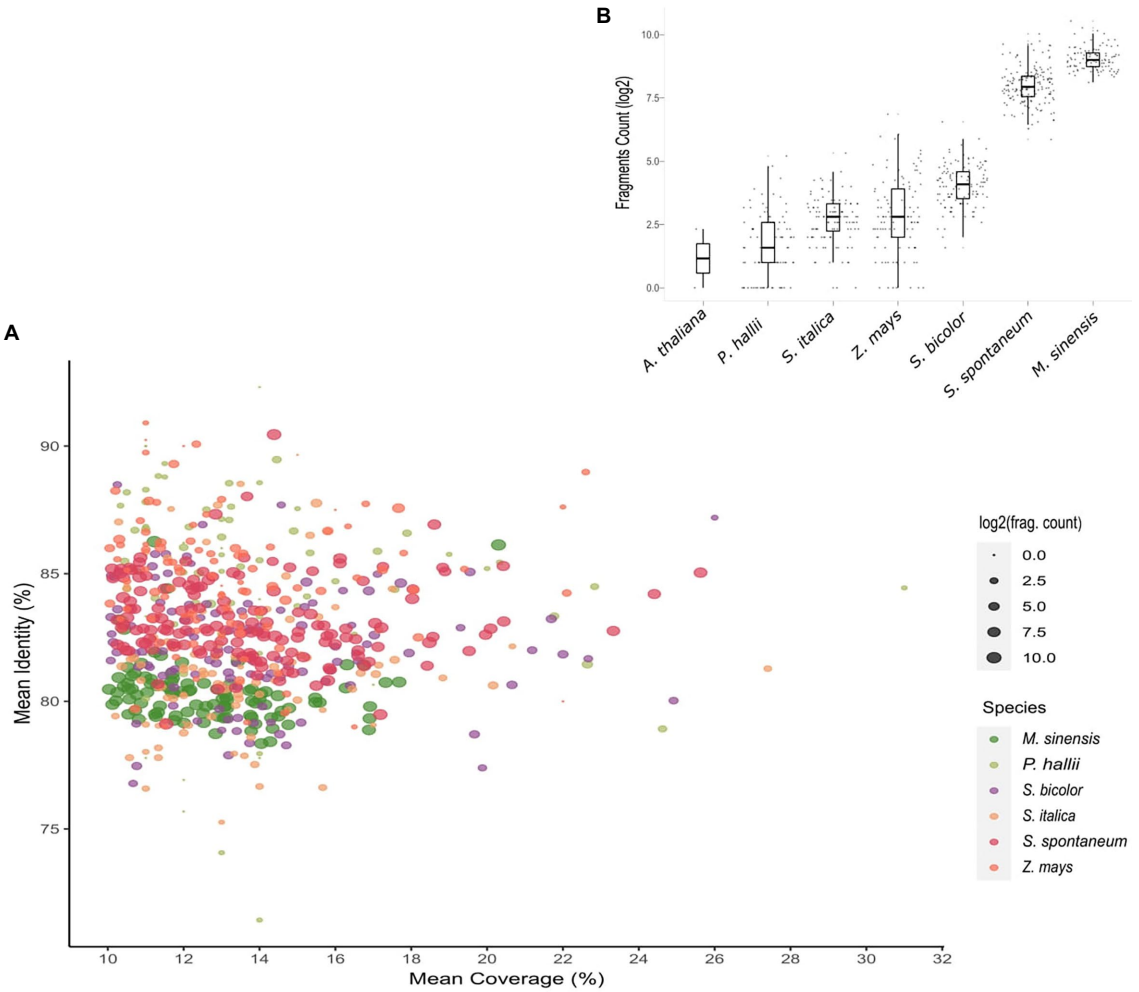


FIGURE 1 | Evidence of orphan gene (OG) vestiges within grass genomes. **(A)** Scatterplot showing the average percentage identity and coverage of each OG in six grass species. Each dot represents an individual OG. The dot sizes represent the number of fragments of each sugarcane OG in other grass species. The number of fragments of each OG in the focal species is also shown in the bar plot **(B)**.

from TE insertion (**Supplementary Table 5**). This search was motivated by previous observations suggesting that 51% of the OGs in rice are derived from TEs (Jin et al., 2019). To investigate this hypothesis more deeply and shed light on the origin of these putative lineage-specific genes, we performed an alignment of the initially selected set of genes (335 OGs) against the Repbase TE library.[1] A total of 153 putative OGs aligned to TEs with significant hits (E-value ≤1E-10). For a subset of these genes (16 OGs), at least 70% of the sequence aligned to TEs. We assumed that these genes were putative TEs, and we did not consider them to be sugarcane OGs. In the remaining subset (319 OGs), some genes had traces of TE insertion into the coding region, albeit with poor alignment. To better understand this result, we estimated the fractions of both OGs and non-OGs in sugarcane with similarity to TEs. Most sugarcane genes had traces of TEs in their coding region, and ~55% of the OGs and ~75% of the non-OGs had at least 10% of their sequence aligning to TEs (**Supplementary Figure 5**).

Evidence of Orphan Gene Expression Across Sugarcane Tissues and Genotypes

We searched for evidence that the 319 OGs were being transcribed across several sugarcane tissues and genotypes by aligning them against reference transcriptomes and RNA-Seq libraries. We selected two representative sugarcane IsoSeq datasets (Hoang et al., 2017; Piriyapongsa et al., 2018), a collection of transcripts from six sugarcane varieties (Cardoso-Silva et al., 2014), and 218 RNA-Seq samples from sugarcane hybrids, *Saccharum officinarum,* and *S. spontaneum* (**Supplementary Table 1**). Evidence of transcription was detected in 89.34% of the OGs (TPM value ≥1), which were expressed in at least one transcriptome experiment or tissue (**Supplementary Table 6**). Almost one-third of the OGs had an expression level considered low (1 ≤ TPM ≤ 10), while only 6% of them had a value greater than 100 TPM. In fact, when

[1]http://www.girinst.org/repbase

we compared the expression levels of OGs and non-OGs in four sugarcane tissues (**Supplementary Figure 6**), we found that OGs had proportionally lower expression levels. However, these differences were almost imperceptible in the meristem tissue (bud).

The expression profile of OGs across tissues and genotypes revealed a clear pattern of sample clustering. Overall, OGs showed similar expression patterns among genotypes when we compared the same tissue (**Figure 2** and **Supplementary Table 6**). Although the expression matrix combined tissues and genotypes, we also observed clustering by genotype to the same degree. For example, samples originating from *S. spontaneum* (Krakatau, IN84_58, and SES205A) were clustered together and separated from those originating from *S. officinarum* (BadilaDeJava, CriollaRayada, and WhiteTransparent) and *Saccharum* hybrids. Notably, samples originating from *Saccharum* hybrids and *S. officinarum* had more similar expression patterns than those originating from *S. spontaneum*. In particular, the expression pattern of OGs in the internodes of *S. spontaneum* was noticeably different from that in *S. officinarum* and hybrids. Furthermore, some OGs presented contrasting expression patterns between samples from *S. officinarum* and *S. spontaneum*. Specifically, a subset of 42 OGs had an average expression level that was three times higher in *S. officinarum* than in *S. spontaneum*. In contrast, the expression level of 35 OGs was higher in *S. spontaneum*. Interestingly, genotypes originating from the same geographical location tended to be clustered together because they had more similar expression profiles. In particular, this pattern was observed in the expression levels of OGs in internode 1 of French hybrids (F36.819 and R570) and the roots of Brazilian hybrids (RB855536, RB855113, and RB867515).

Interestingly, a few OGs seemed to have tissue-specific regulation, as observed for some genes only expressed in roots (Sspon.06G0001310 and Sspon.06G0024430), internode 1 (Sspon.08G0030340), internodes 3 and 8 (Sspon.05G0032460 and Sspon.08G0007980), and all internodes (Sspon.06G.0027080).

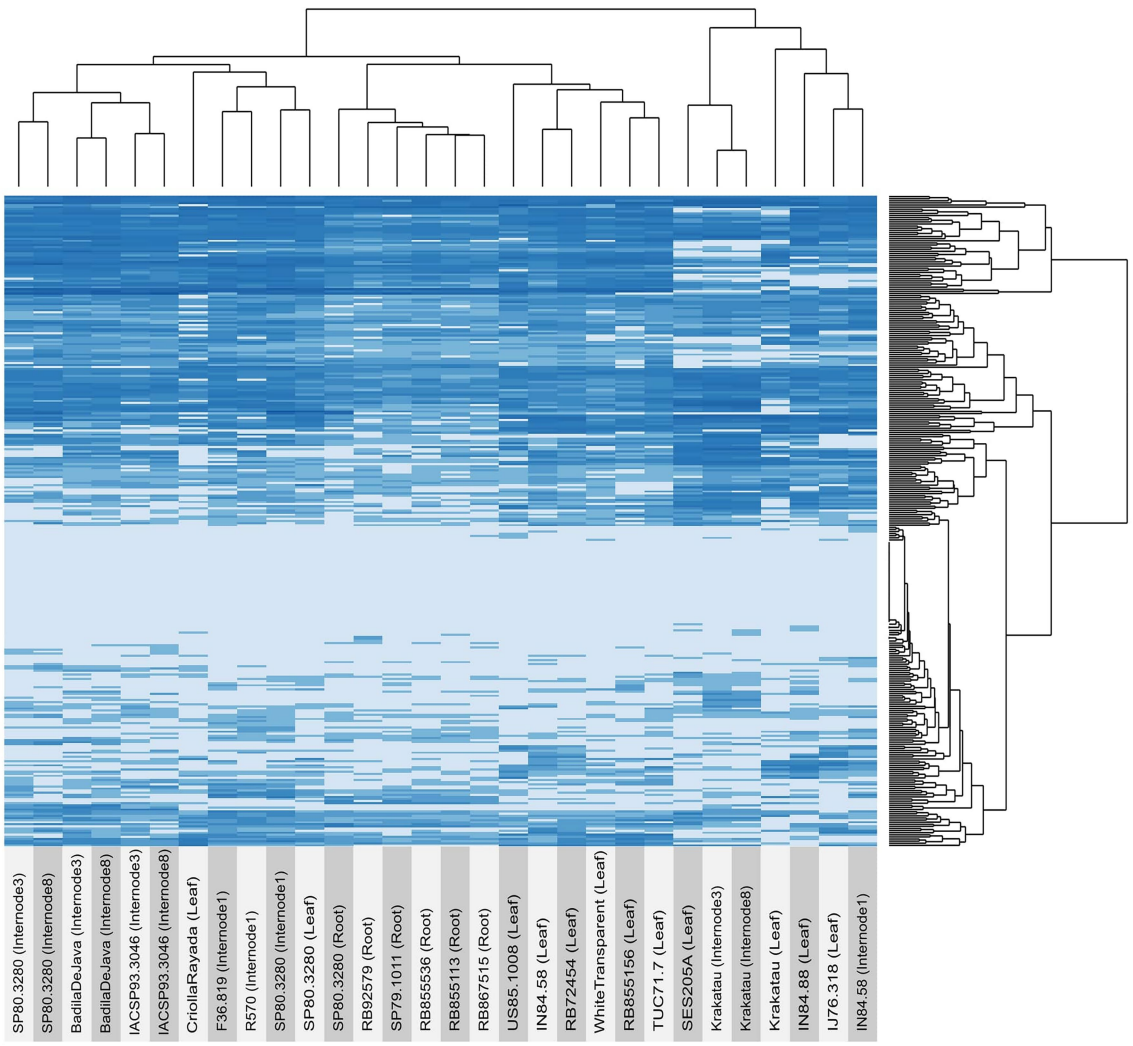


FIGURE 2 | Expression profiles of orphan genes (OGs) in several sugarcane tissues and genotypes. The expression of each gene (TPM) was estimated using a pseudoalignment method implemented in Salmon software.

Orphan Genes Are Differentially Expressed Under Stress Conditions

We performed eight RNA-Seq experiments representing a variety of conditions to test whether OGs change their expression levels (**Table 1**). After filtering the raw data, we selected more than 13 billion high-quality reads to perform gene expression analysis.

We observed at least one DE OG in five of the eight RNA-Seq experiments (**Supplementary Table 7**). We did not detect DE OGs in the RNA-Seq experiments related to developmental stage, sucrose accumulation, and plant infection with *Sporisorium scitamineum*. Generally, more genes were DE in the experiments related to abiotic stress than in those related to biotic stress. For example, we estimated that 6,440 genes were DE under cold stress (in genotypes Guitang08-1,180 and ROC22), while only 1,548 genes were DE after infection with yellow canopy syndrome, and 2,612 genes were DE in plants infected with smut disease. Overall, we identified 66 OGs that were DE under at least one type of stress ($\log_2$FoldChange ≥ 2; padj < 0.05). Most of the genes were regulated by abiotic stresses, while only 4 OGs were regulated by biotic stresses. We detected only two OGs (Sspon.02G0052680-1C and Sspon.06G0009900-2C) that were DE in both the osmotic stress and cold stress experiments, and one gene (Sspon.04G0024550-1B) showed DE in both the cold stress and low-nitrogen experiments.

In the cold stress experiment, we identified 8,921 DE genes in the hybrid Guitang08-1,180 (5,644 upregulated and 3,277 downregulated) and 8,913 DE genes in the hybrid ROC22 (5,596 upregulated and 3,317 downregulated), and 72% of these genes were DE in both genotypes. A total of 50 OGs were determined to be DE in this experiment, with 33 OGs in Guitang08-1,180 (12 upregulated and 21 downregulated) and 36 OGs in the ROC22 hybrid (12 upregulated and 24 downregulated), while 18 OGs were DE in both genotypes (**Figure 3** and **Supplementary Table 7**). An OG that was upregulated in both genotypes (Sspon.06G00100300) has three conserved copies annotated in the *S. spontaneum* genome (**Supplementary Figure 7**).

In the osmotic stress experiment conducted on leaves and root samples from *S. officinarum*, a total of 4,207 genes were DE in the leaves (1,815 upregulated and 2,392 downregulated), while 4,222 genes were DE in the root samples (1,707 upregulated and 2,515 downregulated). Of these genes, 13 OGs were identified as DE, nine in the leaf samples and six in the root samples. Two OGs, Sspon.01G0060030-1D and Sspon.05G0013120-1A, were detected as DE in both the root and leaf tissues (**Figure 4** and **Supplementary Table 7**).

In the sugarcane plants exposed to low-nitrogen conditions, most of the DE genes were detected in the leaf samples (4,524 genes in the Badila variety and 2,345 genes in the ROC22 hybrid), while in the roots, 894 and 726 genes were estimated to be DE in Badila and ROC22, respectively. In the root samples, the number of genes upregulated was three times greater than the number of genes downregulated. This pattern was observed in both genotypes. We did not detect DE OGs in the roots of either genotype; however, in the leaves, we detected six DE genes in Badila (four upregulated and two downregulated) and four DE genes in the ROC22 hybrid (one upregulated and three downregulated; **Figure 5** and **Supplementary Table 7**).

Co-expression Network and Modules With OGs

The co-expression network was built with the full set of sugarcane genes, including 288 OGs with estimated expression values. The sugarcane genes were distributed among 153 modules, of which 78 had at least one OG. More than one-third of the OGs (120 of 319) were assigned to seven modules. An enrichment analysis of these modules containing OGs (Fisher's exact test; value of $p \leq 0.05$) suggested that the genes are associated with several biological processes (**Supplementary Table 8**). Overall, the most frequent Gene Ontology terms observed in modules containing OGs included those associated with responses to several stimuli and defence mechanisms (**Supplementary Figure 8**). We identified a module with a set of co-expressed genes containing 64 OGs, which were enriched in Gene Ontology function terms related to transport (GO:0006810), response to starvation (GO:0042594), and response to external stimulus (GO:0009605). We also found modules enriched with genes associated with protein modification

TABLE 1 | RNA-Seq experiments performed for sugarcane gene expression analysis.

Accession	Condition	Cultivar/Species	Tissue	Rep[1]	References
PRJNA474042	yellow canopy syndrome	Hybrid	leaf	5	Marquardt et al., 2019
PRJNA479814	developmental stages	Q208 and KQ228	root/leaf/ internode	3	Thirugnanasambandam et al., 2019
PRJNA483518*	cold stress	Guitang08-1,180 and ROC22	leaf	4	Tang et al., 2018
PRJNA533093	low nitrogen	Badila	leaf/root	3	Yang et al., 2019
PRJNA291816	smut disease	RB925345	bud	3	Bedre et al., 2019
PRJNA415122	*S. scitamineum* infection	CP74_2005	bud	3	McNeil et al., 2018
PRJNA371469	osmotic stress	*S. officinarum*	root/leaf	2	Pereira-Santana et al., 2017
PRJNA681593	sucrose accumulation	Hybrids	top and bottom internodes	3	Aono et al., 2021

Represented by multiple accessions (see more details in **Supplementary Table 1).*
[1]Number of biological replicates.

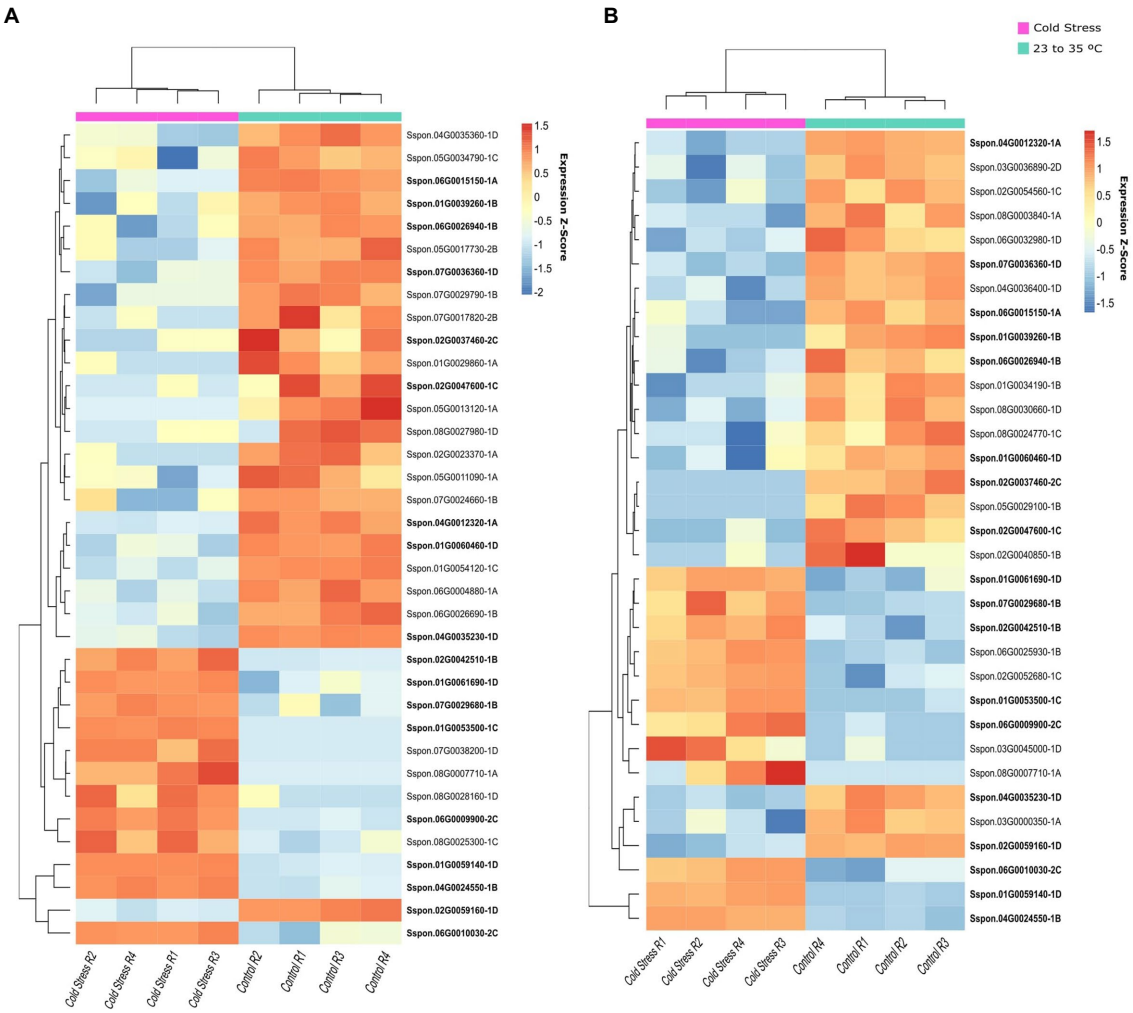


FIGURE 3 | Orphan genes (OGs) differentially expressed (DE) under cold stress. Hierarchical clustering of the genes expressed at normal and cold temperatures with four replicates for each treatment (row). DE analysis was carried out in leaf tissues at an ambient temperature (ranging from 23°C to 35°C) and a cold temperature (4°C in a well-controlled climate chamber) in two sugarcane genotypes: ROC22 **(A)** and Guitang08-1180 **(B)**.

(GO:0032446; module containing 13 OGs) and regulation of DNA methylation (GO:0044030; 10 OGs) as well as two modules with eight OGs each, which were associated with general processes such as macromolecule metabolic process (GO:0043170) and catabolic process (GO:0009056). Interestingly, a total of 24 OGs DE under stress conditions were included in modules enriched in genes functionally associated with response to stimuli, such as response to nutrient levels (GO:0031667), response to fungus (GO:0009620), and response to stress (GO:0006950).

DISCUSSION

Identification and Characterization of Sugarcane OGs

The definition and determination of OGs are context dependent, but generally, a gene that lacks detectable sequence homology in other taxa is typically classified as an OG (Rödelsperger et al., 2019). OGs with no homology to genes in other species are candidates for the *de novo* evolution of genes. However, we cannot reject the hypothesis that, in some cases, these homolog genes were missing in these target species because the genome is incomplete, missing sequences that could be biologically informative, including entire genes. Additionally, most OGs identified in our study did not show sequence similarity to any known functional domain. There are two possible reasons for this, which are linked to OG origin, complete divergence from ancestral sequences or *de novo* emergence from non-genic sequences (Van Oss and Carvunis, 2019; Singh and Syrkin Wurtele, 2020; Vakirlis et al., 2020). In the first scenario, a pair of genes sharing a common ancestor can diverge at some point when similarity is no longer detectable. Because of the lack of similarity with known functional domains, it is also possible that these genes are ncRNAs. Even though we did not observe any OGs similar to ncRNAs deposited in databases, some OGs were predicted

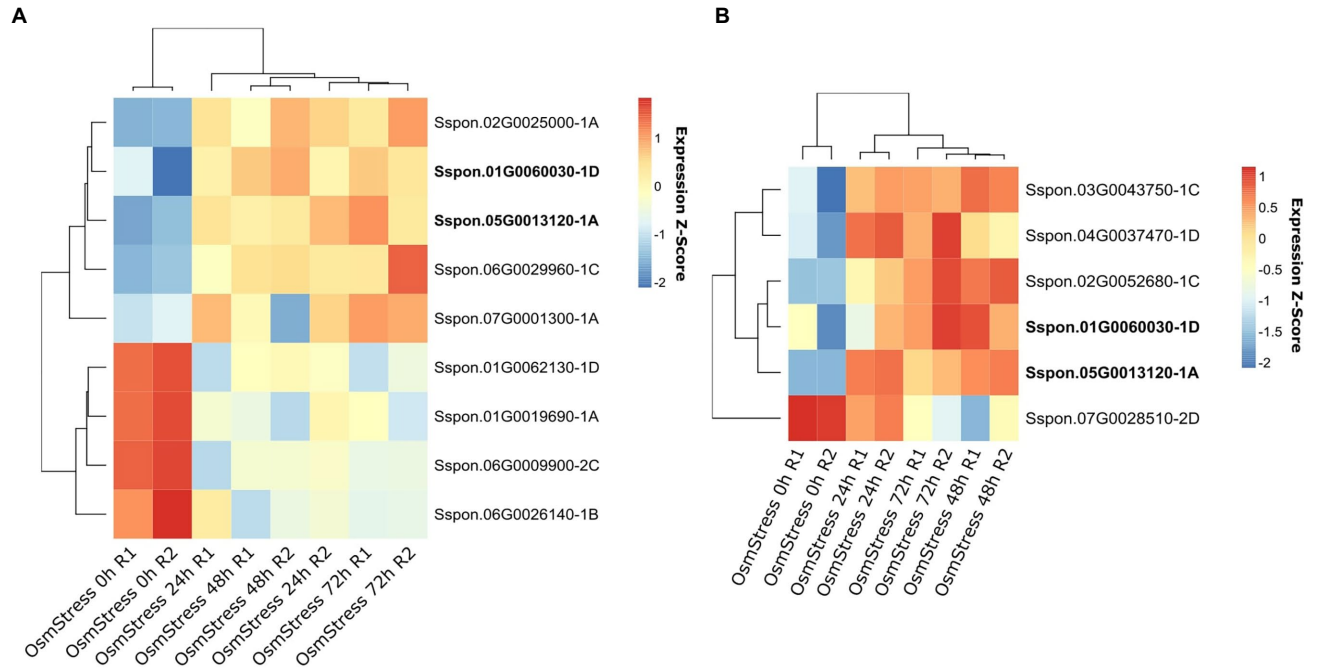


FIGURE 4 | Heatmap of orphan genes (OGs) differentially expressed (DE) under osmotic stress. Plants were subjected to osmotic stress for 24, 48, and 72 h, and other plants were maintained without osmotic stress (0 h). RNA samples were extracted from the leaves **(A)** and roots **(B)** of *S. officinarum*.

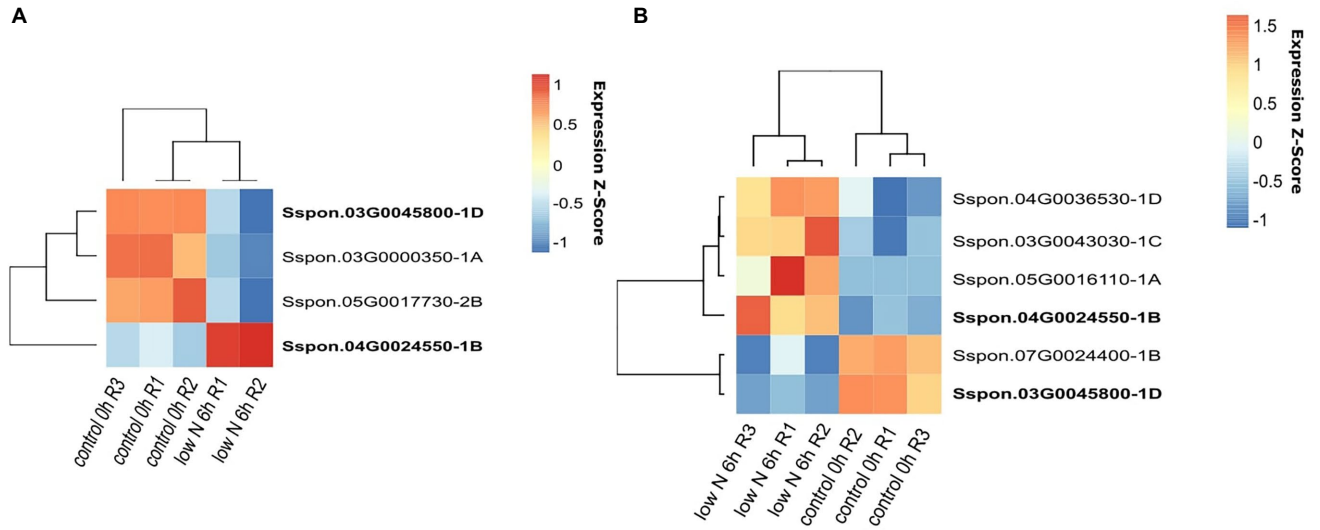


FIGURE 5 | Genes differentially expressed (DE) under low-nitrogen conditions. RNA samples were extracted from leaves of the ROC22 **(A)** and Badila **(B)** genotypes.

to be lncRNAs based on the coding potential calculator. However, this method assumes that a protein-coding RNA is more likely to have a long open reading frame than a non-coding transcript. This likely explains the strong positive correlation that we observed between gene length and the probability of being a coding sequence. Consequently, short OGs tend to be assigned as lncRNAs.

TEs Might Be Involved in OGs Origin in Sugarcane

The presence of TE fragments within a coding region is not surprising. One of the molecular mechanisms by which TEs are recruited to be part of a gene is named exonization. It takes place after TE insertion into an intron, after which parts of the TE can be incorporated, leading to the presence of a

TE exon in a protein-coding gene (Schrader and Schmitz, 2018). TE-derived proteins have been recurrently domesticated during evolution, and they have contributed to adaptive evolutionary innovation (Jangam et al., 2017; Schrader and Schmitz, 2018). In support of this view, a mechanism of OG birth *via* TEs has been reported in rice (Jin et al., 2019). Similarly, most OGs predicted in primates were found to include fragments of TEs in the transcript (Toll-Riera et al., 2008). In the sugarcane genome, we observed that most genes, both orphan and non-orphan, contained TE traces producing poor alignment in the coding region. This may indicate frequent recruitment of TEs as part of novel genes.

OGs Expression Patterns and Responses to Stress Conditions

The expression profiles of OGs across sugarcane hybrids may indicate that these genes have similar expression patterns when we compare genotypes with the same geographical origin. Indeed, it would be expected that hybrids from the same breeding programme would show greater genetic similarity, including regulatory elements, because of the admixture among genotypes sharing a common parentage. We also observed that the expression pattern of the hybrids was more similar to that of *S. officinarum* than to those of other species, suggesting preserved regulatory control of the expression of these genes after hybridization. These findings also suggest that genotype clustering is influenced by parental genome contributions. Indeed, all modern sugarcane varieties originated from hybridization between *S. officinarum* and *S. spontaneum*, followed by several backcrosses using *S. officinarum* to restore a high sucrose content (Price, 1961). This breeding process resulted in unequal contributions of the sub-genomes in *Saccharum* hybrids. For instance, the genome of the R570 hybrid received approximately 80% of its chromosomes from *S. officinarum* (D'Hont, 2005), which could explain the similar expression patterns between hybrids and *S. officinarum*. Furthermore, the observation that some OGs have tissue-specific expression may indicate a sophisticated mechanism controlling their expression, perhaps mediated by TEs. Previous studies revealed that TEs could contribute to the emergence of new regulatory elements in a tissue-specific manner (Feschotte, 2008; Sundaram et al., 2014; Trizzino et al., 2018), which is a theory based on Barbara McClintock's discovery that TEs can control gene expression (McClintock, 1956). This is a plausible conjecture because more than 70% of the sugarcane genome is represented by TEs, and fragments of these elements were found in some OGs.

Evidence of OGs being regulated under stress conditions has been reported previously. Studies have demonstrated the importance of lineage-specific genes in plants subjected to biotic and abiotic stresses. For example, in cowpea (*V. unguiculata*), OGs seem to be more involved than conserved genes in drought adaptation, as OG expression was highly induced compared to conserved gene expression under drought conditions (Li et al., 2019). Similarly, the expression levels of two OGs, CpCRP1 and CpEDR1, are modulated when individuals of a model plant widely studied for understanding the mechanism of desiccation are subjected to dehydration and rehydration processes (Giarola et al., 2014). Here, most sugarcane OGs were DE in experiments in which plants were exposed to abiotic stress. However, we cannot confirm that all these genes changed their expression pattern as an adaptive response to these stresses. Further investigation needs to be carried out to experimentally validate the effectiveness of these genes for minimizing stress effects. Even though these findings were not experimentally validated, to confirm that these genes were DE, we observed a significant number of OGs that were up- and downregulated simultaneously in independent genotypes and analyses in the same experiment.

Modules Containing OGs Are Functionally Enriched With Stress-Response Genes

The functional prediction of OGs based on sequence homology is not possible. Hence, we built a co-expression network to provide some insight into the functionality of these genes. This prediction method follows the guilty-by-association rationale (Zhang and Horvath, 2005), where the functional enrichment of modules containing genes co-expressed with OGs suggests the potential biological roles of those OGs. In general, we observed OGs distributed in several network modules, indicating that these genes are involved in diverse biological activities. This is in accordance with the findings of previous studies that described OGs as being involved in several biological networks (Beike et al., 2014; Giarola et al., 2014; Khraiwesh et al., 2015; Schlötterer, 2015). Accordingly, the main biological function attributed to these genes, the stress response, is itself a complex mechanism involving multiple biological pathways (Shulaev et al., 2008; Mantri et al., 2011). We highlight the evidence that most modules containing OGs were functionally enriched in biological processes related to stimulus responses, including those associated with various stresses. In addition to canonical terms related to stresses, such as defence responses and responses to stimuli, we also detected modules enriched, for example, with genes associated with sulphate transport and responses to auxin. Genes functionally related to these processes play an important role in stress responses (Rahman, 2012; Chan et al., 2013; Shani et al., 2017).

Because sugarcane OGs have no homology with genes in other organisms, we cannot infer their functions based on homology searches. In fact, in most cases, we did not find a known domain in the OG sequences, suggesting that sequence divergence is not the main source of OG origin in sugarcane. Even though we did not obtain strong evidence of their functionality, we cannot discard the relevance of these genes for understanding the unique biological aspects of the *Saccharum* lineage. To advance our knowledge, further investigation needs to be undertaken to better understand how these genes effectively participate in the stress response.

To date, the potential biotechnological applications of a few OGs have been tested. The QQS gene, an *Arabidopsis* gene involved in carbon and nitrogen allocation, was introduced into the soybean genome and increased starch and protein levels in the leaves (Li and Wurtele, 2014; Li et al., 2015; O'Conner et al., 2018). OGs

were also validated *via* gene editing as vital for soluble sugar metabolism in brassicas (Jiang et al., 2020). Despite the biological relevance of the OGs, it is still unknown how many of them are functional and produce stable proteins (Schlötterer, 2015; McLysaght and Hurst, 2016). Although OGs are not essential for survival, they may play an important role in responses to environmental stresses (Arendsee et al., 2014; Ma et al., 2020).

In our study, we developed an approach for the identification of OGs in sugarcane and characterization of their expression patterns. We propose that non-coding regions might provide important genetic raw material for the functional innovation, by which novel ORFs are selected and may evolve into adaptive stress response pathways. Finally, the OGs responsive to abiotic stress in sugarcane might be good candidates for further experiments to investigate their biological functions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

CC-S, MM, and DS conducted the experiments. CC-S and AA analysed the data. CC-S wrote the manuscript. All authors discussed the data, interpreted the results, read and edited the manuscript, and approved the final version.

FUNDING

This work was supported by grants from the Fundação de Amparo à Pesquisa de Estado de São Paulo (FAPESP, 08/52197–4), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES—Computational Biology Program 88882.160095/2013–01). CC-S received a postdoctoral fellowship from FAPESP (2015/16399–5 and BEPE 2017/26781–0); AA received a PhD fellowship from FAPESP (2019/03232–6); and CS and MM received postdoctoral fellowships from FAPESP (CS 2015/24346–9 and MM 2014/11482–9). AS received a Research Fellowship from CNPq (312777/2018–3).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for financial support and fellowships.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.923069/full#supplementary-material

REFERENCES

Aono, A. H., Pimenta, R. J. G., Garcia, A. L. B., Correr, F. H., Hosaka, G. K., Carrasco, M. M., et al. (2021). The wild sugarcane and Sorghum Kinomes: insights Into expansion, diversification, and expression patterns. *Front. Plant Sci.* 12:8623. doi: 10.3389/fpls.2021.668623

Arendsee, Z. W., Li, L., and Wurtele, E. S. (2014). Coming of age: orphan genes in plants. *Trends Genet.* 19, 698–708. doi: 10.1016/j.tplants.2014.07.003

Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6:11. doi: 10.1186/s13100-015-0041-9

Barter, R. L., and Yu, B. (2018). Superheat: an R package for creating beautiful and extendable heatmaps for visualizing complex data. *J. Comput. Graph. Stat.* 27, 910–922. doi: 10.1080/10618600.2018.1473780

Bedre, R., Irigoyen, S., Schaker, P. D. C., Monteiro-Vitorello, C. B., Da Silva, J. A., and Mandadi, K. K. (2019). Genome-wide alternative splicing landscapes modulated by biotrophic sugarcane smut pathogen. *Sci. Rep.* 9:8876. doi: 10.1038/s41598-019-45184-1

Beike, A. K., Lang, D., Zimmer, A. D., Wüst, F., Trautmann, D., Wiedemann, G., et al. (2014). Insights from the cold transcriptome of *Physcomitrella patens*: global specialization pattern of conserved transcriptional regulators and identification of orphan genes involved in cold acclimation. *New Phytol.* 205, 869–881. doi: 10.1111/nph.13004

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Campbell, M. A., Zhu, W., Jiang, N., Lin, H., Ouyang, S., Childs, K. L., et al. (2007). Identification and characterization of lineage-specific genes within the *Poaceae*. *Plant Physiol.* 145, 1311–1322. doi: 10.1104/pp.107.104513

Cardoso-Silva, C. B., Costa, E. A., Mancini, M. C., Balsalobre, T. W. A., Canesin, L. E. C., Pinto, L. R., et al. (2014). De novo assembly and transcriptome analysis of contrasting sugarcane varieties. *PLoS One* 9:e88462. doi: 10.1371/journal.pone.0088462

Carver, T., Harris, S. R., Berriman, M., Parkhill, J., and McQuillan, J. A. (2011). Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28, 464–469. doi: 10.1093/bioinformatics/btr703

Chan, K. X., Wirtz, M., Phua, S. Y., Estavillo, G. M., and Pogson, B. J. (2013). Balancing metabolites in drought: the sulfur assimilation conundrum. *Trends Genet.* 18, 18–29. doi: 10.1016/j.tplants.2012.07.005

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163

D'Hont, A. (2005). Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenet. Genome Res.* 109, 27–33. doi: 10.1159/000082378

Domazet-Lošo, T., Brajković, J., and Tautz, D. (2007). A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* 23, 533–539. doi: 10.1016/j.tig.2007.08.014

Dong, X.-M., Pu, X.-J., Zhou, S.-Z., Li, P., Luo, T., Chen, Z.-X., et al. (2022). Orphan gene PpARDT positively involved in drought tolerance potentially by enhancing ABA response in Physcomitrium (Physcomitrella) patens. *Plant Sci.* 319:111222. doi: 10.1016/j.plantsci.2022.111222

Donoghue, M. T., Keshavaiah, C., Swamidatta, S. H., and Spillane, C. (2011). Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana*. *BMC Evol. Biol.* 11, 1–23. doi: 10.1186/1471-2148-11-47

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405. doi: 10.1038/nrg2337

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367

Florea, L., Hartzell, G., Zhang, Z., Rubin, G. M., and Miller, W. (1998). A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* 8, 967–974. doi: 10.1101/gr.8.9.967

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

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., et al. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat. Commun.* 9:2638. doi: 10.1038/s41467-018-05051-5

Giarola, V., Krey, S., Frerichs, A., and Bartels, D. (2014). Taxonomically restricted genes of *Craterostigma plantagineum* are modulated in their expression during dehydration and rehydration. *Planta* 241, 193–208. doi: 10.1007/s00425-014-2175-2

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

Guo, W.-J., Li, P., Ling, J., and Ye, S.-P. (2007). Significant comparative characteristics between orphan and nonorphan genes in the rice (Oryza sativa L.) genome. *Comp. Funct. Genom* 2007:21676. doi: 10.1155/2007/21676

Hoang, N. V., Furtado, A., Mason, P. J., Marquardt, A., Kasirajan, L., Thirugnanasambandam, P. P., et al. (2017). A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genomics* 18:395. doi: 10.1186/s12864-017-3757-8

Jangam, D., Feschotte, C., and Betrán, E. (2017). Transposable element domestication As an adaptation to evolutionary conflicts. *Trends Genet.* 33, 817–831. doi: 10.1016/j.tig.2017.07.011

Jiang, M., Zhan, Z., Li, H., Dong, X., Cheng, F., and Piao, Z. (2020). *Brassica rapa* orphan genes largely affect soluble sugar metabolism. *Hortic. Res.* 7:181. doi: 10.1038/s41438-020-00403-z

Jin, G. H., Zhou, Y. L., Yang, H., Hu, Y. T., Shi, Y., Li, L., et al. (2019). Genetic innovations: transposable element recruitment and de novo formation lead to the birth of orphan genes in the rice genome. *J. Syst. Evol.* 59, 341–351. doi: 10.1111/jse.12548

Jurka, J., Klonowski, P., Dagman, V., and Pelton, P. (1996). Censor—a program for identification and elimination of repetitive elements from DNA sequences. *Computers & Chemistry* 20, 119–121. doi: 10.1016/s0097-8485(96)80013-1

Kang, Y.-J., Yang, D.-C., Kong, L., Hou, M., Meng, Y.-Q., Wei, L., et al. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* 45, W12–W16. doi: 10.1093/nar/gkx428

Kaur, N., Chen, W., Zheng, Y., Hasegawa, D. K., Ling, K.-S., Fei, Z., et al. (2017). Transcriptome analysis of the whitefly, *Bemisia tabaci* MEAM1 during feeding on tomato infected with the crinivirus, tomato chlorosis virus, identifies a temporal shift in gene expression and differential regulation of novel orphan genes. *BMC Genomics* 18:370. doi: 10.1186/s12864-017-3751-1

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., and Bosch, T. C. G. (2009). More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* 25, 404–413. doi: 10.1016/j.tig.2009.07.006

Khraiwesh, B., Qudeimat, E., Thimma, M., Chaiboonchoe, A., Jijakli, K., Alzahmi, A., et al. (2015). Genome-wide expression analysis offers new insights into the origin and evolution of *Physcomitrella patens* stress response. *Sci. Rep.* 5:17434. doi: 10.1038/srep17434

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* 9:559. doi: 10.1186/1471-2105-9-559

Langfelder, P., Zhang, B., and Horvath, S. (2007). Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. *Bioinformatics* 24, 719–720. doi: 10.1093/bioinformatics/btm563

Li, G., Wu, X., Hu, Y., Muñoz-Amatriaín, M., Luo, J., Zhou, W., et al. (2019). Orphan genes are involved in drought adaptations and ecoclimatic-oriented selections in domesticated cowpea. *J. Exp. Bot.* 70, 3101–3110. doi: 10.1093/jxb/erz145

Li, L., and Wurtele, E. S. (2014). The QQS orphan gene of *Arabidopsis* modulates carbon and nitrogen allocation in soybean. *Plant Biotechnol. J.* 13, 177–187. doi: 10.1111/pbi.12238

Li, L., Zheng, W., Zhu, Y., Ye, H., Tang, B., Arendsee, Z. W., et al. (2015). QQS orphan gene regulates carbon and nitrogen partitioning across species via NF-YC interactions. *Proc. Natl. Acad. Sci. U. S. A.* 112, 14734–14739. doi: 10.1073/pnas.1514670112

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Ma, S., Yuan, Y., Tao, Y., Jia, H., and Ma, Z. (2020). Identification, characterization and expression analysis of lineage-specific genes within *Triticeae*. *Genomics* 112, 1343–1350. doi: 10.1016/j.ygeno.2019.08.003

Mantri, N., Patade, V., Penna, S., Ford, R., and Pang, E. (2011). "Abiotic stress responses in plants: present and future," in *Abiotic Stress Responses in Plants*. eds. P. Ahmad and M. Prasad (New York: Springer), 1–19.

Marquardt, A., Henry, R. J., and Botha, F. C. (2019). Midrib sucrose accumulation and sugar transporter gene expression in YCS-affected sugarcane leaves. *Trop. Plant Biol.* 12, 186–205. doi: 10.1007/s12042-019-09221-7

McClintock, B. (1956). Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* 21, 197–216. doi: 10.1101/sqb.1956.021.01.017

McLysaght, A., and Hurst, L. D. (2016). Open questions in the study of de novo genes: what, how and why. *Nat. Rev. Genet.* 17, 567–578. doi: 10.1038/nrg.2016.78

McNeil, M. D., Bhuiyan, S. A., Berkman, P. J., Croft, B. J., and Aitken, K. S. (2018). Analysis of the resistance mechanisms in sugarcane during *Sporisorium scitamineum* infection using RNA-seq and microscopy. *PLoS One* 13:e0197840. doi: 10.1371/journal.pone.0197840

Ming, R., Liu, S.-C., Lin, Y.-R., da Silva, J., Wilson, W., Braga, D., et al. (1998). Detailed alignment of *Saccharum* and Sorghum chromosomes: comparative Organization of Closely Related Diploid and Polyploid Genomes. *Genetics* 150, 1663–1682. doi: 10.1093/genetics/150.4.1663

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Mitros, T., Session, A. M., James, B. T., Wu, G. A., Belaffif, M. B., Clark, L. V., et al. (2020). Genome biology of the paleotetraploid perennial biomass crop Miscanthus. *Nat. Commun.* 11:5442. doi: 10.1038/s41467-020-18923-6

O'Conner, S., Neudorf, A., Zheng, W., Qi, M., Zhao, X., Du, C., et al. (2018). "From arabidopsis to crops: the arabidopsis QQS orphan gene modulates nitrogen allocation across species," in *Engineering Nitrogen Utilization in Crop Plants*. eds. A. Shrawat, A. Zayed and D. Lightfoot (Cham: Springer), 95–117.

Paterson, A. H., Wang, X., Li, J., and Tang, H. (2012). "Ancient and recent polyploidy in monocots," in *Polyploidy and Genome Evolution*. eds. P. S. Soltis and D. E. Soltis (Berlin: Springer), 93–108.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197

Paytuví Gallart, A., Hermoso Pulido, A., Martínez, A., de Lagrán, I., Sanseverino, W., and Aiese Cigliano, R. (2015). GREENC: a wiki-based database of plant lncRNAs. *Nucleic Acids Res.* 44, D1161–D1166. doi: 10.1093/nar/gkv1215

Pereira-Santana, A., Alvarado-Robledo, E. J., Zamora-Briseño, J. A., Ayala-Sumuano, J. T., Gonzalez-Mendoza, V. M., Espadas-Gil, F., et al. (2017). Transcriptional profiling of sugarcane leaves and roots under progressive osmotic stress reveals a regulated coordination of gene expression in a spatiotemporal manner. *PLoS One* 12:e0189271. doi: 10.1371/journal.pone.0189271

Perochon, A., Jianguang, J., Kahla, A., Arunachalam, C., Scofield, S. R., Bowden, S., et al. (2015). TaFROG encodes a Pooideae orphan protein that interacts with SnRK1 and enhances resistance to the mycotoxigenic fungus Fusarium graminearum. *Plant Physiol.* 169:2895, –2906. doi: 10.1104/pp.15.01056

Piriyapongsa, J., Kaewprommal, P., Vaiwsri, S., Anuntakarun, S., Wirojsirasak, W., Punpee, P., et al. (2018). Uncovering full-length transcript isoforms of sugarcane cultivar Khon Kaen 3 using single-molecule long-read sequencing. *PeerJ* 6:e5818. doi: 10.7717/peerj.5818

Prabh, N., and Rödelsperger, C. (2016). Are orphan genes protein-coding, prediction artifacts, or non-coding RNAs? *BMC Bioinform.* 17:226. doi: 10.1186/s12859-016-1102-x

Price, S. (1961). Cytological studies in *saccharum* and allied genera VII. Maternal chromosome transmission by *S. officinarum* in intra- and interspecific crosses. *Bot. Gaz.* 122, 298–305. doi: 10.1086/336118

Rahman, A. (2012). Auxin: a regulator of cold stress response. *Physiol. Plant.* 147, 28–35. doi: 10.1111/j.1399-3054.2012.01617.x

Renny-Byfield, S., and Wendel, J. F. (2014). Doubling down on genomes: polyploidy and crop plants. *Am. J. Bot.* 101, 1711–1725. doi: 10.3732/ajb.1400119

Rödelsperger, C., Prabh, N., and Sommer, R. J. (2019). New gene origin and deep taxon phylogenomics: opportunities and challenges. *Trends Genet.* 35, 914–922. doi: 10.1016/j.tig.2019.08.007

Schlötterer, C. (2015). Genes from scratch--the evolutionary fate of de novo genes. *Trends Genet.* 31, 215–219. doi: 10.1016/j.tig.2015.02.007

Schrader, L., and Schmitz, J. (2018). The impact of transposable elements in adaptive evolution. *Mol. Ecol.* 28, 1537–1549. doi: 10.1111/mec.14794

Shani, E., Salehin, M., Zhang, Y., Sanchez, S. E., Doherty, C., Wang, R., et al. (2017). Plant stress tolerance requires auxin-sensitive Aux/IAA transcriptional repressors. *Curr. Biol.* 27, 437–444. doi: 10.1016/j.cub.2016.12.016

Shulaev, V., Cortes, D., Miller, G., and Mittler, R. (2008). Metabolomics for plant stress response. *Physiol. Plant.* 132, 199–208. doi: 10.1111/j.1399-3054.2007.01025.x

Singh, U., and Syrkin Wurtele, E. (2020). How new genes are born. *elife* 9:e55136. doi: 10.7554/eLife.55136

Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., et al. (2009). Polyploidy and angiosperm diversification. *Am. J. Bot.* 96, 336–348. doi: 10.3732/ajb.0800079

Souza, G. M., Van Sluys, M.-A., Lembke, C. G., Lee, H., Margarido, G. R. A., Hotta, C. T., et al. (2019). Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop. *GigaScience* 8:giz129. doi: 10.1093/gigascience/giz129

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., et al. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* 24, 1963–1976. doi: 10.1101/gr.168872.113

Szcześniak, M. W., Rosikiewicz, W., and Makałowska, I. (2015). CANTATAdb: A collection of plant long non-coding RNAs. *Plant Cell Physiol.* 57:e8. doi: 10.1093/pcp/pcv201

Tang, S., Yang, L., and Li, Y. (2018). Comparative analysis on Transcriptome Among different sugarcane cultivars Under low temperature stress[J]. *Biotechnol. Bull.* 34, 116–124. doi: 10.13560/j.cnki.biotech.bull.1985.2018-0522

Tautz, D., and Domazet-Lošo, T. (2011). The evolutionary origin of orphan genes. *Nat. Rev. Genet.* 12, 692–702. doi: 10.1038/nrg3053

Thirugnanasambandam, P. P., Hoang, N. V., and Henry, R. J. (2018). The challenge of analyzing the sugarcane genome. *Front. Plant Sci.* 9:616. doi: 10.3389/fpls.2018.00616

Thirugnanasambandam, P. P., Mason, P. J., Hoang, N. V., Furtado, A., Botha, F. C., and Henry, R. J. (2019). Analysis of the diversity and tissue specificity of sucrose synthase genes in the long read transcriptome of sugarcane. *BMC Plant Biol.* 19:160. doi: 10.1186/s12870-019-1733-y

Toll-Riera, M., Bosch, N., Bellora, N., Castelo, R., Armengol, L., Estivill, X., et al. (2008). Origin of primate orphan genes: A comparative genomics approach. *Mol. Biol. Evol.* 26, 603–612. doi: 10.1093/molbev/msn281

Trizzino, M., Kapusta, A., and Brown, C. D. (2018). Transposable elements generate regulatory novelty in a tissue-specific fashion. *BMC Genomics* 19:468. doi: 10.1186/s12864-018-4850-3

Vakirlis, N., Carvunis, A.-R., and McLysaght, A. (2020). Synteny-based analyses indicate that sequence divergence is not the main source of orphan genes. *elife* 9:e53500. doi: 10.7554/eLife.53500

Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 10, 725–732. doi: 10.1038/nrg2600

Van Oss, S. B., and Carvunis, A.-R. (2019). De novo gene birth. *PLoS Genet.* 15:e1008160. doi: 10.1371/journal.pgen.1008160

Wheeler, T. J., and Eddy, S. R. (2013). Nhmmer: DNA homology search with profile HMMs. *Bioinformatics* 29, 2487–2489. doi: 10.1093/bioinformatics/btt403

Wilson, G. A., Bertrand, N., Patel, Y., Hughes, J. B., Feil, E. J., and Field, D. (2005). Orphans as taxonomically restricted and ecologically important genes. *Microbiology* 151, 2499–2501. doi: 10.1099/mic.0.28146-0

Xu, Y., Wu, G., Hao, B., Chen, L., Deng, X., and Xu, Q. (2015). Identification, characterization and expression analysis of lineage-specific genes within sweet orange (*Citrus sinensis*). *BMC Genomics* 16:995. doi: 10.1186/s12864-015-2211-z

Yang, Y., Gao, S., Su, Y., Lin, Z., Guo, J., Li, M., et al. (2019). Transcripts and low nitrogen tolerance: regulatory and metabolic pathways in sugarcane under low nitrogen stress. *Environ. Exp. Bot.* 163, 97–111. doi: 10.1016/j.envexpbot.2019.04.010

Yip, A. M., and Horvath, S. (2007). Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinform.* 8:22. doi: 10.1186/1471-2105-8-22

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:1128. doi: 10.2202/1544-6115.1128

Zhang, X., Xuan, J., Yao, C., Gao, Q., Wang, L., Jin, X., et al. (2022). A deep learning approach for orphan gene identification in moso bamboo (*Phyllostachys edulis*) based on the CNN+transformer model. *BMC Bioinfo.* 23:162. doi: 10.1186/s12859-022-04702-1

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., et al. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* 50, 1565–1573. doi: 10.1038/s41588-018-0237-2

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

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

frontiers | Frontiers in Plant Science

REVIEW
published: 30 June 2022
doi: 10.3389/fpls.2022.932737

Interspecific Hybridization Is an Important Driving Force for Origin and Diversification of Asian Cultivated Rice *Oryza sativa* L.

*Jiawu Zhou[1], Ying Yang[1], Yonggang Lv[1], Qiuhong Pu[1], Jing Li[1], Yu Zhang[1], Xianneng Deng[1], Min Wang[1,2], Jie Wang[1,2] and Dayun Tao[1]**

[1] Yunnan Key Laboratory for Rice Genetic Improvement, Food Crops Research Institute, Yunnan Academy of Agricultural Sciences, Kunming, China, [2] Institute of Plant Resources, Yunnan University, Kunming, China

OPEN ACCESS

Edited by:
Cheng Shihua,
China National Rice Research Institute (CAAS), China

Reviewed by:
Longbiao Guo,
China National Rice Research Institute (CAAS), China
Guanglong Hu,
Beijing Academy of Agricultural and Forestry Sciences, China

***Correspondence:**
Dayun Tao
taody12@aliyun.com

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

Received: 30 April 2022
Accepted: 25 May 2022
Published: 30 June 2022

Citation:
Zhou J, Yang Y, Lv Y, Pu Q, Li J, Zhang Y, Deng X, Wang M, Wang J and Tao D (2022) Interspecific Hybridization Is an Important Driving Force for Origin and Diversification of Asian Cultivated Rice Oryza sativa L. Front. Plant Sci. 13:932737. doi: 10.3389/fpls.2022.932737

As one of the most important crops, Asian cultivated rice has evolved into a complex group including several subgroups adapting various eco-climate-systems around the globe. Here, we pictured a comprehensive view of its original domestication, divergences, and the origin of different subgroups by integrating agriculture, archeology, genetics, nuclear, and cytoplasm genome results. Then, it was highlighted that interspecific hybridization-introgression has played important role in improving the genetic diversity and adaptation of *Oryza sativa* during its evolution process. Natural hybridization-introgression led to the origin of *indica, aus,* and *basmatic* subgroups, which adapted to changing cultivated environments, and produced feral weedy rice coexisting and competing with cultivars under production management. Artificial interspecific hybridization-introgression gained several breakthroughs in rice breeding, such as developing three-line hybrid rice, new rice for Africa (NERICA), and some important pest and disease resistance genes in rice genetic improvement, contributing to the stable increase of rice production to meet the expanding human population. We proposed a series to exploit the virtues of hybridization-introgression in the genetic improvement of Asian cultivated rice. But some key issues such as reproductive barriers especially hybrid sterility should be investigated further, which are conducive to gene exchange between cultivated rice and its relatives, and even is beneficial to exploiting interspecific hybrid vigor. New technologies help introduce favorable genes from distant wild species to Asian cultivated rice, such as transgenic and genome editing systems. Rising introgression lines in a wider range with multi-donor benefits allele mining, understanding genetic network of rice growth and development, yield formation, and environmental adaptation. Then, integration of new tools and interspecific hybridization can be a future direction to develop more usable breeding populations which can make Asian cultivated rice more resilient to the changing climate and world.

Keywords: rice, origin, evolution, interspecific hybridization, introgression, diversity, adaptation

INTRODUCTION

The process of plant domestication often leads to genetic bottlenecks. Hybridization supplied chances to increase plant genetic diversity and adaptation ability with its expansion progress. Hybridization has been an important force in generating angiosperm species diversity (Soltis and Soltis, 2009). At least 25% of plant species, especially the youngest species, are involved in hybridization and potential introgression with other species (Mallet, 2005). It occurs when distinct populations, subspecies, or species come into contact, hybridize, and the gene pools are merged (Mallet, 2007). This often leads to the formation of a hybrid zone or a hybrid swarm (Nolte and Tautz, 2009). Hybridization can result in rapid genomic changes, which may lead to beneficial new phenotypes (Baack and Rieseberg, 2007). Carolus Linnaeus proposed a radical evolutionary hypothesis that new species could arise via hybridization (Larson, 1968). Interspecific hybridization occurred or is carried out in many plants (Kopecký et al., 2022), such as tomato (Lippman et al., 2007), cotton (Zhang et al., 2014), soybean (Wang et al., 2019), rice (Purugganan, 2019), grape (Foria et al., 2022), tobacco (Mino et al., 2022), and oak (Fu et al., 2022), served as an important mechanism for their genetic diversification, environment adaptation, and range expansion, as well as improvement of economical traits. Older crops, such as rice, also have longer time to diversify under cultivation and thus adapt to local environments as their geographic ranges widened (Milla and Osborne, 2021).

Without domesticated crops, very few members of human societies would survive if they had only a field of wild grain and herbs and their wits to sustain them (Doebley et al., 2006); thus, the continuous progress of domesticated crops improvement is the vital foundation for human survival, progress, and prosperity. Asia cultivated rice, *Oryza sativa*, is one of the most important crops to feed the world for thousands of years. It is believed to be domesticated from the wild relative, *Oryza rufipogon*, in East Asia, then evolved and expanded to form different climatic ecotypes over the world. Regarding the domestication of the Asian cultivated rice, one hypothesis is independent domestication of *indica* and *japonica,* the prototype of *O. sativa* evolved primarily along the foothills of the Himalayas in South Asia and its associated mountain ranges in mainland Southeast Asia and Southwest China, and the genetic differentiation developed parallel to the ecologic diversification process (Chang, 1976). Another hypothesis is that the domestication processes of Asian rice occurred only once, and the *japonica* originated from domesticated *indica* (Oka, 1988). The cloning and haplotype analysis of some domestication genes provided evidence to challenge the above hypotheses (Sang and Ge, 2007; Sweeney and McCouch, 2007; Sweeney et al., 2007; Tan et al., 2008), which initiated some questions, such as why are most of the domestication genes identical in dispersed subgroups in various eco-systems? And how did the subgroups originate in distinct geographical regions? Genome-wide molecular markers and sequence data supplied more evidence about the phylogenetic relationship among Asian cultivated varieties, *O. rufipogon* and *Oryza nivara* accession, which provided strong clues to the origin of different subgroups of *O. sativa* when its geographical scope is expanding (Huang et al., 2012; Civán et al., 2015; Kim et al., 2016; Moner et al., 2018; Choi et al., 2020; Gutaker et al., 2020). The recent advances in rice genetics and genomics enable us to better understand the process of improving genetic diversity and environmental adaptability of Asian cultivated rice through hybridization-introgression during its evolution process. In this study, we reviewed the stories of *O. sativa* about its domestication, distribution, introgression, variation, and eco-climatic-economic adaptation, especially focusing on how the crop became better and better to meet the challenges of changing environments through continuous interspecific hybridization-introgression with its wild relatives, and genetically exchange and recombine to improve its genetic diversity and environmental adaptation ability, thus ultimately formatted the most important crop containing several subgroups to adapt various cultivation environments, to meet the increasing food requirement of expanding human population.

ENVIRONMENTAL ADAPTATION AND NATURAL HYBRIDIZATION LED TO THE ORIGIN OF DIFFERENT SUBGROUPS IN *Oryza sativa*

Asian Cultivated Rice Is a Wide Distribution Species With High Diversification

Compared to the hunter-gatherer system, cultivating domesticated crops can feed the more human population. Ten thousand years ago, human societies began to transition from hunting-gathering to agriculture. Both archeological and archeobotanical evidence show that the Asian cultivated rice domestication began in the Yangtze Valley in China approximately 8,000–8,500 years ago (Higham and Lu, 1998; Fuller et al., 2009). Rice is widely cultivated from 55^{o}N in China to 36^{o}S in Chile, and grown under different conditions such as irrigated, rainfed lowland, rainfed upland, and flood-prone ecosystems (Khush, 1997). It was estimated that more than 4,120,000 rice cultivars and germplasm accessions have been recognized worldwide (Song et al., 2021). As one of the earliest domesticated cereal crops in the world, with its spread of cultivation, rice appears more and more abundant in morphological and genetic diversity to adapt to the various diverse and complex cultivation environments.

Cultivars of *O. sativa* were first classified into two major types, namely *indica* and *japonica,* according to morphological and serological characters as well as inter-varietal hybrid fertility (Kato et al., 1928). Based on morphological traits, *O. sativa* was further classified into three types, A, B, and C (Matsuo, 1952), named *indica*, *javanica*, and *japonica* (Morinaga, 1954). Oka (1958) subsequently referred to the above three types as *indica*, *tropical japonica*, and *temperate japonica* groups. Using 15 polymorphic loci coding for eight enzymes, 1688 traditional rice varieties from Asia were classified into six varietal groups (Glaszmann, 1987), and it roughly represents different

ecological types and geographical distribution. Two hundred and thirty-four accessions of rice were classified into five groups, corresponding to *indica, aus, aromatic, temperate japonica*, and *tropical japonica* by 169 nuclear SSR markers and two chloroplast loci (Garris et al., 2005). The classification of five subgroups was confirmed and accepted by subsequent studies that resulted from analysis of nuclear and cytoplast DNA markers, as well as genomic datum (McNally et al., 2009; Wang et al., 2018; Chen et al., 2019). However, some authors used *aromatic* to refer scented rice from *indica*, *temperate japonica*, *tropical japonica*, *aus*, and *aromatic* (Bourgis et al., 2008), while some authors used *aromatic* to refer a special subgroup from Southern Asia (Garris et al., 2005; Kovach et al., 2007). To avoid confusion, we used *basmatic* to refer this subgroup in this paper.

Temperate and Tropical *japonica* Were the Results of Environmental Adaptation and Selection

The *japonica* is the first subgroup of fully domesticated Asian cultivated rice (Huang et al., 2012; Choi et al., 2017), and it was largely confined to China in the early cultivation stage (Khush, 1997). The *temperate japonica* was the first major population separated from tropical landraces (Gutaker et al., 2020). Rice diversified into *temperate* and *tropical japonica* rice during a global cooling event about 4,200 years ago (d'Alpoim Guedes et al., 2015). Standing variation of cold tolerance genes underwent stepwise selection to facilitate cold adaptation to expand rice cultivation from high altitude to high latitude regions (Li et al., 2021). The Hap^{tej} of *COLDF* (*CHILLING-TOLERANCE DIVERGENCE F-box*) is responsible for enhanced cold tolerance in the *temperate japonica* cultivars (Zheng et al., 2022). *Temperate japonica* is the result of adaptation to cooling climatic conditions, and dispersed in North and Northeastern Asia, while *japonica* rice migrated to Southeast Asia, where it rapidly diversified as *tropical japonica* nearly 2,500 years ago (Gutaker et al., 2020).

The *indica*, *aus,* and *basmatic* Varieties Were Produced by Hybridization and Introgression

Some domesticated traits, such as erect growth, loss of grain shattering, shortened awns or awnless, and compact panicles, are essential for the transition from wild rice to the domesticated crop, and these key transitions occurred in the earliest steps of rice domestication. Grain dormancy, grain color, grain size and number, and environmental adaptability were selected during the post-domestication stage (Xu and Sun, 2021). Dissection of allelic variants of essential domestication genes can reveal strong evidence about the origins of several subgroups in *O. sativa*. Most of the widely disseminated domestication alleles appear to have originated in *japonica* or a *japonica*-like ancestor and were subsequently introgressed into the *indica* group (Kovach et al., 2007). *prog1* is a single-origin allele and was selected during Asian rice domestication (Jin et al., 2008; Tan et al., 2008). A 110-kb deletion closely linked to *PROG1,* referred to as *rapd,* also contributed to the transition from prostrate to erect growth in Asian cultivated rice; this 110-kb deletion also represents a single evolutionary event, as all tested Asian cultivated rice carries it (Wu et al., 2018). The identical allele of *prog1* and *rapd* across *O. sativa* strongly suggests that introgression occurred after the spread of cultivated rice. The *sh4* allele originated in the early stages of Asian rice domestication (Li et al., 2006; Lin et al., 2007; Zhang et al., 2009) and introgressed into all the progenies as the crop expanded. A similar example occurred in rice pericarp color. The main mutation *rc* allele originated in the *japonica* and moved into the *indica*, *aus*, and *basmatic*. A fragment of no more than 1 Mb of *japonica* DNA with the *rc* allele hitchhiked into most *indica* varieties (Sweeney et al., 2007).

Evidence combining agricultural production, archeology, genetics, nuclear, and cytoplasm genome datum supplied a strong indication that subgroups *indica*, *aus*, and *basmatic* were descended from distinct lineages (Civán et al., 2015; Carpentier et al., 2019; He et al., 2021; Zhang et al., 2021), accompanying hybridization and introgression events, and both ancient and more recent gene flow continues to dilute the once-distinctive gene pools (Kim et al., 2016; Wang et al., 2017). Thus, the course of origin in *indica*, *aus,* and *basmatic* has been gradually clear (**Figure 1**).

The *indica* rice is with the most abundant genetic diversity, widely geographically distributed, and adapted in tropical and subtropical regions. *O. rufipogon* and *O. nivara* are native to India and well-distributed there today. The archeological data for the Ganges River basin reveals that sedentary agricultural villages were established after 2500 B.C. (Fuller, 2006; Fuller et al., 2010); thus, *proto-indica* rice was of unclear domestication status in South Asia before the domesticated *japonica* was introduced into India. Domesticated *japonica* rice dispersed from China, south then west ward, via Myanmar, Assam, Bangladesh to India Plain (Kovach et al., 2007; Vaughan et al., 2008), or from China, northwest ward, via Hexi Corridor through the Silk Road trade, to Pakistan, to Northwest India, then the Indian Plain (Fuller et al., 2010; Fuller, 2011b; Silva et al., 2018). Once *japonica* was present alongside local unimproved *proto-indica* or wild populations, hybridization, and back-crossing could rapidly induce the origin of *indica* subgroup by selection regimes of farmers. The complete domestication of *indica* was only when the arriving of domesticated *japonica* from China and hybridized with local wild relatives or *proto-indica* about 4,000-5,000 years ago (Fuller, 2011a; Gross and Zhao, 2014). As a result, the *indica* subgroup contained the most genetic diversity, harboring almost all (seven of the eight) chloroplast haplotypes, and encompassing all the chloroplast diversity found in the *temperate* and *tropical japonica* (Garris et al., 2005).

The *aus* rice is mainly distributed across South Asia, mainly concentrated in the Indian subcontinent. The *aus* group is grown under rain-fed upland and lowland conditions, and some are adapted to the irrigated or deepwater cultivated system. It likely originated from the hybridization between *indica* and a local wild population, with the identical recessive bacterial blight resistant gene, *xa-5*. Some *aus* varieties still remain some characters of wild relatives of rice, such as spreading tillers and easy thresh (Khush et al., 2003), and are under incomplete domestication selections (Zhao et al., 2018). Molecular and

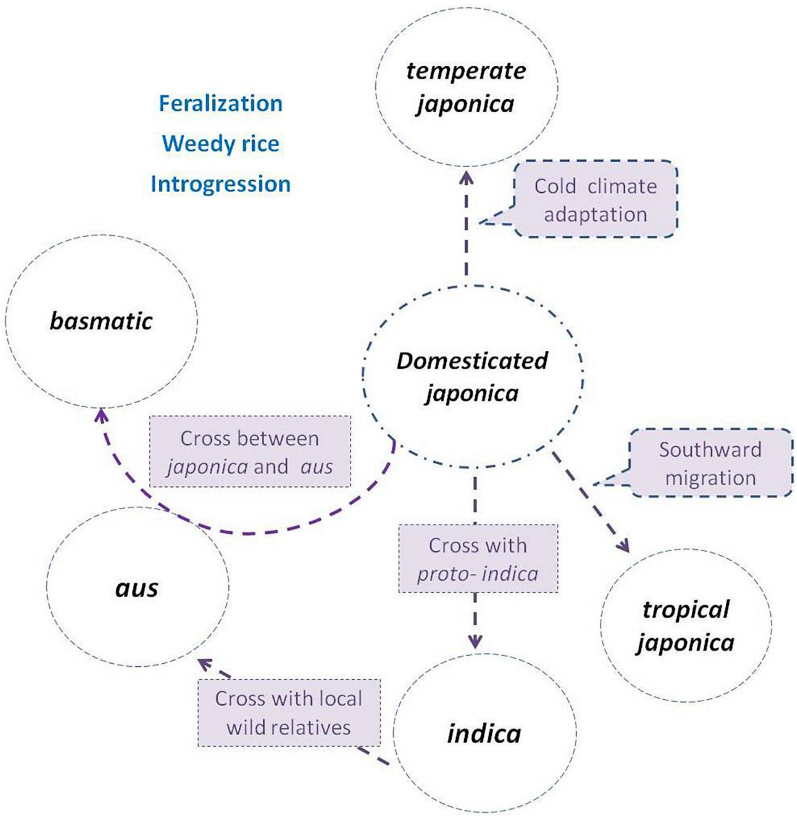


FIGURE 1 | Subgroup differentiation of *Oryza sativa* in the evolution process. The *japonica* rice (circle in the center) was domesticated first from *Oryza rufipogon*, then divided northward as *temperate japonica*, and south ward as *tropical japonica*. The *indica* was descent from hybridization between *japonica* and local wild populations or *proto-indica.* The *aus* was derived from the hybridization between *indica* and local wild populations, while the *basmatic* was derived from hybridization between *japonica* and *aus*. Weedy rice strains de-domesticated from and coexisted with cultivated subgroups, and frequently crossed with wild populations (if present) or landraces during their evolution process.

genomic studies suggested that the *aus* rice was originated from annual *Oryza nivara* (Kim et al., 2016), with distinct lineages of both nuclear and cytoplasm inheritance (Civán et al., 2015; Choi et al., 2017; He et al., 2021) and is relatively close to *indica* (Garris et al., 2005; Moner et al., 2018, 2020; He et al., 2021).

The *basmatic* rice is a small subgroup of *O. sativa*, identified as group V by Glaszmann (1987), including "Basmati" and "Sadri" varieties. The pleasure fragrance of rice genetically results from a recessive allele of *badh2.1* on chromosome 8 (Bourgis et al., 2008). Sequence analysis around *BADH2* indicated that they contain *badh2.1* allele originating from *japonica* (Kovach et al., 2009). The basmati rice, a typical representative of the *basmatic* group, is characterized as extra-long slender grain, elongation on cooking, soft and fluffy texture of cooked rice, and a pleasant aroma. The earliest reference to *basmatic* rice in India can be traced to the documents before 2,400 years ago. The *basmatic* is now a geographical indication, grown in the north-western foothills of the Himalayas in the Indian subcontinent, and is attributed to a unique combination of genetics, soil, water, climate, and cultural practices (Siddiq et al., 2012). Both the nuclear and chloroplast data demonstrate a close relationship to *japonica* (Garris et al., 2005). The *japonica* rice contributed the highest amount of genetic material to circum-*basmatic*, and strong evidence of admixture between the circum-*basmatic* and circum-*aus* groups was detected (Choi et al., 2020). A unique pericarp color haploid, *rc-s,* is shared only in *aus* and *basmatic* subgroups (Sweeney et al., 2007). The *basmatic* rice is a result of a hybridization between *japonica* and *aus* (Civán et al., 2015; Kishor et al., 2020).

Therefore, although a genetic bottleneck occurred in the early domestication stage, subsequent hybridization-introgression enabled *O. sativa* to enlarge its genetic diversity and differentiated into several subgroups to adapt to various cultivation environments. It is believed that Asian cultivated rice has the evolution progress of "One domestication and multiple origins" or "One origin and multiple interspecific introgressions." The *indica*, *japonica*, *aus,* and *basmatic* were parallel in origin, but gene introgressions among each other happened (He et al., 2021). The main domesticated genes were from *japonica* rice. Selective sweeps with reduced diversity were widespread in both *indica* and, in particular, *tropical japonica* (Molina et al., 2011). That is, there was only one domestication event in *O. sativa* (Gross and Zhao, 2014).

Weed Rice Is a Cospecific Form of Cultivar Rice, Resulted From Hybridization, Feralization, and Introgression, Adapting to the Improving Cultivation Environments

Gene flow from a crop plant to its wild progenitor is particularly effective when it helps the latter to mimic the crop (Ladizinsky, 1998). When wild relatives are growing near the farmers' field, they can cross with cultivars and form weedy types, which can be a serious problem to the farmers, particularly when the rice is broadcast sown, such as in deepwater rice fields in Asia river deltas (Vaughan and Sitch, 1991). Genes from cultivated rice can be easily incorporated into the gene pool of weedy rice through recurrent crop-to-weed gene flow and introgression. This progress may promote the persistent and better adaptation of weedy rice to the human-influenced habitats (Harlan, 1965). Typical specimens of *O. rufipogon* and *O. nivara* are rarely obtainable today even in their adapted habitats because of continuous hybridization among the wild, weed, and cultivated races (Chang, 1976). As wild species have been a serious weed in rice fields in India, the purple-leaved variety was introduced to allow the farmers to weed out the spontaneous green-leaved wild form in the field. However, no more than 20 years later, the purple-leaf gene had been introduced to the weedy population to such an extent that weeding according to leaf color was no longer a safe practice for eliminating wild species (Dave, 1943). Weedy rice now occurs worldwide and is prevalent in mechanized, direct-seeded farming systems. Genetic surveys have revealed multiple independent origins of weedy rice around the world. When wild *O. rufipogon* or *O. nivara* were absent, de-domestication from rice varieties or landrace plays a major role in the formation of weedy rice in northern temperate rice-growing regions. Interestingly, in the United States, almost all weedy strains de-domesticated from *aus*, *indica*, and *japonica* which originated and grew outside of the United States. When *O. rufipogon* or *O. nivara* was present in South Asia, and Southeast Asia, weedy rice strains were morphologically diverse, with characteristics ranging from resembling cultivated rice to similarity to wild rice, and showed evidence of recent hybridization between domesticated (or de-domesticated) and wild populations (Li and Olsen, 2020; Vigueira et al., 2020).

ARTIFICIAL INTERSPECIFIC HYBRIDIZATION AND INTROGRESSION LED TO SEVERAL BREAKTHROUGHS IN RICE BREEDING

Introgression of Sterile Cytoplasm From *Oryza rufipogon* to *Oryza sativa* Is the Start Point for the Three-Line Hybrid Rice Revolution

Chinese breeders began researching the utilization of rice heterosis in 1961. However, male-sterile plants found in cultivated rice are difficult to be maintained. In the Winter of 1970, a male-sterile plant was identified in *O. rufipogon* in Hainan Province, China. The cytoplasm of this male-sterile (CMS) *O. rufipogon* plant was introduced into *O. sativa* nuclear background, and successfully bred a series of wild abortive male sterility lines, which led to the famous "green revolution" of hybrid rice. By 2019, more than 7,000 hybrid rice cultivars had been released in China, with a total planting area of about 6×10^8 ha^2 and an increased rice grain of nearly 9×10^8 t, which had fed an additional 2.3 billion people. Chinese hybrid rice has been successfully tested or developed in more than 60 countries with planting areas exceeding 6×10^6 ha^2 (He et al., 2020).

New Rice for Africa Derived From Interspecific Hybridization Between *Oryza sativa* and *Oryza glaberrima*, Adapting Low Input Cultivation Conditions in West Africa

African cultivated rice *O. glaberrima* has been cultivated in West Africa for approximately 3,000 years (Linares, 2002). African rice has some undesirable features, such as seed shattering, brittle grain, lower yield potential, lodging, long seed dormancy, and being replaced by the Asian species, which was introduced into the continent by the Portuguese as early as the mid-sixteenth century (Jones et al., 1997; van Andel et al., 2016). The African rice has been described as the least diverse crop species ever documented (Nabholz et al., 2014). Despite losing substantial genetic diversity during domestication, it still retained important traits such as weed competitiveness, tolerance to drought, resistance to African gall midge, rice mottle virus, stem borer, and nematodes, as well as adaptation to acidic and low phosphorus soils among others (Wambugu et al., 2013). Although modern Asian cultivars performed better than farmers' traditional *O. glaberrima* varieties under relatively higher input conditions, they performed poorly when cultivated under the low input systems which dominate extensive upland farming in West Africa, due to the characteristics of modern Asian cultivars, such as poor competition with weeds and limited resistance to many of the stresses that affect upland rice in the region (Jones et al., 1997). Interspecific hybridization between *O. sativa* and *O. glaberrima*, subsequently backcrossing with *O. sativa*, produced progenies combining advantages of both species (high yield potential, weed-suppressing, and adaptation to low input conditions) and a new rice type to adapt West Africa upland condition was developed, denoted as NERICA (new rice for Africa), and this led to the naming of many rice cultivars for both rainfed upland and lowland irrigated system of West Africa (Jones et al., 1997; Nayar, 2010).

Miscellaneous New Genes Utilization and Breeding Achievements by Interspecific Introgression

During the 1970s, the grassy stunt virus is prevalent in several countries. Severe yield losses occurred under epidemic

conditions. Of the 6,723 cultivated rice accessions and several wild species screened for resistance, only one *O. nivara* accession Acc. 101508 was found to be resistant (Ling et al., 1970). A dominant resistant gene from *O. nivara* was denoted as *Gs* (Khush and Ling, 1974). Resistant varieties with *Gs*, such as IR28, IR29, IR30, IR32, IR34, IR36, and subsequent varieties were released in the prevalent regions, and grassy stunt infected plants are now rarely observed in farmers' fields (Khush and Brar, 2017).

Bacterial blight disease (BB) is one of the most serious diseases of rice in Asia. A dominant gene *Xa21* for resistance to BB was transferred from *O. longistaminata* and was widely used in many varieties in the Philippines, India, China, and other countries.

Recently, the progress of mining favorable genes from wild relatives involves more traits and genes/alleles. Great progress has been made in identifying genes/alleles resistant to biotic stresses. For blast resistance, *Pid3* from *O. rufipogon*, *Pi54rh* from *O. rhizomatis*, *Pi54* from *O. officinalis*, *Pi57*(t) from *O. longistiminata*, *Pi68*(t) from *O. glumaepatula*, and *Pi 69*(t) from *O. glaberrima* were identified (Das et al., 2012; Lv et al., 2013; Devanna et al., 2014; Xu et al., 2015; Devi et al., 2020; Dong et al., 2020). For bacterial blight, one novel resistant gene *xa-45*(t) from *O. glaberrima* was identified (Neelam et al., 2019). For sheath blight resistance, one major QTL *qShB9-2* was mined from *O. meridionalis* (Eizenga et al., 2013). For brown plant hopper (BPH) resistance, *qBph3* and *qBph4* from *O. officinalis*, *Bph34* from *O. nivara*, and *Bph27, Bph35,* and *Bph36* from *O. rufipogon* were identified (Huang et al., 2013; Hu et al., 2015; Kumar et al., 2018; Li et al., 2019; Zhang et al., 2020).

Some elite cultivars, such as Yundao1 [(IRGC102203 (*O. glaberrima*)/Boro5//Dianxi1///Hongza135], Yunlu142 [Yundao1/Acc.104613(*O. barthii*)//Yundao1/3/Yundao1] were bred via interspecific hybridization and released at Yunnan Academy of Agricultural Sciences (YAAS), Yunnan province, China.

INTERSPECIFIC INTROGRESSION LINE IS NOT ONLY A RESOURCE FOR FAVOR ALLELE MINING BUT ALSO AN EFFECTIVE WAY TO ANALYZE THE GENETIC BASIS OF HOMOGENEOUS GENES IN CULTIVATED RICE

The genetic difference between *O. sativa* and wild species has been observed in many studies. The hybrid progenies between cultivated and wild species usually showed abundant phenotypical and genotypical variation in various traits in segregated populations. The developing rice genomics has greatly contributed to the inherent nature of rice genetics and variation. However, focusing on a single trait or gene is convenient to elucidate its genetic nature by excluding disturbs of other genetic variances. Development of introgression libraries containing a single or few segments from wild species of rice into the cultivated background is the benefit to evaluate and dissect the genetic effect and breeding value for each genetic unit one by one. Series introgression libraries have been developed. Some libraries are single substitute lines (CSSLs) or backcross inbred lines (BILs) from a single donor, including *O. glaberrima* (Doi et al., 1997; Shim et al., 2010; Yamagata et al., 2019), *O. rufipogon* (Tian et al., 2005; Furuta et al., 2014; Qin et al., 2018; Ma et al., 2019; Yamagata et al., 2019), *O. longistaminata* (Ramos et al., 2016), *O. nivara* (Furuta et al., 2016; Yamagata et al., 2019), *O. meridionalis* (He et al., 2017), *O. bathii* (Bessho-Uehara et al., 2017; Zhao et al., 2019), and *O. glumaepatula* (Zhao et al., 2019). Compared with the rich genetic diversity of wild relatives of Asian cultivated rice, the introgression library with a single donor only represents a small part of its genetic diversity. Considering the great potential of genetic diversity in wild species, in recent studies, a multi-donor method was employed. In one study, 70 accessions of six AA genome species, *O. glaberrima*, *O. barthii*, *O. nivara*, *O. rufipogon*, *O. longistaminata*, and *O. glumaepatula,* were crossed with two elite cultivars of *O. sativa* L., then 1780 backcross inbred lines (BILs) were generated, and 15 BILs showed >10% yield superiority over the recurrent parents, with yield-related traits (Bhatia et al., 2017). A single-segment substitution lines (SSSLs) library, which includes 2360 SSSLs derived from 43 donors of seven species of rice AA genome in the Huajingxian 74 (HJX74) genetic background, was widely used to detect QTLs for traits of agronomic importance, clone genes of functional importance, and mine alleles of functional variants (Zhang, 2021). Recently, researchers developed 26,763 advanced backcross families using 309 donors of *O. glaberrima*, *O. barthii*, *O. nivara*, *O. rufipogon*, *O. longistaminata*, *O. glumaepatula*, *O. meridionalis,* and *O. sativa* (upland rice varieties). Based on phenotype selection, 6,732 introgression lines (ILs) with agronomic and morphological variation were raised. Using this library, 22 alleles responsible for grain shape were newly found (Zhang et al., 2022).

Steady agronomic and genetic interventions are essential for sustaining productivity in intensive rice cropping. Further progress in rice genetic improvement lies not only in the identification of more favored genes/alleles but also in a deep understanding of the genetic pathways of rice growth and development, yield formation, environmental response, and so on. However, as a result of extreme domestication, many excellent traits/genes in cultivated rice have been highly homogeneous, and the differences among varieties are difficult to be found, which is not conducive to the in-depth dissection of these important traits and their genetic networks. Compared with cultivated rice, wild relatives of the Asian cultivated rice have not been artificially domesticated and selected, they retain more genetic diversity, and the significant genetic difference between wild relatives and the Asian cultivated rice is an effective tool to analyze these genetic networks. Therefore, on one hand, the abundant genetic diversity is the elite resource for favor allele mining; on the other hand, as a result of long-term domestication and selection, many genes in cultivated rice are difficult to be characterized due to the lack of allelic differences. Exploring the huge genetic difference between cultivated rice and wild relatives is an effective way to analyze the genetic basis of these genes. The genetic dissection of *PROG1* gene

using the genetic difference between *O. sativa* and *O. rufipogon* is a good example (Jin et al., 2008; Tan et al., 2008). Thus, interspecific hybridization-introgression is not only useful for Asian cultivated rice to be high diversity and resilience to the changing condition but also useful for new gene discovery and genomics development, which makes rice to be a model crop for genomics development.

PROSPECT

Thousands of years of rice evolution have made great progress in increasing yield, meeting human eating preferences, and adapting to a variety of cultivation environments. In this process, interspecific hybridization-introgression played important impact on the improvement of adapting to different environments and ensuring the genetic diversity of rice (**Figure 1**). During this progress, more and more favor alleles or mutations have been accumulated in cultivars to improve the performance of the Asian cultivated rice. The progress of breeding programs, especially breakthrough progress, lies in further more favorable gene utilization. However, most of the naturally occurring variants in rice are of low frequency (Zhao et al., 2018). Thus, interspecific hybridization will continuously play an important role in the introgression of favor genes/alleles and increase its genetic diversity in the future practices of genetic improvement in rice. With the intensification of modern cropping systems, most of the wild relatives of *O. sativa* have lost their natural habitats, and interspecific hybridization-introgression in genetic diversity improvement of Asian cultivated rice would mainly depend on artificial implementation (**Figure 2**).

Artificial interspecific hybridization has made great progress in the genetic improvement of cultivated rice. However, due to many difficulties in distant interspecific hybridization and introgression, the progress in genetic improvement of Asian cultivated rice by interspecific hybridization is not comprehensive. First, due to strict reproductive barriers, compared with non-AA genome species, wild relatives with the same AA genome as the Asian cultivated rice are easier to obtain interspecific hybridization offspring with Asian cultivated rice. Therefore, more favorable genes/alleles have been found from wild relatives with the AA genome, while few favorable genes/alleles have been found from the non-AA genome. Second, because wild species grew in the natural environment for a long time, resistance/tolerance to many bio- and abio-stresses is a necessary condition for their survival. It is easier for us to find resistant/tolerant genes from them. The yield-related traits of rice are mostly quantitative traits, which are easy to be affected by environment and genetic background. The expression of these yield traits was unstable among individuals/generations. Although there were many reports on rice yield-related genes/QTLs (Gaikwad et al., 2021), the progress is limited. Third, many favorable genes from wild relatives have been discovered and cloned, but few of them are used in breeding practice. Further efforts should focus on several respects in the future.

Understanding and Overcoming Reproductive Barriers Between *Oryza sativa* and Wild Relatives, Including AA and Non-AA Genome Wild Relatives

For non-AA genome wild relatives, overcoming the reproductive barriers between them and the Asian cultivated rice is conducive to exploring a wider range of genetic variation. For AA genome relatives, in addition to exploring their genetic variation, the potential heterosis between the Asian cultivated rice and its relatives of the AA genome can be expected. Hybrids between varieties within subspecies of *indica* or *japonica* have shown great heterosis in the past half-century, and hybrids between subspecies or subgroups can gain more heterosis than the former. However, the hybrid sterility frequently occurred and hampered the utilization of the distant heterosis. Recent research demonstrated an example to overcome the inter-subspecies hybrid sterility (Guo et al., 2016), which greatly enhanced the confidence in the utilization of inter-subspecific heterosis between *indica* and *japonica*. The interspecific genetic difference between *O. sativa* and its relatives (including another cultivated species *O. glaberrima*) is greater than inter- and intra-subspecies, which show stronger hybrid vigor (Hu et al., 2002; Sun et al., 2020). Previous studies demonstrated that genetic knowledge about interspecific hybrid sterility can help us overcome this obstacle between *O. sativa* and its relatives (Deng et al., 2010; Xie et al., 2017; Yu et al., 2018). Future research needs to take efforts for the comprehensive identification of interspecific hybrid sterility genes between Asian cultivated rice and its relatives with potential heterosis, analyze their genetic mechanism, so as to overcome the obstacles of hybrid sterility and promote the utilization of interspecific vigor through the exploration of natural neutral alleles, and the development of bridge parents and gene editing of hybrid sterility genes.

New Techniques Help Wider Interspecific Introgression

Some wild species have favor characteristics, such as salinity tolerance of *O. rufipogon*, high biomass production of CCDD genome species, *O. latifolia, O. alta,* and *O. grandiglumis* (Khush and Brar, 2017). However, several pre- or post-zygotic barriers, such as hybrid sterility, genomic disharmony, nuclear instability, hybrid breakdown, reversion to parental types, limited recombination, presence of deleterious genes, and linkage drag, hinder the production of remote interspecific hybrids and the transfer of genes into crops through interspecific hybridization (Brar, 2004). The new technologies help to introduce favorable genes from distant wild species. In crop breeding, transgenic technology is an effective system to introduce favorable genes from distant species. Thanks to the development of efficient tissue culture, transformation, genome editing and high-quality genome assembly technologies, it is possible that artificial de novo domestication of a wild plant in a few generations. Recently, artificial *de novo* domestication of a wild plant gained great progress, including in wild tomato (Li et al., 2018), orphan crop *Physalis peruviana* (Lemmon et al., 2018), and wild relative

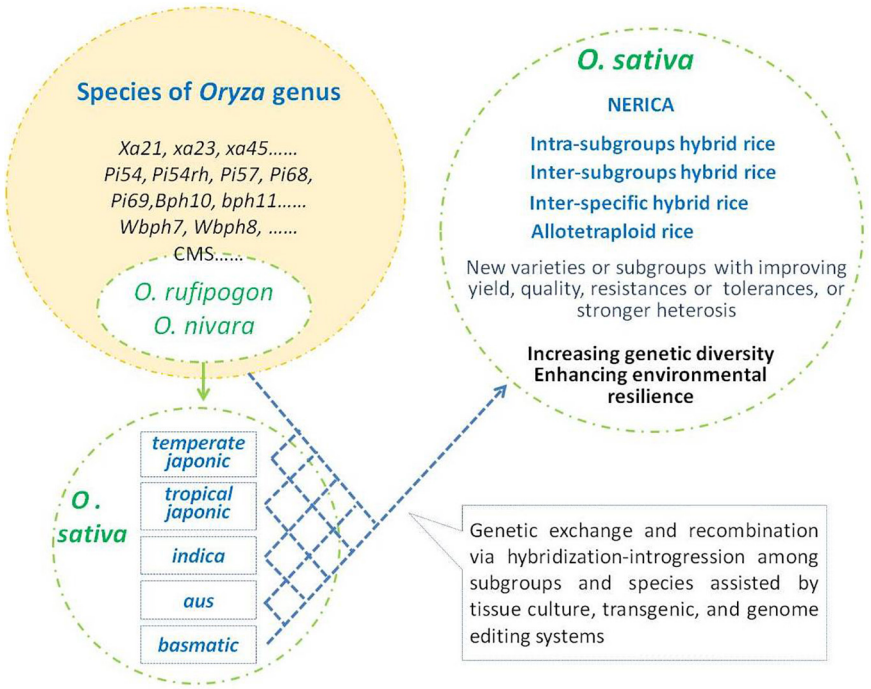


FIGURE 2 | Artificial hybridization-introgression for genetic improvement of *Oryza sativa*. *O. sativa* group was descent from *Oryza rufipogon* and *Oryza nivara*. Species of *Oryza* genus including *O. rufipogon* and *O. nivara* are rich in genetic diversity and favorable traits/alleles. Some elite genes/alleles were transferred into Asian cultivated rice and gained several breakthroughs in rice breeding. Further hybridization-introgression among interspecies and inter-subgroups will provide more chances to enlarge genetic diversity and gain breeding achievements in *O. sativa*.

O. alta of rice (Yu et al., 2021). But, integration of new tools and interspecific hybridization can be mainstream to improve the Asian cultivated rice further in the future.

In Addition to Interspecific Hybridization, Inter-Subgroups Hybridization Can Make the Asian Cultivated Rice More Adaptable

As the Asian cultivated rice differentiated into subgroups in various environments to meet multifarious cultivation conditions and eating preferences, many favorable traits and genes/alleles were raised in distinct populations. Introducing favorable alleles specific to one subgroup (such as quality, biotic- and abiotic-resistance/tolerance, and growth vigor) into other subgroups can enrich and improve their diversity in these traits. For instance, *basmatic* rice is usually with excellent rice qualities, its genetic elements of favor allele can be used to improve the quality traits of other subgroups. The *temperate japonica* rice at high altitude or high latitude usually has strong cold tolerance, and its genetic components of cold tolerance can also improve the adaptability of *indica* rice to cold climates. Similarly, the higher nitrogen-use efficiency (NUE) in *aus* and *basmatic* rice can also improve this trait of *japonica* and *indica* rice (Liu et al., 2021), so as to reduce the application of nitrogen fertilizer and then reduce input and environmental pollution. The premise of achieving the above purpose is to clarify the genetic basis controlling these traits.

Inter-subgroup hybrids can be another field to utilize heterosis in rice. Heterotic loci existed in wild progenitors of cultivated rice, and cultivated rice was divergently selected among rice subgroups. Divergent selection among the subgroups induced various heterotic alleles in different subgroups. Many heterotic loci in the current commercial hybrids were shaped by genome introgressions from different subgroups (Lin et al., 2020). Thus, introgression among subgroups could enhance heterosis in hybrid rice.

Developing a Series of Introgression Lines With Favor Genes/Alleles From Wild Relatives in Various Subgroups Genetic Background Will Contribute to the Breeding Utilization of These Favorable Genes in Various Ecological Subgroups of Rice

In the past decades, many genes/alleles from wild relative were identified and even cloned, and the corresponding introgression lines (ILs) or near-isogenic lines (NILs) were developed in a specific genetic background. A specific genetic background has the adaptability to a given ecological environment; although distant hybridization can increase genetic diversity, when these ILs or NILs were crossed with parents of distinct genetic backgrounds to develop breeding populations, from the perspective of breeders, the offspring of distant hybridization

usually perform poorly under a specific ecological condition. This may be the main reason for the limited utilization of these alleles in rice breeding projects. Yield-related traits usually are quantitative traits, and are often affected by both genetic background and environment, improving them in breeding programs needs to deal with more challenges. Therefore, developing a series of ILs or NILs with these favor gene/alleles in various specific genetic backgrounds will be conductive to reevaluate their genetic effect and breeding value in various genetic backgrounds and environments and to develop adaptive breeding populations.

CONCLUSION

Asian cultivated rice originated from *O. rufipogon* and evolved thousands of years. During the domestication and evolution progress, when it gained favor mutations or was introduced into changing environments, only a few numbers of individuals could be retained and subsequently expanded, then the genetic bottlenecks occurred and genetic diversity decreased. Fully and long domesticated subgroup *japonica* showed a higher degree of differentiation and lower genetic diversity than *indica* (Zhu et al., 2006). Interspecific hybridization between domesticated rice and its wild relatives supplied chances to increase its genetic diversity, expanded and evolved in various subgroups, and adapted to more and more changing environments. Natural hybridization and introgression led to various subgroups of *O. sativa,* which adapted to changing cultivated environments, and produced feral weedy rice coexisting and competing with cultivars under cultivation management. Artificial interspecific hybridization and introgression gained several breakthroughs in rice breeding, such as hybrid rice, NERICA, and biotic stress resistance genes, which are contributed to the stable increase of rice production to meet the expanding human population. As interspecific hybridization-introgression have played important roles in improving genetic diversity and adaptation of *O. sativa*, more efforts need to make to exploit its virtues. Some key issues, such as reproductive barriers especially hybrid sterility should be investigated further, which are conducive to gene exchange between cultivated rice and its relatives, and even is beneficial to exploiting interspecific hybrid vigor. New techniques, such as efficient tissue culture, transgenics, genome editing system, and high-quality genome assembly, will help wide distant interspecific introgression and make the crop better. Compared with the huge genetic diversity of wild relatives, the construction of introgression lines in a wider range with more donors is needed, which benefits allele mining, understanding the genetic network of rice growth and development, yield formation, and environmental adaptation. Thus, developing a series of ILs or NILs with these favor gene/alleles in the various specific genetic background will be conductive to reevaluate their genetic effect and breeding value in various genetic backgrounds and environments and helpful to develop more usable breeding populations in various situations. In addition to interspecific hybridization, hybridization between inter-subgroups can make the Asian cultivated rice wider adaptable and stronger heterosis.

AUTHOR CONTRIBUTIONS

DT proposed the concept. JZ conceived and wrote the manuscript. YY, YL, QP, JL, YZ, XD, MW, and JW reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the National Natural Science Foundation of China (Grant Nos. 31991221, U2002202, 32160489, and 31860372), the Yunnan Provincial Science and Technology Department, China (202003AD150007, 202001AS070003, 202101AS070036, 202101AT070193, and 530000210000000013809), the Yunnan Provincial Government (YNWR-QNBJ-2018-359), the Yunnan Seed Industrialization Laboratory Program, and the Applied Basic Research Foundation of Yunnan Academy of Agricultural Sciences (JZ201801).

REFERENCES

Baack, E. J., and Rieseberg, L. H. (2007). A genomic view of introgression and hybrid speciation. *Curr. Opin. Genet. Dev.* 17, 513–518. doi: 10.1016/j.gde.2007.09.001

Bessho-Uehara, K., Furuta, T., Masuda, K., Yamada, S., Angeles-Shim, R. B., Ashikari, M., et al. (2017). Construction of rice chromosome segment substitution lines harboring *Oryza barthii* genome and evaluation of yield-related traits. *Breed Sci.* 67, 408–415. doi: 10.1270/jsbbs.17022

Bhatia, D., Joshi, S., Das, A., Vikal, Y., Sahi, G. K., Neelam, K., et al. (2017). Introgression of yield component traits in rice (ssp.) through interspecific hybridization. *Crop Sci.* 57:1557. doi: 10.2135/cropsci2015.11.0693

Bourgis, F., Guyot, R., Gherbi, H., Tailliez, E., Amabile, I., Salse, J., et al. (2008). Characterization of the major fragance gene from an *aromatic japonica* rice and analysis of its diversity in Asian cultivated rice. *Theor. Appl. Genet.* 117, 353–368. doi: 10.1007/s00122-008-0780-9

Brar, D. S. (2004). "Interspecific hybridization," in *Encyclopedia of Plant and Crop Science* (New York, NY: Marcel Dekker, Inc), 619–624.

Carpentier, M., Manfroi, E., Wei, F., Wu, H., Lasserre, E., Llauro, C., et al. (2019). Retrotranspositional landscape of Asian rice revealed by 3000 genomes. *Nat. Comm.* 10:24. doi: 10.1038/s41467-018-07974-5

Chang, T. T. (1976). The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. *Eupbytica* 25, 425–441.

Chen, E., Huang, X., Tian, Z., Wing, R. A., and Han, B. (2019). The genomics of Oryza species provides insights into rice domestication and heterosis. *Annu. Rev. Plant Biol.* 70, 639–665. doi: 10.1146/annurev-arplant-050718-100320

Choi, J. Y., Lye, Z. N., Groen, S. C., Dai, X., Rughani, P., Zaaijer, S., et al. (2020). Nanopore sequencing-based genome assembly and evolutionary genomics of circum-basmati rice. *Genome Biology* 21:21. doi: 10.1186/s13059-020-1938-2

Choi, J. Y., Platts, A. E., Fuller, D. Q., Hsing, Y. I., Wing, R. A., and Purugganan, M. D. (2017). The rice paradox: multiple origins but single domestication in Asian rice. *Mol. Biol. Evol.* 34, 969–979. doi: 10.1093/molbev/msx049

Civán, P., Craig, H., Cox, C. J., and Brown, T. A. (2015). Three geographically separate domestications of Asian rice. *Nat. Plants* 1:15164. doi: 10.1038/NPLANTS.2015.164

d'Alpoim Guedes, J., Jin, G., and Bocinsky, R. K. (2015). The impact of climate on the spread of rice to north-eastern China: a new look at the data from Shandong province. *PLoS One* 10:e0130430. doi: 10.1371/journal.pone.0130430

Das, A., Soubam, D., Singh, P., Thakur, S., Singh, N., and Sharma, T. (2012). A novel blast resistance gene, *Pi54rh* cloned from wild species of rice, *Oryza rhizomatis* confers broad spectrum resistance to *Magnaporthe oryzae*. *Funct. Integr. Genomic*. 12, 215–228. doi: 10.1007/s10142-012-0284-1

Dave, B. B. (1943). The wild rice problem in the central provinces and its solution. *Ind. J. Agric. Sci.* 13, 46–53.

Deng, X., Zhou, J., Xu, P., Li, J., Hu, F., and Tao, D. (2010). The role of *S1-g* allele from *Oryza glaberrima* in improving interspecific hybrid sterility between *O. sativa* and *O. glaberrima*. *Breed Sci.* 60, 342–346. doi: 10.1270/jsbbs.60.342

Devanna, N. B., Vijayan, J., and Sharma, T. R. (2014). The blast resistance gene *Pi54* of cloned from *Oryza officinalis* interacts with *Avr-Pi54* through its novel non-LRR domains. *PLoS One* 9:e104840. doi: 10.1371/journal.pone.0104840

Devi, S. J. S. R., Singh, K., Umakanth, B., Vishalakshi, B., Vijaya Sudhakara, Rao, K., et al. (2020). Identification and characterization of a large effect QTL from *Oryza glumaepatula* revealed *Pi68(t)* as putative candidate gene for rice blast resistance. *Rice* 13:17. doi: 10.1186/s12284-020-00378-4

Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321. doi: 10.1016.j.cell.2006.12.006

Doi, K., Iwata, N., and Yoshimura, A. (1997). The construction of chromosome substitution lines of African rice (*Oryza glaberrima* Steud.) in the background of *japonica* rice (*O. sativa* L.). *Rice Genet. Newsl.* 14, 39–41.

Dong, L., Liu, S., Kyaing, M. S., Xu, P., Tharreau, D., Deng, W., et al. (2020). Identification and fine mapping of *Pi69*(t), a new gene conferring broad-spectrum resistance against *Magnaporthe oryzae* from *Oryza glaberrima* Steud. *Front. Plant Sci.* 11:1190. doi: 10.3389/fpls.2020.01190

Eizenga, G., Prasad, B., Jackson, A., and Jia, M. (2013). Identification of rice sheath blight and blast quantitative trait loci in two different O. sativa / O. nivara advanced backcross populations. *Molecul. Breed.* 31, 889–907. doi: 10.1007/s11032-013-9843-y

Foria, S., Magris, G., Jurman, I., Schwope, R., De Candido, M., De Luca, E., et al. (2022). Extent of wild–to–crop interspecific introgression in grapevine (*Vitis vinifera*) as a consequence of resistance breeding and implications for the crop species definition. *Horticulture Research* 9:uhab010. doi: 10.1093/hr/uhab010

Fu, R., Zhu, Y., Liu, Y., Feng, Y., Lu, R., Li, Y., et al. (2022). Genome-wide analyses of introgression between two sympatric Asian oak species. *Nat. Ecol. Evol.* ahead of print]. doi: 10.1038/s41559-022-01754-7

Fuller, D. Q. (2006). Agricultural origins and frontiers in South Asia: a working synthesis. *J. World Prehist.* 20, 1–86. doi: 10.1007/s10963-006-9006-8

Fuller, D. Q. (2011a). Finding plant domestication in the Indian subcontinent. *Curr. Anthropol.* 52, S347–S362. doi: 10.1086/658900

Fuller, D. Q. (2011b). Pathways to Asian civilizations: tracing the origins and spread of rice and rice cultures. *Rice* 4, 78–92. doi: 10.1007/s12284-011-9078-7

Fuller, D. Q., Qin, L., Zheng, Y., Zhao, Z., Chen, X., Hosoya, L. A., et al. (2009). The domestication process and domestication rate in rice: spikelet bases from the Lower Yangtze. *Science* 323, 1607–1610. doi: 10.1126/science.1166605

Fuller, D. Q., Sato, Y., Castillo, C., Qin, L., Weisskopf, A. R., Kingwell-Banham, E. J., et al. (2010). Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeol. Anthropol Sci.* 2, 115–131. doi: 10.1007/s12520-010-0035-y

Furuta, T., Uehara, K., Angeles-Shim, R. B., Shim, J., Ashikari, M., and Takashi, T. (2014). Development and evaluation of chromosome segment substitution lines (CSSLs) carrying chromosome segments derived from *Oryza rufipogon* in the genetic background of *Oryza sativa* L. *Breed Sci.* 63, 468–475. doi: 10.1270/jsbbs.63.468

Furuta, T., Uehara, K., Angeles-Shim, R. B., Shim, J., Nagai, K., Ashikari, M., et al. (2016). Development of chromosome segment substitution lines harboring *Oryza nivara* genomic segments in Koshihikari and evaluation of yield-related traits. *Breed Sci.* 66, 845–850. doi: 10.1270/jsbbs.16131

Gaikwad, K. B., Singh, N., Kaur, P., Rani, S., Babu, H. P., and Singh, K. (2021). Deployment of wild relatives for genetic improvement in rice (*Oryza sativa* L.). *Plant Breed.* 140, 23–52. doi: 10.1111/pbr.1287

Garris, A. J., Tai, T. H., Coburn, J., Kresovich, S., and McCouch, S. (2005). Genetic structure and diversity in *Oryza sativa* L. *Genetics* 169, 1631–1638. doi: 10.1534/genetics.104.035642

Glaszmann, J. C. (1987). Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.* 74, 21–30. doi: 10.1007/bf00290078

Gross, B. L., and Zhao, Z. (2014). Archaeological and genetic insights into the origins of domesticated rice. *Proc. Natl. Acad. Sci. U. S. A.* 111, 6190–6197. doi: 10.1073/pnas.1308942110

Guo, J., Xu, X., Li, W., Zhu, W., Zhu, H., Liu, Z., et al. (2016). Overcoming inter-subspecific hybrid sterility in rice by developing indica-compatible *japonica* lines. *Sci. Rep.* 6:26878. doi: 10.1038/srep26878

Gutaker, R. M., Groen, S. C., Bellis, E. S., Choi, J. Y., Pires, I. S., Bocinsky, R. K., et al. (2020). Genomic history and ecology of the geographic spread of rice. *Nat. Plants* 6, 492–502. doi: 10.1038/s41477-020-0659-6

Harlan, J. R. (1965). Possible role of weed races in evolution of cultivated plants. *Euphytica* 14, 173–176. doi: 10.1007/bf00038984

He, N., Wu, R., Pan, X., Peng, L., Sun, K., Zou, T., et al. (2017). Development and trait evaluation of chromosome single-segment substitution lines of *O. meridionalis* in the background of *O. sativa*. *Euphytica* 213:281. doi: 10.1007/s10681-017-2072-4

He, Q., Deng, H., Sun, P., Zhang, W., Shu, F., Xing, J., et al. (2020). Hybrid Rice. *Engineering* 6, 967–973. doi: 10.1016/j.eng.2020.08.005

He, W., Chen, C., Xiang, K., Wang, J., Zheng, P., Tembrock, L. R., et al. (2021). The history and diversity of rice domestication as resolved from 1464 complete plastid genomes. *Front. Plant Sci.* 12:781793. doi: 10.3389/fpls.2021.781793

Higham, C., and Lu, T. L. (1998). The origins and dispersal of rice cultivation. *Antiquity* 72, 867–877. doi: 10.1017/S0003598X00087500

Hu, F., Tao, D., Yang, Y., Xu, P., Li, J., and Zhou, J. (2002). Studies of vegetative heterosis of interspecific hybrids between *Oryza sativa* and *O. glaberrima*. *J. Anhui Agric. Univ.* 24, 146–150.

Hu, J., Xiao, C., Cheng, M., Gao, G., Zhang, Q., and He, Y. (2015). Fine mapping and pyramiding of brown planthopper resistance genes *QBph3* and *QBph4* in an introgression line from wild rice *O. officinalis*. *Mol. Breed.* 35:3. doi: 10.1007/s11032-015-0228-2

Huang, D., Qiu, Y., Zhang, Y., Huang, F., Meng, J., Wei, S., et al. (2013). Fine mapping and characterization of *BPH27*, a brown planthopper resistance gene from wild rice (*Oryza rufipogon* Griff.). *Theor. Appl. Genet.* 126, 219–229. doi: 10.1007/s00122-012-1975-7

Huang, X., Kurata, N., Wei, X., Wang, Z., Wang, A., Zhao, Q., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497–501. doi: 10.1038/nature11532

Jin, J., Huang, W., Gao, J., Yang, J., Shi, M., Zhu, M., et al. (2008). Genetic control of rice plant architecture under domestication. *Nat. Genet.* 40, 1365–1369. doi: 10.1038/ng.247

Jones, M. P., Dingkuhn, M., Aluko, G. K., and Semon, M. (1997). Interspecific *Oryza sativa* L. x *O. glaberrima* Steud. progenies in upland rice improvement. *Euphytica* 92, 237–246.

Kato, S., Kosaka, H., and Hara, S. (1928). On the affinity of rice varieties as shown by the fertility of rice plants. *Centr. Agric. Inst. Kyushu. Imp. Univ.* 2, 241–276.

Khush, G. S. (1997). Origin, dispersal, cultivation and variation of rice. *Plant Molecul. Biol.* 35, 25–34.

Khush, G. S., and Brar, D. S. (2017). Alien introgression in rice. *Nuleus* 60, 251–261. doi: 10.1007/s13237-017-0222-7

Khush, G. S., Brar, D. S., Virk, P. S., Tang, S., Malik, S. S., Busato, G. A., et al. (2003). "Classifying rice germplasm by isozyme polymorphism and origin of cultivated rice," in *IRRI Discussion Paper Series* No 46, (Los Banos: international Rice research Institute)

Khush, G. S., and Ling, K. C. (1974). Inheritance of resistance to grassy stunt virus and its vector in Rice. *J. Heredity* 65, 135–136. doi: 10.1093/oxfordjournals.jhered.a108483

Kim, H., Jung, J., Singh, N., Greenberg, A., Doyle, J. J., Tyagi, W., et al. (2016). Population dynamics among six major groups of the *Oryza rufipogon* species complex, wild relative of cultivated Asian rice. *Rice* 9:56. doi: 10.1186/s12284-016-0119-0

Kishor, D. S., Seo, J., Chin, J. H., and Koh, H.-J. (2020). Evaluation of whole-genome sequence, genetic diversity, and agronomic traits of *basmati* rice (*Oryza sativa* L.). *Front. Genet.* 11:86. doi: 10.3389/fgene.2020.00086

Kopecký, D., Martín, A., and Smýkal, P. (2022). Interspecific hybridization and plant breeding: from historical retrospective through work of Mendel to current crops. *Czech. J. Genet. Plant Breed.* 1212–1975. doi: 10.17221/19/2022-CJGPB

Kovach, M. J., Calingacion, M. N., Fitzgerald, M. A., and McCouch, S. R. (2009). The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proc. Natl. Acad. Sci. U. S. A.* 106, 14444–14449. doi: 10.1073/pnas.0904077106

Kovach, M. J., Megan, Sweeney, T., and McCouch, S. R. (2007). New insights into the history of rice domestication. *Trends Genet.* 23, 578–587. doi: 10.1016/j.tig.2007.08.012

Kumar, K., Sarao, P. S., Bhatia, D., Neelam, K., Kaur, A., Mangat, G. S., et al. (2018). High-resolution genetic mapping of a novel brown planthopper resistance locus, *Bph34* in *Oryza sativa* L. x *Oryza nivara* (Sharma & Shastry) derived interspecifc F2 population. *Theor. Appl. Genet.* 131, 1163–1171. doi: 10.1007/s00122-018-3069-7

Ladizinskv, G. (1998). *Plant Evolution Under Domestication*. United states: Kluwer.Academic Publishers 86. doi: 10.1007/978-94-011-4429-2

Larson, J. L. (1968). The species concept of Linnaeus. *Isis* 59, 291–299. doi: 10.1086/350398

Lemmon, Z. H., Reem, N. T., Dalrymple, J., Soyk, S., Swartwood, K. E., Rodriguez-Leal, D., et al. (2018). Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat. Plants* 4, 766–770. doi: 10.1038/s41477-018-0259-x

Li, C., Zhou, A., and Sang, T. (2006). Rice domestication by reducing shattering. *Science* 311, 1936–1939. doi: 10.1126/science.1123604

Li, J., Zeng, Y., Pan, Y., Zhou, L., Zhang, Z., Guo, H., et al. (2021). Stepwise selection of natural variations at *CTB2* and *CTB4a* improves cold adaptation during domestication of *japonica* rice. *New Phytol.* 231, 1056–1072. doi: 10.1111/nph.17407

Li, L., and Olsen, K. M. (2020). "Population genomics of weedy crop relatives: insights from weedy rice," in *Population Genomics Crop Plants*, ed. O. P. Rajora (Switzerland AG: Springer Nature), doi: 10.1007/13836_2020_77

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., et al. (2018). Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* 36, 1160–1163. doi: 10.1038/nbt.4273

Li, Z., Xue, Y., Zhou, H., Li, Y., Usman, B., Jiao, X., et al. (2019). High-resolution mapping and breeding application of a novel brown planthopper resistance gene derived from wild rice (*Oryza. rufipogon* Griff). *Rice* 12:41. doi: 10.1186/s12284-019-0289-7

Lin, Z., Griffith, M. E., Li, X., Zhu, Z., Tan, L., Fu, Y., et al. (2007). Origin of seed shattering in rice (*Oryza sativa* L.). *Planta* 226, 11–20. doi: 10.1007/s00425-006-0460-4

Lin, Z., Qin, P., Zhang, X., Fu, C., Deng, H., Fu, X., et al. (2020). Divergent selection and genetic introgression shape the genome landscape of heterosis in hybrid rice. *Proc. Natl. Acad. Sci. U. S. A.* 117, 4623–4631. doi: 10.1073/pnas.1919086117

Linares, O. F. (2002). African rice (*Oryza glaberrima*): history and future potential. *Proc. Natl. Acad. Sci. U. S. A.* 99, 16360–16365. doi: 10.1073/pnas.252604599

Ling, K. C., Aguiero, V. M., and Lee, S. H. (1970). A mass screening method for testing resistance to grassy stunt disease of rice. *Plant Dis. Rep.* 56, 565–569.

Lippman, Z. B., Semel, Y., and Zamir, D. (2007). An integrated view of quantitative trait variation using tomato interspecific introgression lines. *Curr. Opinion Genet. Dev.* 17, 545–552. doi: 10.1016/j.gde.2007.07.007

Liu, Y., Wang, H., Jiang, Z., Wang, W., Xu, R., Wang, Q., et al. (2021). Genomic basis of geographical adaptation to soil nitrogen in rice. *Nature* 590, 600–605. doi: 10.1038/s41586-020-03091-w

Lv, Q., Xu, X., Shang, J., Jiang, G., Pang, Z., Zhou, Z., et al. (2013). Functional analysis of *Pid3-A4*, an ortholog of rice blast resistance gene *Pid3* revealed by allele mining in common wild rice. *Phytopathology* 103, 594–599. doi: 10.1094/phyto-10-12-0260-r

Ma, X., Han, B., Tang, J., Zhang, J., Cui, D., Geng, L., et al. (2019). Construction of chromosome segment substitution lines of Dongxiang common wild rice (*Oryza rufipogon* Griff.) in the background of the *japonica* rice cultivar Nipponbare (*Oryza sativa* L.). *Plant Physiol. Biochem.* 144, 274–282. doi: 10.1016/j.plaphy.2019.09.041

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends Ecol. Evol.* 20, 229–237.

Mallet, J. (2007). Hybrid speciation. *Nature* 446, 279–283.

Matsuo, T. (1952). Genecological studies on cultivated rice (in Japanese). *Bull. Nat. Inst. Agric. Sci. Jpn. D.* 3, 1–111.

McNally, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., Zhao, K., Ulat, V. J., et al. (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12273–12278. doi: 10.1073/pnas.0900992106

Milla, R., and Osborne, C. P. (2021). Crop origins explain variation in global agricultural relevance. *Nat. Plants* 7, 598–607. doi: 10.1038/s41477-021-00905-1

Mino, M., Tezuka, T., and Shomura, S. (2022). The hybrid lethality of interspecific F1 hybrids of *Nicotiana*: a clue to understanding hybrid inviability-a major obstacle to wide hybridization and introgression breeding of plants. *Molecul. Breed.* 42, 1–20. doi: 10.1007/s11032-022-01279-8

Molina, J., Sikora, M., Garud, N., Flowers, J. M., Rubinstein, S., Reynolds, A., et al. (2011). Molecular evidence for a single evolutionary origin of domesticated rice. *Proc. Natl. Acad. Sci. U. S. A.* 108, 8351–8356. doi: 10.1073/pnas.1104686108

Moner, A. M., Furtado, A., and Henry, R. J. (2018). Chloroplast phylogeography of AA genome rice species. *Molecul. Phylogenet. Evol.* 127, 475–487. doi: 10.1016/j.ympev.2018.05.002

Moner, A. M., Furtado, A., and Henry, R. J. (2020). Two divergent chloroplast genome sequence clades captured in the domesticated rice gene pool may have significance for rice production. *BMC Plant Biol.* 20:472. doi: 10.1186/s12870-020-02689-6

Morinaga, T. (1954). Classification of rice varieties on the basis of affinity. *Stud. Rice Breed.* 4, 1–14.

Nabholz, B., Sarah, G., Sabot, F., Ruiz, M., Adam, H., Nidelet, S., et al. (2014). Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*). *Mol. Ecol.* 23, 2210–2227. doi: 10.1111/mec.12738

Nayar, N. M. (2010). The history and genetic transformation of the African rice, *Oryza glaberrima* Steud. (Gramineae). *Curr. Sci.* 99, 1681–1689. doi: 10.1007/978-94-007-3934-5_10024-1

Neelam, K., Mahajan, R., Gupta, V., Bhatia, D., Gill, B. K., Komal, R., et al. (2019). High-resolution genetic mapping of a novel bacterial blight resistance gene *xa-45*(t) identified from *Oryza glaberrima* and transferred to *Oryza sativa*. *Theor. Appl. Genet.* doi: 10.1007/s00122-019-03501-2

Nolte, A. W., and Tautz, D. (2009). Understanding the onset of hybrid speciation. *Trends Genet.* 26, 55–58. doi: 10.1016/j.tig.2009.12.001

Oka, H. I. (1958). Intervarietal variation and classification of cultivated rice. *Indian J. Genet. Plant Breed.* 18, 79–89. doi: 10.1093/aob/mcl210

Oka, H. I. (1988). *Origin of Cultivated Rice*. Tokyo: Japan Scientific Society Press.

Purugganan, M. D. (2019). Evolutionary insights into the nature of plant domestication. *Curr. Biol.* 29, 705–714. doi: 10.1016/j.cub.2019.05.053

Qin, G., Nguyen, H. M., Luu, S. N., Wang, Y., and Zhang, Z. (2018). Construction of introgression lines of *Oryza rufipogon* and evaluation of important agronomic traits. *Theor. Appl. Genet.* doi: 10.1007/s00122-018-3241-0

Ramos, J. M., Furuta, T., Uehara, K., Chihiro, N., Angeles-Shim, R. B., Shim, J., et al. (2016). Development of chromosome segment substitution lines (CSSLs) of *Oryza longistaminata* A. Chev. & Röhr in the background of the elite *japonica* rice cultivar, Taichung 65 and their evaluation for yield traits. *Euphytica* 210, 151–163. doi: 10.1007/s10681-016-1685-3

Sang, T., and Ge, S. (2007). The puzzle of rice domestication. *J. Integr. Plant Biol.* 49, 760–768. doi: 10.1111/j.1744-7909.2007.00510.x

Shim, R. A., Angeles, E. R., Ashikari, M., and Takashi, T. (2010). Development and evaluation of *Oryza glaberrima* Steud. chromosome segment substitution lines (CSSLs) in the background of *O. sativa* L. cv. *Koshihikari*. *Breed Sci.* 60, 613–619. doi: 10.1270/jsbbs.60.613

Siddiq, E. A., Vemireddy, L. R., and Nagaraju, J. (2012). *Basmati* rices: genetics, breeding and trade. *Agric Res* 1, 25–36. doi: 10.1007/s40003-011-0011-5

Silva, F., Weisskopf, A., Castillo, C., Murphy, C., Kingwell-Banham, E., Qin, L., et al. (2018). A tale of two rice varieties: modelling the prehistoric dispersals of *japonica* and *proto-indica* rices. *The Holocene* 28, 1745–1758. doi: 10.1177/0959683618788634

Soltis, P. S., and Soltis, D. E. (2009). The Role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039

Song, J., Arif, M., Zi, Y., Sze, S., Zhang, M., and Zhang, H. (2021). Molecular and genetic dissection of the USDA rice mini-core collection using high-density SNP markers. *Plant Sci.* 308:110910. doi: 10.1016/j.plantsci.2021.110910

Sun, Y., He, W., Xie, Y., Zhao, W., Tan, J., Yang, X., et al. (2020). Exploiting interspecific heterosis between African rice and Asian rice. *Crop Sci.* 60, 2343–2353. doi: 10.1002/csc2.20224

Sweeney, M., and McCouch, S. (2007). The complex history of the domestication of rice. *Annals. Bot.* 100, 951–957. doi: 10.1093/aob/mcm128

Sweeney, M. T., Thomson, M. J., Cho, Y. G., Park, Y. J., Williamson, S. H., Bustamante, C. D., et al. (2007). Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genet.* 3:e133. doi: 10.1371/journal.pgen.0030133

Tan, L., Li, X., Liu, F., Sun, X., Li, C., Zhu, Z., et al. (2008). Control of a key transition from prostrate to erect growth in rice domestication. *Nat. Genet.* 40, 1360–1364. doi: 10.1038/ng.197

Tian, F., Li, D., Fu, Q., Zhu, Z., Fu, Y., Wang, X., et al. (2005). Construction of introgression lines carrying wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield-related traits. *Theor. Appl. Genet.* 112, 570–580. doi: 10.1007/s00122-005-0165-2

van Andel, T. R., Meyer, R. S., Aflitos, S. A., Carney, J. A., Veltman, M. A., Copetti, D., et al. (2016). Tracing ancestor rice of Suriname Maroons back to its African origin. *Nat. Plants* 2:16149. doi: 10.1038/nplants.2016.149

Vaughan, D. A., Lu, B., and Tomooka, N. (2008). Was Asian rice (*Oryza sativa*) domesticated more than once? *Rice* 1, 16–24. doi: 10.1007/s12284-008-9000-0

Vaughan, D. A., and Sitch, L. A. (1991). Gene flow from the jungle to farmers. *Bioscience* 41, 22–28. doi: 10.2307/1311537

Vigueira, P. A., Olsen, K. M., Wagner, C. R., Chittick, Z. B., and Vigueira, C. C. (2020). Weedy rice from South Korea arose from two distinct de-domestication events. *Front. Agron.* 2:602612. doi: 10.3389/fagro.2020.602612

Wambugu, P., Furtado, A., Waters, D., Nyamongo, D., and Henry, R. (2013). Conservation and utilization of African *Oryza* genetic resources. *Rice* 6:29. doi: 10.1186/1939-8433-6-29

Wang, H., Vieira, F. G., Crawford, J. E., Chu, C., and Nielsen, R. (2017). Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. *Genom. Res.* 27, 1029–1038. doi: 10.1101/gr.204800.116

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49. doi: 10.1038/s41586-018-0063-9

Wang, X., Chen, L., and Ma, J. (2019). Genomic introgression through interspecific hybridization counteracts genetic bottleneck during soybean domestication. *Genom. Biol.* 20, 22. doi: 10.1186/s13059-019-1631-5

Wu, Y., Zhao, S., Li, X., Zhang, B., Jiang, L., Tang, Y., et al. (2018). Deletions linked to *PROG1* gene participate in plant architecture domestication in Asian and African rice. *Nat. Commun.* 9:4157. doi: 10.1038/s41467-018-06509-2

Xie, Y., Xu, P., Huang, J., Ma, S., Xie, X., Tao, D., et al. (2017). Interspecific hybrid sterility in rice is mediated by *OgTPR1* at the *S1* locus encoding a peptidase-like protein. *Mol. Plant* 10, 1137–1140. doi: 10.1016/j.molp.2017.05.005

Xu, P., Dong, L., Zhou, J., Li, J., Zhang, Y., Hu, F., et al. (2015). Identification and mapping of a novel blast resistance gene *Pi57*(t) in *Oryza longistaminata*. *Euphytica* 205, 95–102. doi: 10.1007/s10681-015-1402-7

Xu, R., and Sun, C. (2021). What happened during domestication of wild to cultivated rice. *Crop J.* 9, 564–576. doi: 10.1016/j.cj.2021.02.005

Yamagata, Y., Thanda Win, K., Miyazaki, Y., Ogata, C., Yasui, H., and Yoshimura, A. (2019). Development of introgression lines of AA genome *Oryza* species, *O. glaberrima, O. rufipogon*, and *O. nivara,* in the genetic background of *O. sativa* L. cv. Taichung 65. *Breed. Sci.* 69, 359–363. doi: 10.1270/jsbbs.19002

Yu, H., Lin, T., Meng, X., Du, H., Zhang, J., Liu, G., et al. (2021). A route to *de novo* domestication of wild allotetraploid rice. *Cell* 184, 1156–1170.e14. . doi: 10.1016/j.cell.2021.01.013

Yu, X., Zhao, Z., Zheng, X., Zhou, J., Kong, W., Wang, P., et al. (2018). A selfish genetic element confers non-Mendelian inheritance in rice. *Science* 360, 1130–1132. doi: 10.1126/science.aar4279

Zhang, F., Wang, C., Li, M., Cui, Y., Shi, Y., Wu, Z., et al. (2021). The landscape of gene– CDS–haplotype diversity in rice: properties, population organization, footprints of domestication and breeding, and implications for genetic improvement. *Mol. Plant.* 14, 787–804. doi: 10.1016/j.molp.2021.02.003

Zhang, G. (2021). Target chromosome-segment substitution: a way to breeding by design in rice. *Crop J.* 9, 658–668. doi: 10.1016/j.cj.2021.03.001

Zhang, J., Percy, R. G., and McCarty, J. C. (2014). Introgression genetics and breeding between Upland and Pima cotton: a review. *Euphytica* 198, 1–12. doi: 10.1007/s10681-014-1094-4

Zhang, L., Zhu, Q., Wu, Z., Ross-Ibarra, J., Gaut, B. S., Ge, S., et al. (2009). Selection on grain shattering genes and rates of rice domestication. *New Phytol.* 184, 708–720. doi: 10.1111/j.1469-8137.2009.02984.x

Zhang, Y., Qin, G., Ma, Q., Wei, M., Yang, X., Ma, Z., et al. (2020). Identification of major locus *Bph35* resistance to brown planthopper in rice. *Rice Sci.* 27, 237–245. doi: 10.1016/j.rsci.2020.04.006

Zhang, Y., Zhou, J., Peng, X., Li, J., Deng, X., Deng, W., et al. (2022). A Genetic resource for rice improvement: introgression library of agronomic traits for all AA genome *Oryza* species. *Front. Plant Sci.* 13:856514. doi: 10.3389/fpls.2022.856514

Zhao, H., Sun, L., Xiong, T., Wang, Z., Liao, Y., Zou, T., et al. (2019). Genetic characterization of the chromosome single-segment substitution lines of *O. glumaepatula* and *O. barthii* and identification of QTLs for yield-related traits. *Molecul. Breed.* 39:51. doi: 10.1007/s11032-019-0960-0

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., et al. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* 50, 278–284. doi: 10.1038/s41588-018-0041-z

Zheng, X., Pang, H., Wang, J., Yao, X., Song, Y., Li, F., et al. (2022). Genomic signatures of domestication and adaptation during geographical expansions of rice cultivation. *Plant Biotechnol. J.* 20, 16–18. doi: 10.1111/pbi.13730

Zhu, Q., Zheng, X., Luo, J., Gaut, B. S., and Ge, S. (2006). Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Molecul. Biol. Evol.* 24, 875–888. doi: 10.1093/molbev/msm005

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

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

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 01 July 2022
doi: 10.3389/fpls.2022.892805

Expression Patterns Divergence of Reciprocal F_1 Hybrids Between *Gossypium hirsutum* and *Gossypium barbadense* Reveals Overdominance Mediating Interspecific Biomass Heterosis

Tengyu Li[1,2,3], Fuqiu Wang[3], Muhammad Yasir[1], Kui Li[4], Yuan Qin[3], Jing Zheng[1], Kun Luo[1], Shouhong Zhu[2], Hua Zhang[1], Yurong Jiang[1], Yongshan Zhang[2] and Junkang Rong[1]**

[1] The Key Laboratory for Quality Improvement of Agricultural Products of Zhejiang Province, Zhejiang Agriculture and Forestry University, Hangzhou, China, [2] State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, China, [3] National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China, [4] Institute of Food and Nutrition Development, Ministry of Agriculture and Rural Affairs, Chinese Academy of Agricultural Sciences, Beijing, China

Hybrid breeding has provided an impetus to the process and achievement of a higher yield and quality of crops. Interspecific hybridization is critical for resolving parental genetic diversity bottleneck problems. The reciprocal interspecific hybrids and their parents (*Gossypium hirsutum* and *Gossypium barbadense*) have been applied in this study to elucidate the transcription regulatory mechanism of early biomass heterosis. Phenotypically, the seed biomass, plant height over parent heterosis, leaf area over parent heterosis, and fresh and dry biomass were found to be significantly higher in hybrids than in parents. Analysis of leaf areas revealed that the one-leaf stage exhibits the most significant performance in initial vegetative growth vigor and larger leaves in hybrids, increasing the synthesis of photosynthesis compounds and enhancing photosynthesis compound synthesis. Comparative transcriptome analysis showed that transgressive down-regulation (TDR) is the main gene expression pattern in the hybrids (*G. hirsutum* × *G. barbadense*, HB), and it was found that the genes of photosystem I and Adenosine triphosphate (ATP)-binding may promote early growth vigor. Transgressive up-regulation (TUR) is the major primary gene expression pattern in the hybrids (*G. barbadense* × *G. hirsutum*, BH), and photosystem II-related genes mediated the performance of early biomass heterosis. The above results demonstrated that overdominance mediates biomass heterosis in interspecific hybrid cotton and the supervisory mechanism divergence of hybrids with different females. Photosynthesis and other metabolic process are jointly involved in controlling early biomass heterosis

OPEN ACCESS

Edited by:
Dayun Tao,
Yunnan Academy of Agricultural Sciences, China

Reviewed by:
Kailiang Bo,
Institute of Vegetables and Flowers (CAAS), China
Claudio Benicio Cardoso-Silva,
State University of the North Fluminense Darcy Ribeiro, Brazil

***Correspondence:**
Yongshan Zhang
13938698299@163.com
Junkang Rong
junkangrong@126.com

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

Received: 09 March 2022
Accepted: 06 June 2022
Published: 01 July 2022

Citation:
Li T, Wang F, Yasir M, Li K, Qin Y, Zheng J, Luo K, Zhu S, Zhang H, Jiang Y, Zhang Y and Rong J (2022) Expression Patterns Divergence of Reciprocal F_1 Hybrids Between Gossypium hirsutum and Gossypium barbadense Reveals Overdominance Mediating Interspecific Biomass Heterosis. Front. Plant Sci. 13:892805. doi: 10.3389/fpls.2022.892805

in interspecific hybrid cotton. The expression pattern data of transcriptome sequencing were supported using the qRT-PCR analysis. Our findings could be useful in theoretical and practical studies of early interspecific biomass heterosis, and the results provide potential resources for the theoretical and applied research on early interspecific biomass heterosis.

Keywords: cotton, interspecific hybrid, biomass vigor, transcriptome analysis, overdominant

INTRODUCTION

Heterosis, also known as hybrid vigor, refers to the phenomenon in which hybrids outperform their inbred parents in terms of development, biomass, yield, and fertility (Hochholdinger and Baldauf, 2018). So far, extensive research has revealed the mechanism of heterosis in rice and maize at various levels (Huang et al., 2016; Xiao et al., 2021) by multi-omics, which dramatically accelerates the use of hybrids in agriculture to meet human needs. During the last century, various hypotheses such as dominance, overdominance, and epistasis (Williams, 1959; Xiao et al., 1995; Yu et al., 1997; Li et al., 2001; Luo et al., 2001; Hua et al., 2003) were proposed to explain the hybrid vigor. In addition, incomplete dominance of alleles was also believed to be the main reason for the hybrid vigor in the preliminary work (Yang et al., 2017). Recently, epigenetics was found to be another molecular mechanism controlling the performance of heterosis through regulating gene expression, and several breakthroughs have been achieved in this field (Dapp et al., 2015; Song et al., 2017; Lauss et al., 2018; Sinha et al., 2020). However, systematically demonstrating heterosis using one of these hypotheses is difficult.

Gossypium hirsutum is widely cultivated due to its higher fiber yield and acceptable quality, accounting for more than 90% of the world's cotton cultivated acreage (National Cotton Council, 2006[1]; Chen et al., 2007; Teodoro et al., 2019), although fiber quality of island cotton is better than that of upland cotton. Cotton breeders have been trying to enhance the yield (Shahzad et al., 2020a) and quality (Li et al., 2018) of upland cotton by utilizing hybrid vigor, and rapid progress is being made in variety breeding. Despite extensive research into fiber yield and quality heterosis, only a few intraspecific upland cotton hybrids have been successfully used in cotton production. Interspecific hybrids were thought to have a lot of potential in crop production because of their hybrid vigors. The main impediment was most likely the hybrids' massive vegetative growth (Zhang et al., 2014). As a result, interspecific hybrids have received less attention, and only a few have been successfully used in agriculture. Indica-japonica intersubspecific hybrids rice was one of the few intersubspecific hybrids already planted in China. Interspecific hybrids between *G. hirsutum* and *Gossypium barbadense*, meanwhile, received far less attention than intraspecific hybrids, resulting in the discovery of few mechanisms of hybrid vigor.

Biomass heterosis is a widely documented phenomenon at the early seedling development stage of hybrids in multiple species. Previous research has indicated that leaf growth played a pivotal role in biomass heterosis during early vegetative development (Zhu D. et al., 2016), because hybrids own larger leaves, they can provide more photosynthesis area, resulting in more photosynthetic products, such as nutrients, energy, and raw materials for plant growth and development (Liu et al., 2020). *Arabidopsis* hybrids, for example, the process of photosynthesis earlier than their parents during the seedling vegetative development stage, were accompanied by the expression of photosynthesis-related genes such as chloroplast-targeted genes and photosynthetical system genes (Fujimoto et al., 2012). Furthermore, Liu et al. (2021) investigated the transcriptional regulatory network in hybrids together with their parents during early seedling development, and they found that dominant expression complementary capacities between photosynthesis and cell cycle contribute to *Arabidopsis* interspecific heterosis. The two fundamental biological pathways of photosynthesis and auxin biosynthesis play significant roles in biomass heterosis, which were identified in *Brassica napus* hybrids (Zhu et al., 2020). In addition, the balance between primary and secondary metabolism is also likely to serve important function in biomass heterosis in the allodiploid hybrid between *Brassica rapa* and *Raphanus sativus* (Gibum et al., 2020). Similarly, researchers have discovered that genes of the circadian rhythm pathway were down-regulated in leaves and roots in intraspecific hybrids for biomass heterosis of upland cotton (Shahzad et al., 2020b). If mechanisms related to biomass heterosis such as these crucial genes or pathways can be revealed thoroughly after extensive investigation, breeders can probably find a way to inhibit the unnecessary vegetative growth and subsequently enhance the cotton reproductive development to improve the yield and quality.

Genome-wide transcriptome analysis is an efficient method that has been widely used to investigate genes related to biomass heterosis. In this study, we characterized genome-wide transcriptional profiles of reciprocal hybrids (HB and BH) between *G. hirsutum* (GH) and *G. barbadense* (GB) and their parents to explore the function of homologous genes in biomass heterosis during the early vegetative development stage. Genome-wide comparative transcriptome analysis was performed to investigate the molecular bases of regulating biomass heterosis after interspecific hybridization. We presented the dynamic changes of gene expression in hybrids with different females at vegetative development stages to determine the related genes that the polyploid plants express under interspecific hybridization, the processes involved, and how they can be efficiently controlled for plant architecture and rationally solve the problem that vegetative growth is too exuberant to be directly applied in production.

[1]http://www.cotton.org

MATERIALS AND METHODS

Plant Materials and Measurement of Early Biomass Heterosis

In a previous study, fiber quality and yield traits of 53 interspecific F_1 hybrids derived from the crosses between 12 upland cotton and 5 island cotton varieties and their parents were evaluated in Lin'an, Zhejiang and Sanya, Hainan (Li Z. et al., 2020), and the reciprocal F_1 from crossing between the upland cotton variety Luyuan343 and island cotton variety Achang599 was found having the fiber quality better than their parents. The hybrid HB was developed by the female parent (GH) with the male parent (GB), and its reverse cross BH was developed by crossing the female parent (GB) with the male parent (GH). The seeds of this cross combination were used in this study to investigate biomass heterosis at the seedling stage.

After the cotton seeds were delinted using sulfuric acid, 15 healthy and full seeds were selected from each material. Two uniform seeds of every cultivar were sown in each pot and placed in an artificial climate incubator at 25°C for 16 h light (12000 Lx) and at 20°C for 8 h dark (0 Lx). Eleven days after sowing (DAS), the cotyledons were fully opened, and plant height and leaf area were measured every other day. After 15 DAS, the first true leaf was unrolled and the third true leaf emerged at 21 DAS. We stopped measuring phenotypic traits and biomass heterosis after the appearance of the third true leaf. Six plants of each material were randomly selected to determine the plant height, leaf area, and biomass heterosis estimation. Quantitative analysis of the leaf area of the parents and F_1 hybrids at different stages using ImageJ software as Syed and Khan (2018) described. Fresh biomass of the whole plant was measured and dried in a thermostat oven for 2 h (105°C) and thoroughly dried in a thermostatic blast oven at 80°C.

MS Excel and SPSS (22.0) software were used for statistical analysis, and the significant difference analysis was performed using Student's *t*-test. Mid-parent heterosis (MPH) and over-parent heterosis (OPH) of the F_1 hybrid over its parents for biomass were calculated as follows (Sundrish et al., 2010):

MPH = (F1 – mean P)/mean P in percent

OPH = (F1 – best P)/best P in percent

Three cotyledons (cotyledon stage) and the most apical tip of the true leaf (one-leaf stage and trefoil stage) in four materials were taken per sample for the next experiments, frozen in liquid nitrogen after mixing, and stored at –80°C.

RNA Extraction, Library Construction, Transcriptome Sequencing, and Quantitative Real-Time PCR

The total RNA was extracted from mixed samples using a FastPure® plant total RNA isolation kit (polysaccharides and polyphenolics-rich) (RC401-01, Vazyme Biotech, Co., Ltd., Nanjing, China) according to the manufacturer's instructions. Two biological replicates were performed for every sample. RNA purity and concentration were determined using a NanoDrop 2000 spectrophotometer (Thermo, Shanghai, China). One microliter of the extracted RNA was subjected to 1% agarose gel electrophoresis to check the degradation and contamination. In total, 24 RNA-seq libraries with a data volume of 6 G comprising four cotton materials (GH, GB, HB, BH) at three developmental stages (cotyledon stage, one-leaf stage, and trefoil stage) were constructed using an NEBNext Ultra^{TM} II RNA Library Prep Kit (NEB, E7775), and 125 bp/150 bp paired-end sequences were generated using an Illumina Hiseq^{TM} 2500 platform.

A reverse transcription kit (R212-01, Vazyme Biotech Co., Ltd.) was used to perform reverse transcription to synthesize cDNAs. The quantitative real-time PCR (qRT-PCR) was carried out using the two-step RT-PCR with CFX Connect Real-Time System (Bio-Rad, Hercules, CA, United States) and analyzed by the $2^{-\Delta\Delta CT}$ method.

Data Quality Control and Alignment of RNA-Seq Reads to the Reference Genomes

Before being aligned to the reference genome, adapter-contaminated and low-quality reads were removed from all 24 samples' raw reads using Cutadapt (Kechin et al., 2017) and FastQC software (FastQC[2]). The clean reads of parent samples (GH and GB) were mapped by HISAT 2 (Kim et al., 2019) against the updated genome sequences of *G. hirsutum* (TM-1_ZJU v2.1) and *G. barbadense* (H7124_ZJU) (Hu et al., 2019). The clean reading of F_1 hybrids was mapped to the sum of the reference genomes of the parents. The reference genome data were downloaded from COTTONGEN (Yu et al., 2014). The number of reads uniquely mapped to transcribed regions of each gene was extracted for all gene fragments per kilobase per million mapped reads (FPKM), and the reads mapping on multiple positions were discarded (Mortazavi et al., 2008). The differentially expressed genes (DEGs) with $|\log2 \text{(fold change)}| > 1$ and an adjusted (p-value < 0.05) by DEseq2 (Love et al., 2014) were determined.

Homologous Gene Alignment Between Two Reference Genomic

Protein sequences between *G. hirsutum* genome and *G. barbadense* genome were aligned by BLAST (Altschul et al., 1997) to identify the corresponding relationship of homologous gene pairs (E-value, $<10^{-5}$), which were considered to be homologs having at least 90% sequence in two genomes, and only one pair of homologs depending on the above relationship of orthologous genomic regions was screened as a homologous gene.

Expression Patterns Analysis, Functional Annotation, and Classification of Overdominant Genes

The expression level dominance (ELD) analysis in F_1 hybrids compared to their parents was divided into five categories (**Figure 3A**), total of 12 expression patterns including additivity (I and XII), GH-expression level dominance (ELD-H, IV and IX), GB-expression level dominance (ELD-B, II and XI), transgressive

[2]http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

up-regulation (TUR, V, VI, and VIII), and transgressive down-regulation (TDR, III, VII, and X) using STEM (Short Time-series Expression Miner) software as described by Rapp et al. (2009).

The genes annotation from different expression patterns for gene ontology (GO) enrichment analysis using *G. hirsutum* genome annotation was performed by the OmicShare tools, a free online platform for data analysis,[3] with corrected $p < 0.05$ as the significantly enriched and rich factor.

RESULTS

Early Establishment of Growth Vigor in Interspecific Hybrids

It has been reported that the heterosis of seedling growth vigor may be related to the seed mass (East, 1936); however, this hypothesis was not upheld by the research of heterosis in maize inbred lines hybrids (Ko et al., 2016). In addition, our study depicted that the seed mass of reciprocal F_1 hybrids was significantly higher than that of the parents (**Figure 1A**), seedling growth vigor heterosis occurred soon after germination, and it was significantly greater than in the high parents at the cotyledons unrolled stage (**Supplementary Figure 1**) and persisted throughout the seedling stage (**Figures 1A,B** and **Supplementary Figure 1**).

The measurement of plant height and leaf area started 11 DAS when the cotyledons were unrolled in parents and F_1 hybrids. Plant height and leaf area of F_1 hybrids showed significant or highly significant OPH from the first measurement (**Figures 1C,D**). Both biomass traits peaked at 15 DAS (one-leaf stage). The dry weight and fresh weight of the one-leaf stage were measured, and it was found that the reciprocal F_1 hybrids were significantly greater than the HP value (GB, **Figures 1E,F**).

The leaf is the primary photosynthesis organ in plants and the primary source of storage and utilization of essential substances for growth (Boeven et al., 2020). F_1 hybrids had greater leaf area and plant height than their parents, which was highly significant at the first true leaf unrolled stage (**Figures 1C,D**). The OPH of plant height is greater than 40% (**Figure 1C**), and the OPH of leaf area is more than 80% (**Figure 1D**). Moreover, there is a lower peak that occurred at the third true leaf unrolled (Leaf area: 21 DAS) or 2 days prior to the trefoil stage (Plant height: 19 DAS). Then, the heterosis decreased from extremely meaningful to a significant level, the trend essentially flattened and stabilized at the OPH of about 10%.

Homologous Expression Divergence Is Caused by Hybridization During the Early Vegetative Growth Development Stage

We harvested the leaves of three representative stages: cotyledon stage, one-leaf stage, and trefoil stage according to the above results (**Figures 1C,D**), and transcriptome sequencing was performed subsequently. Leaf tissues of the parents (GH and GB) and hybrids (HB and BH) at three stages were used to build 24 cDNA libraries for sequencing. The overall sequencing means of total mapped was 94.74%, and the value of Q30 was 91.49% in our sequencing data. In the leaf of GH, GB, HB, and BH, the mean of clean reads were approximately 43.62, 45.63, 43.60, and 43.85 million, respectively. In the parents, more than 91.87 and 93.70% mean of unique reads were mapped to the *G. hirsutum* and *G. barbadense* reference genome, respectively. In the F_1 hybrids, the mean of unique reads was 41.69 and 42.21%, which is due to mapping to the addition of the two sets of reference genomes. Therefore, the mean number of reads mapped to the gene region of parents and hybrids was ∼85.68, 80.31, 79.50, and 86.35%, respectively (**Supplementary Table 1**).

Analysis of transcript accumulation patterns in parents relative to their reciprocal F_1 hybrids, an average of genes expressed at different stages was investigated by RNA-seq (**Figure 2A**). The average expression level of the F_1 hybrids compared with the parents shows minor differences, but it is not significant (**Figure 2B**).

To study the gene expression patterns after interspecific hybridization, we first established the relationship between homologous gene pairs among two reference genomes by whole-genome alignment and aligned the expression profile according to the homologous gene pairs of *G. hirsutum* and *G. barbadense* to construct the interworking best hits.

The proportion of DEGs between hybrids with different females and their parents was different at three seedling stages (**Figure 2C**). The largest differentially up-regulated genes after hybridization were found in BH when compared with the male parent GH (15.70%), and the largest down-regulated genes were found in BH when compared with GH too (32.73%). The BH with maternal GB had the fewest DEGs that were up-regulated and down-regulated after hybridization, with 4.62 percent and 20.13 percent, respectively (**Figure 2C**). DEGs analysis of hybrids and parents found that the number of down-regulated genes was significantly higher than the up-regulated genes.

To explore the specific biological processes that were enriched in the process at the one-leaf stage, we classified the DEGs in the top 50 and selected the top 15 according to ascending order of FDR value and rich factor. Assessment of biological processes for the up- and down-regulated DEGs in hybrids and parents found that most genes were enriched in the same GO terms, involved in the chloroplast thylakoid membrane, chloroplast part, chloroplast thylakoid, photosystem, cellular biosynthetic process, cellular process, plastid part, plastid thylakoid, plastid thylakoid membrane, thylakoid membrane, and RNA binding. DEGs involved in photosynthesis, intracellular membrane-bounded organelle, membrane-bounded organelle, and protein complex were upregulated, whereas the DEGs involved in the biosynthetic process, cellular nitrogen compound metabolic process, organelle envelope, organic substance biosynthetic process were downregulated in both the GB vs. BH, GB vs. HB, GH vs. BH, and GH vs. HB comparisons (**Figures 2D,E**). Nevertheless, a portion of the allele derived from the parent after hybridization were both up- and down-regulated the genes related to the chloroplast, cellular process, and plastids. Hybrids strengthened the nitrogen metabolism and organelles in the remaining items while weakening the membrane binding.

[3] http://www.omicshare.com/tools

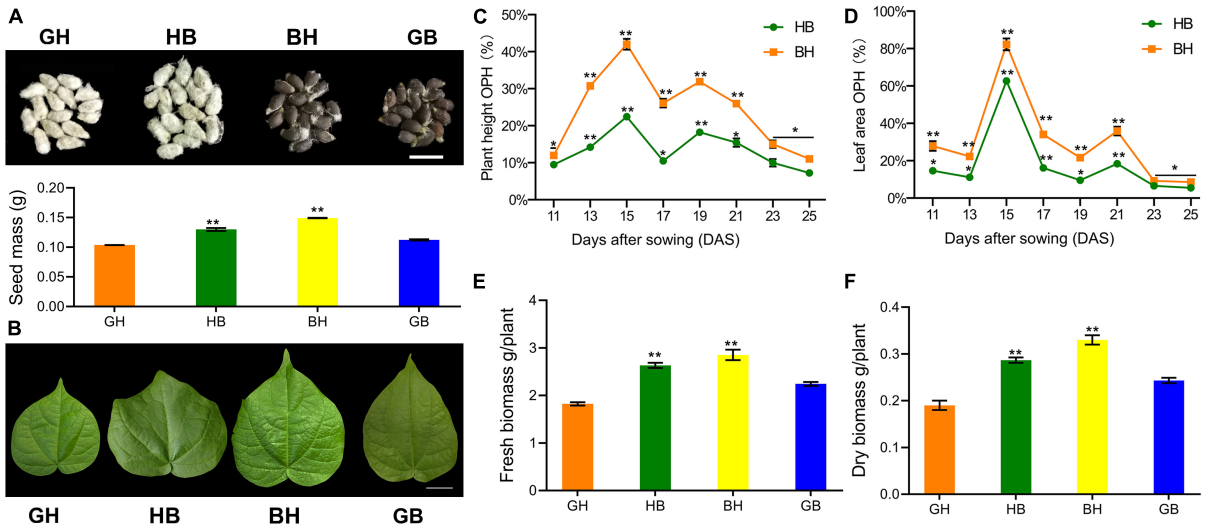


FIGURE 1 | Biomass heterosis exists in interspecific hybrids at the early seedling stage. **(A)** Seed morphology (upper panel) and weight (lower panel, means ± SEM) in parents (GH and GB) and their reciprocal F_1 hybrids (HB: GH × GB and BH: GB × GH). Scale bar = 10 mm. **(B)** Leaf morphology and area at one-leaf stages in GH, GB, and their hybrids. Scale bar = 10 mm. **(C)** Plant height over-parent heterosis (OPH) performance of HB and BH (mean ± SEM, $n = 6$). **(D)** Leaf area OPH performance of HB and BH (mean ± SEM, $n = 6$). **(E)** Fresh biomass observed in hybrids relative to their parents (mean ± SEM, $n = 6$). **(F)** Dry biomass observed in hybrids relative to their parents (mean ± SEM, $n = 6$). Here ** is used for significant difference with high parents (HP, GB) performance at $p < 0.01$ and * at $p < 0.05$.

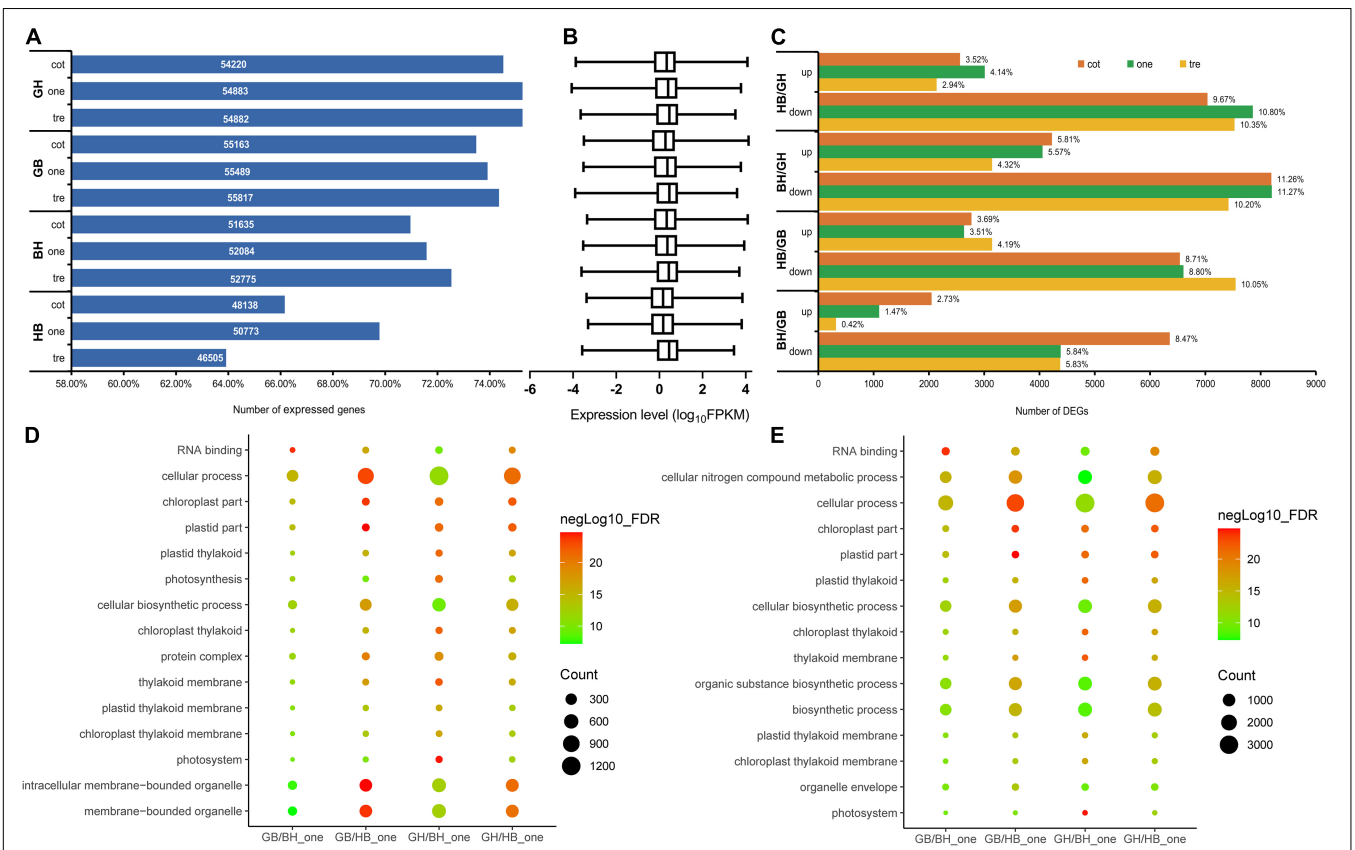


FIGURE 2 | Transcriptome sequencing of the parents (GH and GB) and their reciprocal hybrids (HB and BH). **(A)** The number and percentage of genes expressed in the four materials at three vegetative growth development stages. **(B)** Expression levels in the four materials at three vegetative growth development stages. **(C)** Differential expression genes in F_1 hybrids with their parents at a different stage. **(D,E)** The top 15 terms from GO enrichment analysis of DEGs at one-leaf stage, **(D)** up-regulation, and **(E)** down-regulation. cot, cotyledon stage; one, one-true leaf stage; tre, trefoil stage.

Genome-Wide Homologous Genes Expression Patterns Divergence of Reciprocal F_1 Leads to Different Mechanisms of Biomass Heterosis

Expression level dominance has been discovered in allopolyploids, a phenomenon in which progeny gene expression is statistically similar to that of one parent (Shen et al., 2017; Boeven et al., 2020), including asymmetric expression patterns in the polyploidy process from two diploids to tetraploid cotton (Yoo et al., 2012). In this section, we classified the expression patterns of the genes in parents and hybrids into 12 groups and divided them into five categories, including Additive (I and XII), ELD-H (IV and IX), ELD-B (II and XI), TUR (V, VI, and VIII), and TDR (III, VII, and X) according to the description of Rapp (Rapp et al., 2009; **Figure 3A**).

The results showed the expression pattern of transgressive regulation with the largest number of genes in HB hybrids (**Figure 3B**), among which the pattern of TDR contained more genes than TUR. In the ELD of parents, it can be observed that the number of genes of ELD-B (13.01%) is more than that of ELD-H (11.41%), but the difference was not significant (less than 2%). Among the five categories of expression patterns, additives contained the least number of genes (9.17%). The patterns in BH showed transgressive regulation contained the largest number of genes (61.79%), the same as HB, in which TUR had the largest gene classification, followed by ELD-H (17.04%), which was different from HB (**Figure 3C**). And, the third was TDR (15.40%), followed by ELD-B (10.95%) and additives (10.22%).

Taken together, the interspecific heterosis was possibly due to the transgressive regulation, but the reciprocal F_1 hybrids (HB and BH) exhibited a distinct expression pattern. In F_1 hybrids (both HB and BH), the expression pattern transgressive regulation (including TUR and TDR) exceeds 60% of the total expressed genes. The biomass heterosis of HB with upland cotton as a female parent was generated through the expression pattern of TDR, while the expression pattern of TUR contributed to the heterosis of BH with island cotton as a female parent. Furthermore, the results of expression pattern analysis revealed that overdominance at the genome expression level plays an unusual role in the early biomass heterosis in interspecific hybrid cotton.

The Comparison of Overdominance Genes Between Reciprocal Hybrids

The expression divergence of overdominance genes (TUR and TDR) at different seedling stages was compared to investigate the gene expression pattern in reciprocal hybrids at early biomass vigor. The findings revealed that the total number of transgressive up-regulated genes in BH was significantly higher than in HB, but not in TDR (**Figures 4A,B**).

A comparison of the transgressively up-regulated genes expressed in HB at various seedling stages revealed that the number of expressed genes increased over time, but there was no significant difference between the cotyledon stage and the one-leaf stage (**Figure 4A**). Co-expression analysis of genes at different seedling stages showed that 744 genes were co-expressed at three developmental stages in HB (**Figure 4B**). Differences among BH and HB in the number of transgressive up-regulated genes were observed from cotyledon stages and peaked at the one-leaf stage, followed by the cotyledon stage and the three-leaf stage (**Figure 4A**), and 1796 genes were co-expressed at these three stages in BH (**Figure 4C**). The number of transgressive down-regulated genes expressed at various seedling stages revealed that two hybrids possessed the greatest number of expressed genes at the one-leaf stage (**Figure 4D**). Co-expression analysis of the TDR genes expressed in HB found that 5917 genes were co-expressed in these three stages (**Figure 4E**), whereas 4237 genes were co-expressed in BH (**Figure 4F**).

An overall comparison of TUR and TDR genes expression at different developmental stages and comparative co-expression analysis of genes at all seedling developmental stages suggested a tendency for an increased number of expressed genes at the one-leaf stage, except for TUR genes of HB, which complies with our previous phenotypic analysis (**Figure 1**). These findings indicated that an increase in the number of expressed genes could have been just one of many factors contributing to biomass heterosis. Meanwhile, differentially expressed patterns were performed in hybrids with different females, and co-expressed genes play important roles at different stages.

Functional Annotations of Overdominance Genes

To better understand the functions of the overdominant homeolog expression genes in hybrids and parents during early vegetative development, GO enrichment analysis was performed on these genes, which had previously been identified as commonly expressed (**Figures 4C,E**).

Gene ontology enrichment analysis of transgressive up-regulated (TUR) genes in BH compared with its parents was enriched in 26 terms (p value < 0.05), such as glycerol ether metabolic process, transcription coactivator activity, photosystem II, photosynthesis, and oxidoreductase activity (**Figure 5A**). Heatmaps were created to display the expression patterns and clustering of photosystem II, photosynthesis, and oxidoreductase activity genes in BH and its parents (**Figures 5B–D**). Furthermore, at three different developmental stages, the gene expression representation of BH was both higher and more consistent than that of its parents.

Conversely, the transgressive down-regulated (TDR) gene expression pattern in HB compared with its parents found that match 172 Go terms in summary (p value < 0.05) compared with its parents and selected the top 15 display with the lowest p value (**Figure 6A**). The most significantly enriched terms were the cytoplasm; following enriched terms such as protein folding, unfolded protein binding, and RNA binding, it was found that Adenosine triphosphate (ATP) binding terms enriched the most DEGs. In addition, we found that the terms related to photosynthesis (light-harvesting, oxidoreductase activity, photosystem I, and NAD binding) were also enriched. Heatmap display of gene expression patterns of TDR is related to photosystem I, oxidoreductase activity, light-harvesting, and

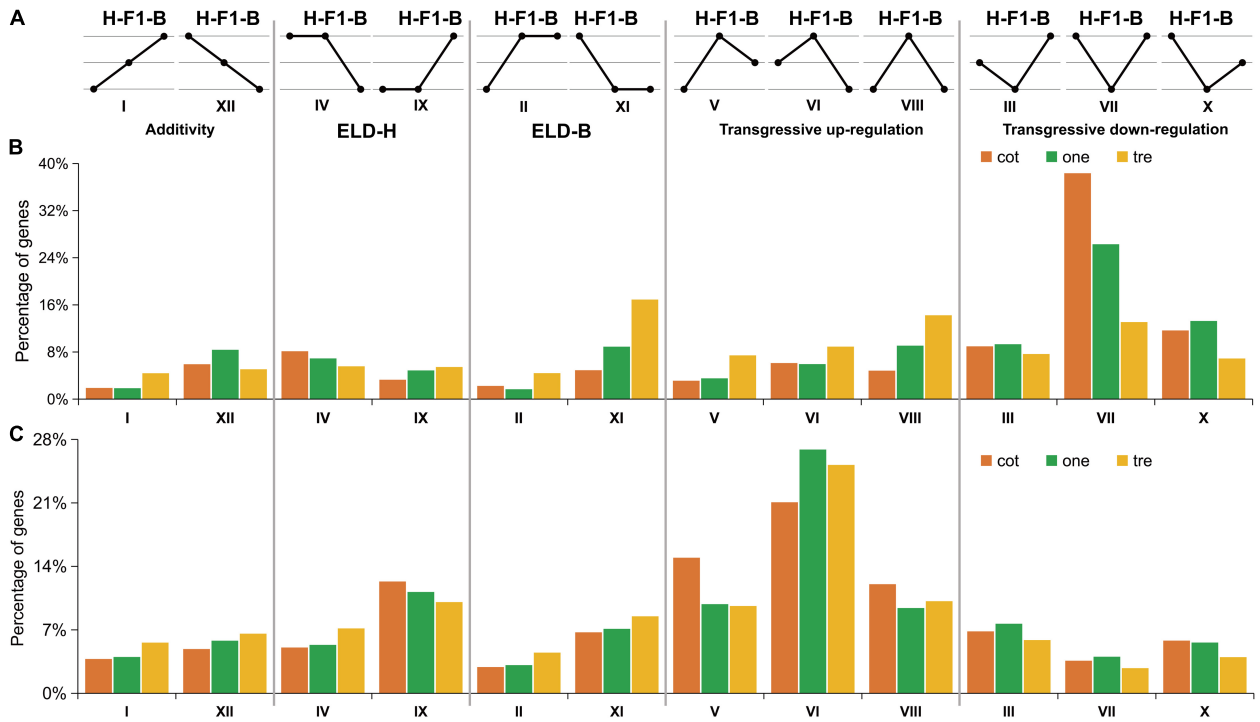


FIGURE 3 | The 12 possible differential expression patterns in F_1 hybrids compared with their parents at different stages. **(A)** Expression patterns of 12 groups, H: *Gossypium hirsutum*; F_1: hybrid, and B: *Gossypium barbadense*. Roman numerals indicate the same categorization as used in Rapp et al. (2009). **(B)** The percentage of genes from the 12 patterns in HB relative GH and/or GB. **(C)** The percentage of genes from the 12 patterns in BH relative GH and/or GB. cot, cotyledon stage; one, one-true leaf stage; tre, trefoil stage.

NAD binding genes in HB and its parents (**Figures 6B–E**). These results suggested that the expression of HB in photosynthesis-related genes was lower than that of parents in three stages.

All these results indicate that up- or down-regulation of photosynthesis and ATP-related genes might be the reason for initial biomass heterosis. Furthermore, genes that regulate heterosis in the photosynthesis system were different; the majority of the up-regulated genes were enriched in photosystem II, such as genes containing Psb-related domains. Contrarily, the majority of the down-regulated genes were enriched in photosystem I, photosystem I reaction center, and light-harvesting, such as CAB, LHC family, and genes containing Psa-related domains. This result was consistent with a study on early biomass heterosis in *B. napus (Canola)* hybrids (Zhu et al., 2020). Additionally, down-regulated genes included the NAD binding genes to reduce hydrogen ions to H_2.

Expression Pattern Verification by Quantitative Real-Time PCR

Validation of RNA-Seq expression pattern data with qRT-PCR analysis at three early seedling stages in parents and hybrids was done, and the results matched well with the RNA-Seq analysis described earlier (**Figure 7**). The qPCR results confirmed the trend in expression of these selected genes including that pathway in the photosystem I (*PSAF*), light-harvesting (*lhcA-P4*), NAD binding (*GPDH*), photosystem II oxygen-evolving complex (*PNSL2*), photosystem II (*psbW*), and oxidoreductase activity (*FRO1*), despite some differences in expression levels. Three genes from HB showed a tendency of down-regulation, whereas three genes from BH, exhibited a trend of up-regulation.

DISCUSSION

Overview of Early Biomass Heterosis and Comparative Transcriptome Analysis

Higher fiber yields and quality can be achieved by reducing vegetative growth in favor of reproductive growth, which has been a major goal of cotton breeding. This may be especially important in elucidating the mechanism of early biomass heterosis in order to better serve breeding efforts in that case. It has been found that hybrids exhibited stronger vegetative growth vigor at the seedling stage for essential crops such as rice (Zhu D. et al., 2016), maize (Li Z. et al., 2020), wheat (Liu et al., 2018), and cotton (Ding et al., 2021) in previous research.

In this study, we found that the maximum OPH at the first true leaf (one-leaf stage), followed by decreasing trend that gradually leveled off, similar findings have been reported in maize (Liu et al., 2018). However, there were no significant differences between hybrids and their parents in the contents of chlorophyll per unit area determined

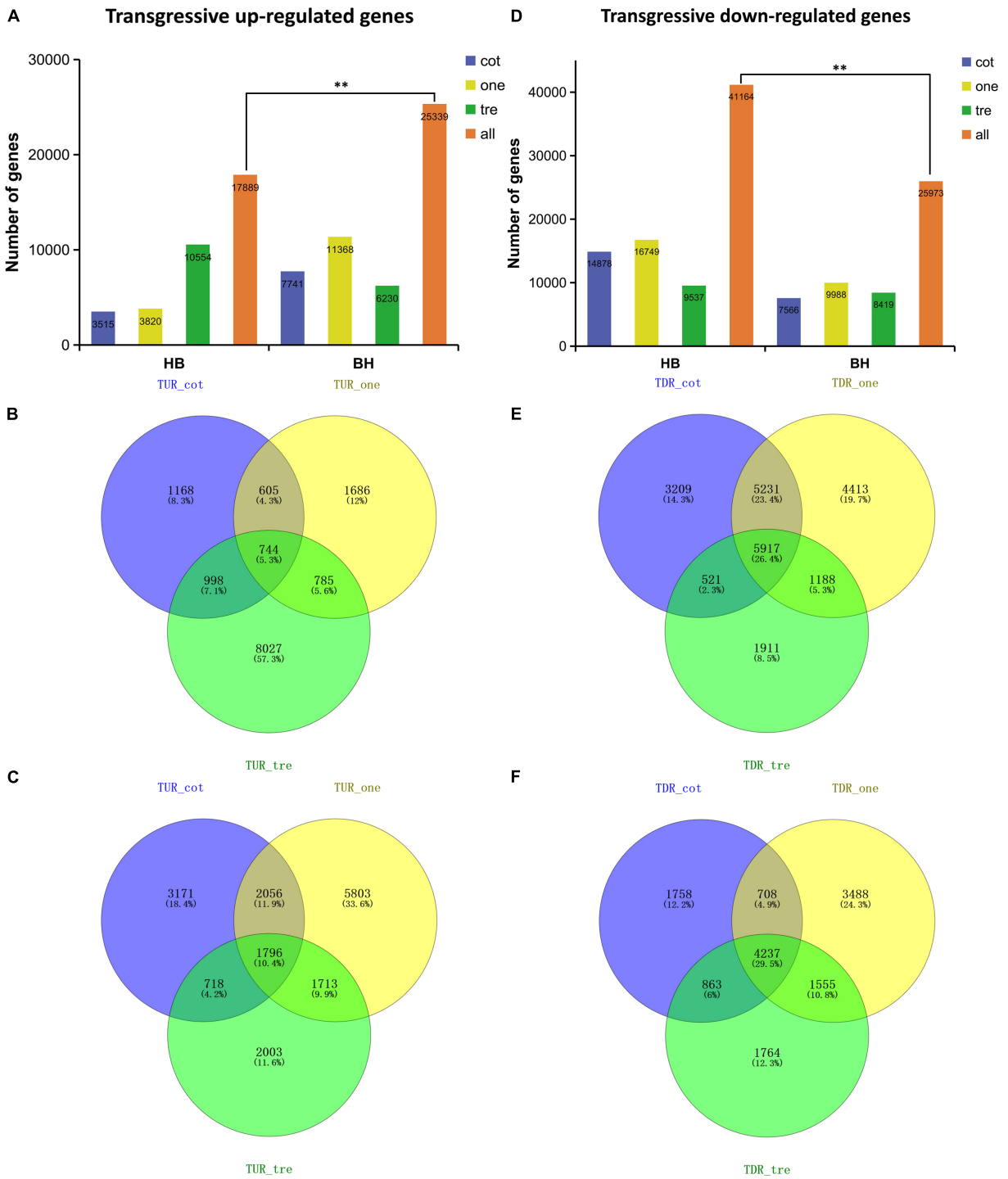


FIGURE 4 | Quantification of genes that showed transgressive regulation patterns from the three different stages. **(A,D)** The number of transgressive up- and down-regulation expression patterns at three different stages. **(B)** Transgressive up-regulation (TUR) expression pattern in HB. **(C)** TUR expression pattern in BH. **(E)** Transgressive down-regulation (TDR) expression pattern in HB. **(F)** TDR expression pattern in BH represents the distribution of unique and common expression genes. Here ** is used for the number of transgressive regulated genes in HB performed significant difference with that in BH at $p < 0.01$.

(**Supplementary Figure 2**), the results comply with previous researches on chlorophyll fluorescence parameters determined in *Arabidopsis* (Zhu A. Y. et al., 2016). We inferred that the one-leaf stage is the critical period in early biomass heterosis performance integrated with biomass-related traits determination and the findings of other research, but this

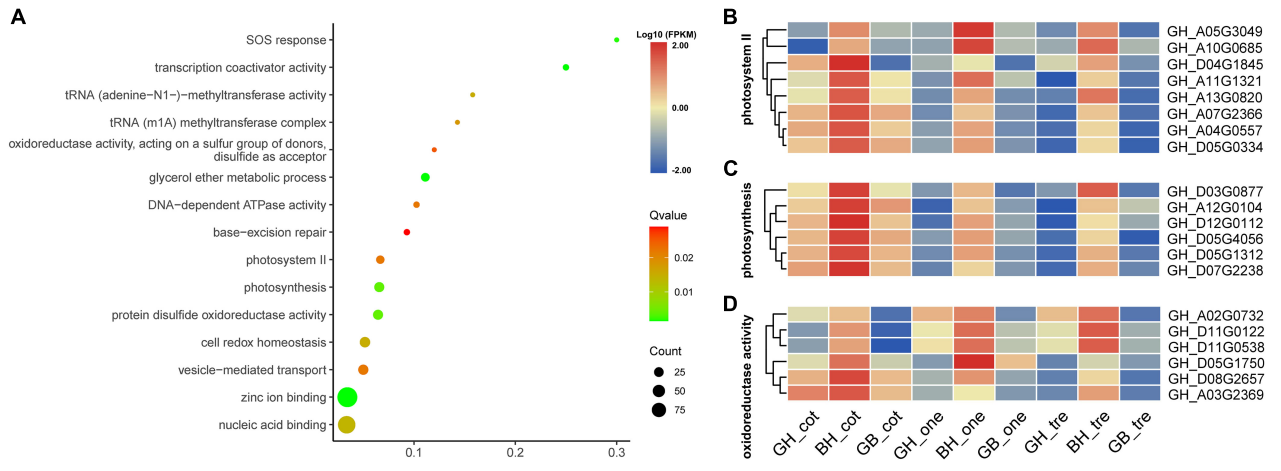


FIGURE 5 | Homolog expression levels in the BH relative to the levels in its parents explain the phenomenon of TDR. **(A)** The significantly enriched GO terms of TDR genes from BH relative to its parents. **(B–D)** The expression heatmap of homologous genes with different stages from enriched GO terms of photosystem II, photosynthesis, and oxidoreductase activity.

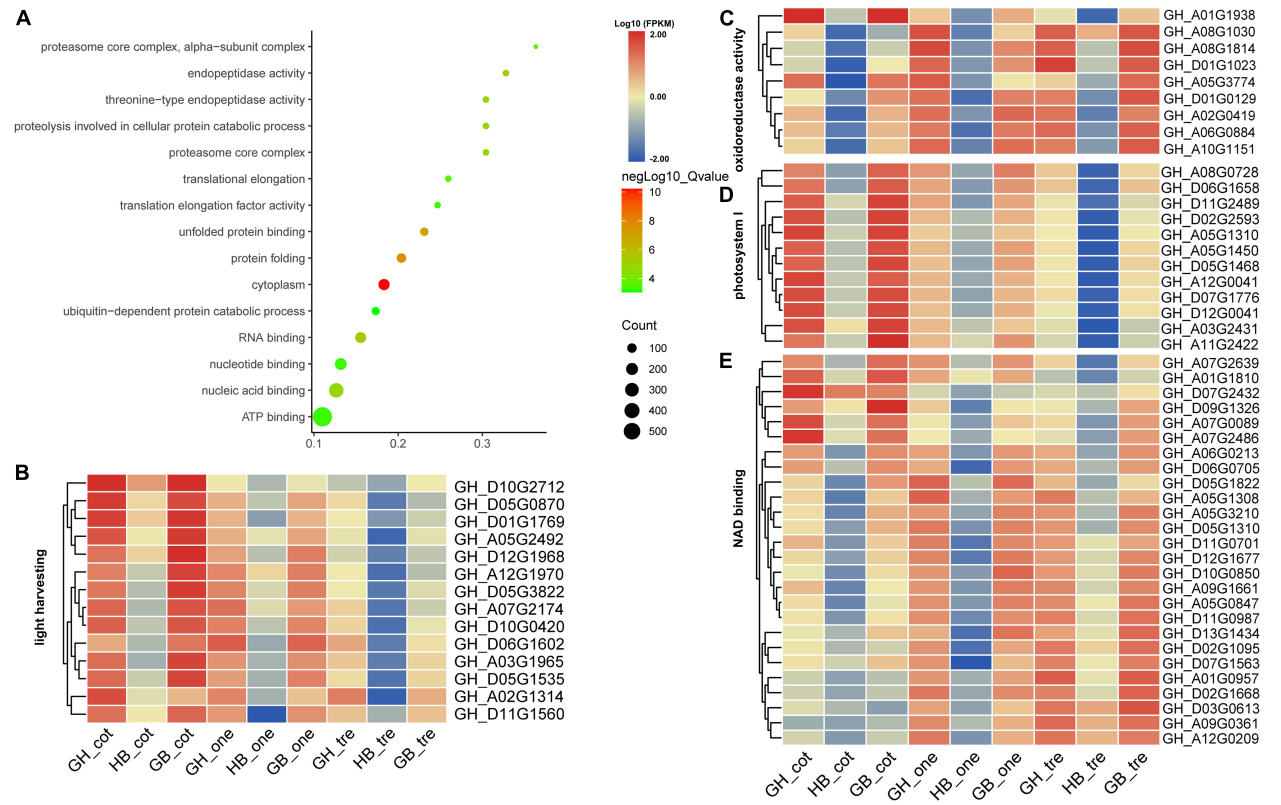


FIGURE 6 | Homolog expression levels in the HB relative to the levels in its parents explain the phenomenon of TUR. **(A)** The significantly enriched GO terms of TUR genes from BH relative to its parents. **(B–E)** The expression heatmap of homologous genes with different stages from enriched GO terms of light-harvesting, oxidoreductase activity, photosystem I, and NAD binding.

phenomenon may not be responsible for the contents of chlorophyll or the intensity of photosynthesis between hybrids and their parents. Certainly, the OPH in the leaf area of reciprocal hybrids compared to their parents causes increased photosynthesis, which exhibits significant interspecific biomass heterosis.

Transcriptome analyses of leaves from reciprocal hybrids and their parents at three seedling stages revealed that the

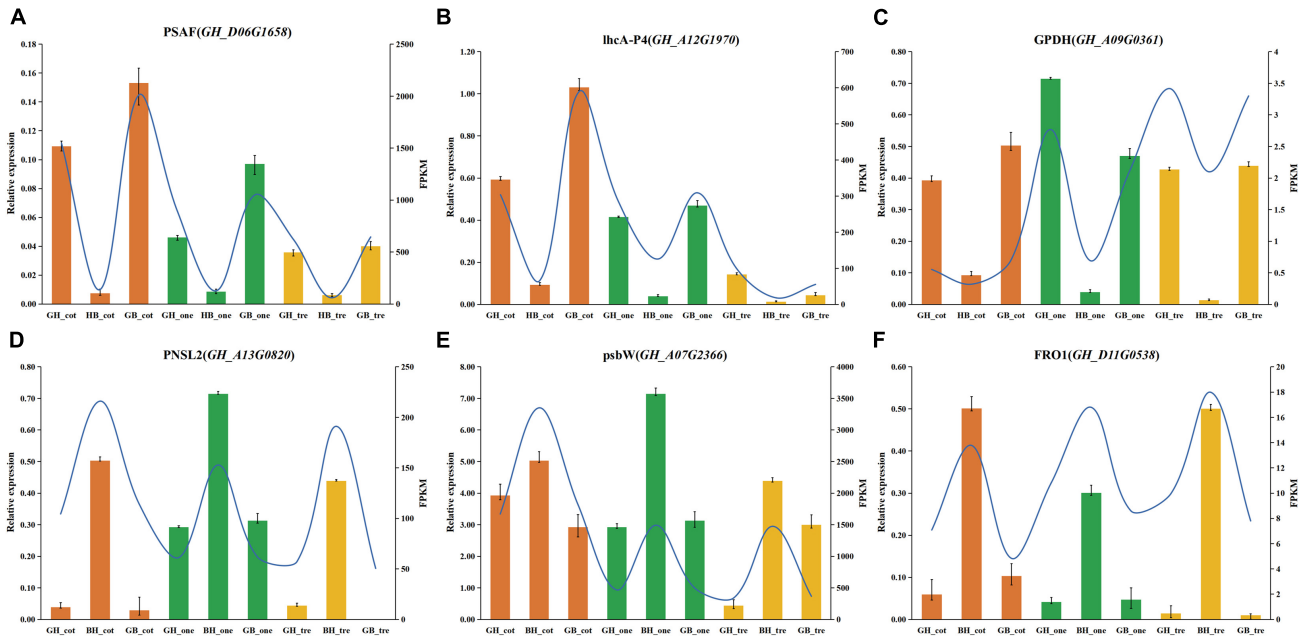


FIGURE 7 | Validation of transcriptional data changes of selected genes from expression patterns of TUR and TDR using the qRT-PCR in hybrids and their parents. **(A–F)** Comparisons of RNA-seq data (FPKM value; blue color lines) with the qRT-PCR (By relative expression data; orange, green, and yellow color columns denote cotyledon stage, one-true leaf stage, and trefoil stage, respectively) of selected genes at three time points during early biomass stages; Panels **(A–C)** indicate genes from HB and panels **(D–F)** indicate genes from BH.

number of expressed genes in hybrids was lower than in parents, particularly in crosses with GB as the female. However, there were no significant differences in gene expression levels between hybrids and parents. This indicates that the performance of biomass heterosis contributed to the inhibition of gene expression after hybridization, similar results have been recorded in maize (Luo et al., 2021), tetraploid *Medicago sativa* (Li et al., 2009), and intraspecific hybridization of upland cotton.

Expression Patterns Divergence of Reciprocal F_1 Hybrids After Hybridization and Both Overdominant Genes Mediate Early Biomass

The evolution of cotton from diploid to tetraploid was a known process of natural hybridization and chromosome doubling. Previous research on the phenomenon of subgenome asymmetry and ELD during hybridization and evolution from diploid to tetraploid in cotton has been published (Peng et al., 2020). Of course, this pattern is largely determined by the up- or down-regulation of homologous genes in non-dominant parents. Furthermore, Yoo et al. (2012) found that new gene expression conditions are selected for evaluation by adjusting the mismatches in natural diploid genomic mergers. Cotton interspecific hybrids exhibited heterosis for yield and quality traits. Breeders struggle to achieve considerable heterosis in intraspecific upland cotton crosses, but it is an intricated task. Early-stage biomass production influences plant yield increases. Keeping this in mind, we used comparative transcriptome analysis to evaluate the genetic basis of early biomass heterosis in hybrid cotton of reciprocal crosses.

Previous studies of comparative transcriptomics analysis in root and leaf found that overdominant gene expression is the major pattern of early biomass heterosis (Shahzad et al., 2020b). These findings comply with the overdominance hypothesis as overdominance was found to be activated in gene expression for upland cotton intraspecific hybrids at different development stages. However, interspecific hybridization has experienced gene flow among different subspecies; whether the hybridization between interspecific and intraspecific represent the exact underlying mechanism remains unknown. It has been reported that overdominant loci are the key genetic factors of lint yield heterosis in interspecific hybrids between *G. hirsutum* and *G. barbadense* (Tian et al., 2019); similarly, Zhao et al. (2009) investigated the mechanism of cotton heterosis in four different stages of growth and development. They detected quantitative and qualitative gene expression differences between hybrids and parents and hypothesized that overdominant gene expressions influenced heterosis (Zhao et al., 2009). The influence of early biomass heterosis was established in our study by differentiating gene expression patterns, inferring the expression of a few genes as additives, similar to that of the male/female parent (ELD) or transgressive up-/down-regulated found by genome-wide expression patterns classification of reciprocal F_1 hybrids with their parents. Overdominance is the main pattern that regulates the performance of interspecific biomass heterosis, and

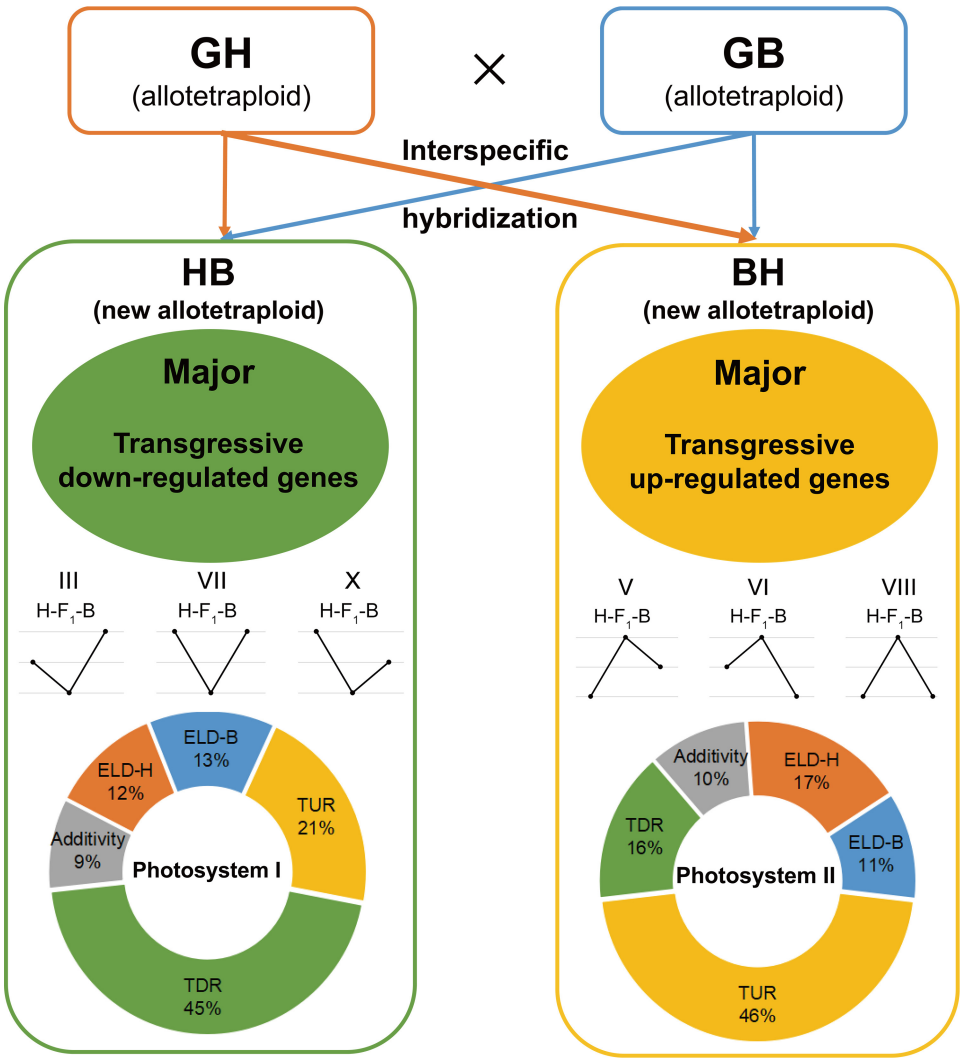


FIGURE 8 | Graphic illustration of early biomass vigor over-parent heterosis (OPH) performance by overdomiant regulated genes in interspecific hybrids and their parents. GH, GB, HB, and BH are the abbreviation of *G. hirsutum*, *G. barbadense*, *G. hirsutum* × *G. barbadense*, and *G. barbadense* × *G. hirsutum*, respectively. The circle proportion refers to the average proportion of different five types [expression level dominance-GH (ELD-H), expression level dominance-GB (ELD-B), transgressive up-regulation (TUR), transgressive down-regulation (TDR), additivity] in three stages. The representation of III, VII, and X (belonging to TDR) and V, VI, and VIII (belonging to TUR) is consistent with that of **Figure 3**.

more than 60% of the genes followed the pattern of transgressive regulation (TUR and TDR). Although the overdominance regulates the performance of interspecific biomass heterosis in both reciprocal F_1 hybrids, the predominance of up-regulation or down-regulation is opposite in hybrids with a different female parent, and related studies have not been reported previously. We found that TDR is the major pattern in HB (female, *G. hirsutum*), and this is consistent with the previous research on intraspecific hybridization of upland cotton. The hybrids BH with island cotton as a female parent are just the opposite, which is regulated by TUR. Moreover, the expression pattern of these two hybrids shows a trend for the ELD of the paternal genome, this trend was not significantly different (**Figure 8**, proportion of ELD-male expression in the torus diagram).

Photosynthesis-Related Genes Play a Critical Role in Early Biomass Interspecific Heterosis

Biomass heterosis is mainly produced by TUR/TDR regulation patterns in reciprocal hybrids (**Figure 4**). It was found that the differential expression of photosynthesis-related genes plays a key role in interspecific heterosis by the further functional analyses of overdominant genes. This is probably attributed to the higher leaf area, rather than the difference in chlorophyll content or other photosynthetic parameters (**Figure 1**). The results are shown in *Arabidopsis* interspecific hybrids between *Col-0* and *Per-1* as well, and it was found that the transcriptional level of hybrids was greater than their parents, even though the internal structure of the leaf is stable (Liu et al., 2021). Likewise, research into the

hybrids between *Col* and *C24* found that mature embryo organ sizes or cell numbers were not altered in a leaf unit area (Meyer et al., 2012). These findings are consistent with our previous studies (Li T. Y. et al., 2020).

Functional analysis of transgressive regulation genes found that it was not uniform in expressed genes enriched in photosynthesis between HB and BH. These items were similar to those described above, except that they are both enriched in photosynthesis and oxidoreductase activity terms. It indicated that hybrids are similar in regulating biomass through photosynthesis-related genes, but there are still some differences that TUR genes are mainly enriched in photosystem I, additionally, there are NAD, light-harvesting, and ATP-binding related genes. In contrast, TDR genes were mainly enriched in photosystem II, as well as in terms of cell redox homeostasis and DNA-dependent ATPase activity. Neither biomass nor yield was found to be associated with photosystem II at heterosis in the previous work, which exemplified our hypothesis too. On the other hand, we did not identify any additional relevant studies that suggest that photosystem I is able to maintain a complementary relationship with photosystem II, nonetheless, the repair system in photosystem II is faster than in photosystem I. This might be due to the photosystem I binding chlorophyll *a* and the photosystem II binding chlorophyll *a* and *b* (Chen et al., 2021). It has been found that improving the ability of the light system in maize can promote the heterosis of biomass (Meena et al., 2021). These findings provide a creditable theoretical basis for the high expression of photosystem II related genes to promote the accumulation of photosynthetic products when photosystem I related genes are inhibited.

CONCLUSION

It is worth noting that cotton hybrids demonstrated biomass vigor throughout the early phases of vegetative growth. Herein, the one-leaf stage is the most significant stage in early biomass heterosis revealed by comparison of hybrids and their parents, and furthermore, it has indicated that larger leaves in hybrids enhanced photosynthesis compound synthesis. As shown in **Figure 8**, comparative transcriptome analysis and expression patterns of reciprocal F_1 hybrids and their parents reveal overdominance mediates interspecific biomass heterosis. However, expression patterns divergence in hybrids with a different female. TDR is the main patterning among the hybrids with upland cotton as a female parent (HB), and it was found that genes of photosystem I and ATP-binding may promote early growth vigor compared to hybrids and their parents. TUR is the major pattern among the hybrids with island cotton as a female parent (BH), and up-regulation of photosystem II-related genes mediated the performance of early biomass heterosis. In parallel, the interaction of the functional gene to exhibit heterosis between photosystem I and photosystem II still needs to be clarified. Our findings add to the body of knowledge in the field of interspecific biomass heterosis between *G. hirsutum* and *G. barbadense*, specifically the mechanism of OPH among hybrids with different female parents. Our findings and hypotheses have a greater potential for application to interspecific hybrid cotton breeding.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the NCBI repository, BioProject accession number: PRJNA851519.

AUTHOR CONTRIBUTIONS

TL: conceptualization, data curation, software, and writing – original draft, review and editing. FW: data curation and software. MY: writing – original draft, review and editing. KLi: software. YQ and HZ: conceptualization. JZ: formal analysis. KLuo and SZ: writing – original draft. YJ: funding acquisition. YZ: writing – original draft, review and editing, and funding acquisition. JR: conceptualization, data curation, writing – review and editing, and funding acquisition. All authors assisted in the critical follow-up of the work and read and approved the manuscript.

FUNDING

This work was supported by the National Key R&D Program for Crop Breeding (2016YFD0101417), State Key Laboratory of Cotton Biology Open Fund (CB2021C11), and the Natural Science Foundation of Zhejiang Province, China (Y21C130006).

ACKNOWLEDGMENTS

We thank Nanjing Personalbio Corporation (http://www.personalbio.cn/) for assistance with RNA-seq.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.892805/full#supplementary-material

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Boeven, P. H. G., Zhao, Y. S., Thorwarth, P., Liu, F., Maurer, H. P., Gils, M., et al. (2020). Negative dominance and dominance-by-dominance epistatic effects reduce grain-yield heterosis in wide crosses in wheat. *Sci. Adv.* 6:eaay4897. doi: 10.1126/sciadv.aay4897

Chen, Y., Yamori, W., Tanaka, A., Tanaka, R., and Ito, H. (2021). Degradation of the photosystem II core complex is independent of chlorophyll degradation mediated by Stay-Green Mg^{2+} dechelatase in *Arabidopsis*. *Plant Sci.* 307:110902. doi: 10.1016/j.plantsci.2021.110902

Chen, Z. J., Scheffler, B. E., Dennis, E., Triplett, B. A., Zhang, T., Guo, W., et al. (2007). Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiol.* 145, 1303–1310. doi: 10.1104/pp.107.107672

Dapp, M., Reinders, J., Bediee, A., Balsera, C., Bucher, E., Theiler, G., et al. (2015). Heterosis and inbreeding depression of epigenetic *Arabidopsis* hybrids. *Nat. Plants*. 1:15092. doi: 10.1038/nplants.2015.92

Ding, Y. H., Zhang, R., Zhu, L. F., Wang, M. J., Ma, Y. Z., Yuan, D. J., et al. (2021). An enhanced photosynthesis and carbohydrate metabolic capability contributes to heterosis of the cotton (*Gossypium hirsutum*) hybrid 'Huaza Mian H318', as revealed by genome-wide gene expression analysis. *BMC Genomics* 22:277. doi: 10.1186/s12864-021-07580-8

East, E. M. (1936). Heterosis. *Genetics* 21, 375–397. doi: 10.1093/genetics/21.4.375

Fujimoto, R., Taylor, J. M., Shirasawa, S., Peacock, W. J., and Dennies, E. S. (2012). Heterosis of *Arabidopsis* hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7009–7014. doi: 10.1073/pnas.1204464109

Gibum, Y., Hosub, S., Hye, R. P., Jeong, E. P., Jong, H. A., Sooyeon, L., et al. (2020). Revealing biomass heterosis in the allodiploid xBrassicoraphanus, a hybrid between *Brassica rapa* and *Raphanus sativus*, through integrated transcriptome and metabolites analysis. *BMC Plant Biol.* 20:252. doi: 10.1186/s12870-020-02470-9

Hochholdinger, F., and Baldauf, J. A. (2018). Heterosis in plants. *Curr. Biol.* 28, R1089–R1092. doi: 10.1016/j.cub.2018.06.041

Hu, Y., Chen, J. D., Fang, L., Zhang, Z. Y., Ma, W., Niu, Y. C., et al. (2019). *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* 51, 739–748. doi: 10.1038/s41588-019-0371-5

Hua, J. P., Xing, Y. Z., Wu, W. R., Xu, C. G., Sun, X. L., Yu, S. B., et al. (2003). Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. U.S.A.* 100, 2574–2579. doi: 10.1073/pnas.0437907100

Huang, X. H., Yang, S. H., Gong, J. Y., Zhao, Q., Feng, Q., Zhan, Q. L., et al. (2016). Genomic architecture of heterosis for yield traits in rice. *Nature* 537, 629–633. doi: 10.1038/nature19760

Kechin, A., Boyarskikh, U., Kel, A., and Filipenko, M. (2017). CutPrimers: a new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J. Comput. Biol.* 24, 1138–1143. doi: 10.1089/cmb.2017.0096

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Ko, D. K., Rohozinski, D., Song, Q. X., Taylor, S. H., Juenger, T. E., Harmon, F. G., et al. (2016). Temporal shift of circadian-mediated gene expression and carbon fixation contributes to biomass heterosis in maize hybrids. *PLoS Genet.* 12:e1006197. doi: 10.1371/journal.pgen.1006197

Lauss, K., Wardenaar, R., Oka, R., Hulten, M. H. A., Guryev, V., Keurentjes, J. J. B., et al. (2018). Parental DNA methylation states are associated with heterosis in epigenetic hybrids. *Plant Physiol.* 176, 1627–1645. doi: 10.1104/pp.17.01054

Li, C., Yu, H. R., Li, C., Zhao, T. L., Dong, Y. T., Deng, X. L., et al. (2018). QTL mapping and heterosis analysis for fiber quality traits across multiple genetic populations and environments in upland cotton. *Front. Plant Sci.* 9:1364. doi: 10.3389/fpls.2018.01364

Li, T. Y., Chao, X., Li, Y. M., Gou, C. F., Hong, Z., Ding, M. Q., et al. (2020). Identification and analysis of fiber quality and yield related traits of interspecific (*Gossypium hirsutum* L. × *G. barbadense* L.) hybrids. *Cotton Sci.* 32, 348–359. doi: 10.11963/1002-7807.ltyscd.20200612

Li, X. H., Wei, Y. L., Nettleton, D., and Brummer, E. C. (2009). Comparative gene expression profiles between heterotic and non-heterotic hybrids of tetraploid *Medicago sativa*. *BMC Plant Biol.* 9:107. doi: 10.1186/1471-2229-9-107

Li, Z., Zhu, A., Song, Q., Chen, H. Y., Harmon, F. G., and Chen, Z. J. (2020). Temporal regulation of the metabolome and proteome in photosynthetic and photorespiratory pathways contributes to maize heterosis. *Plant Cell.* 32, 3706–3722. doi: 10.1105/tpc.20.00320

Li, Z. K., Luo, L. J., Mei, H. W., Wang, D. L., Shu, Q. Y., Tabien, R., et al. (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics* 158, 1737–1753. doi: 10.1093/genetics/158.4.1737

Liu, P. C., Peacock, W. J., Wang, L., Furbank, R., Larkum, A., and Dennies, E. S. (2020). Leaf growth in early development is key to biomass heterosis in *Arabidopsis*. *J. Exp. Bot.* 71, 2439–2450. doi: 10.1093/jxb/eraa006

Liu, W. W., He, G. M., and Deng, X. W. (2021). Biological pathway expression complementation contributes to biomass heterosis in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2023278118. doi: 10.1073/pnas.2023278118

Liu, Y. J., Gao, S. Q., Tang, Y. M., Gong, J., Zhang, X., Wang, Y. B., et al. (2018). Transcriptome analysis of wheat seedling and spike tissues in the hybrid Jingmai 8 uncovered genes involved in heterosis. *Planta* 247, 1307–1321. doi: 10.1007/s00425-018-2848-3

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Luo, J. H., Wang, M., Jia, G. F., and He, Y. (2021). Transcriptome-wide analysis of epitranscriptome and translational efficiency associated with heterosis in maize. *J. Exp. Bot.* 72, 2933–2946. doi: 10.1093/jxb/erab074

Luo, L. J., Li, Z. K., Mei, H. W., Shu, Q. Y., Tabien, R., Zhong, D. B., et al. (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. *Genetics* 158, 1755–1771. doi: 10.1093/genetics/158.4.1755

Meena, R. K., Reddy, K. S., Gautam, R., Maddela, S., Reddy, A. R., and Gudipalli, P. (2021). Improved photosynthetic characteristics correlated with enhanced biomass in a heterotic F_1 hybrid of maize (*Zea mays* L.). *Photosynth Res.* 147, 253–267. doi: 10.1007/s11120-021-00822-6

Meyer, R. C., Witucka-Wall, H., Becher, M., Blacha, A., Boudichevskaia, A., Dörmann, P., et al. (2012). Heterosis manifestation during early *Arabidopsis* seedling development is characterized by intermediate gene expression and enhanced metabolic activity in the hybrids. *Plant J.* 71, 669–683. doi: 10.1111/j.1365-313X.2012.05021.x

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Peng, Z., Cheng, H., Sun, G. F., Pan, Z. E., Wang, X., Geng, X. L., et al. (2020). Expression patterns and functional divergence of homologous genes accompanied by polyploidization in cotton (*Gossypium hirsutum* L.). *Sci. China Life Sci.* 63, 1565–1579. doi: 10.1007/s11427-019-1618-7

Rapp, R. A., Udall, J. A., and Wendel, J. F. (2009). Genomic expression dominance in allopolyploids. *BMC Biol.* 7:18. doi: 10.1186/1741-7007-7-18

Shahzad, K., Zhang, X. X., Gou, L. P., Qi, T. X., Bao, L. S., Zhang, M., et al. (2020a). Comparative transcriptome analysis between inbred and hybrids reveals molecular insights into yield heterosis of upland cotton. *BMC Plant Biol.* 20:239. doi: 10.1186/s12870-020-02442-z

Shahzad, K., Zhang, X. X., Guo, L. P., Qi, T. X., Tang, H. N., Zhang, M., et al. (2020b). Comparative transcriptome analysis of inbred lines and contrasting hybrids reveals overdominance mediate early biomass vigor in hybrid cotton. *BMC Genomics* 21:140. doi: 10.1186/s12864-020-6561-9

Shen, Y., Sun, S., Hua, S., Shen, E., Ye, C. Y., Cai, D., et al. (2017). Analysis of transcriptional and epigenetic changes in hybrid vigor of allopolyploid *Brassica napus* uncovers key roles for small RNAs. *Plant J.* 91, 874–893. doi: 10.1111/tpj.13605

Sinha, P., Singh, V. K., Saxena, R. K., Kale, S. M., Li, Y. Q., Garg, V., et al. (2020). Genome-wide analysis of epigenetic and transcriptional changes associated with heterosis in pigeonpea. *Plant Biotechnol. J.* 18, 1697–1710. doi: 10.1111/pbi.13333

Song, Q. X., Zhang, T. Z., Stelly, D. M., and Chen, Z. J. (2017). Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* 18:99. doi: 10.1186/s13059-017-1229-8

Sundrish, S., Darleen, A. D., Bahman, E., Adam, J. L., and Giles, W. J. (2010). Dosage effect of the short arm of chromosome 1 of rye on root morphology and anatomy in bread wheat. *J. Exp. Bot.* 61, 2623–2633. doi: 10.1093/jxb/erq097

Syed, H. B. S., and Khan, M. Y. (2018). Quantitative analysis of the area of the Apical Ectodermal Ridge in Chick Appendages Using Image-J. *J. Coll Physic. Surg. Pak.* 28, 419–422. doi: 10.29271/jcpsp.2018.06.419

Teodoro, P. E., Azevedo, C. F., Farias, F. J. C., Alves, R. S., and Bhering, L. L. (2019). Adaptability of cotton (*Gossypium hirsutum*) genotypes analysed using a Bayesian AMMI model. *Crop. Pasture Sci.* 70, 615–621. doi: 10.1071/CP18318

Tian, S. H., Xu, X. L., Zhu, X. F., Wang, F., Song, X. L., and Zhang, T. Z. (2019). Overdominance is the major genetic basis of lint yield heterosis in interspecific

hybrids between *G. hirsutum* and *G. barbadense*. *Heredity* 123, 384–394. doi: 10.1038/s41437-019-0211-5

Williams, W. (1959). Heterosis and the genetics of complex characters. *Nature* 184, 527–530. doi: 10.1038/184527a0

Xiao, J., Li, J., Yuan, L., and Tanksley, S. D. (1995). Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. *Genetics* 140, 745–754. doi: 10.1093/genetics/140.2.745

Xiao, Y. J., Jiang, S. Q., Cheng, Q., Wang, X. Q., Yan, J., Zhang, R. Y., et al. (2021). The genetic mechanism of heterosis utilization in maize improvement. *Genome Biol.* 22:148. doi: 10.1186/s13059-021-02370-7

Yang, J. L., Mezmouk, S., Baumgarten, A., Buckler, E. S., Guill, K. E., McMullen, M. D., et al. (2017). Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. *PLoS Genet*. 13:e1007019. doi: 10.1371/journal.pgen.1007019

Yoo, M. J., Szadkowski, E., and Wendel, J. F. (2012). Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110, 171–180. doi: 10.1038/hdy.2012.94

Yu, J., Jung, S., Cheng, C. H., Ficklin, S. P., Lee, T., Zheng, P., et al. (2014). CottonGen: a genomics, genetics and breeding database for cotton research. *Nucleic Acids Res.* 42, D1229–D1236. doi: 10.1093/nar/gkt1064

Yu, S. B., Li, J. X., Xu, C. G., Tan, Y. F., Gao, Y. J., Li, X. H., et al. (1997). Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. U.S.A.* 94, 9226–9231. doi: 10.1073/pnas.94.17.9226

Zhang, J. F., Percy, R. G., and Mccarty, J. C. (2014). Introgression genetics and breeding between Upland and Pima cotton: a review. *Euphytica* 198, 1–12. doi: 10.1007/s10681-014-1094-4

Zhao, Y. L., Yu, S. X., Xing, C. Z., Fan, S. L., Song, M. Z., and Ye, W. W. (2009). Differential gene expression between hybrids and their parents during the four crucial stages of cotton growth and development. *Agric. Sci. China* 8, 144–153. doi: 10.1016/S1671-2927(09)60021-3

Zhu, A. Y., Greaves, I. K., Liu, P. C., Wu, L. M., Dennis, E. S., and Peacock, W. J. (2016). Early changes of gene activity in developing seedlings of *Arabidopsis* hybrids relative to parents may contribute to hybrid vigour. *Plant J.* 88, 597–607. doi: 10.1111/tpj.13285

Zhu, A. Y., Wang, A. H., Zhang, Y., Dennis, E. S., Peacock, W. J., et al. (2020). Early establishment of photosynthesis and auxin biosynthesis plays a key role in early biomass heterosis in *Brassica napus*(Canola) hybrids. *Plant Cell Physiol.* 61, 1134–1143. doi: 10.1093/pcp/pcaa038

Zhu, D., Zhou, G., Xu, C. G., and Zhang, Q. F. (2016). Genetic components of heterosis for seedling traits in an elite rice hybrid analyzed using an immortalized F_2 population. *J. Genet. Genomics* 43, 87–97. doi: 10.1016/j.jgg.2016.01.002

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

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

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 05 August 2022
DOI 10.3389/fpls.2022.940787

OPEN ACCESS

EDITED BY
Andrew H. Paterson,
University of Georgia,
United States

REVIEWED BY
Ming Zheng,
Oil Crops Research Institute (CAAS), China
Xueyong Yang,
Institute of Vegetables and Flowers (CAAS), China

*CORRESPONDENCE
Renxiang Liu
rxliu@gzu.edu.cn

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

RECEIVED 10 May 2022
ACCEPTED 01 July 2022
PUBLISHED 05 August 2022

CITATION
Mo Z, Luo W, Pi K, Duan L, Wang P, Ke Y, Zeng S, Jia R, Liang T, Huang Y and Liu R (2022) Comparative transcriptome analysis between inbred lines and hybrids provides molecular insights into K^+ content heterosis of tobacco (*Nicotiana tabacum* L.). *Front. Plant Sci.* 13:940787. doi: 10.3389/fpls.2022.940787

Comparative transcriptome analysis between inbred lines and hybrids provides molecular insights into K^+ content heterosis of tobacco (*Nicotiana tabacum* L.)

Zejun Mo[1,2], Wen Luo[1,3], Kai Pi[1,3], Lili Duan[1,2], Pingsong Wang[1,2], Yuzhou Ke[1,3], Shuaibo Zeng[1,3], Rongli Jia[1,3], Ting Liang[1,2], Ying Huang[1,3] and Renxiang Liu[1,3]*

[1]College of Agriculture, Guizhou University, Guiyang, China, [2]Key Laboratory of Tobacco Quality in Guizhou Province, Guiyang, China, [3]College of Tobacco, Guizhou University, Guiyang, China

Potassium (K^+) is essential for crop growth. Increasing the K^+ content can often directly promote the improvement of crop yield and quality. Heterosis plays an important role in genetic improvement and leads to genetic gains. We found that the K^+ content of tobacco showed significant heterosis, which is highly significant for cultivating tobacco varieties with high K^+ content. However, the mechanism by which K^+ content heterosis occurs in tobacco leaves is not clear. In this study, a comprehensive comparative transcriptome sequencing analysis of root samples from the hybrid G70×GDH11 and its parental inbred lines G70 and GDH11 was performed to elucidate the importance of the root uptake capacity of K^+ in the formation of heterosis. The results showed that 29.53% and 60.49% of the differentially expressed genes (DEGs) exhibited dominant and over-dominant expression patterns, respectively. These non-additive upregulated DEGs were significantly enriched in GO terms, such as metal ion transport and reaction, ion balance and homeostasis, ion channel activity, root meristem growth, and regulation of root hairs. The KEGG annotation results indicated that these genes were mainly involved in the pathways such as energy metabolism, carbohydrate formation, amino acid metabolism, and signal transduction. Further analysis showed that probable potassium transporter 17 (*NtKT17*) and potassium transporter 5-like (*NtKT5*), associated with potassium ion absorption, glutamate receptor 2.2-like and glutamate receptor 2.8-like, associated with ion channel activity, LOC107782957, protein detoxification 42-like, and probable glutamate carboxypeptidase 2, associated with root configuration, showed a significantly higher expression in the hybrids. These results indicated that the over-dominant expression pattern of DEGs played a key role in the heterosis of K^+ content in tobacco leaves, and the overexpression of the genes related to K^+ uptake, transport, and root development in hybrids helped to improve the K^+ content of plants, thus showing the phenomenon of heterosis.

KEYWORDS

potassium, heterosis, transcriptomics, *Nicotiana tabacum*, quality

Introduction

The new "green revolution" includes crop genetic improvement, healthy plant production, and efficient utilization of water and fertilizer (Beddington, 2010). Among them, the genetic improvement of crops ensures sustainable agricultural development through the cultivation of new varieties and the production of genetic gains (Wan, 2018). For the genetic improvement of crops, heterosis plays an important role in breeding desirable varieties and has led to remarkable achievements (Wu et al., 2021). Since it can greatly improve the production and quality of animals and plants, it significantly benefits society and the economy (Zhang et al., 2020), and also plays an important role in solving issues regarding global food security (Li et al., 2016).

Heterosis is a common biological phenomenon in which the hybrid offspring have better phenotypic traits than their parents (Hochholdinger and Baldauf, 2018). In China, it was found that the hybridization between horses and donkeys produced stronger mules (Yu et al., 2021). However, the study of plant heterosis began in Europe. In the mid-18th century, Kolreute performed tobacco hybridization between different species to cultivate early maturing tobacco hybrids with excellent qualities (Murad et al., 2002). Dominance (Davenport, 1908), overdominance (East, 1936), and epistasis (Williams, 1959) can explain heterosis partially, and their effects have been studied to some extent. Crop heterosis is not only different among different crops and traits, but the same trait of the same crop may be different due to different hybrid combinations, hybridization methods, and the cultivation environment (Schnable and Springer, 2013). Therefore, the mechanism of heterosis formation lacks a general explanation; specifically, heteropolyploid heterosis is relatively under-studied (Tian et al., 2018).

Tobacco (*Nicotiana tabacum* L.) is a special economic crop that is widely cultivated and has the characteristics of a short growth cycle, clear genetic background, and easy genetic transformation. It is also a model plant to study and elucidate various biological mechanisms (Xu et al., 2020; Zhang et al., 2022). Potassium ion (K^+) is a macroelement required for plant growth and affects the yield, quality, and stress resistance of plants by participating in activities such as protein synthesis, cell osmotic regulation, photosynthesis, stomatal movement, and enzyme activation (Pettigrew, 2008; Ragel et al., 2019). Tobacco is a potassium-loving crop, and potassium can not only significantly improve the color, combustion power, and firepower of tobacco leaves but also increase their flexibility and softness (Wang et al., 2022). Therefore, high potassium content in tobacco leaves increases the yield and quality of the leaves.

Tobacco directly absorbs K^+ from the soil through roots and subsequently accumulates K^+ in the leaves. Potassium ion is transported to leaves through the xylem due to the expression of potassium channel genes and potassium transporter genes (Sano et al., 2009). The potassium absorption capacity of tobacco roots is closely related to the physiological characteristics (root active absorption area, root activity, ATPase activity, etc.) and root morphology (Li and Li, 2001; Li and Ma, 2003), and there are also considerable differences in the capacity to absorb K^+ among different genotypes (Chen et al., 2010). Various genes related to K^+ absorption and transport have been isolated and cloned. These genes play various roles in different stages of life (Chai et al., 2019) and include influx potassium channel genes: *AKT1* (Hirsch et al., 1998), *KAT1* (Sottocornola et al., 2008), *KC1* (Reintanz et al., 2002), etc.; outflow potassium channel genes: *SKOR* (Gaymard et al., 1998); and potassium transporter genes: *KUP1* (Santa-María et al., 2018), *HAK* (Greiner et al., 2011), *TPK1* (Gobert et al., 2007), etc. The mining of these key genes that regulate K^+ absorption and transport is important for the directional improvement of potassium content in plants.

The potassium content of tobacco leaves increases in the topping period. After topping, the potassium content decreases rapidly, resulting in the low potassium content of mature tobacco leaves (Dai et al., 2009). Studies on improving tobacco quality focus on the strategies to effectively enhance the potassium content and select tobacco varieties with a high potassium content. In the early stages of research, we found that the heterosis of potassium content in tobacco leaves was prominent, and through heterosis utilization breeding, we obtained the hybrids Guiyan 202 and Guiyan 4 with high potassium content in tobacco leaves (Liu et al., 2006b, 2016). However, the mechanism of development of potassium content heterosis in tobacco leaves is unclear. Therefore, to determine the molecular components that might act on potassium heterosis, we performed transcriptome analysis of root samples of hybrids and their parents of the model plant common tobacco (*N. tabacum* L.) to elucidate the underlying mechanism of root potassium uptake capacity in the formation of potassium heterosis.

Materials and methods

Plant materials and planting

The tobacco hybrid F_1 (G70 × GDH11) and two parental inbred lines (G70 and GDH11) selected for this study were provided by the Guizhou Key Laboratory of Tobacco Quality Research. We ensured that the collection of plant material and experimental research and field studies on plants complied with relevant institutional, national, and international guidelines and legislation. The seeds were sown in a special substrate for flue-cured tobacco and raised by floating the seedlings in a greenhouse. The seedlings were grown there till they had five true leaves, and then they were transplanted to the field. The field experiment was conducted in the tobacco scientific research and experimental base of Guizhou University in 2021, with three replicates and a row plant spacing of 110 cm × 55 cm. All plants were topped after 50% of the center flowers of the plants were open.

Samples were collected every 10 days between 60 and 90 days after transplantation. Three tobacco plants with the same growth performance were randomly selected from each plot, and their young fibrous roots and root tips were sampled and mixed together. The samples were first washed with clean water and then with PBS. Next, they were placed in separate sterilized centrifuge tubes, frozen with liquid nitrogen, and stored in a low-temperature refrigerator at −80°C at the earliest. The leaf samples were killed at 105°C for 30 min, then turned to 75°C to dry, ground into powder, bagged, and sealed for storage.

Determination of K^+ content and analysis of heterosis

To determine the content of K^+ in leaves, 0.5 g of the sample was soaked in 0.5 mol/L of dilute hydrochloric acid and filtered. Then, the K^+ content was determined using a flame spectrophotometer (Munns et al., 2010). The K^+ content was calculated according to the following formula:

$$K^+\% = \frac{C \times V}{G \times 10^6} \times 100$$

Here, C indicates the K^+ concentration (ppm) obtained from the standard curve; V indicates the volume of the liquid to be measured; G indicates the dry smoke sample weight (g); 10^6 is the weight conversion coefficient.

The values of over high-parent heterosis (OPH), mid-parent heterosis (MPH), and below low-parent heterosis (BPH) were calculated according to the following formulae: $\mathrm{OPH}(\%) = \left(\frac{F1 - HP}{HP}\right) \times 100$, $\mathrm{MPH}(\%) = \left(\frac{F1 - MP}{MP}\right) \times 100$, $\mathrm{BPH}(\%) = \left(\frac{F1 - LP}{LP}\right) \times 100$, where F_1 represents the first hybrid generation, HP represents the high-value parent, MP represents the average parent value $\left(\frac{parent1 + parent2}{2}\right)$, and LP represents the low-value parent.

RNA isolation and library preparation

The biological samples of 70 days after transplanting were used as materials for RNA sequencing analysis. Total RNA from the root of nine samples was extracted using the TRIzol reagent following the manufacturer's instructions (OE Biotech Co., Ltd., Item No.: T105096). The purity and quantity of RNA were evaluated using the NanoDrop 2000 spectrophotometer (Thermo Scientific, United States). The integrity of the RNA was assessed using the Agilent 2,100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States). Then, the libraries were constructed using TruSeq Stranded mRNA LT Sample Prep Kit (Illumina, San Diego, CA, United States) following the manufacturer's instructions.

RNA sequencing and analysis of differentially expressed genes

The libraries were sequenced on an Illumina HiSeq X Ten platform, and 150 bp paired-end reads were generated. An average of 8.67 million reads was obtained per sample. Raw data (raw reads) in the fastq format were initially processed using Trimmomatic (Bolger et al., 2014), and the low-quality reads were removed to obtain clean reads. Then, about 8.60 million clean reads for each sample were retained for subsequent analyses.

The clean reads were mapped to the tobacco genome[1] using HISAT2 (Kim et al., 2015). The Fragments Per Kilobase of exon model per million mapped fragments (FPKM; Roberts et al., 2011) of each gene was calculated using Cufflinks (Trapnell et al., 2011), and the read counts of each gene were obtained using the HTSeq-count tool (Anders et al., 2015). Differential expression analysis was performed using the DESeq2 (Anders and Huber, 2012); the value of $p < 0.05$ and foldchange >2 or foldchange <0.5 were set as the threshold for significant differential expression. Hierarchical cluster analysis of differentially expressed genes (DEGs) was performed to determine the expression pattern of genes in different groups and samples. The GO enrichment and KEGG pathway enrichment analyses of the DEGs were performed using the R software, based on the hypergeometric distribution.

Real-time fluorescence quantitative PCR

To confirm the level of gene expression, based on the initial results of RNA sequencing, six genes were randomly selected for real-time fluorescence quantitative PCR (RT-qPCR) experiments. The experiments were performed using the SYBR Premix Ex Taq kit (Takara) and the Applied Biosystems 7500 Real-Time PCR system (Life Technologies Corporation, Beverly, MA, United States). The genes and corresponding primers used for the qPCR test are listed in Supplementary File 4. To calculate the relative expression level of each gene, the $2^{-\Delta\Delta Ct}$ method was used (Livak and Schmittgen, 2001).

Statistical analyses

Duncan's new multiple range tests were conducted to analyze the variation in the K^+ content ($p < 0.05$) using the SPSS (version 16.0) software. Use Origin 2018 (95_64) and Adobe Illustrator CS6 for figure drawing.

1 https://solgenomics.net/organism/Nicotiana_tabacum/genome

10.3389/fpls.2022.940787

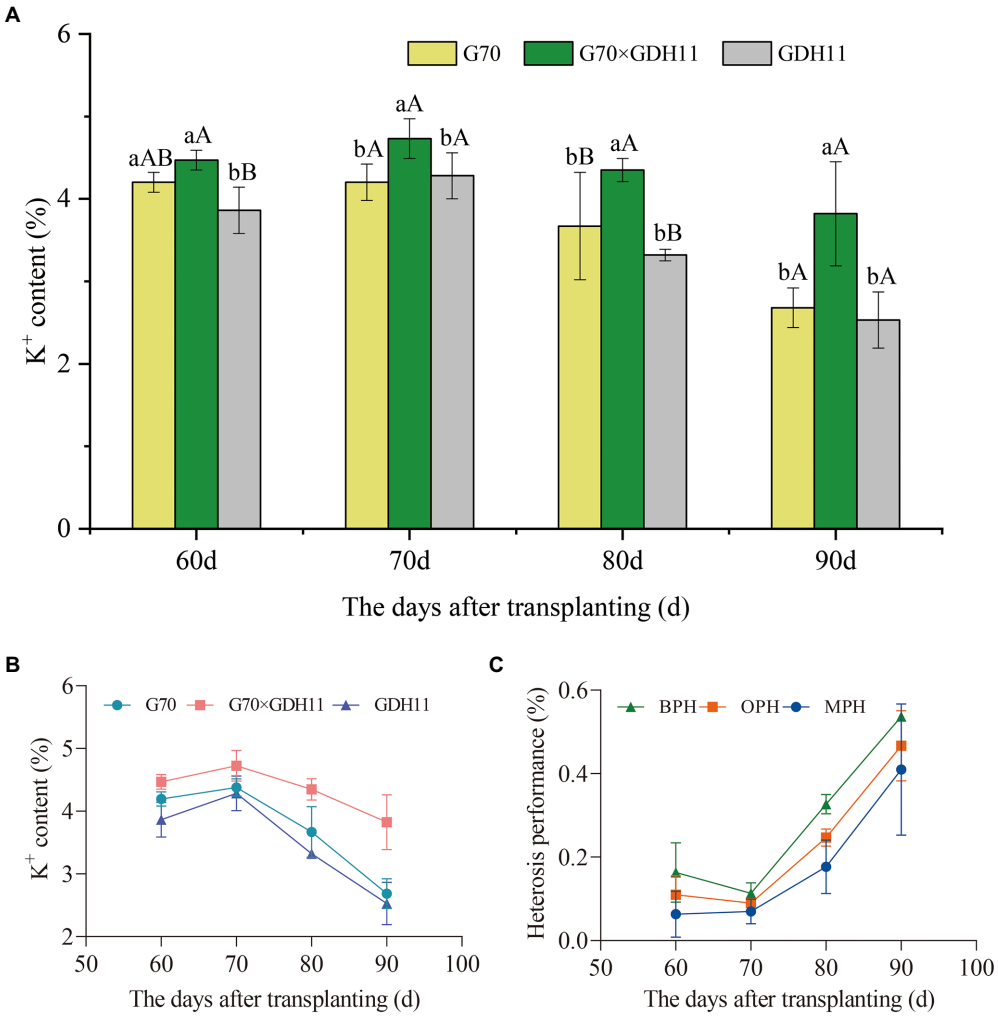


FIGURE 1
Potassium content of tobacco hybrids and their parents. **(A)** The K^+ content of the hybrid G70×GDH11 and its parents G70 and GDH11. **(B)** The changes in the K^+ content in hybrid F_1 and its parents between 60 and 90 days after transplantation. **(C)** The heterosis performance values of the hybrid G70×GDH11. The average value of three biological replicates per material was used for mapping. Error bars represent significant differences among the three replicates. Significant differences in the K^+ content at $p<0.05$ and $p<0.01$ were determined using Duncan's new multiple range tests. The lowercase alphabets represent a significant difference ($p<0.05$), while the uppercase alphabets represent a highly significant difference ($p<0.01$).

Results

The performance of K^+ content and its heterosis in tobacco leaves

To determine the difference in the K^+ content between the hybrids and their parents in different growth periods, we took samples every 10 days between 60 and 90 days after transplantation and evaluated the K^+ content in parents and their hybrids (Supplementary File 1). In the four growth stages studied, the K^+ content of the hybrid G70 × GDH11 was significantly higher than that of the female parent G70 and the male parent GDH11 (Figure 1A). Between 60 and 90 days after transplantation, the K^+ content of the three tested plant samples first increased and then decreased, and the shift (increase to decrease) in the K^+ content in the plants occurred 70 days after transplantation (Figure 1B). Moreover, 70 days after transplantation, the K^+ content in the two-parent samples decreased significantly more than that in the hybrid. Subsequently, we investigated the field agronomic traits of the hybrid and their parents (Supplementary Table S1), and found that the hybrid G70 × GDH11 performed significantly better over the two parents, indicating that the hybrids with strong K^+ content advantage had more robust stalks, longer and wider leaves, and larger biomass weight. These results showed that the hybrid had a stronger K^+ absorption capacity and a higher ability to retain K^+ in the leaves.

The heterosis performance of G70 × GDH11 in different periods showed that K^+ heterosis increased rapidly 70 days after transplantation (Figure 1C). Therefore, we took the plant material

TABLE 1 Statistical table of sequencing data of three materials.

Sample	Raw reads	Raw bases	Clean reads	Clean bases	Error rate (%)	Q20 (%)	Q30 (%)	GC content (%)
G70_1	83,981,574	12,681,217,674	83,075,818	12,324,430,722	0.0244	98.29	94.76	43.47
G70_2	89,563,688	13,524,116,888	88,790,214	13,068,587,523	0.0241	98.4	95.03	43.09
G70_3	94,593,544	14,283,625,144	93,595,804	13,764,269,216	0.0242	98.35	94.94	43.08
GDH11_1	74,585,534	11,262,415,634	74,033,572	10,963,664,867	0.0246	98.21	94.5	43.27
GDH11_2	83,101,872	12,548,382,672	82,620,736	12,250,716,518	0.0243	98.33	94.78	42.9
GDH11_3	90,796,556	13,710,279,956	90,038,852	13,299,062,270	0.0243	98.31	94.81	43.16
G70×GDH11_1	96,273,928	14,537,363,128	95,551,060	14,099,671,149	0.0243	98.31	94.77	43.15
G70×GDH11_2	92,506,634	13,968,501,734	91,988,384	13,556,441,950	0.0243	98.34	94.86	43.57
G70×GDH11_3	75,007,222	11,326,090,522	74,555,956	11,064,020,575	0.0245	98.26	94.57	42.6

Raw reads: the total number of items of the original sequencing data (reads, representing the sequencing reading segment, one read is one); Raw bases: the total amount of raw sequencing data (i.e., the number of raw reads multiplied by the read length); Clean reads: total number of items of sequencing data after quality control; Clean bases: the total amount of sequencing data after quality control (i.e., the number of clean reads multiplied by the length of reads); Error rate (%): average error rate of sequencing base corresponding to quality control data; Q20 (%) and Q30 (%): evaluate the quality of sequencing data after quality control. Q20 and Q30, respectively, refer to the percentage of bases with sequencing quality of more than 99 and 99.9% in the total bases; GC content (%): the percentage of the sum of G and C bases corresponding to quality control data in the total base.

TABLE 2 Comparison results of root sequences on the reference genome.

Sample Names	Total reads	Total mapped	Multiple mapped	Uniquely mapped
G70_1	83,075,818	54,759,364 (65.91%)	1,166,278 (1.4%)	53,593,086 (64.51%)
G70_2	88,790,214	58,537,912 (65.93%)	1,492,868 (1.68%)	57,045,044 (64.25%)
G70_3	93,595,804	61,695,124 (65.92%)	1,529,132 (1.63%)	60,165,992 (64.28%)
GDH11_1	74,033,572	49,019,700 (66.21%)	1,021,789 (1.38%)	47,997,911 (64.83%)
GDH11_2	82,620,736	54,311,755 (65.74%)	1,455,276 (1.76%)	52,856,479 (63.97%)
GDH11_3	90,038,852	59,960,465 (66.59%)	1,735,791 (1.93%)	58,224,674 (64.67%)
G70×GDH11_1	95,551,060	62,606,990 (65.52%)	1,345,318 (1.41%)	61,261,672 (64.11%)
G70×GDH11_2	91,988,384	60,960,400 (66.27%)	1,395,315 (1.52%)	59,565,085 (64.75%)
G70×GDH11_3	74,555,956	48,469,163 (65.01%)	1,364,515 (1.83%)	47,104,648 (63.18%)

Total reads: Statistics on the number of sequenced sequences after filtering (i.e., clean reads); Total mapped: Number of clean reads that can be located on the genome; Multiple mapped: The number of clean reads with multiple alignment positions on the reference sequence; Unique mapped: The number of clean reads that have unique alignment positions on the reference sequence.

70 days after transplantation as the experimental sample to investigate the key role of the absorptive capacity of tobacco roots in the formation of K^+ content heterosis. Ninety days after transplantation, the OPH, MPH, and BPH of G70×GDH11 were more than 40%, indicating that the K^+ content of tobacco was higher in the hybrid offspring and heterosis breeding might produce hybrid plants with significantly higher K^+ content than that in either parent plant.

Differences in transcriptome expression between the roots of parents and hybrids

Using the hybrid G70×GDH11 and its parents G70 and GDH11, RNA sequencing was performed using IlluminaHiSeq2000 to obtain a comprehensive reference transcriptome of nine tobacco root samples. The sequencing results were subjected to quality control analysis and sequence alignment. Information on the quality control analysis of the sequencing data is presented in Table 1, and the results of the comparison between the original data and the reference genome after quality control are presented in Table 2. With an average error rate of 0.02% across all samples, Q20 > 98%, Q30 > 94%, and GC content >42%, the clean sequences accounted for more than 99% of the total sequences (Table 1). The total sequence numbers and the results of genome mapping of the three cDNA libraries showed that, on average, 65.9% of clean sequences were mapped to the tobacco K326 reference genome (Table 2). These results demonstrated that the sequencing data were reliable and could be used for further analyzing differentially expressed genes associated with target traits.

To evaluate the quality of the transcriptomics and gene differential expression level data, six genes were randomly selected for qRT-PCR analysis. The actin gene was used as an internal reference gene to analyze the level of expression of each gene. The genes and the corresponding primers used for qRT-PCR determination were listed in Supplementary Table S2. The changes in the expression of the selected genes based on the relative expression level followed a pattern similar to that found in the transcriptomics data (Figure 2), which indicated that the transcriptome profiling data were reliable. Therefore, we can carry out the next step of differential gene screening and expression analysis.

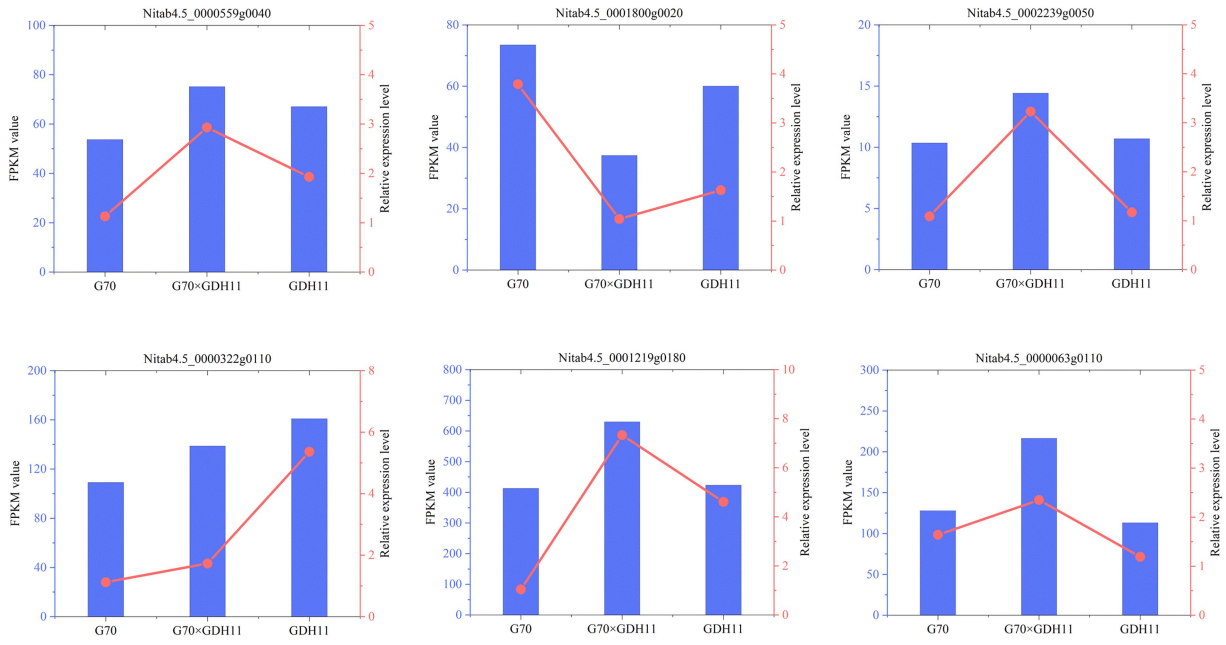


FIGURE 2
Validation of transcriptomics data through qRT-PCR analysis.

To analyze the causes of tobacco heterosis at the transcriptional level, we compared the transcriptional information of the F_1 hybrid with that of G70 and GDH11, as well as the transcriptional information between the parents. In the subsequent analysis, we considered 18,245 genes that were simultaneously identified in the three genotypes (Supplementary File 2, Figure 3A). Using the threshold of $p \leq 0.05$ and $|\log2 \text{ Fold Change}| \geq 2$, we identified 226 upregulated and 367 downregulated transcripts between the hybrid and G70. Similarly, 400 upregulated and 287 downregulated transcripts were observed between the hybrid and GDH11, and 215 upregulated and 424 downregulated transcripts were observed between G70 and GDH11 (Figure 3C). The results of a Venn diagram analysis showed that only 24 DEGs co-existed in the three groups (Figure 3B), which indicated that there were significant differences in the transcriptional ubiquitous expression of the three genotypes. Therefore, we speculated that these DEGs might be the reason that tobacco hybrids show heterosis in some traits.

Identification and analysis of expression patterns of DEGs in the hybrid

To further analyze the effects of these DEGs in the formation of K^+ content heterosis in tobacco leaves, we divided the differential genes into 12 expression patterns (P1–P12, Figure 4A) and determined the effect of additive and non-additive expression of the genes on heterosis of K^+ content. Genes in the P1 and P2 patterns were additively expressed. The genes in the P3–P6 pattern showed dominant expression, in which P3 and P4 showed partial paternal dominant expression, and P5 and P6 showed partial maternal dominant expression. P7–P9 showed low parental dominant expression, and P10–P12 showed over-high parental dominant expression. We divided these 12 expression patterns into five categories (Figure 4B), and different colors were used to indicate an expression pattern type.

In the hybrid G70 × GDH11, only 9.97% (P1 and P2) of the gene expression levels were between the parents, showing an additive expression pattern (Figure 4C). The remaining 90.03% of the genes showed a non-additive expression pattern, which indicated that the formation of K^+ content heterosis was more affected by the non-additive effect of the genes. Among the non-additively expressed genes, 29.53% (P3-P6) of the genes showed a dominant expression pattern, and 60.49% of the genes (P7–P12) showed an over-dominant expression pattern. The results indicated that the over-dominant expression pattern of DEGs played a major role in the formation of K^+ content heterosis.

Functional enrichment analysis of non-additively expressed genes

The degree of heterosis was found to depend on the non-additive effect of genes. Moreover, the non-additive expression of the genes affected the performance of hybrid offspring. To understand the biological functions of these non-additivity expressed genes, we performed the GO functional enrichment analysis on the gene sets with dominant (120 DEGs) and over-dominant upregulated (318 DEGs) expression. The GO enrichment analysis results of downregulated genes in non-additivity expression are shown in Supplementary Figures S1, S2. The results showed that

10.3389/fpls.2022.940787

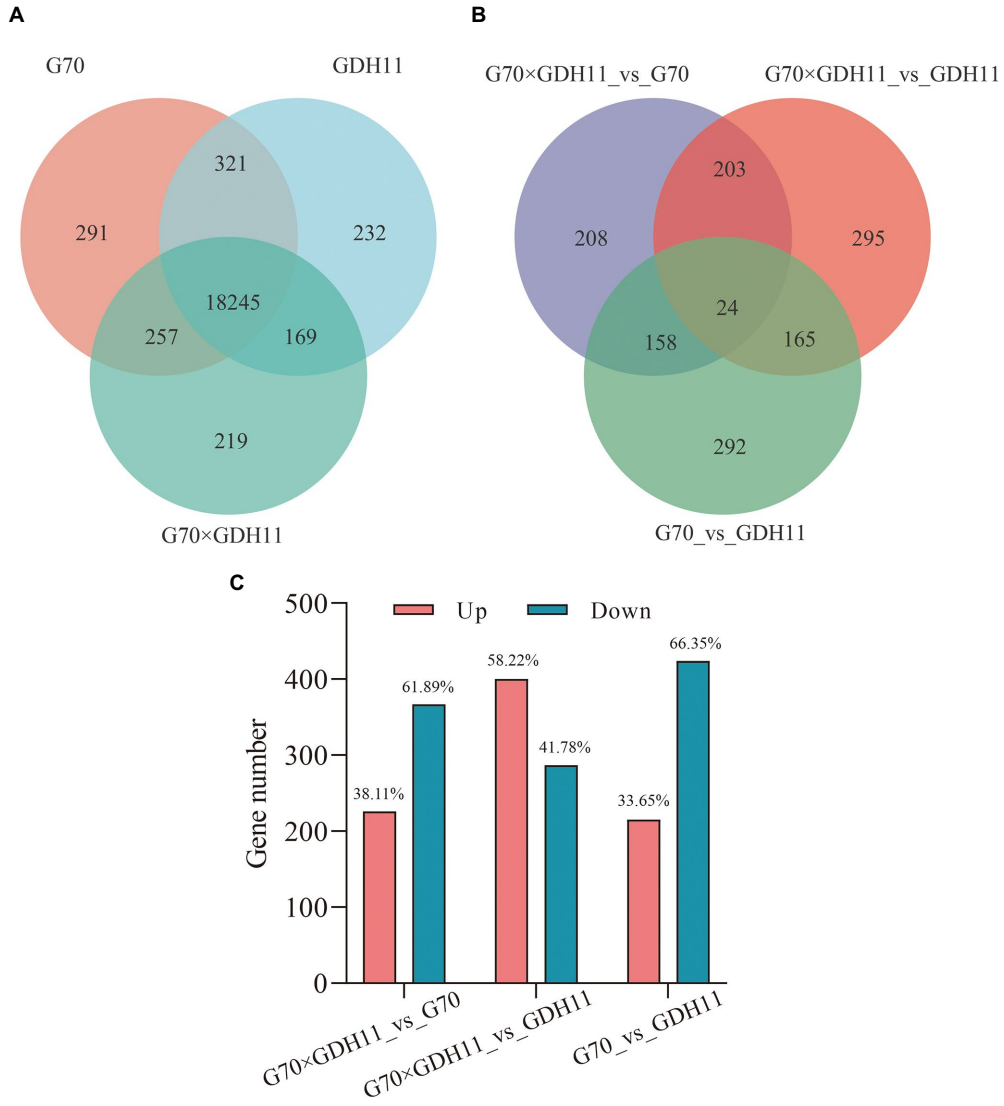
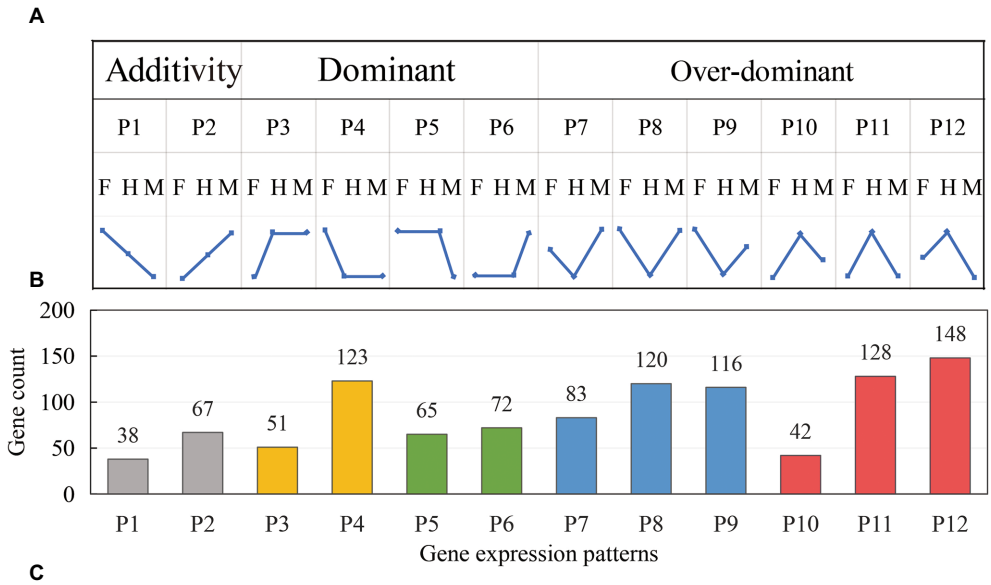


FIGURE 3
The transcriptional profiling statistics of different materials and comparison groups. **(A)** Venn diagram of the identified genes for the three tobacco varieties. **(B)** Venn diagram of the DEGs in the comparison groups G70_vs_GDH11, G70×GDH11_vs_G70, and G70×GDH11_vs_GDH11. **(C)** Statistics of up and down regulated differentiallyexpressed genes between hybrid and parental inbredlines.

the upregulated genes were mainly enriched in metal ion transport, calcium ion binding, potassium/sodium transporter activity, regulation of root morphogenesis, and ion channel activity (Figure 5A). The upregulated genes with over-dominant expression could be enriched in more GO functional terms, mainly including root defense response, metal ion transport and response, ion balance and homeostasis, ion channel activity, root meristem growth, and the regulation of root hair (Figure 5B). The results of the functional analysis indicated that these non-additively upregulated differential genes in the hybrid G70×GDH11 were good sources of potassium uptake in the hybrid. Their overexpression facilitates root growth and regulation of the meristem. These genes are also involved in ion transport and the regulation of ion channel activity, which enables K^+ in the soil to be more efficiently absorbed by the roots and transported to the tobacco leaves, thus causing the potassium content in the leaves of the hybrids to show heterosis.

To understand the biological pathways associated with the non-additive upregulated DEGs, we performed the KEGG pathway enrichment analysis for these genes. The results showed that the DEGs of the dominant and over-dominant gene sets were enriched in 36 and 62 pathways, respectively (Supplementary Files 3, 4). The first-level classification results of these pathways showed that most of them belonged to the categories of metabolism, genetic information processing, cellular processes, and environmental information processing (Figures 6A,B). The non-additive upregulated DEGs were mainly involved in energy metabolism, starch and sucrose metabolism,

10.3389/fpls.2022.940787



C

Additivity	Partial paternal dominant	Partial maternal dominant	Under low-parental dominant	Over high-parental dominant
105	174	137	319	318
9.97%	16.52%	13.01%	30.29%	30.20%

FIGURE 4
Identification and classification of differential gene expression patterns in the hybrids. **(A)** The division of 12 expression patterns; F: female parent, H: hybrid, and M: negative parent. **(B)** The distribution of DEGs in the hybrids among 12 expression patterns. **(C)** The number and proportion of DEGs in the five expression patterns.

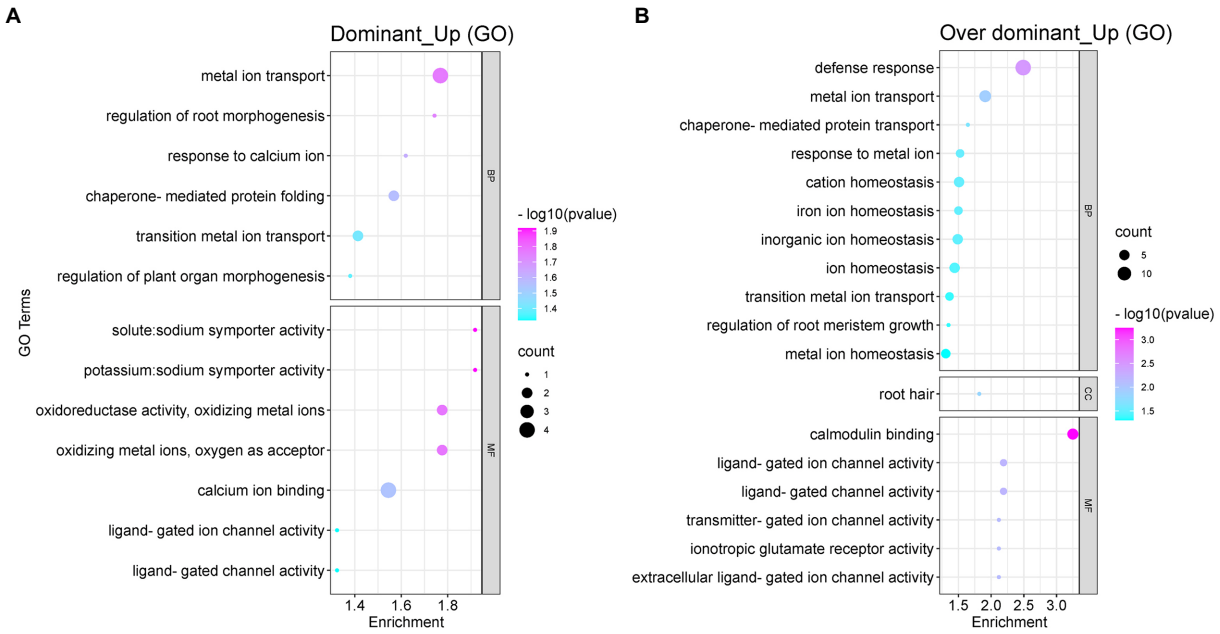


FIGURE 5
The GO functional enrichment analysis of the non-additive upregulated DEGs. **(A)** The GO functional enrichment analysis of the upregulated dominant DEGs. **(B)** The GO functional enrichment analysis of the upregulated over-dominant DEGs.

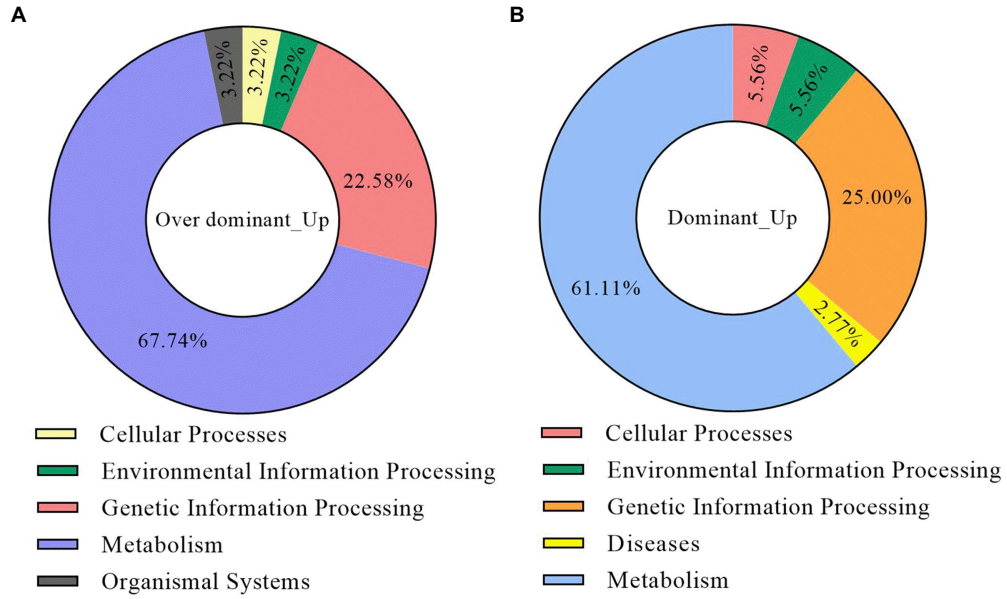


FIGURE 6
The KEGG pathway enrichment analysis of the non-additive upregulated DEGs. **(A)** The classification of the KEGG enrichment results for dominant upregulated DEGs. **(B)** The classification of the KEGG enrichment results for the over-dominant upregulated DEGs.

photosynthesis, carbohydrate formation, amino acid metabolism, plant hormone signal transduction, and signaling pathway, indicating that the significant expression of these biological processes in hybrids provides energy for the development of the roots and the absorption and transport of K^+, which is helpful to the accumulation of K^+ and the production of K^+ content superiority in hybrids.

Mining of key genes regulating potassium absorption in hybrids

To further analyze the basis of the formation of tobacco K^+ heterosis, we performed homology alignment and annotation analysis of the genes associated with the important GO entries (Table 3). The results showed that the probable potassium transporter 17 (*NtKT17*) and potassium transporter 5-like (*NtKT5*), associated with potassium transport, showed over-dominant upregulation and dominant expression in the hybrid G70 × GDH11, respectively. They play an important role in metal ion transport in plants. The glutamate receptor 2.2-like and glutamate receptor 2.8-like associated with ligand-gated channel activity showed an over-dominant upregulation expression in the hybrids. Some calcium-dependent protein kinase 28-like and calmodulin-binding protein 60 D-like related to the potassium ion absorption mechanism also showed a over-dominant upregulation. Furthermore, tabacum protein detoxification 42-like, involved in root hair development, and LOC107782957, involved in root morphology regulation, also showed over-dominant and dominant upregulation expression in the hybrids, respectively. We selected six genes related to K+ absorption and transport for the Real-Time Fluorescence Quantitative PCR detection (Figure 7). The genes and the corresponding primers are listed in Supplementary Table S3. The results showed that the expression level of these genes in hybrid G70 × GDH11 was significantly higher than that in two parental inbred lines, which indicated that these genes indeed have significant over-dominant and dominant expression patterns in hybrids. These results suggested that the over-dominant and dominant upregulation of key genes associated with K^+ uptake and transport leads to the generation of K^+ heterosis.

Discussion

The growth and development of plants are inseparable from the action of various essential elements. Among them, K^+ is a major element that affects various aspects of plant growth (Sano et al., 2009). Especially for food crops and cash crops, an increase in the K^+ content can often directly improve the yield and quality (Yang and Wang, 2017). Thus, the absorption and transport of K^+ have gained attention in recent years (Pettigrew, 2008). The quality of the tobacco plant is affected by K^+. However, studies on K^+ channels and transporters in tobacco are limited, especially on the genes regulating K^+ heterosis. Our study showed that the heterosis expression of K^+ content in tobacco was prominent and affected by the non-additive effect of the genes, of which the over-dominant effect played a major role. The results showed that breeding new tobacco varieties with high K^+ content through heterosis is feasible.

TABLE 3 Statistics of genes related to potassium content heterosis.

Pattern	Gene ID	Description	F_1	MP	FC	H_FPKM
Over-dominant	LOC107784880	Laccase-12-like	17.60	8.10	2.17	117.24%
	LOC107779045	Sodium/hydrogen exchanger 3-like	0.21	0.10	2.03	103.23%
	LOC107787287	Monocopper oxidase-like protein SKU5	1.34	0.94	1.43	42.96%
	LOC107760542	Probable potassium transporter 17	12.60	3.74	3.37	236.99%
	LOC107759178	Solute carrier family 40 member 3	124.56	65.82	1.89	89.26%
	LOC107758984	Tabacum protein detoxification 42-like	4.83	2.82	1.71	71.39%
	LOC107795534	Triosephosphate isomerase	22.25	12.21	1.82	82.28%
	LOC107783543	Oxalate–CoA ligase-like	1.23	0.83	1.49	48.59%
	LOC107791078	Calcium uniporter protein 5	4.01	1.96	2.05	104.94%
	LOC107790202	Probable glutamate carboxypeptidase 2	1.10	0.54	2.02	102.45%
	LOC107762391	Glutamate receptor 2.2-like	7.86	5.48	1.44	43.58%
	LOC107796561	Glutamate receptor 2.8-like	1.51	0.84	1.79	79.37%
	LOC107810988	Calmodulin-binding protein 60 D-like	10.83	7.22	1.50	50.03%
	LOC107769099	MLO-like protein 6	1.33	0.73	1.82	82.19%
	LOC107800917	Calcium-transporting ATPase 2	6.17	3.25	1.90	89.74%
	LOC107771176	Calmodulin-binding protein 60 B-like	0.12	0.08	1.64	64.44%
	LOC107765993	Uncharacterized LOC107765993	56.00	19.93	2.81	180.95%
	LOC107825481	Calcium-dependent protein kinase 28-like	5.39	2.21	2.45	144.60%
Dominant	LOC107812584	Potassium transporter 5-like	6.55	2.96	2.21	121.05%
	LOC107777935	L-ascorbate oxidase homolog	99.27	50.36	1.97	97.15%
	LOC107792406	L-ascorbate oxidase homolog	1.15	0.46	2.49	149.10%
	LOC107782957	Uncharacterized LOC107782957	0.74	0.28	2.64	163.91%
	LOC107803414	Heat shock cognate 70 kDa protein 1-like	77.18	38.92	1.98	98.30%
	LOC107809995	Calcineurin subunit B-like	0.51	0.32	1.59	58.55%
	LOC107769668	Alpha-amylase-like	2.39	1.84	1.30	29.84%
	LOC107793486	Uncharacterized LOC107793486	21.86	13.28	1.65	64.65%
	LOC107783728	Peroxygenase-like	6.09	4.32	1.41	41.16%
	LOC107800970	Uncharacterized LOC107800970	51.60	18.09	2.85	185.21%

MP represents the average value of gene expression of two parents; FC represents the foldchange of hybrid (G70 × GDH11), and MP. H_ MP represents the heterosis value of gene expression.

10.3389/fpls.2022.940787

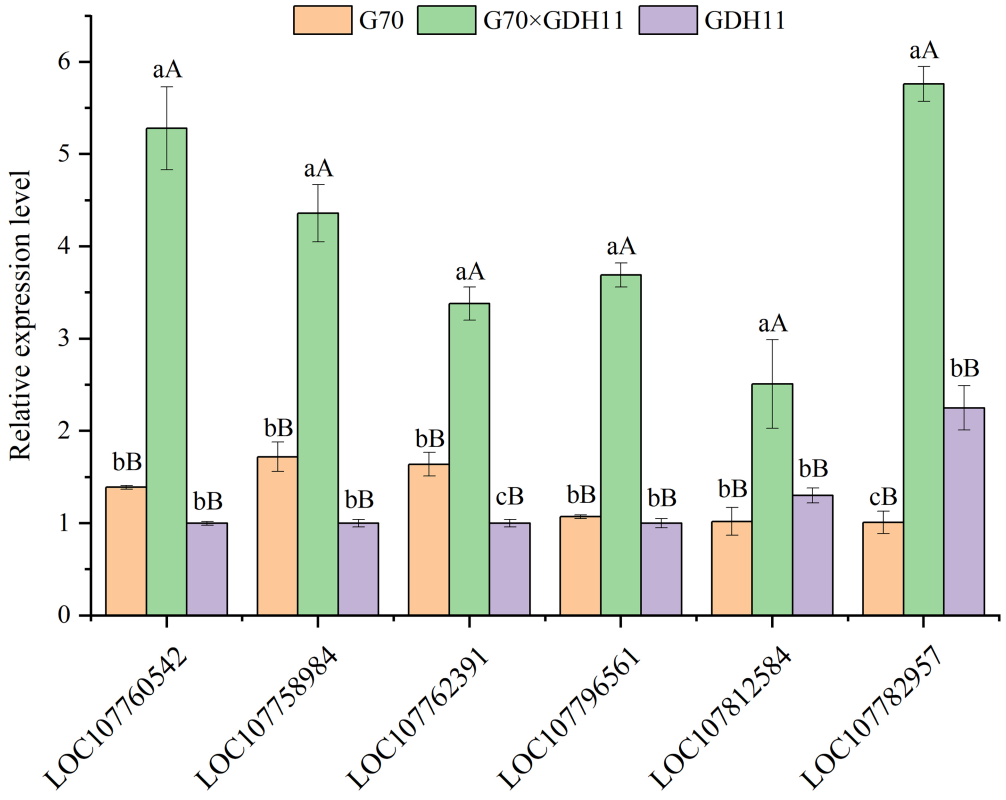


FIGURE 7
Expression level analysis of K⁺ related genes by qRT-PCR. The lowercase alphabets represent a significant difference (p<0.05), while the uppercase alphabets represent a highly significant difference (p<0.01).

With the development of modern molecular biology and advancement in quantitative genetics, research on heterosis has progressed. Transcriptomics is commonly used to analyze the mechanism of crop heterosis formation and satisfactory results have also been achieved in many studies (Palareti et al., 2016; Jiang et al., 2017; Birdseye et al., 2021; Liu et al., 2021, 2022). A correlation between heterosis and gene expression levels was shown in many studies (Wei et al., 2009; He et al., 2010). However, studies on the heterosis of important traits in tobacco are limited. Our transcriptomics results revealed that 35,520 genes were transcribed in at least one genotype (Supplementary File 5), which showed the active genes that regulate K^+ uptake and transport in tobacco.

Potassium ion is absorbed from the soil by root epidermal cells and cortical cells in plants, and ion channels and transport carriers play an important role in this process (Sánchez-Caldero et al., 2009). Liu et al. (2019) showed that the expression of *NtTPK1* and *NtORK1* might be related to the K^+ content and the formation of heterosis. Previous studies found that the influx K^+ channel genes *AKT1*, *NtKC1*, *KAT1*, and K^+ transporter gene *NtTPK1* showed dominant and over-dominant expression in the roots and leaves of the hybrids, indicating that these genes played a key role in the formation of K^+ content heterosis in hybrid tobacco leaves (Luo et al., 2022). Some studies have shown that the K+ transporter plays a major role in absorbing K^+ from the rhizosphere and can also be used as a low-affinity K^+ transporter at high K^+ concentrations (Kim et al., 1998; Ahn et al., 2004). In this study, probable potassium transporter 17 (*NtKT17*) and potassium transporter 5-like (*NtKT5*) showed upregulated over-dominant and dominant expression in the hybrids, respectively, with advantages in expression. The results further showed that these two genes were also involved in the absorption of K^+ in the roots, and the absorption rate in the hybrids was higher than that in the parents.

By opening and closing, ion channels regulate the speed of movement of specific substances in and out of cells, required for various cell functions (Hedrich, 2012). K^+ channel parenchyma is a transmembrane protein that is present on the cytoplasmic membrane or the inner membrane of plant cells (Hosy et al., 2003; Dreyer and Uozumi, 2011). Among these, ligand-gated channels, also known as chemical-gated ion channels, open after certain transmitters bind to the channel protein receptor molecule, thus allowing Na^+, Ca^{2+}, or K^+ to pass through (Liu et al., 2006a; Xie et al., 2020). We found that the glutamate receptor 2.2-like and glutamate receptor 2.8-like associated with ligand-gated channel activity showed an upregulated over-dominant expression in the hybrid. Therefore, we speculated that these two types of glutamate receptors might also play an important role in the formation of heterosis in tobacco K^+

content, thus contributing to K^+ transport by the roots. However, their mechanism of action needs to be further analyzed and confirmed.

Plant root architecture plays an important role in nutrient absorption, and there are considerable differences in root architecture between different species of plants or different varieties of the same plant species (Hao, 2015). Studies showed that the number of roots was positively correlated with the K^+ content of leaves in cotton (Hsu et al., 1978). Root hairs can absorb nutrients in the soil more effectively by increasing the contact area with the soil (Jungk, 2001; Wang et al., 2016). From the perspective of molecular biology, ion channels and transporters in root hairs greatly facilitate nutrient absorption (Libault et al., 2010). Our results showed that the uncharacterized LOC107782957 gene involved in the regulation of tobacco root morphogenesis upregulated dominant expression in the hybrids. Furthermore, protein detoxification 42-like, involved in root hair development, and probable glutamate carboxypeptidase 2, involved in the regulation of root meristem growth, showed an upregulated over-dominant expression pattern. The results indicated that the upregulated expression of these key genes in the hybrids promotes root growth and development, which is crucial for nutrient uptake by the roots. Many researchers argue that energy metabolism strongly affects the growth and development of plant roots (Yokota et al., 2003; Libault et al., 2010; Wang et al., 2016). Similar to the results of those studies, the KEGG enrichment results of this study also indicated that many upregulated non-additive DEGs are involved in energy metabolism and carbohydrate formation, thus enhancing the root absorption capacity of tobacco hybrids.

Crop heterosis formation is an extremely complex process. Our study provided a new perspective on the mechanism of K^+ content heterosis at the transcriptional level and might provide new information for cultivating tobacco varieties with high K^+ content. Potassium exists in plants in an ionic state, and its accumulation is jointly influenced by the growth and development of the root, the expression of K^+ uptake genes and transporters, K^+ channel activity, and the homeostasis of ions. Overexpression of the genes associated with these physiological functions in hybrids contributes to K^+ absorption and accumulation, thus exhibiting K^+ content heterosis.

Data availability statement

The data presented in the study are deposited in the NCBI Sequence Read Archive (SRA) repository, accession number PRJNA836042.

Author contributions

ZM, WL, and RL conceived and designed the research. WL, PW, and YK conducted field experiments. KP, SZ, and TL conducted the molecular biology experiments. ZM, LD, and RJ analyzed the data. ZM, KP, and YH wrote the manuscript. ZM and RL revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Science Foundation of China (no. 32060510), the Talent Cultivation Project for High-Level Renovation Plan of Guizhou Province (no. [2016]5663), the Guizhou Provincial Key Foundation (no. [2019]1405), and Tobacco Company of Guizhou Province (nos. 2020XM07 and 201904).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.940787/full#supplementary-material

SUPPLEMENTARY FILE 1
The original data of Figure 1.

SUPPLEMENTARY FILE 2
The number of genes and information on gene expression identified for the three genotype materials.

SUPPLEMENTARY FILE 3
The KEGG annotation results of the dominant upregulated DEGs.

SUPPLEMENTARY FILE 4
The KEGG annotation results of the over-dominant upregulated DEGs.

SUPPLEMENTARY FIGURE S1
The GO functional enrichment analysis of the downregulated dominant DEGs.

SUPPLEMENTARY FIGURE S2
The GO functional enrichment analysis of the downregulated over-dominant DEGs.

References

Ahn, S. J., Shin, R., and Schachtman, D. P. (2004). Expression of KT/KUP genes in Arabidopsis and the role of root hairs in K^+ uptake. *Plant Physiol.* 134, 1135–1145. doi: 10.1104/pp.103.034660

Anders, S., and Huber, W. (2012). Differential expression of RNA-Seq data at the gene level–the DESeq package. *Biotechnol. Bull.* 28, 1–28. doi: 10.13560/j.cnki.biotech.bull.1985.2018-0814

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

Beddington, J. (2010). Food security: contributions from science to a new and greener revolution. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 61–71. doi: 10.1098/rstb.2009.0201

Birdseye, D., De Boer, L. A., Bai, H., Zhou, P., Shen, Z., Schmelz, E. A., et al. (2021). Plant height heterosis is quantitatively associated with expression levels of plastid ribosomal proteins. *Proc. Natl. Acad. Sci. U. S. A.* 118, 1–7. doi: 10.1073/pnas.2109332118

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Chai, W. W., Wang, W. Y., Cui, Y. N., and Wang, S. M. (2019). Research progress of function on KUP/HAK/KT family in plants. *Plant Physiol. J.* 55, 1747–1761. doi: 10.13592/j.cnki.ppj.2019.0133

Chen, P., Shi, J., Zhao, L., Li, Y., Wen, D., and Deng, J. (2010). Study on the potassium of flue-cured tobacco among different variety and locations of Yunna. *J. Yunnan Univ. Sci. Ed.* 32, 42–46.

Dai, X. Y., Su, Y. R., Wei, W. X., Wu, J. S., and Fan, Y. K. (2009). Effects of top excision on the potassium accumulation and expression of potassium channel genes in tobacco. *J. Exp. Bot.* 60, 279–289. doi: 10.1093/jxb/ern285

Davenport, C. (1908). Degeneration, albinism and inbreeding. *Science* 28, 454–455. doi: 10.1126/science.28.718.454.c

Dreyer, I., and Uozumi, N. (2011). Potassium channels in plant cells. *FEBS J.* 278, 4293–4303. doi: 10.1111/j.1742-4658.2011.08371.x

East, E. M. (1936). Heterosis. *Genetics* 21, 375–397. doi: 10.1093/genetics/21.4.375

Gaymard, F., Pilot, G., Lacombe, B., Bouchez, D., Bruneau, D., Boucherez, J., et al. (1998). Identification and disruption of a plant shaker-like outward channel involved in K^+ release into the xylem sap. *Cell* 94, 647–655. doi: 10.1016/S0092-8674(00)81606-2

Gobert, A., Isayenkov, S., Voelker, C., Czempinski, K., and Maathuis, F. J. M. (2007). The two-pore channel TPK1 gene encodes the vacuolar K^+ conductance and plays a role in K^+ homeostasis. *Proc. Natl. Acad. Sci. U. S. A.* 104, 10726–10731. doi: 10.1073/pnas.0702595104

Greiner, T., Ramos, J., Alvarez, M. C., Gurnon, J. R., Kang, M., Van Etten, J. L., et al. (2011). Functional HAK/KUP/KT-like potassium transporter encoded by chlorella viruses. *Plant J.* 68, 977–986. doi: 10.1111/j.1365-313X.2011.04748.x

Hao, Y. (2015). Physiological mechanisms of high potassium efficiency and fertilization cffects of cotton.

He, G., Zhu, X., Elling, A. A., Chen, L., Wang, X., Guo, L., et al. (2010). Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* 22, 17–33. doi: 10.1105/tpc.109.072041

Hedrich, R. (2012). Ion channels in plants. *Physiol. Rev.* 92, 1777–1811. doi: 10.1152/physrev.00038.2011

Hirsch, R. E., Lewis, B. D., Spalding, E. P., and Sussman, M. R. (1998). A role for the AKT1 potassium channel in plant nutrition. *Science* 280, 918–921. doi: 10.1126/science.280.5365.918

Hochholdinger, F., and Baldauf, J. A. (2018). Heterosis in Plants. *Curr. Biol.* 28, R1089–R1092. doi: 10.1016/j.cub.2018.06.041

Hosy, E., Vavasseur, A., Mouline, K., Dreyer, I., Gaymard, F., Porée, F., et al. (2003). The Arabidopsis outward K^+ channel GORK is involved in regulation of stomatal movements and plant transpiration. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5549–5554. doi: 10.1073/pnas.0733970100

Hsu, H. H., Lancaster, J. D., and Jones, W. F. (1978). Potassium concentration in leaf-blades and petioles as affected by potassium fertilization and stage of maturity of cotton 1. *Commun. Soil Sci. Plant Anal.* 9, 265–277. doi: 10.1080/00103627809366807

Jiang, Y., Schmidt, R. H., Zhao, Y., and Reif, J. C. (2017). Quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. *Nat. Genet.* 49, 1741–1746. doi: 10.1038/ng.3974

Jungk, A. (2001). Root hairs and the acquisition of plant nutrients from soil. *J. Plant Nutr. Soil Sci.* 164, 121–129. doi: 10.1002/1522-2624(200104)164:2<121::AID-JPLN121>3.0.CO;2-6

Kim, E. J., Kwak, J. M., Uozumi, N., and Schroeder, J. I. (1998). AtKUP1: An arabidopsis gene encoding high-affinity potassium transport activity. *Plant Cell* 10, 51–62. doi: 10.1105/tpc.10.1.51

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements Daehwan HHS public access. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Li, D., Huang, Z., Song, S., Xin, Y., Mao, D., Lv, Q., et al. (2016). Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase. *Proc. Natl. Acad. Sci. U. S. A.* 113, E6026–E6035. doi: 10.1073/pnas.1610115113

Li, Z., and Li, J. (2001). Study on difference in morpholofical and physiological characters of wheat varieties to potassium. *J. Plant Nutr. Fertil.* 7, 36–43.

Li, T., and Ma, G. (2003). Physiological and morphological characteristics of roots in grain amaranth genotypes enrichment in potassium. Zuo wu xue bao. 30, 1145–1151. Available at: http://europepmc.org/abstract/cba/470214 (Accessed November 2004).

Libault, M., Brechenmacher, L., Cheng, J., Xu, D., and Stacey, G. (2010). Root hair systems biology. *Trends Plant Sci.* 15, 641–650. doi: 10.1016/j.tplants.2010.08.010

Liu, W., He, G., and Deng, X. W. (2021). Biological pathway expression complementation contributes to biomass heterosis in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 118, 1–9. doi: 10.1073/pnas.2023278118

Liu, R., Huang, N., Deng, G., Feng, Y., and Ding, W. (2006b). The study on technique of nitrogen rates, density and leaves of tobacco remained Guiyan 4. *J. Mt. Agric. Biol.* 25, 95–100.

Liu, G., Wang, Y., Sun, Y., and Wang, W. (2006a). Proteins for Transpor t of potassium in higher plants. *Biotechnol. Bull.* 21, 13–18. doi: 10.13560/j.cnki.biotech.bull.1985.2006.05.003

Liu, M., Zhang, W., Chi, X., Ding, F., Xu, X., Zeng, L., et al. (2019). Effect of potassium supply level on potassium content and formation of heterosis in tobacco leaves. *J. Henan Agric. Univ.* 53, 839–846. doi: 10.16445/j.cnki.1000-2340.2019.06.001

Liu, W., Zhang, Y., He, H., He, G., and Deng, X. W. (2022). From hybrid genomes to heterotic trait output: challenges and opportunities. *Curr. Opin. Plant Biol.* 66:102193. doi: 10.1016/j.pbi.2022.102193

Liu, R., Zhu, J., Yu, Q., Ding, F., Nie, Q., Huang, Y., et al. (2016). Breeding of a novel tobacco variety Guiyan 202 and its characteristics. *Chinese Tob. Sci.* 37, 8–13. doi: 10.13496/j.issn.1007-5119.2016.06.002

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

Luo, W., Jiang, C., Yu, Q., Xiong, J., Duan, L., Mo, Z., et al. (2022). Analysis of Heterosis of potassium content in tobacco leaf and breeding. *Mol. Plant* 2, 1–24. doi: 10.13271/j.mpb.020.001788

Munns, R., Wallace, P. A., Teakle, N. L., and Colmer, T. D. (2010). Measuring soluble ion concentrations (Na(+), K(+), cl(−)) in salt-treated plants. *Methods Mol. Biol.* 639, 371–382. doi: 10.1007/978-1-60761-702-0_23

Murad, L., Yoong Lim, K., Christopodulou, V., Matyasek, R., Lichtenstein, C. P., Kovarik, A., et al. (2002). The origin of tobacco's T genome is traced to a particular lineage within Nicotiana tomentosiformis (*Solanaceae*). *Am. J. Bot.* 89, 921–928. doi: 10.3732/ajb.89.6.921

Palareti, G., Legnani, C., Cosmi, B., Antonucci, E., Erba, N., Poli, D., et al. (2016). Comparison between different D-dimer cutoff values to assess the individual risk of recurrent venous thromboembolism: analysis of results obtained in the DULCIS study. *Int. J. Lab. Hematol.* 38, 42–49. doi: 10.1111/ijlh.12426

Pettigrew, W. T. (2008). Potassium influences on yield and quality production for maize, wheat, soybean and cotton. *Physiol. Plant.* 133, 670–681. doi: 10.1111/j.1399-3054.2008.01073.x

Ragel, P., Raddatz, N., Leidi, E. O., Quintero, F. J., and Pardo, J. M. (2019). Regulation of K^+ nutrition in plants. *Front. Plant Sci.* 10:281. doi: 10.3389/fpls.2019.00281

Reintanz, B., Szyroki, A., Ivashikina, N., Ache, P., Godde, M., Becker, D., et al. (2002). AtKC1, a silent Arabidopsis potassium channel α-subunit modulates root hair K^+ influx. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4079–4084. doi: 10.1073/pnas.052677799

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L., and Pachter, L. (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol.* 12, R22–R14. doi: 10.1186/gb-2011-12-3-r22

Sánchez-Caldero, L., Chacó-López, A., Alatorre-Cobos, F., Leyva-González, M. A., and Herrera-Estrella, L. (2009). Signaling and Communication in Plants. Transporters and Pumps in Plant Signaling.

Sano, T., Kutsuna, N., Becker, D., Hedrich, R., and Hasezawa, S. (2009). Outward-rectifying K^+ channel activities regulate cell elongation and cell division of tobacco BY-2 cells. *Plant J.* 57, 55–64. doi: 10.1111/j.1365-313X.2008.03672.x

Santa-María, G. E., Oliferuk, S., and Moriconi, J. I. (2018). KT-HAK-KUP transporters in major terrestrial photosynthetic organisms: a twenty years tale. *J. Plant Physiol.* 226, 77–90. doi: 10.1016/j.jplph.2018.04.008

Schnable, P. S., and Springer, N. M. (2013). Progress toward understanding heterosis in crop plants. *Annu. Rev. Plant Biol.* 64, 71–88. doi: 10.1146/annurev-arplant-042110-103827

Sottocornola, B., Gazzarrini, S., Olivari, C., Romani, G., Valbuzzi, P., Thiel, G., et al. (2008). 14-3-3 proteins regulate the potassium channel KAT1 by dual modes. *Plant Biol.* 10, 231–236. doi: 10.1111/j.1438-8677.2007.00028.x

Tian, M., Nie, Q., Li, Z., Zhang, J., Liu, Y., Long, Y., et al. (2018). Transcriptomic analysis reveals overdominance playing a critical role in nicotine heterosis in *Nicotiana tabacum* L. *BMC Plant Biol.* 18, 48–10. doi: 10.1186/s12870-018-1257-x

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2011). Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621.Transcript

Wan, J. (2018). Genetic crop improvement: a guarantee for sustainable agricultural production. *Engineering* 4, 431–432. doi: 10.1016/j.eng.2018.07.019

Wang, H., Lan, P., and Shen, R. F. (2016). Integration of transcriptomic and proteomic analysis towards understanding the systems biology of root hairs. *Proteomics* 16, 877–893. doi: 10.1002/pmic.201500265

Wang, X., Wang, B., Song, Z., Zhao, L., Ruan, W., Gao, Y., et al. (2022). A spatial–temporal understanding of gene regulatory networks and NtARF-mediated regulation of potassium accumulation in tobacco. *Planta* 255, 9–15. doi: 10.1007/s00425-021-03790-2

Wei, G., Tao, Y., Liu, G., Chen, C., Luo, R., Xia, H., et al. (2009). A transcriptomic analysis of superhybrid rice LYP9 and its parents. *Proc. Natl. Acad. Sci. U. S. A.* 106, 7695–7701. doi: 10.1073/pnas.0902340106

Williams, W. (1959). Heterosis and the genetics of complex characters. *Heredity (Edinb).* 15, 327–328. doi: 10.1038/hdy.1960.88

Wu, X., Liu, Y., Zhang, Y., and Gu, R. (2021). Advances in research on the mechanism of Heterosis in plants. *Front. Plant Sci.* 12:745726. doi: 10.3389/fpls.2021.745726

Xie, Z. Q., Tian, X. T., Zheng, Y. M., Zhan, L., Chen, X. Q., Xin, X. M., et al. (2020). Antiepileptic geissoschizine methyl ether is an inhibitor of multiple neuronal channels. *Acta Pharmacol. Sin.* 41, 629–637. doi: 10.1038/s41401-019-0327-4

Xu, J., Zhou, Y., Xu, Z., Chen, Z., and Duan, L. (2020). Combining physiological and metabolomic analysis to unravel the regulations of coronatine alleviating water stress in tobacco (*Nicotiana tabacum* L.). *Biomol. Ther.* 10, 1–16. doi: 10.3390/biom10010099

Yang, X., and Wang, Y. (2017). Current Progress of potassium nutrition on quality in Chinese herbs. *J. Zhejiang Chinese Med. Univ.* 41, 711–714. doi: 10.16466/j.issn1005-5509.2017.08.019

Yokota, E., Izeki, T., and Shimmen, T. (2003). Possible involvement of energy metabolism in the change of cytoplasm organization induced by a protein phosphatase inhibitor, calyculin A, in root hair cells of *Limnobium stoloniferum*. *Protoplasma* 221, 217–226. doi: 10.1007/s00709-002-0055-2

Yu, D., Gu, X., Zhang, S., Dong, S., Miao, H., Gebretsadik, K., et al. (2021). Molecular basis of heterosis and related breeding strategies reveal its importance in vegetable breeding. *Hortic. Res.* 8, 120–117. doi: 10.1038/s41438-021-00552-9

Zhang, D., Zhang, Y., Li, L., Song, D., Shi, S., Lei, C., et al. (2020). The formation mechanism of Heterosis and its application in animal genetic breeding. *Genomics Appl. Biol.* 39, 104–108. doi: 10.13417/j.gab.039.000104

Zhang, X., Zhang, L., Zhang, J., Jia, M., Cao, L., Yu, J., et al. (2022). Haploid induction in allotetraploid tobacco using DMPs mutation. *Planta* 255, 98–11. doi: 10.1007/s00425-022-03877-4

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 September 2022
DOI 10.3389/fpls.2022.927001

OPEN ACCESS

EDITED BY
Dayun Tao,
Yunnan Academy of Agricultural Sciences, China

REVIEWED BY
Igor Pacheco,
Universidad de Chile, Chile
Hui Yuan,
Shenyang Agricultural University, China

*CORRESPONDENCE
Yushan Qiao
qiaoyushan@njau.edu.cn
Mizhen Zhao
njzhaomz@163.com

†These authors have contributed equally to this work

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

RECEIVED 23 April 2022
ACCEPTED 17 August 2022
PUBLISHED 14 September 2022

CITATION
Du J, Ge C, Wang T, Wang J, Ni Z, Xiao S, Zhao F, Zhao M and Qiao Y (2022) Combined transcriptomic and proteomic analysis reveals multiple pathways involved in self-pollen tube development and the potential roles of FviYABBY1 in self-incompatibility in *Fragaria viridis*. *Front. Plant Sci.* 13:927001. doi: 10.3389/fpls.2022.927001

Combined transcriptomic and proteomic analysis reveals multiple pathways involved in self-pollen tube development and the potential roles of FviYABBY1 in self-incompatibility in *Fragaria viridis*

Jianke Du[1,2†], Chunfeng Ge[1,3†], Tao Wang[1†], Jing Wang[1], Zhiyou Ni[1], Shiwei Xiao[1], Fengli Zhao[1], Mizhen Zhao[4*] and Yushan Qiao[1,4*]

[1]Laboratory of Fruit Crop Biotechnology, College of Horticulture, Nanjing Agricultural University, Nanjing, China, [2]Institute of Horticulture Research, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [3]Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China, [4]Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, Institute of Pomology, Jiangsu Academy of Agricultural Sciences, Nanjing, China

Fragaria viridis exhibits *S*-RNase-based gametophytic self-incompatibility, in which *S*-RNase is the major factor inhibiting pollen tube growth. However, the pathways involved in and the immediate causes of the inhibition of pollen tube growth remain unknown. Here, interactive RNA sequencing and proteome analysis revealed changes in the transcriptomic and proteomic profiles of *F. viridis* styles harvested at 0 and 24 h after self-pollination. A total of 2,181 differentially expressed genes and 200 differentially abundant proteins were identified during the pollen development stage of self-pollination. Differentially expressed genes and differentially abundant proteins associated with self-incompatible pollination were further mined, and multiple pathways were found to be involved. Interestingly, the expression pattern of the transcription factor FviYABBY1, which is linked to polar growth, differed from those of other genes within the same family. Specifically, *FviYABBY1* expression was extremely high in pollen, and its expression trend in self-pollinated styles was consistent with that of *S-RNase*. Furthermore, FviYABBY1 interacted with *S*-RNase in a non-S haplotype way. Therefore, FviYABBY1 affects the expression of polar growth-related genes in self-pollen tubes and is positively regulated by *S*-RNase.

KEYWORDS

self-incompatibility, FviYABBY1, transcriptome, proteome, *Fragaria viridis*

Introduction

Self-incompatibility (SI) is a widespread mating system that avoids inbreeding and promotes outcrossing in angiosperms (Allen and Hiscock, 2008; Vieira et al., 2019; Kumar et al., 2021). In the family Rosaceae, the mechanisms of *S*-RNase-based gametophytic self-incompatibility (GSI) are diverse (Franklin-Tong and Franklin, 2003; McClure and Franklin-Tong, 2006; Sassa et al., 2010; Wu et al., 2013). In particular, studies on self- and non-self-recognition genera belonging, respectively, to tribes Amygdaleae (e.g., *Prunus*) and Maleae (e.g., *Malus* and *Pyrus*) have offered key insights into SI mating systems in the subfamily Amygdaloideae (Kubo et al., 2010; Iwano and Takayama, 2012; Fujii et al., 2016; Chen et al., 2020; Muñoz-Sanz et al., 2020; Harkness and Brandvain, 2021; Vekemans and Castric, 2021). In addition, an increasing number of modifiers have been explored, enabling comprehensive analyses of mechanisms underlying *S*-RNase-based SI (Wu et al., 2013; Claessen et al., 2019; Muñoz-Sanz et al., 2020). Within the genus *Fragaria* of the subfamily Rosoideae (family Rosaceae), some diploid species, such as *Fragaria viridis*, *F. nubicola*, *F. pentaphylla*, and *F. nipponica*, exhibit self-incompatibility as the major pollination barrier (Evans and Jones, 1967; Hancock, 1999; Sargent et al., 2004; Staudt, 2009; Bošković et al., 2010). The germplasm of low-ploidy wild strawberries contains abundant genetic resources controlling valuable traits, and it is a potential resource for improving cultivated strawberries (Hancock, 1999; Marta et al., 2004; Sargent et al., 2004; Staudt, 2009; Liston et al., 2014). However, as a pollination barrier between styles and pollen, SI has considerably hampered strawberry breeding (Evans and Jones, 1967; Li et al., 2000; Marta et al., 2004; Sargent et al., 2004; Bošković et al., 2010; Liston et al., 2014). As such, SI greatly limits the exploitation of valuable traits in diploid *Fragaria* species, such as enhanced disease resistance, stress tolerance, aromaticity, and solid content, for breeding (Evans, 1982; Hancock and Luby, 1993; Maas et al., 1997; Hancock et al., 2002; Ulrich et al., 2007). Therefore, elucidating molecular mechanisms underlying SI is expected to aid the development of strategies aimed at improving molecular breeding in strawberries. The mechanisms of SI in *Fragaria* have been studied for decades (Evans and Jones, 1967; Bošković et al., 2010; Aguiar et al., 2015). Specifically, GSI in *Fragaria* is known to be mediated by *S*-RNase (Du et al., 2021), similar to that in *Rosa* (Vieira et al., 2021); however, molecular mechanisms underlying the *S*-RNase-mediated inhibition of self-pollen tubes remain unclear.

High-throughput sequencing technologies, such as RNA sequencing (RNA-Seq) and proteomic analysis, have been used as efficient tools for studying SI mechanisms. In *Petunia inflate*, pollens of haplotypes S_2 and S_3 were analyzed using *de novo* transcriptome sequencing, and 17 pollen *SLF* determinants were identified (Williams et al., 2014; Wu et al., 2020). Moreover, comparative transcriptomics among unpollinated, self-pollinated, and cross-pollinated pistils of "Xiangshui" lemon revealed a series of genes related to pollen tube growth, programmed cell death (PCD), signal transduction, and transcription. In addition, a novel putative self-pollen rejection-associated *S*-RNase was detected, building the foundation for further research on SI mechanisms and the breeding of seedless varieties in lemons (Zhang et al., 2015). Additionally, proteomic analysis, as a tool for transcriptomics, has been widely applied in studies exploring SI mechanisms. Specifically, seven *S*-RNase-associated proteins have been identified in Japanese pears using two-dimensional gel electrophoresis (2-DE) (Ishimizu et al., 1996). Similarly, using 2-DE, Wang et al. (2013) discovered 22 proteins, some of which may be related to pollen maturation, fertility, germination, growth, and cell death. In a study exploring SI mechanisms in *Solanum pennellii*, proteomic analysis with isobaric tags for relative and absolute quantitation (iTRAQ) revealed that *S*-RNase, HT-A, cell wall-loosening proteins, and defense response-related proteins play key roles in an interspecific reproductive barrier between wild and domesticated species (Chalivendra et al., 2012). Overall, these cutting-edge techniques can not only directly identify potential pollen or style S-determinants but also reveal large-scale changes at the transcriptomic and proteomic levels in various SI-related biological processes during pollination.

Through ABCF transport proteins, *S*-RNase enters the pollen tube and functions non-specifically (Meng et al., 2014a,b). *S*-RNase is cytotoxic to self-pollen tubes, while its cytotoxicity is attenuated in non-self-pollen tubes (Claessen et al., 2019). Initially, the cytotoxicity of *S*-RNase was attributed to RNA degradation, and *S*-RNase was believed to enter the self-pollen tube and degrade RNA, explaining *S*-RNase-based GSI (Franklin-Tong and Franklin, 2003; Kubo et al., 2010; Tao and Iezzoni, 2010; Wu et al., 2013). *S*-RNase can break the gradient of reactive oxygen species (ROS) and calcium ions at the tip of the pollen tube, inducing a series of physiological and biochemical changes and PCD responses. Multiple signal transduction cascades are involved in *S*-RNase-based GSI (Wang et al., 2009, 2010; Wu et al., 2013, 2021; Li et al., 2018; Chen et al.,

Abbreviations: DEG, differentially expressed gene; DAP, differentially abundant protein; FC, fold change; GSI, gametophytic self-incompatibility; SLF, *S-locus* F-box protein; PCD, programmed cell death; SI, self-incompatibility; SC, self-incompatibility; iTRAQ, isobaric tags for relative and absolute quantitation; Nt, NCBI nucleotide sequences; Nr, NCBI non-redundant protein; COG, Cluster of Ortholog Genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; RNA-Seq, RNA sequencing; FDR, false discovery rate; 2-DE, two-dimensional gel electrophoresis; RT-qPCR, real-time quantitative PCR; FPKM, fragments per kilobase per million mapped reads; TCA, trichloroacetic acid; AGP, arabinogalactan proteins; CDPK, calcium-dependent protein kinase; ROS, reactive oxygen species.

2020). Based on the evidence, the cytotoxicity of *S*-RNase in self-pollen tubes can be attributed to a multitude of reactions and signals, and the *S*-RNase-mediated degradation of RNA in pollen tubes may be the result of SI response or RNA homeostasis. The pollen tube is a classic model of polar growth, and the polarity of the self-pollen tube breaks following SI (Feijó, 2010; Wu et al., 2013; Del Duca et al., 2014). The *YABBY* gene family belongs to the zinc finger superfamily and plays pivotal roles in the development of vegetative and reproductive tissues in angiosperms (Kumaran et al., 2002; Bartholmes et al., 2012; Finet et al., 2016; Filyushin et al., 2017). As transcription factors, an important function of *YABBY*s is related to the establishment of polarity (Kumaran et al., 2002; Stahle et al., 2009); however, their link to SI and their involvement in breaking the polar growth of incompatible pollen tubes remain unknown.

The green strawberry *F. viridis* is a self-incompatible diploid (2n = 2x = 14) species. The fruits have an apple-like aroma, and the plants exhibit potent cold hardness and high soil pH tolerance (Labokas and Bagdonaitë, 2005; Staudt, 2009; Gruner et al., 2017). *F. viridis* is the germplasm for strawberry breeding (Gruner et al., 2017), and this species is a suitable model for further research on SI in the genus *Fragaria* (Bošković et al., 2010). In our experiment, at 24 h after self-pollination, most of the *F. viridis* pollen tubes stopped growing at 2/3 of the style. Thus, the receptacle with gynoecium from *F. viridis* line #10-42 (SI) at 0 and 24 h after self-pollination was selected for subsequent proteome sequencing. Integrative analysis with previous RNA-Seq data provided valuable information on the development of self-pollen tubes in the styles of *F. viridis*. In addition, the interaction between FviYABBY1 and *S*-RNase in *F. viridis* implicated polar growth regulators in self-incompatibility.

Materials and methods

Observation of pollen tube growth

Styles of *F. viridis* were self-pollinated and harvested at 1, 6, 12, 18, 24, and 48 h after pollination (HAP). The styles were fixed in FAA solution (5% formalin, 5% acetic acid, and 65% ethanol) for 2 h at 65°C. Next, the styles were washed with distilled water three times, treated with 4 M NaOH for 2 h at 65°C, and washed with distilled water three times before staining with 0.1% (w/v) aniline blue for 12 h at 4°C in the dark. Self-pollinated pistils of *Fragaria vesca* after 24 h were used as controls. The stained styles of *F. viridis* were placed on a glass slide and mounted in a drop of glycerol for observation under a fluorescence microscope (BX53; Olympus, Tokyo, Japan) at a wavelength of 356 μm.

Transcriptomic data annotation

The self-pollinated receptacles with gynoecium from *F. viridis* lines #10–42 were harvested after 0 and 24 h and stored in liquid nitrogen for transcriptome sequencing and data analysis. Total RNA extraction, cDNA preparation, Illumina sequencing, and *de novo* assembly were performed as previously described (Du et al., 2021). The obtained non-redundant unigenes were annotated against the Nr, Nt, Swiss-Prot, COG, and KEGG databases using BLAST (Altschul et al., 1990), GO annotation was performed using Blast2GO, and InterPro annotation was performed using InterProScan5. ESTScan (Iseli et al., 1999) was used to predict the directions of unigenes that could not be annotated.

Differential gene expression analysis

Unigene expression was estimated as fragments per kilobase of transcript per million fragments mapped reads (FPKM) (Mortazavi et al., 2008). Differentially expressed genes (DEGs) were detected using DEseq2 (Love et al., 2014), and *p*-values related to DEGs were modulated for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) method (Benjamini and Hochberg, 1995). Unigenes deemed differently expressed were screened with an FDR-modulated *p*-value of $\leq$0.05 and $|\log_2 FC|$ of $\geq$1. DEGs were subjected to GO functional and KEGG pathway analyses.

Protein extraction, digestion, iTRAQ labeling, and strong cation exchange

iTRAQ analysis was performed at the Beijing Genomics Institute (BGI, Shenzhen, China). Plant material was the same as that for the RNA-Seq analysis described earlier. Two biological replicates were set per sample. Proteins were extracted using the trichloroacetic acid (TCA)/acetone method (Gu et al., 2013). Protein concentration was measured using the Bradford method. For each sample, 100 μg protein was digested with Trypsin Gold (Promega, Madison, WI, USA) at a protein-to-trypsin ratio of 20:1 for 4 h at 37°C. Following trypsin digestion, the peptides were vacuum centrifuged to dryness and dissolved in 0.5 M TEAB. iTRAQ labeling of peptide samples was performed using the iTRAQ® Reagents-8plex Applications Kit according to the manufacturer's protocol. After labeling, the samples were combined and lyophilized, and the fractions were then analyzed using LC-ESI-MS/MS.

Protein identification and quantification

MS/MS raw data were converted to MGF files using the ProteoWizard tool (msConvert, http://proteowizard.sourceforge.net), and proteins were searched using the Mascot search engine (Matrix, Science, London, UK, vision2.3.02) against the *de novo* transcriptomic data of *F. viridis* style (Du et al., 2021). Proteins were quantified using IQuant (BGI, Shenzhen, China), as previously described (Wen et al., 2014). Permutation tests were used for statistical and protein quantification analyses. Proteins were identified and quantified separately for each biological replicate (Tolin et al., 2013). FDR correction was adopted with a threshold of $<1\%$ to reduce the false identification of peptides. Proteins with a $|\log_2 FC|$ of ≥ 0.6 and a p-value of ≤ 0.05 were considered differentially expressed.

Analysis of YABBY family members in *F. viridis* and *F. vesca*

According to the hidden Markov model (HMM) of the YABBY domain, the *hmmsearch* program in HMMER (Finn et al., 2011) was used to search the whole-genome protein database of *F. viridis* and *F. vesca* for obtaining the *YABBY* gene family members. The selected proteins were input into the CDD database on NCBI (https://www.ncbi.nlm.nih.gov/cdd/?term) to verify the integrity of the YABBY domain, and proteins that did not contain the YABBY domain were filtered out. YABBY HMM (PF04690) can be downloaded directly from the Pfam database (http://pfam.xfam.org/). The protein database for *F. viridis* was derived and predicted based on the style transcript database after self-pollination (Du et al., 2021). The proteome database of *F. vesca* was downloaded from the Rosaceae Genome Database (https://www.rosaceae.org/), and *F. vesca* v4.0. a1 was selected.

Molecular weights and isoelectric points (pI) of *YABBY* gene family members were analyzed using ExPaSy (https://web.expasy.org/protparam/) (Artimo et al., 2012), and signal peptides were analyzed using SignalP (http://www.cbs.dtu.dk/services/SignalP/index.php) (Armenteros et al., 2019). The amino acid sequences of genes and proteins were compared using the alignment tool DNAMAN8.0 (Lynnon, Quebec, Canada). The online tool MEME (https://meme-suite.org/meme/tools/meme) was used for the motif analysis of YABBY family members, and TBtools was used for visualization. Using *Arabidopsis* and tomato as references (Bowman, 2000; Huang et al., 2013), multiple sequence alignment of YABBY protein sequences between *F. viridis* and *F. vesca* was performed with the MUSCLE program in MEGA 7.0 (Mello et al., 2017).

Real-time quantitative PCR (RT-qPCR) validation of DEGs and *YABBY* family genes

Receptacles with gynoecium from *F. viridis* were collected at 0, 6, 12, 24, 48, and 72 HAP. Total RNA extraction and cDNA synthesis were performed as previously described (Du et al., 2021). Based on reference unigene sequences, 21 primer pairs (Supplementary Table S1) were designed using an online real-time PCR tool (https://sg.idtdna.com/scitools/Applications/RealTimePCR/). EF-1α was used as the internal control (Du et al., 2021). In addition, for spatiotemporal expression analysis, pollen, leaves, peduncles, calyxes, and petals of *F. viridis* were collected for tissue-specific expression analysis of *YABBY* family genes. The SYBR Premix Ex Taq^{TM} kit (TaKaRa) was used for RT-qPCR. The reactions were conducted with three biological replicates using the ABI 7300 Real-Time System (PE Applied Biosystems, CA, USA). The $2^{-\Delta\Delta CT}$ algorithm was used to analyze quantitative variation.

Cloning of *FviYABBY1* in *F. viridis*

Primers were designed based on the open reading frame of the putative *FviYABBY1* unigene (Unigene16298_All) (Du et al., 2021) (see Supplementary Table S1). rTaq was used for PCR. The reaction conditions were as follows: 94°C for 4 min; 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min; and final extension at 72°C for 10 min. PCR products were evaluated on a 1% agarose gel. The target fragments were purified using the AxyPrep DNA gel extraction kit (Axygen, CA, USA), cloned into the pMD^{TM} 19-T vector (Takara Bio, Dalian, China), and transformed into *Escherichia coli* DH5α competent cells. Positive clones were sequenced as described previously (Li et al., 2015).

Subcellular localization assay

The *FviYABBY1* coding sequence (CDS) lacking stop codon was obtained from the plasmid containing *FviYABBY1* using primers with *Nco* I–*Spe* I restriction sites. *FviYABBY1* was ligated to the pCAMBIA1302-GFP expression vector using T4 ligase to construct the FviYAB1-GFP recombinant plasmid. Tobacco leaves were transiently transformed as described previously (Du et al., 2019). Following culture in the dark for ∼60 h, the fluorescent signal of GFP in tobacco leaves was detected using a laser confocal fluorescence microscope.

Yeast two-hybrid assay

The recombinant reporter vectors pGBKT7-Sa and pGBKT7-Sb were constructed, and yeast AH109 cells were

transformed with the pGADT7 plasmid. To test the toxicity and self-activation of S_a-RNase and S_b-RNase, as well as to obtain the optimal 3-AT concentration for the background expression of eliminating reporter genes in yeast cells, pGADT7-largeT+pGBKT7-p53 and pGADT7-largeT+pGBKT7-laminC combinations were used as the positive and negative control, respectively. SD/-Trp/-Leu and SD/-Trp/-Leu/-His/Ade media with 3-AT concentration gradients of 0, 5, 10, and 15 mM were used, respectively.

The full-length CDSs of *S_a-RNase* and *S_b-RNase* were stored in the previously described vector as templates (Du et al., 2021), and primers with *Nde* I–*Bam* HI restriction sites were used to obtain the CDSs of mature S_a-RNase and S_b-RNase peptides. Enzyme digestion was used to construct the recombinant vectors pGBKT7-Sa and pGBKT7-Sb. The same method was used to recombine the full-length CDS of *F. viridis YABBY1* into pGADT7 for obtaining pGADT7-FviYAB1 with *Nde* I–*Bam* HI restriction sites. The primers used are listed in Supplementary Table S1.

The recombinant plasmids pGBKT7-Sa (pGBKT7-Sb) and pGADT7-FviYAB1 were co-transformed into the yeast strain AH109. The yeast cells were spread on the SD/-Trp-Leu medium and cultured for 3–5 days at 28°C in an incubator. Single colonies were selected, transferred to SD/-Trp-Leu liquid medium, and shaken. After shaking, the yeast suspension was straked on SD/-Trp-Leu, SD/-Trp-Leu-His-Ade, and SD/-Trp-Leu-His-Ade/X-α-gal plates and cultivated at 28°C in an incubator for 3–5 days to observe colony growth. pGADT7-largeT+pGBKT7-p53 and pGADT7-largeT+pGBKT7-laminC were used as the positive and negative controls, respectively.

Validation of protein interaction using bimolecular fluorescence complementation (BiFC)

The target genes *FviYABBY1* and *S-RNase* (*S_a*-RNase and *S_b*-RNase) from the pMD19 (Simple) T-vector containing the corresponding gene were amplified using primers with restriction sites for the gene-specific primers used (Supplementary Table S1). Using restriction digestion, the CDS of *FviYABBY1* lacking the stop codon was cloned into the 35S:YCE vector with *Asc* I–*Bam* HI, and the full-length CDS of *S_a*-RNase (*S_b*-RNase) was cloned into the 35S:YNE vector with *Asc* I–*Kpn* I to obtain the recombinant vectors FviYABBY1-35S:YCE and Sa(Sb)-35S:YNE, respectively. Tobacco cells were transiently transfected with *Agrobacterium* following the same method as that for subcellular localization assay.

Results

Self-pollen tube growth in pistils

Aniline blue staining showed the growth of *F. viridis* pollen tubes in pistils at 1, 6, 12, 18, 24, and 48 HAP (Figures 1A–F). Pollen grains gradually attached to stigma papilla cells at 1 HAP, started germinating at 6 HAP (Figure 1B), and continued to elongate until reaching two-thirds of the pistil at 24 HAP (Figure 1E) refer to self compatible *F. vesca* (Figure 1G); no further distinct growth was observed at 48 HAP (Figure 1F). Thus, genes related to the inhibition of self-pollen tube growth may be expressed around 24 HAP based on the differential expression of pollen tube development. Accordingly, self-pollinated receptacles with gynoecia were harvested after 0 and 24 h and analyzed for differential expression.

Transcriptomic analysis of styles at 0 and 24 h after self-pollination

Six pools generated a total of 40.23 Gb of raw data (Supplementary Table S2). After removing low-quality, adaptor-polluted, and high-content unknown base (N) reads, 268.16 Mb of clean reads were generated (Q30 > 87.14%). A total of 71,756 unigenes were annotated using the Nr, Nt, Swiss-Prot, COG, KEGG, GO, and Interpro databases (Supplementary Figure S1A), with a mean length of 1,209 bp. The identified unigenes of *F. viridis* were highly homologous to those of *F. vesca* (85.5%), *Prunus mume* (1.59%), *Prunus persica* (1.31%), and *Malus domestica* (1.01%) (Supplementary Figure S1B, respectively). Global gene expression profiling between samples collected after 0 and 24 h revealed 2,181 DEGs, including 1,355 downregulated and 826 upregulated genes (Supplementary Table S3). These results indicate that many biological and molecular processes are altered during self-pollination. DEGs identified in the NR database were classified using GO functional analysis (Figures 2A–C), which is an important reference for pollen tube development in self-pollinated *Fragaria* species according to previous reports (Zhao et al., 2015; Zhang et al., 2016, 2017).

Furthermore, KEGG pathway analysis was performed to elucidate the biological pathways activated by pollen–style interaction. A total of 1,571 DEGs were annotated in 125 KEGG pathways (Supplementary Table S4). "Metabolic pathway" (ko01100), "biosynthesis of secondary metabolites" (ko01100), and "starch and sucrose metabolism" (ko00500) were the three major pathways. These results are consistent with the findings of transcriptome analysis of SI and SC pollination in tomatoes (Vieira et al., 2008). Interestingly, many genes were annotated to "plant–pathogen interaction" (ko04626), "phenylpropanoid biosynthesis" (ko00940), "plant hormone signal transduction" (ko04075), and "ubiquitin-mediated proteolysis" (ko04120),

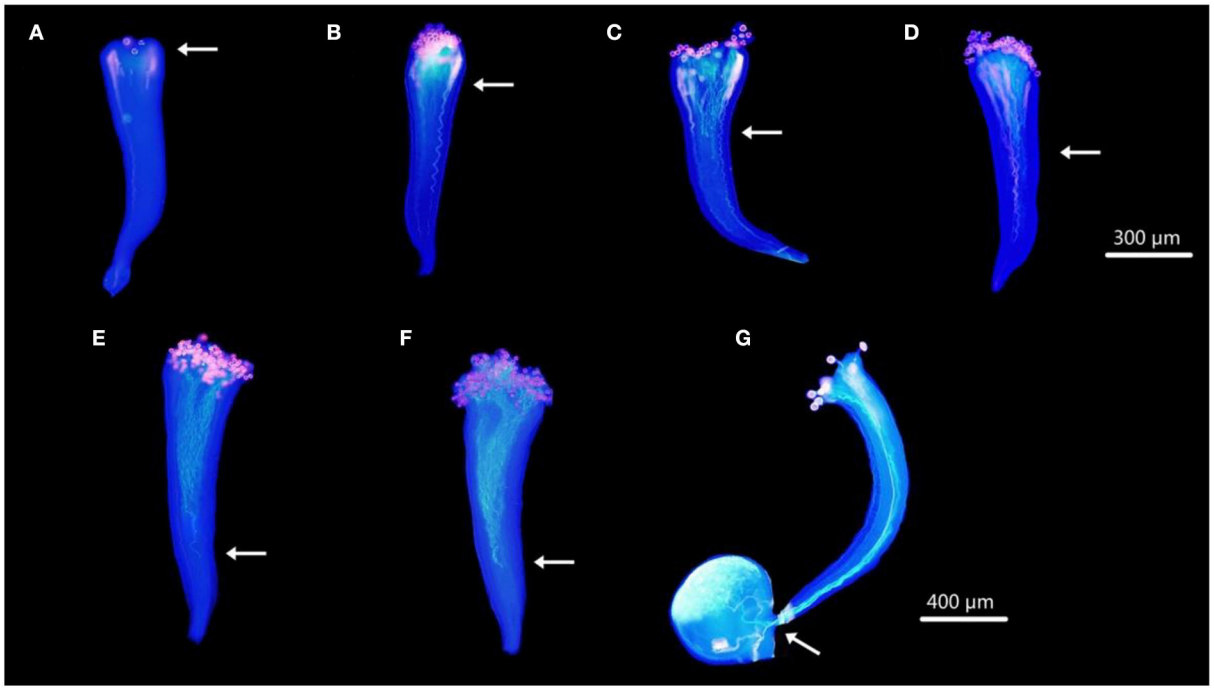


FIGURE 1
Dynamic development of pollen tubes in pistils under SP. **(A–F)** Styles of *F. viridis* at 1, 6, 12, 18, 24, and 48 HAP, respectively. **(G)** Pistil of *F. vesca* at 24 HAP. White arrows indicate the positions at which the majority of the pollen tubes were arrested in pistils. Bars **(A–F)** = 300 μm, **(G)** = 400 μm.

which are all notable pathways related to self-pollination (Elleman and Dickinson, 1999; Williams et al., 2015).

Validation of gene expression using RT-qPCR

The quality of transcriptomic data was confirmed using RT-qPCR. Fourteen genes related to *S-RNases*, cell death, microtubule binding, pathogen interaction, hormone signaling, CDPK signaling, enzyme inhibitor activity, cell wall, pollen development, RNA degradation, and ubiquitin-mediated proteolysis were detected (Figures 3A–N). The results showed good concordance ($R^2 = 0.9409$) between the RT-qPCR and RNA-Seq data (Figure 3O), confirming the reliability of our transcriptomic data. Consistent with previous reports, S_a*-RNase* and S_b*-RNase* showed an initial uptrend, followed by a peak at 12 HAP and then a downtrend. Interestingly, the expression trends of four genes encoding a PR protein, an auxin-induced protein, a Glu S.griseus protease inhibitor, and CAF1 were similar to that of *S-RNase*. The expression levels of these four genes (Pollen-specific LRX protein, CDPK34, PRK2, and SKP1) increased after self-pollination. Conversely, the expression levels of three genes (UBP1 associated protein, PE/PEI 17, GsSRK) first decreased and then increased.

Proteomic analysis and transcriptome–proteome integrative matching

Comparative proteomic analysis between the two samples collected at 0 and 24 HAP was performed as a parallel complement to transcriptomic analysis. We identified 7,105 proteins from 385,662 spectra (data not shown), including 200 DAPs (FC > 1.5, $p < 0.05$), of which 109 were upregulated and 91 were downregulated (Supplementary Table S5). According to the transcriptomic data, the results of GO functional analysis of DAPs and DEGs were highly consistent during self-pollination (Figure 2). In addition, 151 DAPs were annotated to 74 KEGG pathways (Supplementary Table S6), and the majority of these pathways were related to pollen tube development.

Integrative analysis of DEGs and DAPs offered comprehensive information on key genes involved in pollen–style interactions. In the present study, we identified 3,273 genes (Supplementary Table S7) at both transcript and protein levels and used these for correlation analysis. The distributions of the corresponding transcript-to-protein ratios (colored plots) were presented as scatter plots of $\log_2$-transformed data (Figure 4). The expression of most genes (black, blue, and green plots) did not significantly vary at the transcript ($|\log_2 FC| \geq 1$) and protein ($|\log_2 FC| \geq 0.6$) levels. Next, we focused on genes

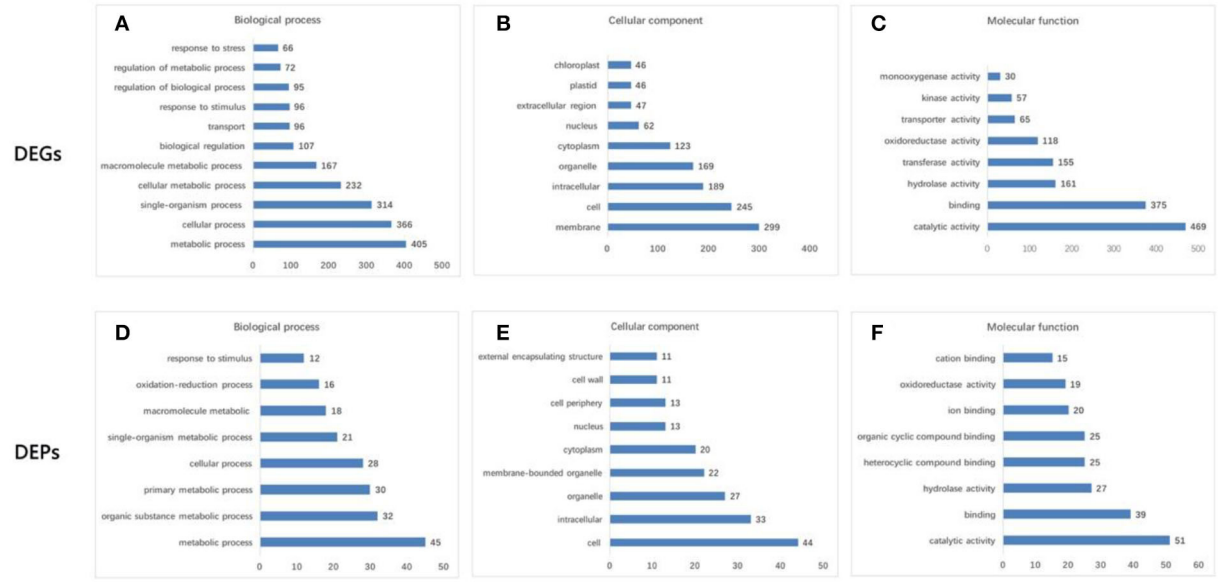


FIGURE 2
GO functional analysis of differentially expressed genes (DEGs) and differentially abundant proteins (DAPs). **(A,D)** Biological process. **(B,E)** Cellular component. **(C,F)** Molecular function. DEGs were mainly enriched in "metabolic process" (405 DEGs) and "cellular process" (366) for the biological process; "membrane" (299) and "cell" (245) for the cellular component; and "catalytic activity" (469) and "binding" (375) for the molecular function. The transcriptome data were significantly enriched in several GO categories, such as "cell wall organization and biogenesis," "defense response," "response to hormone," "gametophyte development," "pollen development," "pollen germination," and "pollination," which are pollen–style interaction-associated biological processes. Moreover, GO terms annotation of DAPs revealed that "cell" (44) and "intracellular" (33) were the major cellular components; "catalytic activity" (51) and "binding" (39) were the major molecular functions, and "metabolic process" (45) and "cellular process" (28) were the major biological processes.

(red plot) in four sections (A–D). Genes in sections B and C exhibited similar expression trends at the transcript and protein levels (both upregulated and downregulated), whereas genes in sections A and D showed negative expression trends. Twelve DAPs ($|\log_2 FC| \geq 1$ and $p \leq 0.05$) that were differentially expressed at the transcript level ($|\log_2 FC| \geq 1$ and $p \leq 0.05$) are listed in Supplementary Table S8. Eleven genes showed consistent trends at both levels, including eight upregulated and three downregulated genes. Only one gene (beta-glucosidase 3) showed a negative expression trend, being downregulated at the transcript level but upregulated at the protein level at 24 HAP compared with that at 0 HAP. Beta-glucosidase is involved in various biological processes, including the formation and degradation of cell walls and response to biotic and abiotic stresses. The difference in the transcript and protein levels of this gene may result from many factors rather than its involvement in the self-incompatibility reaction alone.

Data mining for DEGs and DAPs involved in SI

S-RNase-mediated GSI involves complex genetic mechanisms (Franklin-Tong and Franklin, 2003; Wu et al., 2013), and the integrity of SI depends on both the S locus (pistil S gene and pollen S gene) and the modifier genes (Sassa and Hirano, 2006; Wu et al., 2013; Claessen et al., 2019). These modifiers are closely related to PCD (Thomas and Franklin-Tong, 2004), plant-pathogen interactions (Elleman and Dickinson, 1999; Sanabria et al., 2008), Ca^{2+} signaling (Wu et al., 2013; Gu et al., 2015; Qu et al., 2016), ROS (Wang et al., 2010), hormone signaling (Shi et al., 2017), cell wall formation (Graaf et al., 2003), protein ubiquitination (McClure and Franklin-Tong, 2006), and pollen development (Zhang et al., 2016; Shi et al., 2017). Here, we clustered the SI-related DEGs and DAPs based on GO functional annotation, KEGG pathway classification, and Nr database annotation, as seen in Supplementary Table S9.

The pollen tube grows smoothly in a compatible style, which is regulated by apical polarity (Feijó, 2010; Del Duca et al., 2014). A gene encoding a transcription factor related to the polar growth of the pollen tube, the axial regulator *YABBY1* (Unigene16298_All), named *FviYABBY1*, was detected in the style at 0 HAP, and its expression was remarkably downregulated ($\log_2 FC < -1$) at 24 HAP. In the present study, *FviYABBY1* was annotated to a GO term but not to a protein or a KEGG pathway. The tip of self-incompatible pollen tubes cannot extend normally, which limits their polar growth (Franklin-Tong and

10.3389/fpls.2022.927001

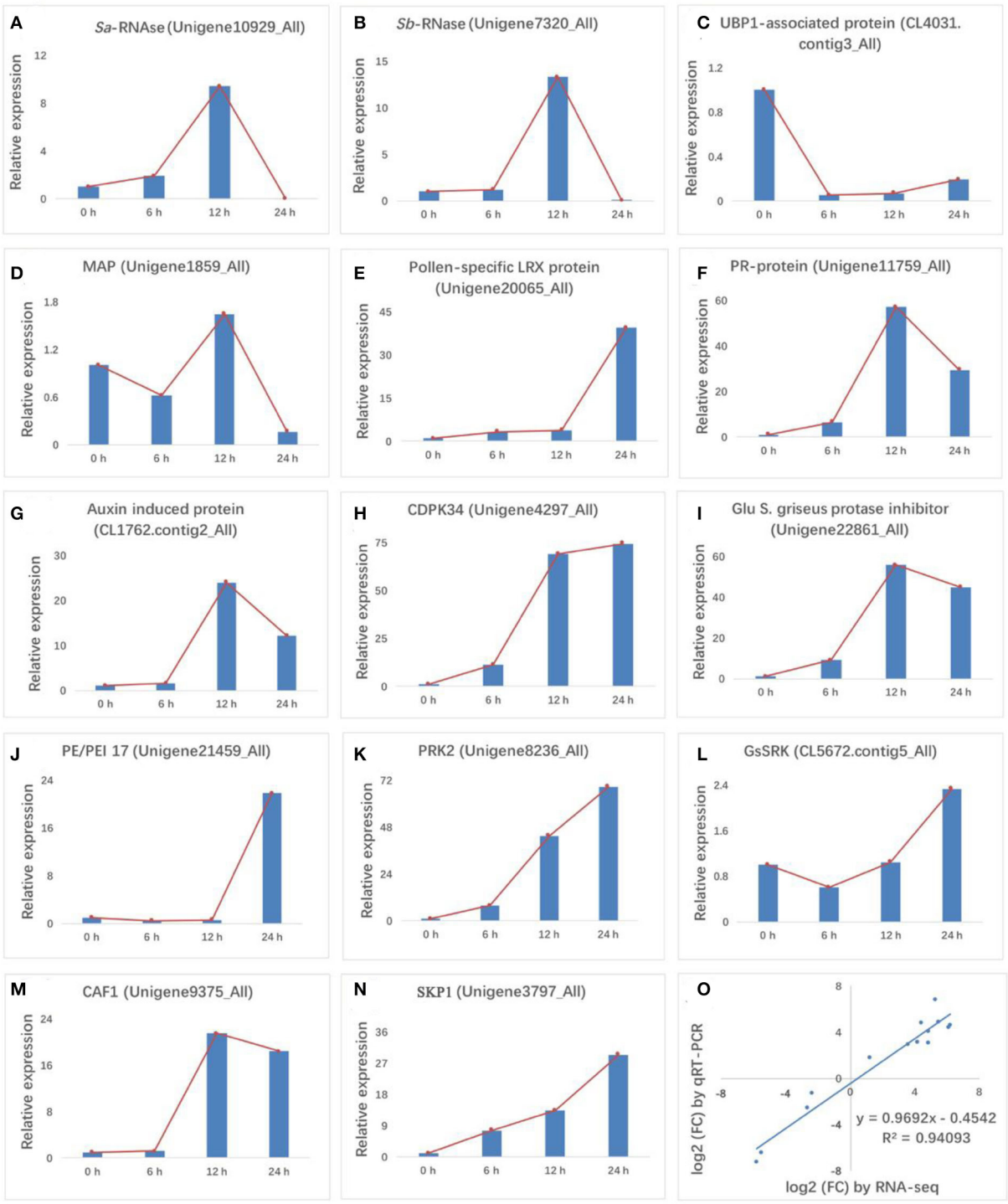


FIGURE 3
RT-qPCR validation of 14 differentially expressed genes (DEGs). **(A,B)** *S-RNase*, **(C)** cell death, **(D)** microtubule binding, **(E)** pathogen response, **(F,G)** hormone signaling, **(H)** CDPK signaling, **(I)** enzyme inhibitor activity, **(J)** cell wall, **(K,L)** pollen development, **(M)** RNA degradation, and **(N)** ubiquitin-mediated proteolysis were validated for gene expression levels at 0, 6, 12, and 24 HAP. **(O)** Comparison of fold-changes in gene expression (24 HAP/0 HAP) between RNA-Seq (x-axis) and RT-qPCR (y-axis) data. Both data were $\log_2$-transformed. In charts, red lines in **(A–N)** indicate the expression trend of each gene at the four pollination stages.

10.3389/fpls.2022.927001

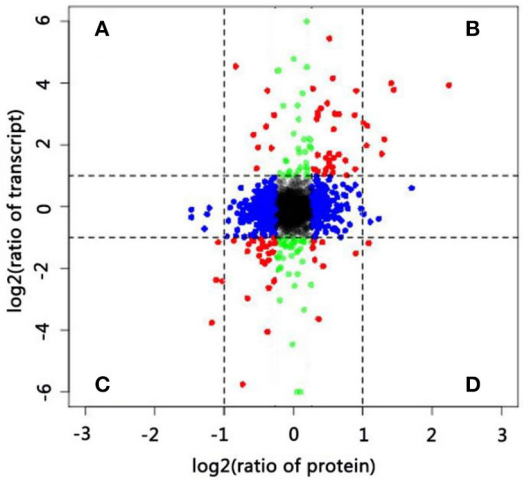


FIGURE 4
Correlations between protein and transcript levels. **(A–D)** Differentially expressed genes at both transcript and protein levels.

Franklin, 2003; McClure and Franklin-Tong, 2006; Wu et al., 2013; Claessen et al., 2019). Therefore, we focused only on the gene related to the polar growth of pollen tubes and explored its link to self-incompatibility through the following series of experiments.

Screening, basic characteristics, and homology analysis of *YABBY* genes in *F. viridis*

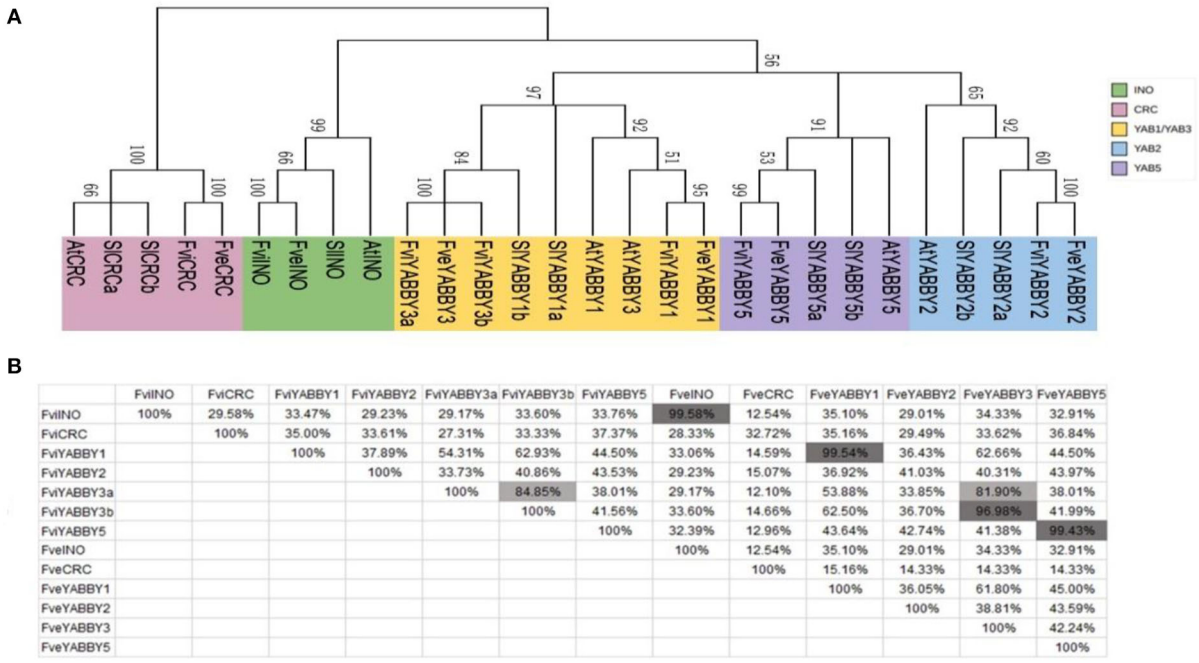
The *YABBY* gene family comprises plant-specific transcription factors playing vital roles in many biological processes during plant growth and development. These genes are differentially expressed after self-pollination. To gain a comprehensive understanding of their roles and homologs in SI, the *YABBY* family genes expressed in the style and pollen interaction system of *F. viridis* were screened using *F. vesca* as the control. Through screening and verification, 13 *YABBY* family members were finally obtained—seven and six from the style and pollen interaction system of *F. viridis* and the whole genome of *F. vesca,* respectively. These genes were named according to their homology with the *Arabidopsis YABBY* family genes (Bowman, 2000), and their basic characteristics were analyzed (Supplementary Table S10). The identified YABBY family members harbor two conserved domains—a C_2C_2 zinc finger structure and a YABBY domain—and the degree of amino acid conservation in the YABBY domain is greater than that in the zinc finger structure (Supplementary Figure S2). In addition to the conserved motifs, the various members of the YABBY family contain unique motifs (Supplementary Figure S3), implying that these genes share certain commonalities and functional specificities.

Consistent with previous reports in Arabidopsis, tomato, rice, wheat, and pomegranate (Bowman, 2000; Toriba et al., 2007; Huang et al., 2013; Zeeshan et al., 2020; Zhao et al., 2020), the *YABBY* genes in *F. viridis* and *F. vesca* were divided into five subfamilies: INO, CRC, YAB2, YAB1/YAB3, and YAB5 (Figure 5A). Similar to those in *F. vesca*, INO, CRC, YAB2, and YAB5 in *F. viridis* comprise one member, while the remaining subfamily YAB1/YAB3 comprises three members, including one in YAB1 and two in YAB3, respectively. Meanwhile, in *F. vesca*, *Arabidopsis* (Bowman, 2000), and tomato (Huang et al., 2013), YAB1/YAB3 comprises two members, one each in YAB1 and YAB3.

Homology analysis revealed that the functions of INO and YAB5 subfamily members are not differentiated between *F. viridis* and *F. vesca*, with similarity exceeding 99.43%; however, the YABBY2 and CRC subfamily members share a low similarity at 41.03 and 32.72%, respectively (Figure 5B). The YABBY1 and YABBY3 proteins in strawberries shared high homology, both within and between species, which explains the clustering of members in these two groups and their classification in the YAB1/3 subfamily. In addition, among the members of the strawberry YABBY family, only two YAB3 members (FviYABBY3a and FviYABBY3b) are present in *F. viridis*, both sharing high similarity, which suggests the occurrence of a gene duplication event in YAB3 throughout the evolution. The amino acid similarity was 99.54% between FviYABBY1 and FveYABBY1, 96.98% between FveYABBY3 and FviYABBY3a, and 81.90% between FveYABBY3 and FviYABBY3b, suggesting that FviYABBY3b has undergone functional diversification following the gene duplication event.

Analysis of *YABBY* gene expression patterns in *F. viridis*

The expression levels of *YABBY* family genes in the pistil, pollen, leaves, peduncle, calyx, and petals of *F. viridis* were analyzed using RT-qPCR. As seen in Figure 6A, the expression levels of *FviYABBY* genes were relatively high in the pistil, suggesting their important roles in pistil development. Unlike other family members, *FviYABBY1* showed the highest expression level in pollens; thus, the key function of FviYABBY1 may be in the pollen. We further analyzed the expression levels of these genes in pistils at different stages after self-pollination. As seen in Figure 6B, except for *FviYABBY1*, the other *YABBY* family genes reached their peak expression levels at 0 HAP. Further, while the expression of all *FviYABBY* genes significantly decreased at 6 h, their expression levels increased at 12 h, following another downward trend at 24 and 48 HAP.



	FviINO	FviCRC	FviYABBY1	FviYABBY2	FviYABBY3a	FviYABBY3b	FviYABBY5	FveINO	FveCRC	FveYABBY1	FveYABBY2	FveYABBY3	FveYABBY5
FviINO	100%	29.58%	33.47%	29.23%	29.17%	33.60%	33.76%	99.58%	12.54%	35.10%	29.01%	34.33%	32.91%
FviCRC		100%	35.00%	33.61%	27.31%	33.33%	37.37%	28.33%	32.72%	35.16%	29.49%	33.62%	36.84%
FviYABBY1			100%	37.89%	54.31%	62.93%	44.50%	33.06%	14.59%	99.54%	36.43%	62.66%	44.50%
FviYABBY2				100%	33.73%	40.86%	43.53%	29.23%	15.07%	36.92%	41.03%	40.31%	43.97%
FviYABBY3a					100%	84.85%	38.01%	29.17%	12.10%	53.88%	33.85%	81.90%	38.01%
FviYABBY3b						100%	41.56%	33.60%	14.66%	62.50%	36.70%	96.98%	41.99%
FviYABBY5							100%	32.39%	12.96%	43.64%	42.74%	41.38%	99.43%
FveINO								100%	12.54%	35.10%	29.01%	34.33%	32.91%
FveCRC									100%	15.16%	14.33%	14.33%	14.33%
FveYABBY1										100%	36.05%	61.80%	45.00%
FveYABBY2											100%	38.81%	43.59%
FveYABBY3												100%	42.24%
FveYABBY5													100%

FIGURE 5
Homology analysis of YABBY family members in *Fragaria viridis* and *Fragaria vesca*. **(A)** Phylogenetic tree of YABBY family members in *F. viridis* and *F. vesca*. **(B)** Amino acid similarity between YABBY family members in *F. viridis* and *F. vesca*.

The YABBY family members were clustered on the phylogenetic tree and harbored the same conserved domains (Figure 5 and Supplementary Figure S3). *FviYABBY3a* and *FviYABBY3b* as well as *FviYABBY2* and *FviYABBY5* were highly similar in terms of their tissue-specific expression patterns and expression patterns in pistils at different stages after pollination, implying that they may serve similar functions. Although the expression level of *FviYABBY1* in pistils decreased at 6 HAP, it reached the peak value at 12 HAP and then declined again. The expression trend of *FviYABBY1* was consistent with that of *S-RNase* (Du et al., 2019), suggesting that *FviYABBY1* is positively regulated by *S*-RNase.

FviYABBY1 cloning and subcellular localization

Using *F. viridis* pistil cDNA as the template, the CDS-specific target band was obtained by PCR amplification (Figure 7A), with a length of 657 bp. Monoclonal sequencing results were consistent with the transcriptome data (Du et al., 2021). *FviYABBY1* was cloned into pCAMBIA1302-GFP, and the recombinant vector was transferred into tobacco leaves through *Agrobacterium*-mediated transformation. The cellular distribution of FviYABBY1 protein was determined based on the GFP signal. Under 488 nm excitation wavelength, green fluorescent signals were observed in the nucleus and the cytoplasm (Figure 7B). Therefore, FviYABBY1 lacks a signal peptide or transmembrane structure (Supplementary Table S10) and is a non-secreted protein that cannot be transferred between cells.

Interaction of FviYABBY1 with S_a-RNase and S_b-RNase

S_a-RNase and S_b-RNase are non-toxic and non-self-activated in yeast cells. Thus, yeast two-hybrid assays were conducted on SD/-Trp-Leu-His-Ade medium without 3-AT to verify whether other proteins interact with the target RNases (Supplementary Figure S4).

The recombinant plasmids FviYABBY1-pGADT7+S_a-pGBKT7 and FviYABBY1-pGADT7+S_b-pGBKT7 were co-transformed into the yeast strain AH109. The positive yeast strains on the SD/-Trp-Leu medium were multiplied and inoculated on prefabricated plates containing SD/-Trp-Leu, SD/-Ade-His-Trp-Leu, and SD/-Trp-Leu-His-Ade/X-α-gal media. As seen in Figure 8A, all yeast strains grew normally on SD/-Trp-Leu medium, indicating that yeast co-transformation was successful. Yeast strains containing FviYABBY1-pGADT7+S_a-pGBKT7 grew stably on thr SD/-Ade-His-Trp-Leu medium, while those containing

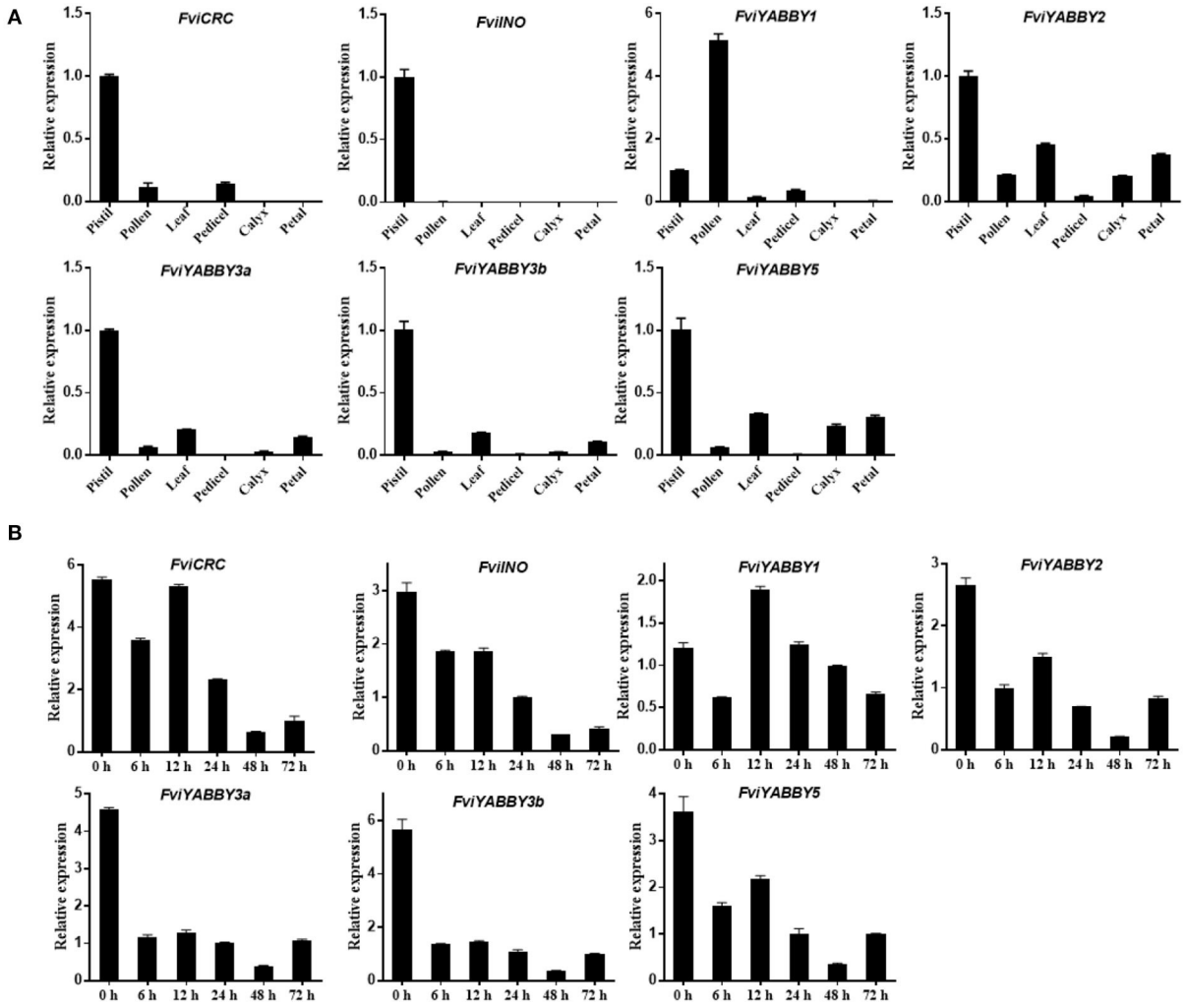


FIGURE 6
Analysis of gene expression pattern using RT-qPCR. **(A)** Tissue-specific expression patterns of *YABBY* family genes in *Fragaria viridis*. **(B)** Spatiotemporal expression after self-pollination of *YABBY* family genes in *F. viridis*.

FviYABBY1-pGADT7+S_b-pGBKT7 grew stably on the SD/-Trp-Leu-His-Ade medium and grew blue colonies on the SD/-Trp-Leu-His-Ade/X-α-gal medium. These results indicate the interaction of the FviYABBY1 protein with S_a-RNase and S_b-RNase without *S*-RNase haplotype specificity.

We further verified the protein interaction results of FviYABBY1 with S_a-RNase and S_b-RNase and the location of the interaction. The recombinant vectors were co-transferred into tobacco cells using *Agrobacterium* harboring FviYABBY1-35S:YCE+S_a-35S:YNE and FviYABBY1-35S:YCE+S_b-35S:YNE. As seen in Figure 8B, the tobacco cells containing FviYABBY1-35S:YCE+S_b-35S:YNE and FviYABBY1-35S:YCE+S_b-35S:YNE exhibited fluorescence, whereas control cells harboring YCE+35S:YNE did not. The results of the BiFC assay confirmed that FviYABBY1 interacts with S_a-RNase/S_b-RNase and that these interactions are not S-haplotype specific. Further, based on fluorescence signals, the cytoplasm is the main location of FviYABBY1–*S*-RNase interaction.

Discussion

S-RNase-based GSI in *F. viridis*

Fragaria viridis is a typical self-incompatible gametophyte. In this species, SI primarily manifests in the pollen tube, which is strongly inhibited in the styles (Evans and Jones, 1967). In a previous study, a pair of style determinants (S_a-RNase and S_b-RNase) was discovered based on transcriptomic data at 0 and 24 HAP, proving the occurrence of *S*-RNase-based GSI in

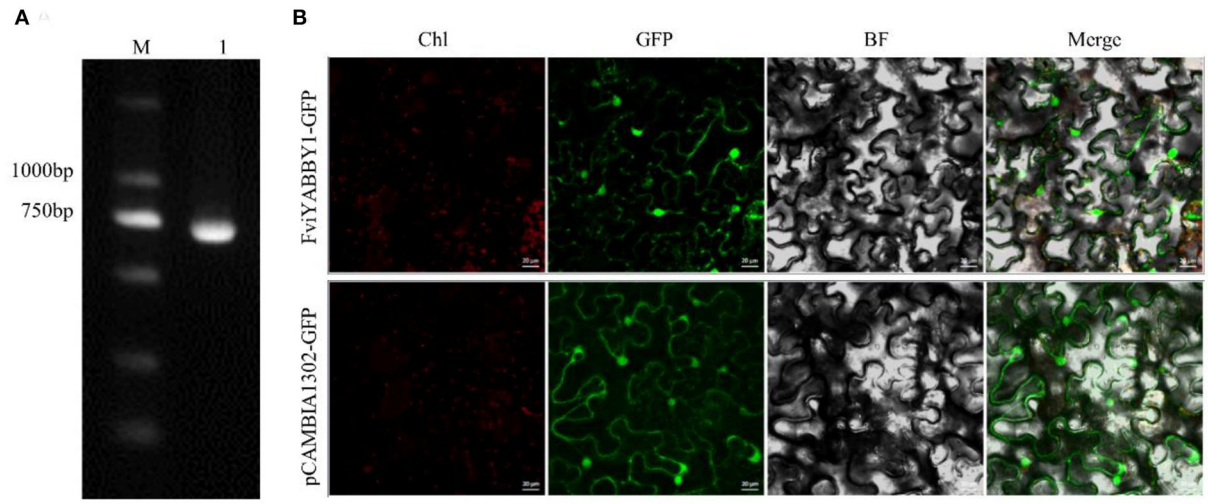


FIGURE 7
Gene cloning and subcellular localization of FviYABBY1. **(A)** *FviYABBY1* cloning. M: 2000 DNA Marker; 1: Fragment of *FviYABBY1* CDS. **(B)** Subcellular localization of FviYABBY1. Chl, chlorophyl II autofluorescence; GFP, green fluorescence protein; BF, bright field; Merge, merged image of Chl, GFP, and BF. Bar = 20 μm.

Fragaria (Du et al., 2021). Consistent with previous reports, most self-incompatible pollen tubes eventually stop growing at 2/3 of the style in the genus *Fragaria* (Evans and Jones, 1967; Du et al., 2019, 2021). Contrary to that in other species, such as *Solanum* and *Pyrus*, the self-incompatible pollen tube in *F. viridis* showed a greater extension degree in styles, suggesting differences in the intermediate mechanisms or SI modifiers among self-incompatible species (Franklin-Tong and Franklin, 2003; Zhang et al., 2009; Baek et al., 2015; Claessen et al., 2019). Therefore, we further analyzed the previously obtained transcriptomic data and combined these with proteomic data to unveil the mechanisms of pollen tube development and pollen tube growth cessation due to SI in *F. viridis*. FviYABBY1, a putative transcription factor related to the polar growth of pollen tubes, has been a research hotspot, and its roles in *F. viridis* SI are discussed in this article.

DEGs and DAPs related to SI

In *S*-RNase-based GSI systems, a type of pollen–pistil interaction, pollen grains germinate and grow normally on the stigma, and the rejection of self-pollen tube is mainly determined by the style (Franklin-Tong and Franklin, 2003; Hiscock and Allen, 2008; Shi et al., 2017; Seth et al., 2019; Aloisi et al., 2020). In the present study, the majority of the *F. viridis* pollen tubes were arrested in the style at 24 HAP (Figure 1E), implying that certain time-specific genes (24 HAP) regulating pollen tube growth are the key targets of the *Fragaria* SI system. Successful manifestation of SI entails the involvement of modifiers other than the *S-locus* determinants (McClure et al., 2000; Sassa and Hirano, 2006). Here, we identified 2,181 genes and 200 proteins that were differentially expressed between 0 and 24 HAP in receptacles with gynoecia, and many SI-related genes were further predicted and classified (Supplementary Table S9). Interestingly, many additional DEGs and DAPs were enriched in GO terms, such as "pollen development," "recognition of pollen," "pollen germination," and "pollen–style interaction." Arabinogalactan proteins (AGPs) and the AGP extension hybrid glycoprotein TTS are closely linked to pollen development and control pollen tube elongation during *S*-RNase-based GSI (Majewska-Sawka and Nothnagel, 2000; Seifert and Roberts, 2007). Furthermore, genes encoding GDSL esterase/lipase protein, fasciclin-like arabinogalactan protein, stigma-specific stig1-like protein, and glycosyltransferase are involved in pollen development and germination (Goubet et al., 2003; Updegraff et al., 2009; Zhang et al., 2022). In addition, pollen receptor-like kinase (RLK) is a key factor in pollen–style recognition (Escobar-Restrepo et al., 2007), and RALF-like proteins play critical roles in the regulation of pollen tube growth (Covey et al., 2010; Murphy and De Smet, 2014; Zhang et al., 2020). All these genes were classified into the category "pollen development" and most of them were significantly upregulated at 24 HAP (Supplementary Table S9). These results suggest that the expression of the above genes is closely related to self-pollen development in styles. Pollen is a typical model of cell polar growth. Thus, a polar growth regulator of DEGs was further investigated according to functional annotation. Although the potential functions of these genes in *Fragaria* SI require further confirmation, our results

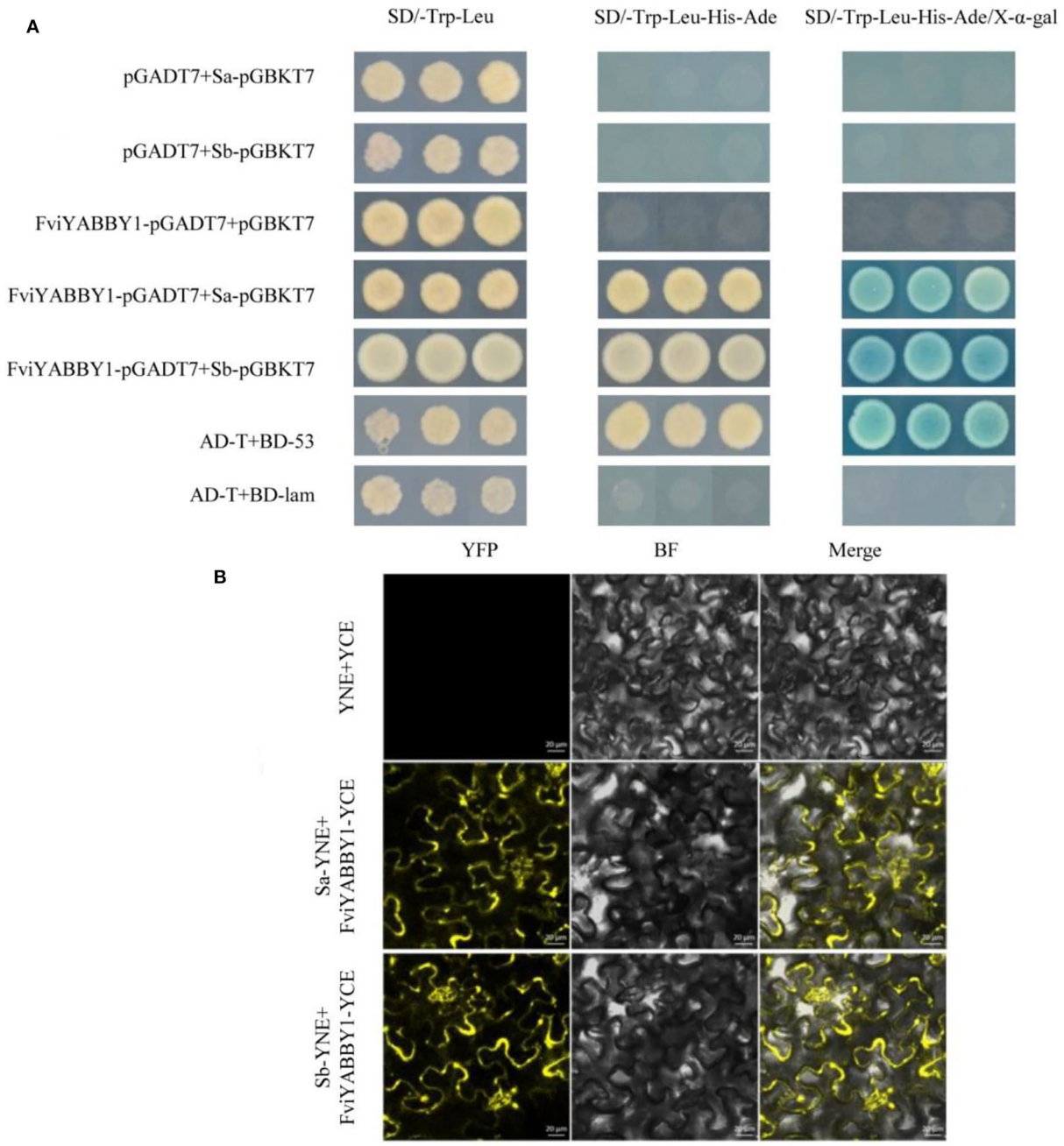


FIGURE 8
Interaction between FviYABBY1 and S_a-RNase/S_b-RNase. **(A)** Yeast two-hybrid assay. **(B)** Bimolecular fluorescence complementation (BiFC) assay. YFP, yellow fluorescence protein; BF, bright field; Merge, merged image of YFP and BF. Bar = 20 μm.

indicate that some regulators related to pollen growth are involved in the SI response of *F. viridis*.

SI is an intricate biological process closely associated with ubiquitin-mediated proteolysis, which eventually leads to PCD and RNA degradation (Thomas and Franklin-Tong, 2004; McClure and Franklin-Tong, 2006; Zhang et al., 2009). In the present study, we found that SKP1 (Unigene3797_All), as a component of the ubiquitin–ligase complex, was gradually upregulated after self-pollination (FC > 29) at 24 h compared with the expression level at 0 h in the validation data (Figure 3N). Two DAPs (CL4737_Contig1_All and CL7114_Contig_All) and one DEG (Unigene8982_All) were annotated as E3 ubiquitin-protein ligases. A series of PCD-related DEGs and DAPs, such as MAPKKKs, 18.2 kDa class I heat shock protein

(Unigene11193_All), and cytochrome P450, were identified (Supplementary Table S9), all of which are important regulators of PCD according to previous reports (Ludovico et al., 2002; del Pozo et al., 2004; Li et al., 2012). As the activity of *S*-RNase is protected in self-pollen tubes, *S*-RNase degrades cytoplasmic RNA, arrests protein synthesis, and finally inhibits pollen tube growth (Kao and Tsukamoto, 2004; Wang et al., 2009; Zhang et al., 2015). In the present study, 15 DEGs and 2 DAPs were enriched in the "RNA degradation" pathway (Supplementary Tables S4, S6), including one CCR4-associated factor (CAF), four NAC domain-containing proteins, three chaperonin 60 subunit beta 4, and two uncharacterized proteins (Supplementary Table S9). Many of these genes were annotated to "RNA degradation" in self-pollination for the first time in this work, and further research may reveal their necessity in the pollen–style interaction pathways.

Biological processes, such as "cell wall modification," "accumulation extracellular calcium," "synthesis of β-1,3 glucan callose," and "presence of phenolic derivatives," are common between SI and pathogen resistance (Zhou et al., 2014). In addition, enzymes, such as pectin lyase, cellulases, and glucanases, are typically secreted by pathogens for cell wall degradation (Wing et al., 1990; Yoder et al., 1990; Ham et al., 1997; Aich et al., 2017). Similar to that during pathogen defense (Wing et al., 1990; Marín-Rodríguez et al., 2002), cell wall modifiers, such as pectate lyase 5, beta-glucosidase 3, and glucan endo-1,3-beta-glucosidase, were downregulated at the transcript and protein levels (Supplementary Table S9). Ca^{2+} is vital in pathogen interactions, and the transient influx of Ca^{2+} across the plasma membrane plays a major role during the early stages of pathogen response (Zimmermann et al., 1997). In our transcriptome data, several Ca^{2+} signaling-related genes were markedly upregulated; as such, these genes were enriched in the KEGG pathway "plant–pathogen interaction" (Supplementary Figure S5 and Supplementary Table S9). As for pathogen response, hormones are essential for incompatibility response (Graaf et al., 2003; Robert-Seilaniantz et al., 2007; López et al., 2008; Song et al., 2013; Zhang et al., 2016). Here, we detected 56 DEGs and 8 DAPs that were enriched in the KEGG pathway "plant hormone signal transduction" (Supplementary Tables S4, S6). These findings support that the mechanism of plant–pathogen interactions are remarkably similar to that of pollen tube development and pollen–style recognition (Hodgkin et al., 1988; Thomas and Franklin-Tong, 2004).

Functional role of *YABBY* genes in *F. viridis* SI

YABBY family genes are seed-plant-specific transcription factors with critical roles in vegetative and reproductive development (Kumaran et al., 2002; Bartholmes et al., 2012; Finet et al., 2016; Filyushin et al., 2017; di Rienzo et al., 2021; Romanova et al., 2021; Luo et al., 2022). *S*-RNase-based SI occurs when pollen tubes stop growing and polar growth is hindered in self-styles (Wu et al., 2013; Claessen et al., 2019). Using transcriptomic data as clues, we focused on *YABBY* family genes, particularly *YABBY1*, which may be involved in the regulation of pollen tube polar growth. YABBY1 is primarily expressed in pollen and lacks a signal peptide, suggesting that it functions in the pollen. Yeast two-hybrid and BiFC assays confirmed that YABBY1 interacts with *S*-RNase, albeit without haplotype specificity. In species exhibiting SI, *S*-RNase enters the pollen tube non-specifically (Meng et al., 2014a,b) and exerts a toxic effect in the self-pollen tube; meanwhile, the toxicity of non-self *S*-RNase is eliminated (Wu et al., 2013; Sassa, 2016; Claessen et al., 2019). Therefore, self-*S*-RNase likely interacts with YABBY1 in the pollen tube, transmits the *S*-RNase cytotoxicity signal, and affects the polar growth of the pollen tube. In the non-self-pollen tube, the cytotoxicity of *S*-RNase is relieved by pollen determinants. In self-compatible *F. vesca*, given the loss of the *S-RNase* gene, *YABBY1* executes the normal function of regulating the polar growth of the pollen tube tip. *YABBY1* is positively correlated with *S-RNase* in an incompatible style–pollen interaction system after self-pollination. Therefore, *S*-RNase may regulate the expression of SI-related genes by positively regulating the expression of *YABBY1*.

SI is a system in which style interacts with pollen (Bedinger et al., 2017). In addition to *YABBY1*, six *YABBY* family members from *F. viridis* were highly expressed. Some *YABBY* family genes exhibited similar expression patterns and may be correlated to some extent. Genes in the same family may serve similar functions (Li et al., 2016; Lin et al., 2018); thus, other genes in the *YABBY* family may also be related to SI. However, whether they are involved in the SI response of *F. viridis*, similar to *YABBY1*, warrants further experimental exploration.

Data availability statement

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

Author contributions

JD, CG, and TW co-performed the experiment design, the major experiment operation, data handling, and manuscript writing. JW, ZN, SX, and FZ performed partial experiments. YQ and MZ proposed the research idea and experimental design and assisted with manuscript editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (Nos. 32072540 and 31872056), the Fundamental Research Funds for the Central Universities (No. KYZZ2022004), Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX21_0612), and the Jiangsu Province Agricultural Science and Technology Innovation Fund Projects [No. CX(21)2019].

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.927001/full#supplementary-material

References

Aguiar, B., Vieira, J., Cunha, A. E., Fonseca, N. A., Iezzoni, A., van Nocker, S., et al. (2015). Convergent evolution at the gametophytic self-incompatibility system in Malus and Prunus. *PLoS ONE* 10, e0126138. doi: 10.1371/journal.pone.0126138

Aich, S., Singh, R. K., Kundu, P., Pandey, S. P., and Datta, S. (2017). Genome-wide characterization of cellulases from the hemi-biotrophic plant pathogen, *Bipolaris sorokiniana*, reveals the presence of a highly stable GH7 endoglucanase. *Biotechnol. Biofuels* 10, 1–14. doi: 10.1186/s13068-017-0822-0

Allen, A. M., and Hiscock, S. J. (2008). "Evolution and phylogeny of self-incompatibility systems in angiosperms," in *Self Incompatibility in Flowering Plants. Evolution, Diversity, and Mechanisms*, ed V. E. Tong (Berlin Heidelberg: Springer-Verlag), 73–101.

Aloisi, I., Distefano, G., Antognoni, F., Potente, G., Parrotta, L., Faleri, C., et al. (2020). Temperature-dependent compatible and incompatible pollen-style interactions in *Citrus clementina* Hort. ex Tan. show different transglutaminase features and polyamine pattern. *Front. Plant Sci.* 11, 1018. doi: 10.3389/fpls.2020.01018

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Armenteros, J. J. A., 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

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 40, W597–W603. doi: 10.1093/nar/gks400

Baek, Y. S., Covey, P. A., Petersen, J. J., Chetelat, R. T., McClure, B., and Bedinger, P. A. (2015). Testing the SI × SC rule: pollen-pistil interactions in interspecific crosses between members of the tomato clade (Solanum section Lycopersicon, Solanaceae). *Am. J. Bot.* 102, 302–311. doi: 10.3732/ajb.1400484

Bartholmes, C., Hidalgo, O., and Gleissberg, S. (2012). Evolution of the YABBY gene family with emphasis on the basal eudicot eschscholzia californica (Papaveraceae). *Plant Biol.* 14, 11–23. doi: 10.1111/j.1438-8677.2011.00486.x

Bedinger, P. A., Broz, A. K., Tovar-Mendez, A., and McClure, B. (2017). Pollen-pistil interactions and their role in mate selection. *Plant Physiol.* 173, 79–90. doi: 10.1104/pp.16.01286

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bošković, R. I., Sargent, D. J., and Tobutt, K. R. (2010). Genetic evidence that two independent S-loci control RNase-based self-incompatibility in diploid strawberry. *J. Exp. Bot.* 61, 755–763. doi: 10.1093/jxb/erp340

Bowman, J. L. (2000). The YABBY gene family and abaxial cell fate. *Curr. Opin. Plant Biol.* 3, 17–22. doi: 10.1016/S1369-5266(99)00035-7

Chalivendra, S. C., Lopez-Casado, G., Kumar, A., Kassenbrock, A. R., Royer, S., Tovar-Mèndez, A., et al. (2012). Developmental onset of reproductive barriers and associated proteome changes in stigma/styles of *Solanum pennellii*. *J. Exp. Bot.* 64, 265–279. doi: 10.1093/jxb/ers324

Chen, X., Li, J., Cheng, T., Zhang, W., Liu, Y., Wu, P., et al. (2020). Molecular systematics of Rosoideae (Rosaceae). *Plant Syst. Evol.* 306, 1–12. doi: 10.1007/s00606-020-01629-z

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

Covey, P. A., Subbaiah, C. C., Parsons, R. L., Pearce, G., Lay, F. T., Anderson, M. A., et al. (2010). A pollen-specific RALF from tomato that regulates pollen tube elongation. *Plant Physiol.* 153, 703–715. doi: 10.1104/pp.110.155457

Del Duca, S., Serafini-Fracassini, D., and Cai, G. (2014). Senescence and programmed cell death in plants: polyamine action mediated by transglutaminase. *Front. Plant Sci.* 5, 120. doi: 10.3389/fpls.2014.00120

del Pozo, O., Pedley, K. F., and Martin, G. B. (2004). MAPKKKα is a positive regulator of cell death associated with both plant immunity and disease. *EMBO J.* 23, 3072–3082. doi: 10.1038/sj.emboj.7600283

di Rienzo, V., Imanifard, Z., Mascio, I., Gasser, C. S., Skinner, D. J., Pierri, C. L., et al. (2021). Functional conservation of the grapevine candidate gene INNER NO OUTER for ovule development and seed formation. *Hortic Res.* 8, 29. doi: 10.1038/s41438-021-00467-5

Du, J., Ge, C., Li, T., Wang, S., Gao, Z., Sassa, H., et al. (2021). Molecular characteristics of S-RNase alleles as the style determinant of self-incompatibility in *Fragaria viridis*. *Hortic Res.* 8, 185. doi: 10.1038/s41438-021-00623-x

Du, J., Lv, Y., Xiong, J., Ge, C., Iqbal, S., and Qiao, Y. (2019). Identifying genome-wide sequence variations and candidate genes implicated in self-incompatibility by resequencing *Fragaria viridis*. *Int. J. Mol. Sci.* 20, 1039. doi: 10.3390/ijms20051039

Elleman, C. J., and Dickinson, H. G. (1999). Commonalities between pollen/stigma and host/pathogen interactions: calcium accumulation during stigmatic penetration by Brassica oleracea pollen tubes. *Sex. Plant Reprod.* 12, 194–202. doi: 10.1007/s004970050192

Escobar-Restrepo, J. M., Huck, N., Kessler, S., Gagliardini, V., Gheyselinck, J., Yang, W. C., et al. (2007). The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. *Science* 317, 656–660. doi: 10.1126/science.1143562

Evans, W. D. (1982). The production of multispecific octoploids from *Fragaria* species and the cultivated strawberry. *Euphytica* 31, 901–907. doi: 10.1007/BF00039230

Evans, W. D., and Jones, J. K. (1967). Incompatibility in Fragaria. *Can. J. Genet. Cytol.* 9, 831–836. doi: 10.1139/g67-088

Feijó, J. A. (2010). The mathematics of sexual attraction. *J. Biol.* 9, 18. doi: 10.1186/jbiol233

Filyushin, M. A., Slugina, M. A., Shchennikova, A. V., and Kochieva, E. Z. (2017). YABBY3-orthologous genes in wild tomato species: structure, variability, and expression. *Acta Naturae* 9, 101–109. doi: 10.32607/20758251-2017-9-4-101-109

Finet, C., Floyd, S. K., Conway, S. J., Zhong, B., Scutt, C. P., and Bowman, J. L. (2016). Evolution of the YABBY gene family in seed plants. *Evol. Dev.* 18, 116–126. doi: 10.1111/ede.12173

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367

Franklin-Tong, N. V. E., and Franklin, F. C. H. (2003). Gametophytic self-incompatibility inhibits pollen tube growth using different mechanisms. *Trends Plant Sci.* 8, 598–605. doi: 10.1016/j.tplants.2003.10.008

Fujii, S., Kubo, K., and Takayama, S. (2016). Non-self-and self-recognition models in plant self-incompatibility. *Nat. Plants* 2, 16130. doi: 10.1038/nplants.2016.130

Goubet, F., Misrahi, A., Park, S. K., Zhang, Z., Twell, D., and Dupree, P. (2003). AtCSLA7, a cellulose synthase-like putative glycosyltransferase, is important for pollen tube growth and embryogenesis in Arabidopsis. *Plant Physiol.* 131, 547–557. doi: 10.1104/pp.014555

Graaf, B. H., Knuiman, B. A., Derksen, J., and Mariani, C. (2003). Characterization and localization of the transmitting tissue-specific PELPIII proteins of Nicotiana tabacum. *J. Exp. Bot.* 54, 55–63. doi: 10.1093/jxb/erg002

Gruner, P., Ulrich, D., Neinhuis, C., and Olbricht, K. (2017). *Fragaria viridis* Weston: diversity and breeding potential of an underutilised strawberry species. *Acta Hortic.* 1156, 203–208. doi: 10.17660/ActaHortic.2017.1156.31

Gu, X., Gao, Z., Zhuang, W., Qiao, Y., Wang, X., Mi, L., et al. (2013). Comparative proteomic analysis of rd29A: RdreB1BI transgenic and non-transgenic strawberries exposed to low temperature. *J. Plant Physiol.* 170, 696–706. doi: 10.1016/j.jplph.2012.12.012

Gu, Z., Meng, D., Yang, Q., Yuan, H., Wang, A., Li, W., et al. (2015). A CBL gene, MdCBL5, controls the calcium signal and influences pollen tube growth in apple. *Tree Genet. Genomes* 11, 27. doi: 10.1007/s11295-015-0853-2

Ham, K. S., Wu, S. C., Darvill, A. G., and Albersheim, P. (1997). Fungal pathogens secrete an inhibitor protein that distinguishes isoforms of plant pathogenesis-related endo-β-1, 3-glucanases. *Plant J.* 11, 169–179. doi: 10.1046/j.1365-313X.1997.11020169.x

Hancock, J. F. (1999). "The strawberry species," in *Strawberries* (Wallingford: CABI Publishing), 25–46.

Hancock, J. F., and Luby, J. J. (1993). Genetic resources at our doorstep: the wild strawberries. *Bioscience* 43, 141–147. doi: 10.2307/1312017

Hancock, J. F., Luby, J. J., Dale, A., Callow, P. W., Serce, S., and El-Shiek, A. (2002). Utilizing wild *Fragaria virginiana* in strawberry cultivar development: inheritance of photoperiod sensitivity, fruit size, gender, female fertility and disease resistance. *Euphytica* 126, 177–184. doi: 10.1023/A:1016309724998

Harkness, A., and Brandvain, Y. (2021). Non-self recognition-based self-incompatibility can alternatively promote or prevent introgression. *New Phytol.* 231, 1304–1307. doi: 10.1111/nph.17249

Hiscock, S. J., and Allen, A. M. (2008). Diverse cell signalling pathways regulate pollen-stigma interactions: the search for consensus. *New Phytol.* 179, 286–317. doi: 10.1111/j.1469-8137.2008.02457.x

Hodgkin, T., Lyon, G. D., and Dickinson, H. G. (1988). Recognition in flowering plants: a comparison of the Brassica self-incompatibility system and plant pathogen interactions. *New Phytol.* 110, 557–569. doi: 10.1111/j.1469-8137.1988.tb00296.x

Huang, Z., Van Houten, J., Gonzalez, G., Xiao, H., and van der Knaap, E. (2013). Genome-wide identification, phylogeny and expression analysis of SUN, OFP and YABBY gene family in tomato. *Mol. Genet. Genom.* 288, 111–129. doi: 10.1007/s00438-013-0733-0

Iseli, C., Jongeneel, C. V., and Bucher, P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *ISMB* 99, 138–148.

Ishimizu, T., Sato, Y., Saito, T., Yoshimura, Y., Norioka, S., Nakanishi, T., et al. (1996). Identification and partial amino acid sequences of seven S-RNases associated with self-incompatibility of Japanese pear, *Pyrus pyrifolia* Nakai. *J. Biochem.* 120, 326–334. doi: 10.1093/oxfordjournals.jbchem.a021417

Iwano, M., and Takayama, S. (2012). Self/non-self discrimination in angiosperm self-incompatibility. *Curr. Opin. Plant Biol.* 15, 78–83. doi: 10.1016/j.pbi.2011.09.003

Kao, T., and Tsukamoto, T. (2004). The molecular and genetic bases of S-RNase-based self-incompatibility. *Plant Cell.* 16, S72–S83. doi: 10.1105/tpc.016154

Kubo, K., Entani, T., Takara, A., Wang, N., Fields, A. M., Hua, Z., et al. (2010). Collaborative non-self recognition system in S-RNase-based self-incompatibility. *Science* 330, 796–799. doi: 10.1126/science.1195243

Kumar, S., Deng, C. H., Hunt, M., Kirk, C., Wiedow, C., Rowan, D., et al. (2021). Homozygosity mapping reveals population history and trait architecture in self-incompatible pear (*Pyrus* spp.). *Front. Plant Sci.* 11, 590846. doi: 10.3389/fpls.2020.590846

Kumaran, M. K., Bowman, J. L., and Sundaresan, V. (2002). YABBY polarity genes mediate the repression of KNOX homeobox genes in Arabidopsis. *Plant Cell* 14, 2761–2770. doi: 10.1105/tpc.004911

Labokas, J., and Bagdonaitė, E. (2005). Phenotypic diversity of *Fragaria vesca* and *F. viridis* in Lithuania. *Biologija* 3, 19–22. Available online at: https://www.lmaleidykla.lt/ojs/index.php/biologija/article/view/583

Li, H., Yang, Q., Li, J., Gao, H., Li, P., and Zhou, H. (2015). The impact of temperature on microbial diversity and AOA activity in the Tengchong Geothermal Field China. *Sci Rep.* 5, 17056. doi: 10.1038/srep17056

Li, Q., Yan, W., Chen, H., Tan, C., Han, Z., Yao, W., et al. (2016). Duplication of OsHAP family genes and their association with heading date in rice. *J. Exp. Bot.* 67, 1759–1768. doi: 10.1093/jxb/erv566

Li, W., Meng, D., Gu, Z., Yang, Q., Yuan, H., Li, Y., et al. (2018). Apple S-RNase triggers inhibition of tRNA aminoacylation by interacting with a soluble inorganic pyrophosphatase in growing self-pollen tubes *in vitro*. *New Phytol.* 218, 579–593. doi: 10.1111/nph.15028

Li, Y., Hou, X., Lin, L., Jing, S., and Deng, M. (2000). Abnormal pollen germination and embryo abortion in the interspecific cross, Fragaria × ananassa × *F. vesca*, as related to cross-incompatibility. *J. Jap. Soc. Hort. Sci.* 69, 84–89. doi: 10.2503/jjshs.69.84

Li, Z., Yue, H., and Xing, D. (2012). MAP Kinase 6-mediated activation of vacuolar processing enzyme modulates heat shock-induced programmed cell death in Arabidopsis. *New Phytol.* 195, 85–96. doi: 10.1111/j.1469-8137.2012.04131.x

Lin, Y., Wang, K., Li, X., Sun, C., Yin, R., Wang, Y., et al. (2018). Evolution, functional differentiation, and co-expression of the RLK gene family revealed in Jilin ginseng, Panax ginseng CA Meyer. *Mol. Genet. Genom.* 293, 845–859. doi: 10.1007/s00438-018-1425-6

Liston, A., Cronn, R., and Ashman, T. L. (2014). Fragaria: a genus with deep historical roots and ripe for evolutionary and ecological insights. *Am. J. Bot.* 101, 1686–1699. doi: 10.3732/ajb.1400140

López, M. A., Bannenberg, G., and Castresana, C. (2008). Controlling hormone signaling is a plant and pathogen challenge for growth and survival. *Curr. Opin. Plant Biol.* 11, 420–427. doi: 10.1016/j.pbi.2008.05.002

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21. doi: 10.1186/s13059-014-0550-8

Ludovico, P., Rodrigues, F., Almeida, A., Silva, M. T., Barrientos, A., and Côrte-Real, M. (2002). Cytochrome c release and mitochondria involvement in programmed cell death induced by acetic acid in *Saccharomyces cerevisiae*. *Mol. Biol. Cell.* 13, 2559–2976. doi: 10.1091/mbc.e01-12-0161

Luo, J., Zhou, Y., Pan, Q., Mu, Q., and Gu, T. (2022). Characterization of YABBY genes and the correlation between their transcript levels and histone modifications in strawberry. *Sci. Hortic.* 295, 110815. doi: 10.1016/j.scienta.2021.110815

Maas, J. L., Pooler, M. R., and Galletta, G. J. (1997). Bacterial angular leafspot disease strawberry: present status and prospects for control. *Adv Strawb.* 14, 18–24.

Majewska-Sawka, A., and Nothnagel, E. A. (2000). The multiple roles of arabinogalactan proteins in plant development. *Plant Physiol.* 122, 3–10. doi: 10.1104/pp.122.1.3

Marín-Rodríguez, M. C., Orchard, J., and Seymour, G. B. (2002). Pectate lyases, cell wall degradation and fruit softening. *J. Exp. Bot.* 53, 2115–2119. doi: 10.1093/jxb/erf089

Marta, A. E., Camadro, E. L., Díaz-Ricci, J. C., and Castagnaro, A. P. (2004). Breeding barriers between the cultivated strawberry, Fragaria× ananassa, and related wild germplasm. *Euphytica* 136, 139–150. doi: 10.1023/B:EUPH.0000030665.95757.76

McClure, B. A., Cruz-Garcia, F., Beecher, B., and Sulaman, W. (2000). Factors affecting inter-and intra-specific pollen rejection in Nicotiana. *Ann. Bot.* 85, 113–123. doi: 10.1006/anbo.1999.1061

McClure, B. A., and Franklin-Tong, V. (2006). Gametophytic self-incompatibility: understanding the cellular mechanisms involved in "self" pollen tube inhibition. *Planta* 224, 233–245. doi: 10.1007/s00425-006-0284-2

Mello, B., Tao, Q., Tamura, K., and Kumar, S. (2017). Fast and accurate estimates of divergence times from big data. *Mol. Biol. Evol.* 34, 45–50. doi: 10.1093/molbev/msw247

Meng, D., Gu, Z., Li, W., Wang, A., Yuan, H., Yang, Q., et al. (2014a). Apple MdABCF assists in the transportation of S-RNase into pollen tubes. *Plant J.* 78, 990–1002. doi: 10.1111/tpj.12524

Meng, D., Gu, Z., Yuan, H., Wang, A., Li, W., Yang, Q., et al. (2014b). The microtubule cytoskeleton and pollen tube Golgi vesicle system are required for *in vitro* S-RNase internalization and gametic self-incompatibility in apple. *Plant Cell Physiol.* 55, 977–989. doi: 10.1093/pcp/pcu031

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Muñoz-Sanz, J. V., Zuriaga, E., Cruz-García, F., McClure, B., and Romero, C. (2020). Self-(in)compatibility systems: target traits for crop-production, plant breeding, and biotechnology. *Front. Plant Sci.* 11, 195. doi: 10.3389/fpls.2020.00195

Murphy, E., and De Smet, I. (2014). Understanding the RALF family: a tale of many species. *Trends Plant Sci.* 19, 664–671. doi: 10.1016/j.tplants.2014.06.005

Qu, H., Zhang, Z., Wu, F., and Wang, Y. (2016). The role of Ca2+ and Ca2+ channels in the gametophytic self-incompatibility of *Pyrus pyrifolia*. *Cell Calcium* 60, 299–308. doi: 10.1016/j.ceca.2016.06.006

Robert-Seilaniantz, A., Navarro, L., Bari, R., and Jones, J. D. (2007). Pathological hormone imbalances. *Curr. Opin. Plant Biol.* 10, 372–379. doi: 10.1016/j.pbi.2007.06.003

Romanova, M. A., Maksimova, A. I., Pawlowski, K., and Voitsekhovskaja, O. V. (2021). YABBY genes in the development and evolution of land plants. *Int. J. Mol. Sci.* 22, 4139. doi: 10.3390/ijms22084139

Sanabria, N., Goring, D., Nürnberger, T., and Dubery, I. (2008). Self/nonself perception and recognition mechanisms in plants: a comparison of self-incompatibility and innate immunity. *New Phytol.* 178, 503–514. doi: 10.1111/j.1469-8137.2008.02403.x

Sargent, D. J., Geibel, M., Hawkins, J. A., Wilkinson, M. J., Battey, N. H., and Simpson, D. W. (2004). Quantitative and qualitative differences in morphological traits revealed between diploid *Fragaria* species. *Ann. Bot.* 94, 787–796. doi: 10.1093/aob/mch217

Sassa, H. (2016). Molecular mechanism of the S-RNase-based gametophytic self-incompatibility in fruit trees of Rosaceae. *Breed. Sci.* 66, 116–121. doi: 10.1270/jsbbs.66.116

Sassa, H., and Hirano, H. (2006). Identification of a new class of pistil-specific proteins of Petunia inflata that is structurally similar to, but functionally distinct from, the self-incompatibility factor HT. *Mol. Genet. Genom.* 275, 97–104. doi: 10.1007/s00438-005-0067-7

Sassa, H., Kakui, H., and Mai, M. (2010). Pollen-expressed F-box gene family and mechanism of S-RNase-based gametophytic self-incompatibility (GSI) in Rosaceae. *Sex. Plant Reprod.* 23, 39–43. doi: 10.1007/s00497-009-0111-6

Seifert, G. J., and Roberts, K. (2007). The biology of arabinogalactan proteins. *Annu. Rev. Plant Biol.* 58, 137–161. doi: 10.1146/annurev.arplant.58.032806.103801

Seth, R., Bhandawat, A., Parmar, R., Singh, P., Kumar, S., and Sharma, R. K. (2019). Global transcriptional insights of pollen-pistil interactions commencing self-incompatibility and fertilization in tea [*Camellia sinensis* (L.) O. *Kuntze*]. *Int. J. Mol. Sci.* 20, 539. doi: 10.3390/ijms20030539

Shi, D., Tang, C., Wang, R., Gu, C., Wu, X., Hu, S., et al. (2017). Transcriptome and phytohormone analysis reveals a comprehensive phytohormone and pathogen defence response in pear self-/cross-pollination. *Plant Cell Rep.* 36, 1785–1799. doi: 10.1007/s00299-017-2194-0

Song, S., Qi, T., Fan, M., Zhang, X., Gao, H., Huang, H., et al. (2013). The bHLH subgroup IIId factors negatively regulate jasmonate-mediated plant defense and development. *PLoS Genet.* 9, e1003653. doi: 10.1371/journal.pgen.1003653

Stahle, M. I., Kuehlich, J., Staron, L., von Arnim, A. G., and Golz, J. F. (2009). YABBYs and the transcriptional corepressors LEUNIG and LEUNIG_HOMOLOG maintain leaf polarity and meristem activity in Arabidopsis. *Plant Cell.* 21, 3105–3118. doi: 10.1105/tpc.109.070458

Staudt, G. (2009). Strawberry biogeography, genetics and systematics. *Acta Hortic.* 842, 71–84. doi: 10.17660/ActaHortic.2009.842.1

Tao, R., and Iezzoni, A. F. (2010). The S-RNase-based gametophytic self-incompatibility system in Prunus exhibits distinct genetic and molecular features. *Sci. Hortic.* 124, 423–433. doi: 10.1016/j.scienta.2010.01.025

Thomas, S. G., and Franklin-Tong, V. E. (2004). Self-incompatibility triggers programmed cell death in Papaver pollen. *Nature* 429, 305–309. doi: 10.1038/nature02540

Tolin, S., Arrigoni, G., Moscatiello, R., Masi, A., Navazio, L., Sablok, G., et al. (2013). Quantitative analysis of the naringenin-inducible proteome in *Rhizobium leguminosarum* by isobaric tagging and mass spectrometry. *Proteomics* 13, 1961–1972. doi: 10.1002/pmic.201200472

Toriba, T., Harada, K., Takamura, A., Nakamura, H., Ichikawa, H., Suzaki, T., et al. (2007). Molecular characterization the YABBY gene family in *Oryza sativa* and expression analysis of OsYABBY1. *Mol. Genet. Genom.* 277, 457–468 doi: 10.1007/s00438-006-0202-0

Ulrich, D., Komes, D., Olbricht, K., and Hoberg, E. (2007). Diversity of aroma patterns in wild and cultivated Fragaria accessions. *Genet. Resour. Crop Evol.* 54, 1185–1196. doi: 10.1007/s10722-006-9009-4

Updegraff, E. P., Zhao, F., and Preuss, D. (2009). The extracellular lipase EXL4 is required for efficient hydration of Arabidopsis pollen. *Sex. Plant Reprod.* 22, 197–204. doi: 10.1007/s00497-009-0104-5

Vekemans, X., and Castric, V. (2021). When the genetic architecture matters: evolutionary and ecological implications of self versus nonself recognition in plant self-incompatibility. *New Phytol.* 231, 1630–1643. doi: 10.1111/nph.17471

Vieira, J., Fonseca, N. A., and Vieira, C. P. (2008). An S-RNase-based gametophytic self-incompatibility system evolved only once in eudicots. *J. Mol. Evol.* 67, 179–190. doi: 10.1007/s00239-008-9137-x

Vieira, J., Pimenta, J., Gomes, A., Laia, J., Rocha, S., Heitzler, P., et al. (2021). The identification of the Rosa S-locus and implications on the evolution of the Rosaceae gametophytic self-incompatibility systems. *Sci. Rep.* 11, 3710. doi: 10.1038/s41598-021-83243-8

Vieira, J., Rocha, S., Vázquez, N., López-Fernández, H., Fdez-Riverola, F., Reboiro-Jato, M., et al. (2019). Predicting specificities under the non-self gametophytic self-incompatibility recognition model. *Front. Plant Sci.* 10, 879. doi: 10.3389/fpls.2019.00879

Wang, C., Wu, J., Xu, G., Gao, Y., Chen, G., Wu, J., et al. (2010). S-RNase disrupts tip-localized reactive oxygen species and induces nuclear DNA degradation in incompatible pollen tubes of *Pyrus pyrifolia*. *J. Cell Sci.* 123, 4301–4309. doi: 10.1242/jcs.075077

Wang, C., Xu, G., Jiang, X., Chen, G., Wu, J., Wu, H., et al. (2009). S-RNase triggers mitochondrial alteration and DNA degradation in the incompatible pollen tube of *Pyrus pyrifolia in vitro*. *Plant J.* 57, 220–229. doi: 10.1111/j.1365-313X.2008.03681.x

Wang, D., Skibbe, D. S., and Walbot, V. (2013). Maize Male sterile 8 (Ms8), a putative β-1, 3-galactosyltransferase, modulates cell division, expansion, and differentiation during early maize anther development. *Plant Reprod.* 26, 329–338. doi: 10.1007/s00497-013-0230-y

Wen, B., Zhou, R., Feng, Q., Wang, Q., Wang, J., and Liu, S. (2014). IQuant: an automated pipeline for quantitative proteomics based upon isobaric tags. *Proteomics* 14, 2280–2285. doi: 10.1002/pmic.201300361

Williams, J. S., Der, J. P., dePamphili, C. W., and Kao, T.-h. (2014). Transcriptome analysis reveals the same 17 S-locus F-box genes in two haplotypes of the self-incompatibility locus of Petunia inflate. *Plant Cell.* 26, 2873–2888. doi: 10.1105/tpc.114.126920

Williams, J. S., Wu, L., Li, S., Sun, P., and Kao, T. H. (2015). Insight into S-RNase-based self-incompatibility in Petunia: recent findings and future directions. *Front. Plant Sci.* 6, 41. doi: 10.3389/fpls.2015.00041

Wing, R. A., Yamaguchi, J., Larabell, S. K., Ursin, V. M., and McCormick, S. (1990). Molecular and genetic characterization of two pollen-expressed genes that have sequence similarity to pectate lyases of the plant pathogen Erwinia. *Plant Mol. Biol.* 14, 17–28. doi: 10.1007/BF00015651

Wu, C., Gu, Z., Li, T., Yu, J., Liu, C., Fan, W., et al. (2021). The apple MdPTI1L kinase is phosphorylated by MdOXI1 during S-RNase-induced reactive oxygen species signaling in pollen tubes. *Plant Sci.* 305, 110824. doi: 10.1016/j.plantsci.2021.110824

Wu, J., Gu, C., Khan, M. A., Wu, J., Gao, Y., Wang, C., et al. (2013). Molecular determinants and mechanisms of gametophytic self-incompatibility in fruit trees of Rosaceae. *Crit. Rev. Plant Sci.* 32, 53–68. doi: 10.1080/07352689.2012.715986

Wu, L., Williams, J. S., Sun, L., and Kao, T. H. (2020). Sequence analysis of the Petunia inflata S-locus region containing 17 S-Locus F-Box genes and the S-RNase gene involved in self-incompatibility. *Plant J.* 104, 1348–1368. doi: 10.1111/tpj.15005

Yoder, M. D., DeChaine, D. A., and Jurnak, F. (1990). Preliminary crystallographic analysis of the plant pathogenic factor, pectate lyase C from *Erwinia chrysanthemi*. *J. Biol. Chem.* 265, 11429–11431. doi: 10.1016/S0021-9258(19)38415-7

Zeeshan, A. B., Yang, Y., Sharif, R., Wu, S. N., Xie, Y., and Wang, C. (2020). Genome wide identification, characterization, and expression analysis of YABBY-gene family in wheat (*Triticum aestivum* L.). *Agronomy* 10, 1189. doi: 10.3390/agronomy10081189

Zhang, C., Wang, L., Wei, K., Wu, L., Li, H., Zhang, F., et al. (2016). Transcriptome analysis reveals self-incompatibility in the tea plant (*Camellia sinensis*) might be under gametophytic control. *BMC Genom.* 17, 359. doi: 10.1186/s12864-016-2703-5

Zhang, H., Jing, X., Chen, Y., Liu, Z., Xin, Y., and Qiao, Y. (2020). The genome-wide analysis of RALF-Like genes in strawberry (wild and cultivated) and five other plant species (Rosaceae). *Genes* 11, 174. doi: 10.3390/genes11020174

Zhang, P., Zhang, H., Du, J., and Qiao, Y. (2022). Genome-wide identification and co-expression analysis of GDSL genes related to suberin formation during fruit russeting in pear. *Hortic Plant J.* 8, 153–170. doi: 10.1016/j.hpj.2021.11.010

Zhang, S., Ding, F., He, X., Luo, C., Huang, G., and Hu, Y. (2015). Characterization of the 'Xiangshui'lemon transcriptome by de novo assembly to discover genes associated with self-incompatibility. *Mol. Genet. Genom.* 290, 365–375. doi: 10.1007/s00438-014-0920-7

Zhang, S., Li, C., Cao, J., Zhang, Y., Zhang, S., Xia, Y., et al. (2009). Altered architecture and enhanced drought tolerance in rice via the down-regulation of indole-3-acetic acid by TLD1/OsGH3. 13 activation. *Plant Physiol.* 151, 1889–1901. doi: 10.1104/pp.109.146803

Zhang, T., Gao, C., Yue, Y., Liu, Z., Ma, C., Zhou, G., et al. (2017). Time-course transcriptome analysis of compatible and incompatible pollen-stigma interactions in *Brassica napus* L. *Front. Plant Sci.* 8, 682. doi: 10.3389/fpls.2017.00682

Zhao, P., Zhang, L., and Zhao, L. (2015). Dissection of the style's response to pollination using transcriptome profiling in self-compatible (*Solanum pimpinellifolium*) and self-incompatible (*Solanum chilense*) tomato species. *BMC Plant Biol.* 15, 119. doi: 10.1186/s12870-015-0492-7

Zhao, Y., Liu, C., Ge, D., Yan, M., Ren, Y., Huang, X., et al. (2020). Genome-wide identification and expression of YABBY genes family during flower development in *Punica granatum* L. *Gene* 752, 144784. doi: 10.1016/j.gene.2020.144784

Zhou, Q., Jia, J., Huang, X., Yan, X., Cheng, L., Chen, S., et al. (2014). The large-scale investigation of gene expression in *Leymus chinensis* stigmas provides a valuable resource for understanding the mechanisms of Poaceae self-incompatibility. *BMC Genom.* 15, 399. doi: 10.1186/1471-2164-15-399

Zimmermann, S., Nürnberger, T., Frachisse, J. M., Wirtz, W., Guern, J., Hedrich, R., et al. (1997). Receptor-mediated activation of a plant Ca2+-permeable ion channel involved in pathogen defense. *Proc. Natl. Acad. Sci. U.S.A.* 94, 2751–2755. doi: 10.1073/pnas.94.6.2751

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership